Entropy measures, entropy estimators, and their performance in quantifying complex dynamics: Effects of artifacts, nonstationarity, and long-range correlations

Wanting Xiong,[1,2] Luca Faes,[3] and Plamen Ch. Ivanov[2,4,5,*]

[1]*School of Systems Science, Beijing Normal University, Beijing 100875, People's Republic of China*

[2]*Keck Laboratory for Network Physiology, Department of Physics, Boston University, Boston, Massachusetts 02215, USA*

[3]*Bruno Kessler Foundation and BIOtech, University of Trento, Trento 38123, Italy*

[4]*Harvard Medical School and Division of Sleep Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA*

[5]*Institute of Solid State Physics, Bulgarian Academy of Sciences, Sofia 1784, Bulgaria*

(Received 25 January 2017; revised manuscript received 27 April 2017; published 12 June 2017)

Entropy measures are widely applied to quantify the complexity of dynamical systems in diverse fields. However, the practical application of entropy methods is challenging, due to the variety of entropy measures and estimators and the complexity of real-world time series, including nonstationarities and long-range correlations (LRC). We conduct a systematic study on the performance, bias, and limitations of three basic measures (entropy, conditional entropy, information storage) and three traditionally used estimators (linear, kernel, nearest neighbor). We investigate the dependence of entropy measures on estimator- and process-specific parameters, and we show the effects of three types of nonstationarities due to artifacts (trends, spikes, local variance change) in simulations of stochastic autoregressive processes. We also analyze the impact of LRC on the theoretical and estimated values of entropy measures. Finally, we apply entropy methods on heart rate variability data from subjects in different physiological states and clinical conditions. We find that entropy measures can only differentiate changes of specific types in cardiac dynamics and that appropriate preprocessing is vital for correct estimation and interpretation. Demonstrating the limitations of entropy methods and shedding light on how to mitigate bias and provide correct interpretations of results, this work can serve as a comprehensive reference for the application of entropy methods and the evaluation of existing studies.

DOI: 10.1103/PhysRevE.95.062114

I. INTRODUCTION

The growing awareness that many real-world systems exhibit complex dynamics that are challenging to quantify has initiated extensive interest in developing measures and approaches for time series analysis to characterize these systems. In this context, the utilization of tools taken from information theory has become extremely popular for the assessment of the degree of complexity of physical, biological, physiological, social, and econometric systems. A variety of measures rooted in the concept of entropy and implemented according to several estimation approaches have been proposed, including approximate entropy [1], sample entropy [2], corrected conditional entropy [3], fuzzy entropy [4], compression entropy [5], permutation entropy [6,7], distribution entropy [8], multiscale entropy [9–12], self entropy and information storage [13,14]. These measures have emerged as a less ambitious but more practical alternative to classical techniques for the analysis of nonlinear dynamical systems, like correlation dimension [15], Lyapunov exponents [16], and nonlinear prediction methods [17,18]. In fact, the popularity of entropy measures stems from their applicability to short and noisy processes with important stochastic components such as those describing the dynamical activity of real-world systems. These measures have been applied with great success to numerous research fields, including heart rate variability [3,19–23], cardiovascular control [3,24–26], cerebrovascular dynamics [27,28], cardiac arrhythmias [29], financial time series analysis [30,31], gait and posture [32–36], climatology [37], earth sciences [38], cellular automata [39,40], electromyography [41], electroencephalography [42–44], magnetoencephalography [45], functional neuroimaging [14,46,47], and others [48–51].

Despite the broad relevance and application of entropy measures for various systems and fields of science, a number of theoretical, computational, and practical issues exist that often prevent a fair evaluation of the performance of these measures, as well as a correct interpretation of the measured complexity of the observed dynamics.

First, since there are many entropy measures with a variety of entropy estimators that are not always independent of each other, it is not straightforward to associate a given measure to the complexity of the dynamical system under analysis and to compare the variety of entropy measures obtained by different estimators. In addition, the crucial but often elusive term of "complexity" is also related to several other concepts in physics and biology, such as the existence of long-range correlations [52–56], nonlinear multifractal properties [57–61], and/or chaotic dynamics [15,16], which are not univocally linked to the signal features reflected by entropy measures [62]; even within the family of entropy measures, different working definitions of complexity have been proposed, e.g., in terms of "randomness" [63], "unpredictability" [3], or "regularity" [64].

Second, given that the practical computation of information-theoretic measures from real-world time series is not a trivial task, several approaches exist for the estimation of these measures [65]. Entropy estimators differ in the assumptions made about the properties of the investigated process and follow different approaches to approximate the probability density function utilized in the computation of entropy measures. Thus, entropy estimates are often highly

*plamen@buphy.bu.edu

 ©2017 American Physical Society

dependent on method-specific parameters. In the absence of a comparative evaluation of the different estimators, assessing their performances and interpreting results obtained using different estimation strategies has become a subjective task. Furthermore, an incorrect or unaware setting of the estimation parameters may easily lead to wrong inference about the properties of the observed dynamics.

Third, even though stationarity of time series is a prerequisite for the estimation of entropy measures for the target dynamical system, entropy estimators are often blindly applied without checking the fulfillment of this prerequisite. The presence of nonstationarity is often due to artifacts of various nature and exists in diversified forms such as trends, spikes, and changes in local variance. Due to the differences in entropy measures and estimators, the effects of nonstationarity vary for different entropy measures and for different estimation approaches. Therefore, a comprehensive investigation of the limitations and biases of entropy-based methods in the presence of nonstationarity is not only vital for reducing the biases in the estimation of entropy measures, but also important for the evaluation and comparison of results from different studies.

Fourth, an unaddressed issue with the computation of entropy measures is the effect of long-range correlations. It is well known that a broad class of dynamic processes in physics, biology, and econometrics exhibit long-range power-law correlations that result in scaling properties observed across multiple temporal scales [53,55,57,58,66–74]. Despite the fact that these intrinsic properties of dynamic processes are manifested even at the shorter scales and within the shorter time windows typically used in the assessment of information theoretic measures, their effects on the estimated values of entropy measures are not comprehensively investigated and not taken into account in the majority of empirical studies.

Due to the theoretical and practical issues related to the variety of entropy measures, entropy estimators, and the complexity of real-world time series mentioned above, it is therefore difficult to compare and evaluate the results from existing literature, which are often not consistent or even contradicting because different studies are based on data with different types of nonstationarity and long-range correlations, and researchers adopt different entropy measures and estimators as well as different data preprocessing and filtering procedures, which affect the outcome of information theoretic analyses.

To address the problems and challenges mentioned above, here we present a systematic study on the performance of entropy measures and estimators in various situations with both simulated and empirical time series. We aim to answer three questions: (1) to what extent entropy measures adequately characterize the dynamics of complex systems; (2) what are the limitations and biases of entropy estimators in approximating entropy measures from time series with various types of nonstationarity and presence of long-range correlations; (3) how to perform credible estimations and provide appropriate interpretations.

We present a unifying framework for the definition of entropy measures and corresponding estimation methods from time series data, which serves to clarify their theoretical meanings and assess their practical significance in the evaluation of the complexity of dynamic processes measured from physical systems. We show that a range of information theoretic measures can be subsumed by the three basic measures of entropy, conditional entropy, and information storage, and three of most widely used approaches for the quantification of these measures—linear estimator, kernel estimator, and nearest-neighbor estimator.

Further, we provide a detailed systematic analysis of the most basic frequently encountered dynamic processes and perform a comparative assessment of entropy measures and entropy estimators on multiple realizations of these processes. We study the dependence of entropy measures on estimator-specific parameters, as well as the effects of three types of nonstationarities due to artifacts that are commonly encountered in real data (i.e., slow trends, random spikes, and local variance changes). Importantly, we present for the first time a systematic quantitative assessment of the impact on entropy measures of trends originating from the intrinsic dynamics of systems exhibiting multifractal scaling properties, both in the case of long-range power-law correlations and in the more complicated and realistic situation in which long-range correlations and short-term autoregression coexist.

Finally, we consider a practical case of study that subsumes all the issues treated in the simulations, i.e., the study of human heartbeat fluctuations in different physiological states (wake and sleep) and pathological conditions (healthy and congestive heart failure). These analyses evidence advantages and pitfalls of entropy measures and estimators, as well as provide indications for their optimal use in the study of real-world time series, including recommendations about which measure to adopt depending on the purpose of the analysis, which estimator to implement in different conditions, how to deal with nonstationarities and artifacts, and how to interpret the values obtained from complex systems with different coexisting types of dynamics.

II. METHODS

A. Entropy measures

In the analysis of dynamical systems, entropy measures are used to characterize the temporal statistical structure of a system evolving in time. In an information-theoretic framework, the "information" contained in a dynamical system varies at each time step. When the system transits from past states to a new state, new information is produced in addition to the information that is already carried by the past states. This process is reflected by entropy measures: the entropy quantifies the information carried by the present state of the system, 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 system.

To introduce the notation, we consider a dynamical system $\mathcal{X}$ and assume that the states visited by this system are described by the stochastic process X. Let us further denote X_n as the random variable obtained by sampling the process X at the present time n, and $X_n^- = [X_1, \ldots, X_{n-2}, X_{n-1}]$ as the vector variable describing the past of X. The probability distribution for an individual variable $X_i, i =$

$1,\ldots,n$, is $p(x_i) = \Pr\{X_i = x_i\}, x_i \in \mathcal{A}_i$, where $\mathcal{A}_i$ is the set of all possible values that may be taken by X_i. Then, the process X is fully characterized by the joint probability distributions $p(x_1,\ldots,x_n) = \Pr\{X_1 = x_1,\ldots,X_n = x_n\}, \forall n \geqslant 1$, with $(x_1,\ldots,x_n) \in \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$. An important property of dynamic processes is stationarity, which defines the time-invariance of the joint probabilities extracted from the process: $\Pr\{X_1 = x_1,\ldots,X_n = x_n\} = \Pr\{X_{1+m} = x_1,\ldots,X_{n+m} = x_n\} = p(x_1,\ldots,x_n), \forall n,m \geqslant 1$. Note that all random variables that can be obtained sampling a stationary process take values inside the same set, i.e., $\mathcal{A}_i = \mathcal{A}\ \forall i \geqslant 1$.

In the following, we provide definitions and illustrations of entropy, conditional entropy, and information storage computed for a stationary stochastic process. Note that the present study considers exclusively univariate stochastic processes describing the activity of individual dynamical systems; the reader is referred to the abundant literature in the field [13,75–78] for an extension to multivariate analysis.

1. Entropy

The central concept for the derivation of entropy measures is the definition of the *Shannon information content* of a random variable V [63]: the information contained in a specific outcome v of a random variable V is the quantity $h(v) = -\log p(v)$, where $p(v) = \Pr\{V = v\}$ is the probability that V takes the value v. The units of information depend on the base of the logarithm, being usually bits (base 2) for discrete random variables, and nats for continuous variables where the natural logarithm is used. According to this definition, the information content will be low for highly probable outcomes of the observed random variable and high for unlikely outcomes. Then, if the variable is continuous, the *differential entropy* expresses the amount of information carried by V intended as its average information content:

$$H(V) = -\int_{\mathcal{A}} p(v)\log p(v)dv, \tag{1}$$

where the integral is computed over a continuous range of values $\mathcal{A}$. When the probability $p(v)$ is discrete rather than continuous, the entropy of the variable is defined as

$$H(V) = -\sum_{v\in\mathcal{A}} p(v)\log p(v), \tag{2}$$

where $\mathcal{A}$ is in this case the finite alphabet of values that can be taken by V. Using a notation that subsumes both Eqs. (1) and (2), entropy can be defined as the expected value of the Shannon information content:

$$H(V) = \mathbb{E}[h(v)] = -\mathbb{E}[\log p(v)], \tag{3}$$

where $\mathbb{E}[\cdot]$ is the expectation operator. Entropy quantifies information as the average uncertainty about the outcomes of the variable: if all observations of the variable take the same value, 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 concept of entropy above defined relies on the seminal work of Shannon performed in the field of communication theory [63]. The relevant measure has been extended to the definition of many alternative measures of information such as the Renyi entropy [79] and the Tsallis entropy [80], of which the Shannon entropy constitutes a limiting case that possesses all the desired properties of an information measure. Moreover, there are close parallels between these information-theoretic entropy measures and the fundamental thermodynamic entropy investigations of Boltzmann and Gibbs [81,82].

The entropies defined in Eqs. (1) and (2) are "static" measures, in the sense that they do not take any temporal information into account when describing an observed probability distribution. "dynamic" measures of entropy can be introduced by studying the information content of a stochastic process that represents the activity of a system evolving in time such as conditional entropy and information storage explained below. Specifically, the *entropy* of the process X is defined as the average information contained in its present state:

$$E(X) = H(X_n) = -\mathbb{E}[\log p(x_n)], \tag{4}$$

where x_n is the value taken by the process X at the present time n. Equation (4) presupposes stationarity of the process, so that it carries the same entropy at all times and dependence on time is dropped in the definition of $E(X)$. The past information contained in the system up to time $n-1$ is defined as the joint entropy of the past variables X_n^-:

$$H(X_n^-) = H(X_1,\ldots,X_{n-1}) = -\mathbb{E}[\log p(x_1,\ldots,x_{n-1})]. \tag{5}$$

Likewise, the total information contained the the system up to time n is the joint entropy of the present and past variables, as given by

$$H(X_n^-,X_n) = H(X_1,\ldots,X_n) = -\mathbb{E}[\log p(x_1,\ldots,x_n)]. \tag{6}$$

The simple ideas of separating the present from the past and of incorporating the temporal information into the definition of entropy as done in Eqs. (4)–(6) form the basis of the studies of Kolmogorov [83] and Sinai [84], who first formalized information-theoretic concepts for the analysis of dynamical systems. As further studied by Ebeling [85] and discussed in the next subsections, dynamic entropies are closely related to the notion of predictability defined for a dynamical system evolving in time.

2. Conditional entropy

In general, the present state of the observed process is partially determined by its past history, but also carries a certain amount of new information that cannot be inferred from the past. The average rate of creation of new information is given by the *conditional entropy*, also known as the Kolmogorov-Sinai entropy [86]:

$$\begin{aligned} C(X) &= H(X_n|X_n^-) = H(X_n^-,X_n) - H(X_n^-) \\ &= -\mathbb{E}[\log p(x_n|x_1,\ldots,x_{n-1})], \end{aligned} \tag{7}$$

where $p(x_n|x_1,\ldots,x_{n-1})$ is the conditional probability that X takes the value x_n at time n given that the values taken previously were $x_1,\ldots,x_{n-1}$.

Thus, the conditional entropy quantifies the amount of information contained in the present of the process that cannot be explained by its past history: if the process is fully

random, the system produces information at the maximum rate, yielding maximum conditional entropy; if, on the contrary, the process is fully predictable, the system does not produce new information and the conditional entropy is zero. When the process is stationary, the system produces new information at a constant rate; i.e., the conditional entropy does not change over time.

The notion of conditional entropy subsumes a wide range of entropy measures and estimates that have been proposed in the recent past to quantify the complexity of a time series intended as the degree of predictability of the underlying process. These measures, which include approximate entropy [1], sample entropy [2], fuzzy entropy [4], corrected conditional entropy [3], and permutation entropy [6], are extremely popular for the estimation of conditional entropy in several fields ranging from applied physics to neuroscience, physiology, econometrics, climatology, earth sciences, and others [24,25,29–32,37,38,87,88].

3. Information storage

Another relevant entropy measure is the so-called information storage, which quantifies the amount of information shared between the present and the past observations of the considered stochastic process. For a generic process X, the information storage is defined as

$$S(X) = I(X_n; X_n^-) = \mathbb{E}\left[\log \frac{p(x_1, \ldots, x_n)}{p(x_1, \ldots, x_{n-1})p(x_n)}\right], \quad (8)$$

where $I(X_n; X_n^-)$ denotes the mutual information between X_n and X_n^-.

The information storage reflects the degree to which information is preserved in a time-evolving system [14]. As such, it measures how much of the uncertainty about the present can be resolved by knowing the past: if the process is fully random, the past gives no knowledge about the present, so that the information storage is zero; if, on the contrary, the process is fully predictable, the present can be fully predicted from the past, which results in maximum information storage. If the process X is stationary, the information shared between the present and the past is constant.

Although information storage has been long recognized as an important aspect of the dynamics of many physical and biological processes, it has been formalized only recently as in Eq. (8) as the information contained in the past of a process that can be used to predict its future [39]. This quantitative definition is gaining more and more relevance and has been used with great success to analyze complex dynamics in physiology [26,27,89], neuroscience [90,91], collective behaviors [92], and artificial systems [40,93].

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 into two parts: the new information that cannot be inferred from the past, which is measured by the conditional entropy and the information that can be explained by its past, which is measured by the information storage. Consequently, entropy, conditional entropy, and information storage are related to each other by the equation $S(X) = E(X) - C(X)$.

4. Illustrative example

In this section we demonstrate the properties of the entropy measures defined above using an exemplary stationary binary Markov process of order one as depicted in Fig. 1. The binary process takes values in the alphabet $\mathcal{A} = \{0,1\}$ and is defined in a way such that the two outcomes are equiprobable; i.e., $p(X_n = 0) = p(X_n = 1) = 0.5$. Moreover, according to the Markov property, the present state of the process depends on the past at only one time lag: $p(x_n|x_1, \ldots, x_{n-1}) = p(x_n|x_{n-1})$. We further assume that the conditional probability for the process to take the same value at times $n-1$ and n, $\Pr\{X_n = x|X_{n-1} = x\}$, is inversely modulated by a parameter $\delta \in [0,1]$ in a way such that δ quantifies the strength of the internal dynamics in the system: the higher δ is, the more the present state is dependent on the past states [Fig. 1(a)].

The exact theoretical values of entropy, conditional entropy and information storage computed as a function of δ are reported in the bottom panel of Fig. 1(a), while the remaining panels depict exemplary realizations of the process and values of the entropy measures for the cases of fully random dynamics [$\delta = 0$, Fig. 1(b)], fully predictable dynamics [$\delta = 1$, Fig. 1(c)] and partially predictable dynamics [$\delta = 0.5$, Fig. 1(d)]. As seen in Fig. 1(a), the entropy of the process is constant and equal to 1 bit, because it only depends on the marginal probabilities, which do not change with δ ($p(X_n = 0) = p(X_n = 1) = 0.5$). When δ moves from 0 to 1, the conditional entropy decreases and the information storage increases, reflecting the increasing predictability of the process. The entropy measures the present information contained the dynamic system, $H(X_n) = E(X)$, represented by the solid line with triangle or the red oval. The conditional entropy measures the rate of increase of the total information of the system, which is represented by the slope of the solid line with squares or the part of the red oval not overlapped with the blue. The information storage measures the shared amount of the present information $H(X_n)$ and the past information $H(X_n^-)$, which is represented by the overlap of the red and blue ovals. The fully random dynamical system described in Fig. 1(b) produces new information at the maximum rate, yielding $C(X) = E(X)$ and $S(X) = 0$ (no overlap of the red and blue ovals). The fully predictable system in Fig. 1(c) produces no new information at any time, yielding $C(X) = 0$ and $S(X) = E(X)$ (superimposition of the red and blue ovals). The partially predictable system in Fig. 1(d) produces new information but also maintains past information, yielding $C(X) \in (0, E(X))$ and $S(X) \in (0, E(X))$ (partial overlap of the red and blue ovals).

B. Entropy estimators

In practical analysis, entropy measures are computed from realizations of the observed process that are available in the form of time series data. In general, the estimation of information-theoretic measures from time series is a difficult task. A major issue is the so-called "curse of dimensionality" [94], which refers to the fact that numerical computation is possible only for entropies of finite order. Specifically, when the dimension of the observed variables increases, the conditional entropy estimated from time series of finite length decays towards zero [3]. Therefore, in the practice of short time

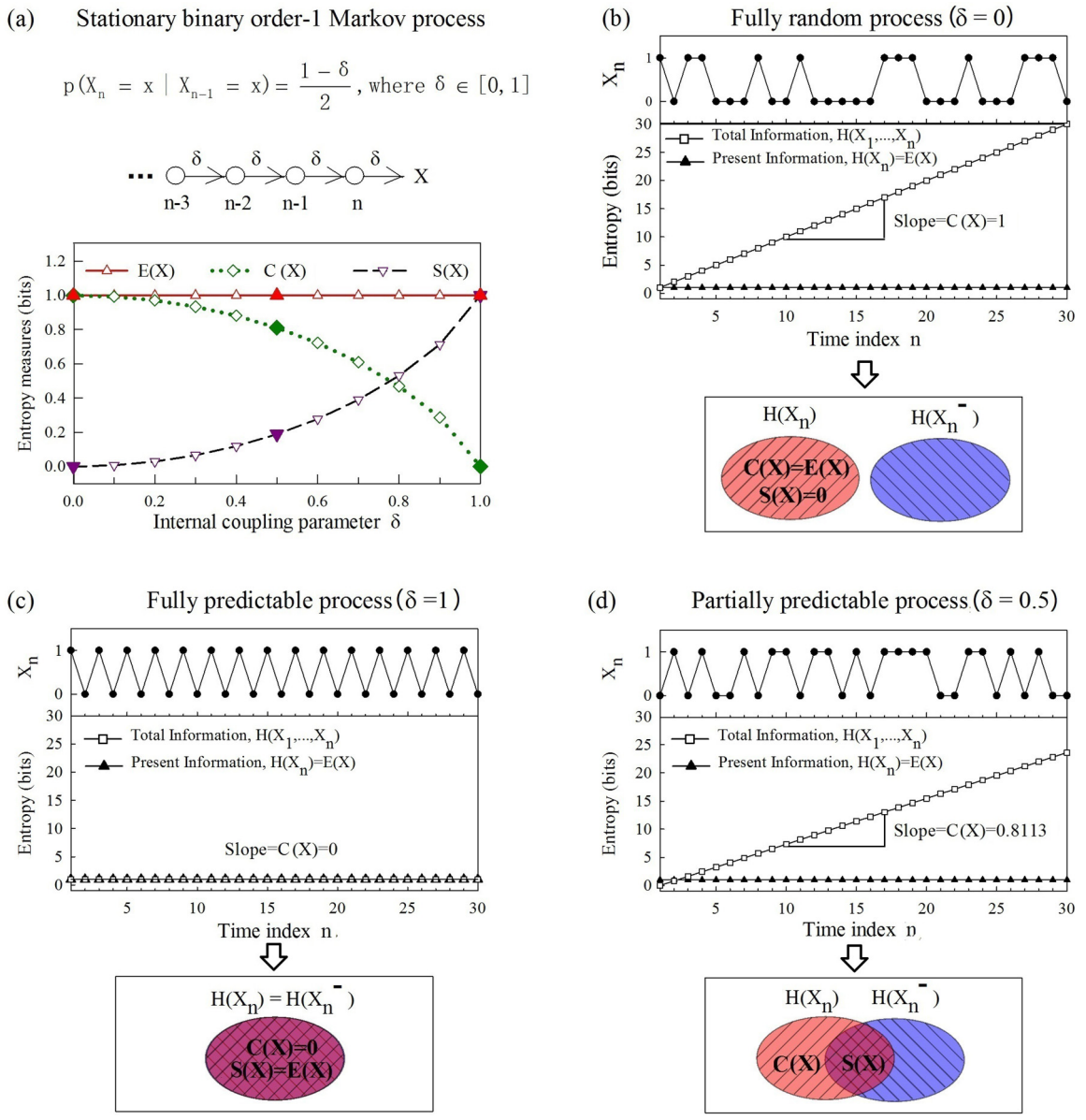

FIG. 1. Computation of entropy measures for a stationary binary order-1 Markov process. Since this process takes discrete values, entropies are computed using the base 2 logarithm and measured in bits. (a) Dependence of the transition probabilities and the entropy measures on the internal coupling parameter δ. When δ rises from 0 to 1, the probability that a state transition keeps the process in the same state ($p(X_n = x|X_{n-1} = x)$) moves from 1/2 to 0; accordingly, conditional entropy decreases from 1 to 0 (green dotted line with diamond) and information storage increases from 0 to 1 (purple dashed line with down-triangle); note that, since for this process the marginal probabilities are unaffected by δ, the entropy of the process does not change (red solid line with up-triangle). The values of entropy measures for $\delta = 0,0.5,1$ are marked with full symbols. (b), (c), (d) Entropy analysis for representative parameter values, showing a realization of the process (solid line with full circles), the total and present system information as a function of time, and the Venn diagram of the entropy measures (present information, red oval; past information, blue oval). For this stationary process, the present information is the same at all times and measures the entropy of the process (E), while the total information increases at a constant rate measured by the conditional entropy (C); the information storage (S) is the information shared between the present and the past (overlap of ovals), while the C is the part of the present information not shared with the past. When the process is fully random ($\delta = 0$), the total information increases at the maximum rate ($C = 1$) and there is no stored information ($S = 0$). On the contrary, a fully predictable process ($\delta = 1$) does not produce new information ($C = 0$) and stores the whole information contained in its present state ($S = 1$). Any intermediate parameter configuration ($0 < \delta < 1$) yields a partially predictable process with presence of both information production and information storage ($0 < S < 1, 0 < C < 1$).

series analysis, conditional entropy and information storage are estimated using a finite number of samples in the past, i.e., X_n^- is approximated by $X_n^m = [X_{n-1}, X_{n-2}, \ldots, X_{n-m}]$ when computing $H(X_n|X_n^-)$ and $I(X_n; X_n^-)$. While optimization techniques such as graphical models [94] or nonuniform embedding [89,95] exist to limit the detrimental effects of the curse of dimensionality, yet in this study we stick to the uniform embedding scheme, which selects m consecutive

past samples, so as to compare the performances of different estimators under the "standard" conditions that are commonly studied in the existing literature.

Various entropy estimators that follow different approaches to compute the probability distribution are available in the literature [75]. The estimators can be categorized into two groups: model-based estimators and model-free estimators. If the probability distribution of the data can be faithfully represented by a known parametric distribution (e.g., Gaussian), entropy measures can be computed using model-based estimators as functions of the parameters of the presumed probability distribution [13,75,96]. On the other hand, when no assumptions can be made about the data distribution, model-free approaches which approximate the probability distribution directly from the data should be followed. The most intuitive method is to build the histogram distribution of the quantized time series amplitudes. However, this method is proved to have serious bias problems and its estimates are strongly dependent on the size of the quantization levels [97,98]. This situation can be improved to some extent by using binless density estimators such as the kernels estimator [2,30,99] or the nearest neighbor estimator [100,101]. In this paper, we consider the linear model-based estimation method and the two model-free methods which employ kernel and nearest-neighbor entropy estimators. Details of these three estimators are presented in the following.

1. Linear estimator

The linear estimator is a model-based approach for the estimation of entropy measures. It adopts the assumption of a joint Gaussian distribution for the observed variables and exploits the exact expressions that hold in this case for the entropy measures. Specifically, the assumed Gaussian probability distribution is given by

$$p(x_n) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{\frac{-x_n^2}{\sigma_X^2}}, \tag{9}$$

where σ_X^2 is the variance of X_n. Then, by plugging Eq. (9) into Eq. (3), the entropy of the present state of the observed process is obtained as

$$E(X) = H(X_n) = \tfrac{1}{2}\ln 2\pi e\sigma_X^2. \tag{10}$$

Note that the entropy of a stationary Gaussian process is a function of its variance only.

Moreover, the linear method estimates the conditional entropy from the variance of the prediction error of the linear regression of the present of the process on its past [96]. Specifically, the linear regression of the present X_n on the past $X_n^m = [X_{n-1}, \ldots, X_{n-m}]$ is performed as

$$X_n = \sum_{i=1}^{m} a_i X_{n-i} + U_n, \tag{11}$$

where m is the order of the regression, $a_i, i = 1, \ldots, m$, are the regression coefficients, and U_n is a zero-mean white Gaussian noise. A paradigmatic example for the linear regression of X_n on X_{n-1} is given in Fig. 2. Then, the linear estimate of the conditional entropy is obtained from the variance of U_n, σ_U^2, as follows:

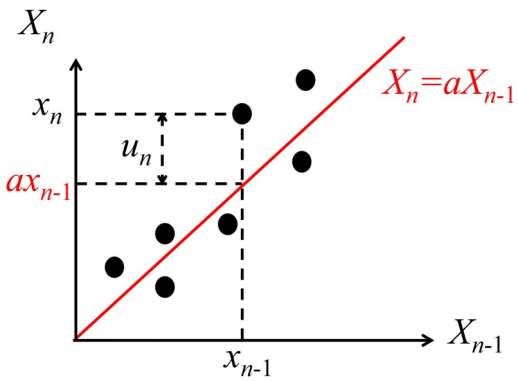

FIG. 2. Schematic illustration of the linear estimation of entropy measures. In this paradigmatic example in which a time series $\{x_1, \ldots, x_8\}$ is considered as a realization of the process X, and the past of the process is approximated with $m = 1$ lags ($X_n^- \approx X_n^m = X_{n-1}$), seven realizations of (X_n, X_{n-1}) are found and used to fill a two-dimensional space. Then, the linear regression of X_n on X_{n-1} is performed yielding the regression line $X_n = aX_{n-1}$ (red line) and, for a given observation x_n of X_n, the prediction error u_n is taken as the difference between the true and the predicted value, $u_n = x_n - ax_{n-1}$. The estimates $\{u_2, \ldots, u_8\}$ are finally used to compute the variance of U_n and the conditional entropy according to Eq. (12).

$$C(X) = H\left(X_n|X_n^m\right) = \tfrac{1}{2}\ln 2\pi e\sigma_U^2. \tag{12}$$

Subtracting Eq. (12) from Eq. (10), the estimation of information storage is obtained as

$$S(X) = I\left(X_n; X_n^m\right) = \frac{1}{2}\ln\frac{\sigma_X^2}{\sigma_U^2}. \tag{13}$$

Hence, under the assumption of Gaussianity, the information storage is determined by the ratio of the variance of the present state of the process to the partial variance of the present given the past.

The formulations above exploit a central result relating the conditional entropy to the prediction error of a multivariate regression [96], which is here adapted to univariate regression and extended to the computation of information storage as proposed recently in both theoretical and empirical studies [13,25,25]. 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 system, it may be extended in a straightforward way to nonlinear representations when non-Gaussian parametric distributions are assigned [75].

2. Kernel estimator

The Kernel entropy estimator is a model-free approach which reconstructs the probability distribution of an observed variable by centering kernel functions at each outcome of the variable and then exploits the estimated probabilities to derive the relevant entropy measures. Kernels are used to weight the distance of each point in the time series to the reference point depending on the kernel function. For instance, the entropy of the present state of the process X is estimated, starting from a realization of length N available in the form of the time

series $\{x_1, x_2, \ldots, x_N\}$, first computing the kernel estimate of the probability distribution:

$$p(x_n) = \frac{1}{N}\sum_{i=1}^{N} K(\|x_n - x_i\|), \tag{14}$$

where K is the kernel function and $\|\cdot\|$ is an appropriate norm, and then plugging Eq. (14) into Eq. (4):

$$E(X) = H(X_n) = -\ln\langle p(x_n)\rangle, \tag{15}$$

where $\langle\cdot\rangle$ denotes the average taken over all possible x_n. Similarly, Eq. (14) can be used to estimate the joint probability distributions $p(x_n^m) = p(x_{n-1}, \ldots, x_{n-m})$ and $p(x_n, x_n^m)$ in the m-dimensional and $(m+1)$-dimensional spaces spanned by the realizations of X_n^m and (X_n, X_n^m), from which the conditional entropy is obtained as

$$C(X) = H\left(X_n|X_n^m\right) = -\ln\frac{\left\langle p\left(x_n, x_n^m\right)\right\rangle}{\left\langle p\left(x_n^m\right)\right\rangle}. \tag{16}$$

Given the expressions of Eqs. (15) and (16) for the kernel estimates of entropy and conditional entropy, the kernel estimate of the information storage is then obtained as follows:

$$S(X) = I\left(X_n; X_n^m\right) = \ln\frac{\left\langle p\left(x_n, x_n^m\right)\right\rangle}{\left\langle p(x_n)p\left(x_n^m\right)\right\rangle}. \tag{17}$$

The most common metric to compute distances using the kernel estimator, which is adopted also in this study, is the so-called Chebyshev distance or maximum norm, which is obtained as the maximum of the absolute differences between scalar components, i.e., $\|x_n - x_i\| = |x_n - x_i|$ and $\|x_n^m - x_i^m\| = \max_{1\leqslant k\leqslant m} |x_{n-k} - x_{i-k}|$. As to the kernel function, the most popular is the Heaviside kernel, which sets a threshold r to weight the distance of each point to the reference point. Its expression is

$$K = \Theta(\|x_n - x_i\|) = \begin{cases} 1, & \|x_n - x_i\| \leqslant r \\ 0, & \|x_n - x_i\| > r \end{cases}. \tag{18}$$

Substituting Eq. (18) into Eq. (14), one can see that the Heaviside Kernel estimator approximates the probability density at the reference point x_n with the fraction of time series points that falls within the distance r from x_n. The threshold r is the width of the Heaviside kernel function, which controls the precision of the density estimation: smaller values of r give more detailed estimates yet requiring more data points to be accurate, while too large values of r yield very coarse probability estimates because too many points are included in the neighborhood of the reference point. In practical computation, the threshold r is usually set to be a fraction of the data variance so as to remove the dependence of entropy measures on the amplitude of the observed process [1,2]. An illustrative example is depicted in Fig. 3 for the estimation of the probabilities $p(x_n, x_{n-1})$ and $p(x_{n-1})$ in a paradigmatic case ($m = 1$).

The kernel estimation for conditional entropy defined in Eq. (16), when implemented applying the heaviside kernel function and using the maximum norm to compute distances, is equivalent to the sample entropy (SampEn) [2], a well-known measure of dynamical complexity proposed to reduce the bias

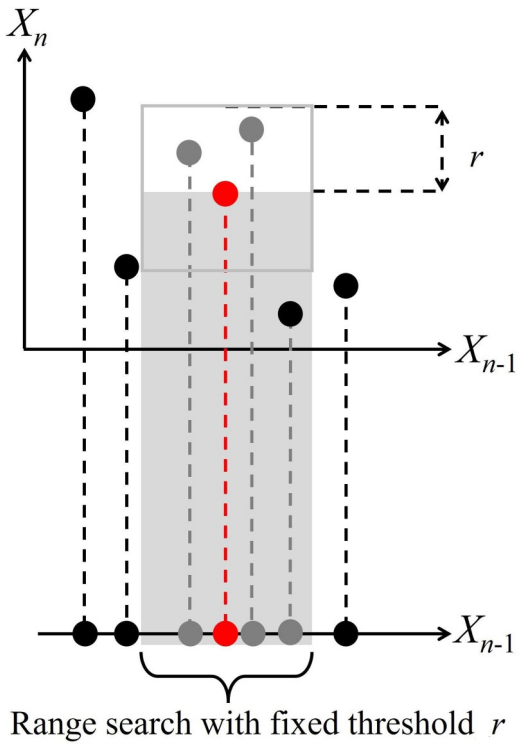

FIG. 3. Schematic illustration of the kernel estimation of entropy measures. In this example, the past of the process X is approximated using $m = 1$ samples, and the Heaviside kernel with fixed threshold r is used. In the $(m+1)$-dimensional space spanned by the realizations of (X_n, X_{n-1}), the probability of a given reference point (x_n, x_{n-1}) (red dot) is estimated as the fraction of points whose distance are less than r (gray dots) from it: $p(x_n, x_{n-1}) = 2/7$. The distance between two points is computed as the maximum between the horizontal and the vertical distance between the two points. Similarly, in the m-dimensional space spanned by the realizations of X_{n-1}, the probability of x_{n-1} is approximated using the same threshold r, yielding $p(x_{n-1}) = 3/7$. This procedure is repeated varying the reference point and the conditional entropy is estimated by Eq. (16).

of the first introduced kernel-based measure of conditional entropy, i.e., the approximate entropy (ApEn) [1]. These measures, and more generally all kernel-based estimators of conditional entropy and information storage, are ubiquitously used to assess the dynamical complexity of time series in several fields ranging from physics to engineering, biology, and medicine [25,29–31,33,37,102,103]. Therefore, it is of utmost importance to investigate how these estimates behave in the conditions typical of real-world time series analysis, as well as to understand their range of applicability and limitations.

3. *Nearest-neighbor estimator*

The k-nearest-neighbor estimator (knn) is another model-free approach that approximates the probability distribution from multiple observed realizations of the considered variable, and then plugs this probability into the entropy definition to yield the entropy estimate. The knn estimator approximates the probability distribution from the statistics of the distances between neighboring points in the multidimensional spaces spanned by the observed variables [100]. Compared to the

kernel estimator, which fixes the neighborhood size for the reference point according to a given threshold distance, the knn 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 kth nearest neighbor. Specifically, the method builds on the central results, published in Refs. [100,101], stating that the average Shannon information content of a generic d-dimensional random variable V can be estimated from a set of realizations $\{v_1, v_2, \ldots, v_N\}$ of the variable as

$$-\mathbb{E}[\ln p(v_n)] = \psi(N) - \psi(k) + d\mathbb{E}[\ln \varepsilon_n], \tag{19}$$

where ψ is the digamma function and ε_n is twice the distance between the outcome v_n and its kth nearest neighbor computed according to the maximum norm (i.e., taking the maximum distance of the scalar components).

Exploiting Eq. (19), one can easily derive the expression for the knn estimate of the entropy of the present state of the process X computed for the time series $\{x_1, x_2, \ldots, x_N\}$:

$$E(X) = H(X_n) = \psi(N) - \psi(k) + \langle \ln \varepsilon_n \rangle. \tag{20}$$

Then, according to Eq. (7), the conditional entropy can be computed as the difference between the joint entropy of the present and the past, $H(X_n^m, X_n)$, and that of only the past of the process, $H(X_n^m)$. The information storage can be computed as the difference between entropy and conditional entropy. However, since $H(X_n)$, $H(X_n^m)$, $H(X_n^m, X_n)$ are computed in spaces with different dimensions (respectively, 1, m and $m+1$), 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, which would introduce different estimation biases that cannot be compensated by taking the entropy differences. Therefore, in order to keep the same distance length in all explored spaces, we adopt the solution [101] of performing a neighbor search only in the highest-dimensional space and projecting the distances found in this space to the lower-dimensional spaces, keeping these distances as the range within which neighbors are counted. An example is depicted Fig. 4 for the paradigmatic case of $m = 1$. Specifically, the knn estimate of $H(X_n^m, X_n)$ is computed through the neighbor search:

$$H\left(X_n, X_n^m\right) = \psi(N) - \psi(k) + (m+1)\langle \ln \varepsilon_n \rangle, \tag{21}$$

where ε_n is twice the distance from (x_n, x_n^m) to its kth nearest neighbor, and then, given the distances ε_n, the entropies in the lower-dimensional spaces are estimated through a range search:

$$H(X_n^m) = \psi(N) - \left\langle \psi\left(N_{X_n^m}\right) \right\rangle + m\langle \ln \varepsilon_n \rangle, \tag{22}$$

$$H(X_n) = \psi(N) - \left\langle \psi\left(N_{X_n}\right) \right\rangle + \langle \ln \varepsilon_n \rangle, \tag{23}$$

where N_{X_n} and $N_{X_n^m}$ are the number of points whose distance from X_n and X_n^m, respectively, is smaller than $\varepsilon_n/2$. Finally, the conditional entropy is obtained by subtracting

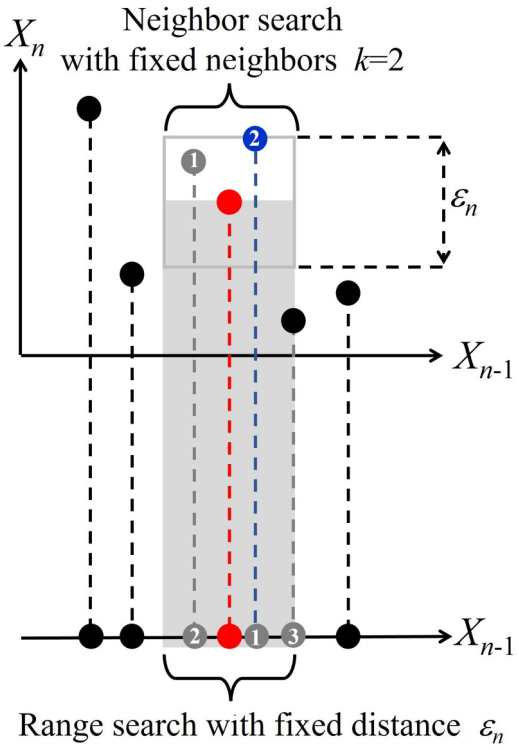

FIG. 4. Schematic illustration of the knn estimation of entropy measures. In this example, the past of the process X is approximated with $m = 1$ samples and $k = 2$ neighbors are used in the search for neighbors. In the $(m+1)$-dimensional space spanned by the realizations of (X_n, X_{n-1}), a neighbor search is performed using the maximum norm to find the kth nearest neighbor (blue dot) of the assigned reference point (red dot). Then, the distance between these two points, $0.5\varepsilon_n$ is used in the projected one-dimensional space spanned by the realizations of of $X_n^1 = X_{n-1}$ as threshold distance to find the neighbors of the reference point x_{n-1}; in this example, $N_{X_n^1} = 3$ neighbors are counted. This procedure is repeated varying the reference point and the obtained values of ε_n and $N_{X_n^1}$ are then used in Eq. (24) to compute the conditional entropy.

Eq. (22) from Eq. (21):

$$C(X) = H\left(X_n | X_n^m\right) = -\psi(k) + \left\langle \psi\left(N_{X_n^m}\right) \right\rangle + \langle \ln \varepsilon_n \rangle, \tag{24}$$

and the information storage is obtained subtracting Eq. (21) from the sum of Eqs. (22) and (23) [26]:

$$\begin{aligned} S(X) &= I\left(X_n; X_n^m\right) \\ &= \psi(N) + \psi(k) - \left\langle \psi\left(N_{X_n^m}\right) \right\rangle - \left\langle \psi\left(N_{X_n}\right) \right\rangle. \end{aligned} \tag{25}$$

Since the nearest-neighbor technique results in an adaptive resolution as it changes the distance scale according to the underlying probability distribution [97,99], and may also achieve bias compensation when implemented through distance projection [101], this approach has gained in recent years increasing popularity for the estimation of entropy measures in time series analysis. While the utilization of this estimator has been directed up to now mostly to the computation of entropy measures for multivariate time series where the issue of dimensionality is more serious [14,25,104–106], in this study we consider its implementation for the computation of

entropy measures for individual time series, as first proposed in [106].

C. Simulation model of stochastic processes

In this section, we introduce the models to simulate four different types of stochastic processes: stationary autoregressive process, autoregressive process with nonstationarities, fractionally correlated process, and fractionally integrated autoregressive process.

We start with the stationary autoregressive process (AR process), which constitutes the basic process on which entropy measures can be applied. For this type of process, techniques to compute the exact theoretical values of the various entropy measures are available [13] and are here reviewed in the Appendix. We use these theoretical values as a reference to evaluate the performance of different entropy estimators. Results of this basic process will serve as a baseline for more complicated processes that are studied later.

Second, we superimpose three types of nonstationarities (i.e., sinusoidal trends, spikes, and local variance changes) on the stationary AR signal. These nonstationarities are selected as they are commonly encountered in real-world time series as factors corrupting the underlying dynamics [57,58,107–110]. In our simulations, by comparing the estimated values of entropy measures for AR signals with nonstationarity and their corresponding theoretical values for original stationary AR signals, we aim to understand the effects of nonstationarity on entropy estimation and figure out potential solutions to mitigate or remove consequent biases.

In addition to the autoregressive process, we also investigate processes with long-range power-law correlations, a property exhibited by many empirical time series such as the human heart rate or the price index of the stock market [58,109]. Unlike the autoregressive process which is considered to be short memory, these processes, usually referred to as fractionally integrated processes, often exhibit long-range or medium-range dependence [111]. In other words, a fractionally integrated process has an autocorrelation function that damps hyperbolically, more slowly than the geometric damping of an autoregressive process. Despite the fact that entropy measures are typically applied to time series with long-range power-law correlation [2,3,24,27,30,30,31,37,89,106], it is not well understood how these measures relate to this type of long-memory dynamics and how their estimation is affected by properties of correlations including its sign and strength. To fill in this knowledge gap, we first extend the approach used in Ref. [13] to compute the theoretical values of entropy measures from given simulation parameters for fractionally integrated processes, as shown in the Appendix. In addition, we compare these theoretical values with the estimated values of different entropy measures and estimators to evaluate their estimation bias. In this way, our work provides a reference for the application of entropy measures and estimators to power-law long-range correlated processes.

Last, we consider more general cases of processes with both autoregression and power-law long-range correlations and follow the same procedures to evaluate the performance of entropy measures and estimators by computing and comparing their estimates with the corresponding theoretical values.

1. Stationary AR process

The AR process is simulated as the output process of a linear univariate AR model driven by a stochastic uncorrelated noise. Using the polynominal notation, an autoregressive process of order p can be expressed as

$$A(L)X_n = U_n, \tag{26}$$

where $A(L) = 1 - \sum_{i=1}^{m} A_i L^i$ is an autoregressive polynomial of order m, L is the backward shift operator ($L^i X_n = X_{n-i}$), and U is a white Gaussian innovation process with zero mean and unit variance.

In this study we simulate an AR process of order $m = 2$ by placing two complex-conjugate poles [roots of $A(L)$] in the complex plane, with modulus ρ and phase $\pm 2\pi f$, in a way such that the coefficients of the AR polynomial become

$$A_1 = 2\rho\cos(2\pi f), \quad A_2 = -\rho^2. \tag{27}$$

With this setting, the parameters ρ and f determine, respectively, the amplitude and frequency of a stochastic oscillation that is imposed for the process. Note that the process is stationary when $\rho \in [0,1)$ and that the AR amplitude ρ determines the regularity of the stochastic oscillation: the process is a fully unpredictable white noise when $\rho = 0$, and becomes a highly predictable stochastic process exhibiting a marked oscillatory behavior around the frequency f when ρ approaches 1.

Stationary realizations of the AR process described above, generated with different values set for the AR amplitude ρ and frequency f, are given in Fig. 5. Comparing Figs. 5(a)–5(c) by column one can see that, for an assigned frequency f, the process is more regular for higher values of ρ, confirming the expected increase in the predictability of the process with the AR amplitude. On the other hand, variations in the predictability of the process are more difficult to appreciate when f is varied by keeping fixed the AR amplitude ρ. To investigate this dependence in more detail, Figs. 5(d) and 5(e) report, respectively, the autocorrelation function of the process, and the two-dimensional (2D) and three-dimensional (3D) phase plots of the temporal relation between the present and the two past samples (X_n versus X_{n-1}, X_{n-2}), computed for the realizations of Fig. 5(c). These plots indicate that the process exhibits longer memory, as well as a much stronger linear dependence of the present on the past values, when the AR frequency is very low ($f = 0.01$) or very high ($f = 0.49$) compared to the intermediate value ($f = 0.25$). This suggests that, besides the pole modulus ρ, also the frequency f of the stochastic oscillation of an AR process plays a role in determining its degree of regularity.

2. AR process with nonstationarity

Stationarity is a prerequisite for the computation of the entropy measures from an individual realization of the process under investigation. In fact, if the process is nonstationary, the joint probability distribution of its present and past values changes over time, which precludes the possibility of pooling observations across time for estimating of such probabilities. If observations are pooled across time in the presence of nonstationarities, the estimated probability distribution is unreliable and the resulting entropy measures deviate from

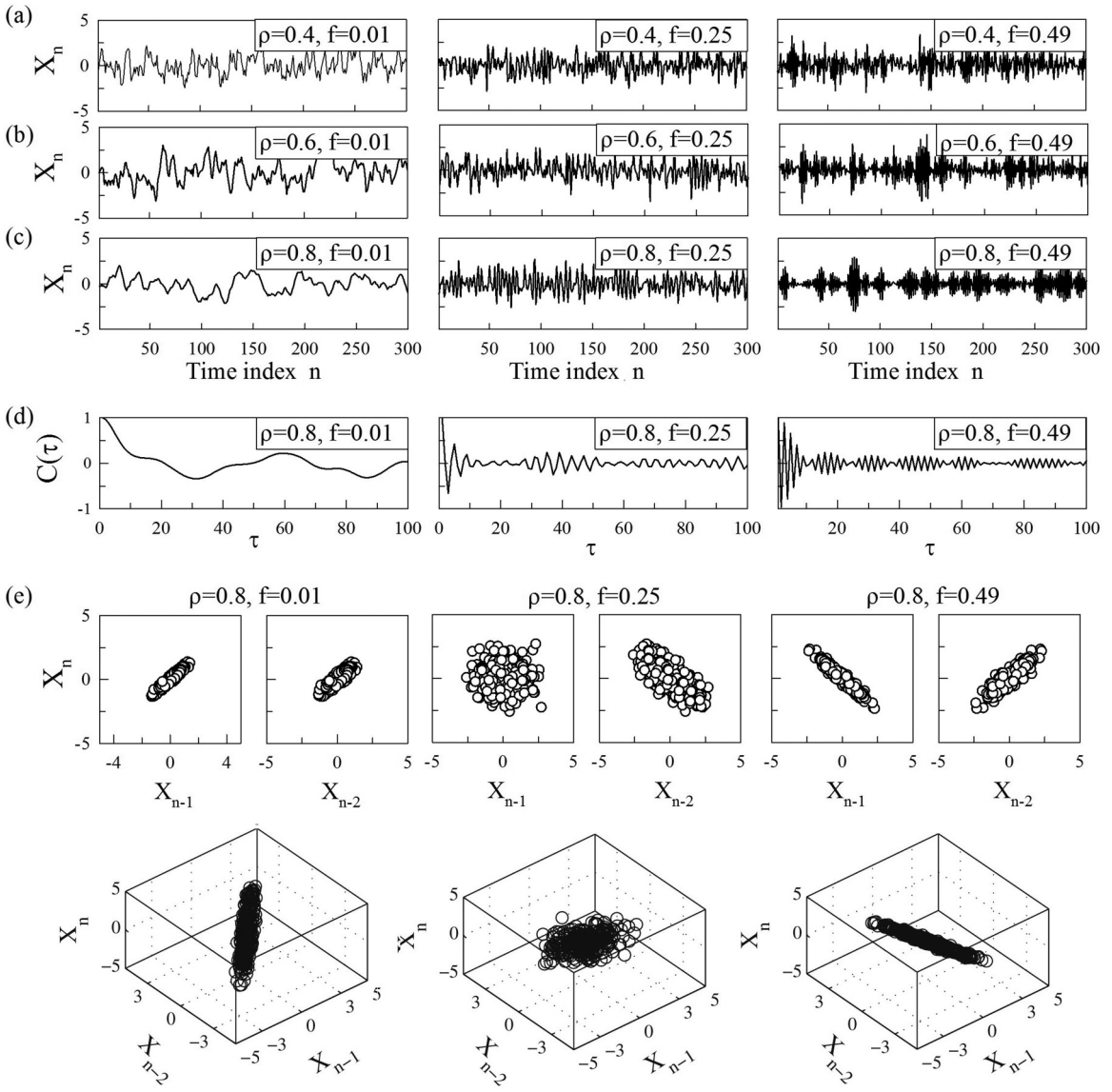

FIG. 5. Autoregressive processes: characterization of a stationary order-2 AR process for different values of the AR amplitude ρ and frequency f. (a)–(c) Exemplary realizations of the process obtained varying ρ (columns) and f (rows). Results: When f is fixed, the process appears more regular for higher values of ρ. (d) Autocorrelation of the process as a function of the lag τ for the realizations in (c). (e) 2D phase plots of (X_n, X_{n-1}) and (X_n, X_{n-2}), and 3D phase plots of (X_n, X_{n-1}, X_{n-2}), for the realizations in (c). When ρ is fixed, the process exhibits shorter memory and weaker dependence of the present on the past for intermediate frequency f.

the value assumed for a stationary distribution to an extent depending on the type and strength of the nonstationary behavior.

Here we study the effects of three types of nonstationarities due to common artifacts, including trends, spikes, and local changes in the signal variance, on the entropy, conditional entropy, and information storage of the AR process of order 2 described above. To reproduce these situations, we superimpose the chosen type of nonstationary behavior to stationary realizations of the AR process generated according to Eqs. (26) and (27). Nonstationary AR signals with sinusoidal trends are obtained by adding to the original stationary AR signals a sine wave of period T and amplitude A. Signals with random spikes with amplitude A and percentage $P\%$ are generated by replacing random points of the original time series with random numbers uniformly distributed in the interval$(-A\sigma_X^2, A\sigma_X^2)$, where σ_X^2 is the variance of the original signal. To simulate local changes in variance, we choose random segments from the original time series and inflate these segments by multiplying their original values by a factor of σ. Each inflated segment contains 20 points and the total number of inflated points covers $P\%$ of the original signal length. The resulting realizations of the AR process with superimposed nonstationary behavior were always normalized to zero mean and unit variance before computing the entropy measures. Exemplary realizations of the analyzed nonstationary AR processes are depicted in Figs. 11, 13, and 15.

3. *Fractionally integrated white noise process*

Stochastic processes with power-law long-range correlations are generated as fractionally integrated white noise [112],

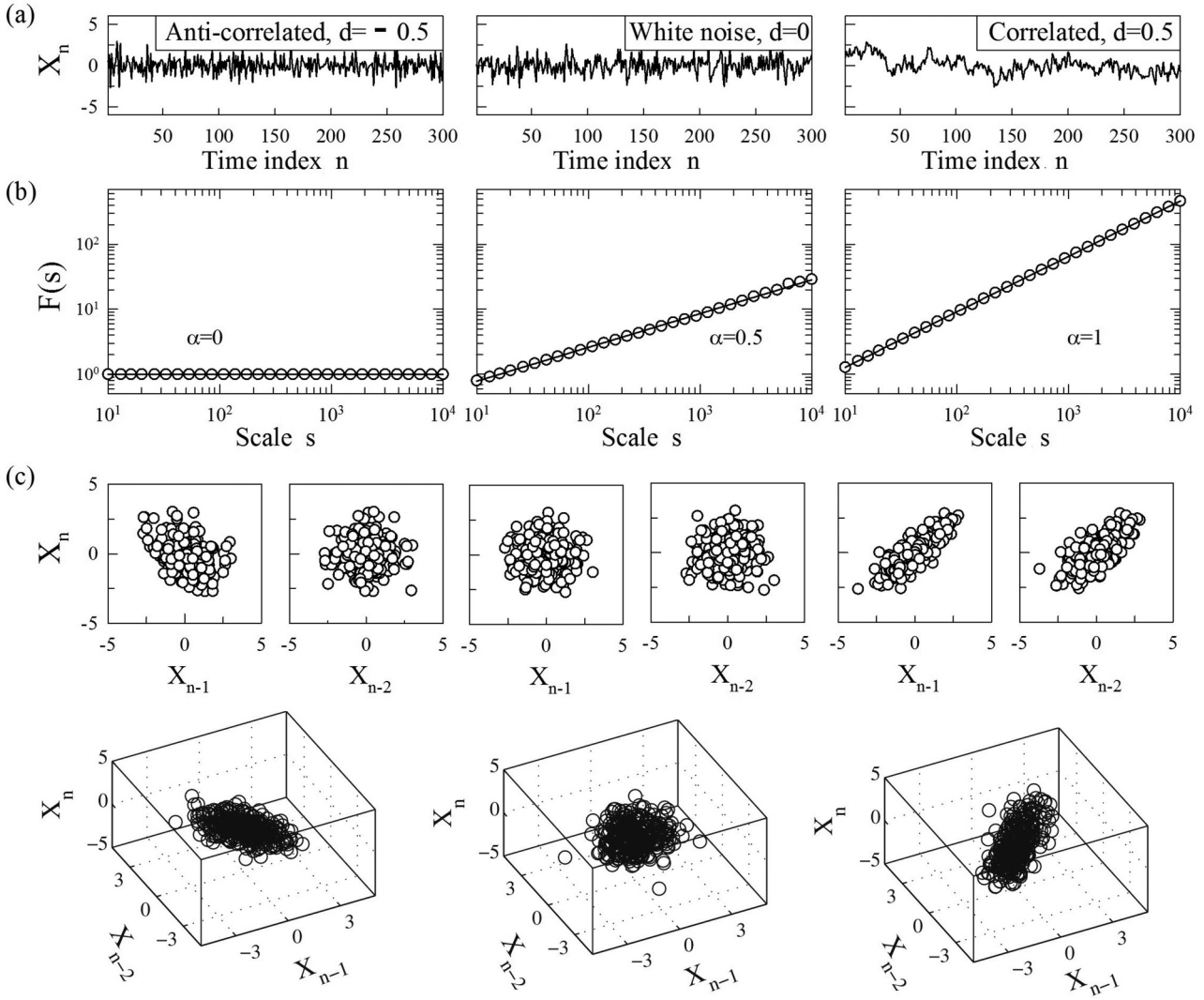

FIG. 6. Processes with long-range correlations: characterization of a fractionally integrated process with long-range power-law correlations for different values of the differencing parameter d, which controls the correlations. (a) Exemplary realizations of the process with $d = -0.5,0,0.5$. (b) Results of detrended fluctuation analysis (DFA) applied to longer realizations (2^{20} data points) of the process in (a). For $d \in [-0.5,0.5]$, the DFA exponent is $\alpha = (2d+1)/2$. (c) 2D phase plots of (X_n, X_{n-1}) and (X_n, X_{n-2}), and 3D phase plots of (X_n, X_{n-1}, X_{n-2}), for the realizations in (a). Results: For fixed modulus of the differencing parameter, positively correlated processes exhibits stronger dependence of the present on the past than anti-correlated processes.

defined by

$$(1-L)^d X_n = U_n, \tag{28}$$

where U is a Gaussian white noise with zero mean and unit variance, $d \geqslant 0$ is the so-called differencing parameter and $(1-L)^d$ is the fractional differencing operator defined by

$$(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)L^k}{\Gamma(-d)\Gamma(k+1)}, \tag{29}$$

with $\Gamma(\cdot)$ denoting the gamma (generalized factorial) function. In this study, computation of Eq. (29) is approximated by $(1-L)^d = \sum_{k=0}^{100} \frac{\Gamma(k-d)L^k}{\Gamma(-d)\Gamma(k+1)}$. The differencing parameter d controls the sign and degree of the correlations imposed in the process. It is related to the Hurst exponent, α, by the relation $\alpha = (2d+1)/2, d \in [-0.5,0.5]$ [113]. Within this range of values for d, the fractionally integrated process is considered as stationary [112]. For $d \in (0,0.5]$, the process is long-range correlated, showing long-range positive dependence, while for $d \in [-0.5,0)$, it is anti-correlated, showing long-range negative dependence. The case $d = 0$ reduces to uncorrelated white noise.

Figure 6(a) shows exemplary realizations of fractionally integrated white noise with differencing parameter set to $d = -0.5, d = 0$, and $d = 0.5$. The corresponding multifractal behavior obtained through detrended fluctuation analysis is depicted Fig. 6(b), confirming the relation between the fractional differencing parameter d and the Hurst exponent α. Figure 6(c) depicts the two-dimensional phase plots of (X_n, X_{n-1}) and (X_n, X_{n-2}) and three-dimensional phase plots of (X_n, X_{n-1}, X_{n-2}) for anticorrelated, uncorrelated, and positively correlated time series. The plots evidence a cloud distribution of the points reflecting the absence of a dependence of the present on the past for the uncorrelated

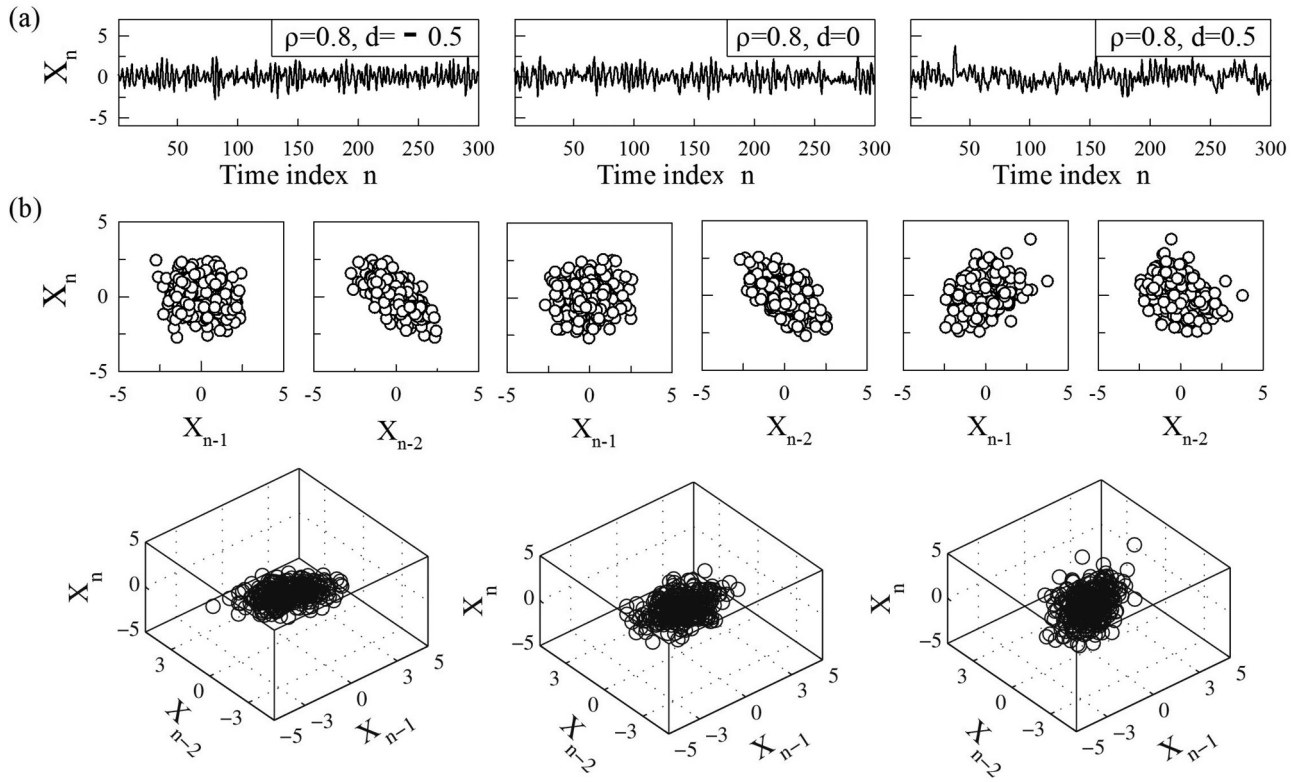

FIG. 7. Autoregressive processes with long-range correlations: characterization of a stationary fractionally integrated AR process with both autoregression and power-law long-range correlations for different values of the differencing (correlation) parameter d and fixed values of the AR parameters ρ and f. (a) Exemplary realizations of the process with $d = -0.5,0,0.5$ and fixed $\rho = 0.8, f = 0.25$. (b) 2D phase plots of (X_n, X_{n-1}) and (X_n, X_{n-2}), and 3D phase plots of (X_n, X_{n-1}, X_{n-2}) for the realizations in (a). Results: The fractionally integrated AR process displays weaker dependence of the present on the past than the corresponding pure autoregressive process with the same AR parameters in the presence of positive long-range correlations and stronger dependence in the presence of negative long-range correlations.

case (middle); moreover, when the degree of correlation is the same, a much stronger dependence of the present on the past is exhibited for a process with positive correlation (right) than for a process with anticorrelation (left).

4. Fractionally integrated AR process

The combination of the autoregressive process and the fractionally integrated processes defined in Eqs. (26) and (28) results in a more general univariate process exhibiting both stochastic oscillations and long memory. The resulting process, which belongs to the class of fractionally integrated autoregressive moving average processes (ARFIMA) [112], is defined as follows:

$$A(L)(1-L)^d X_n = U_n. \tag{30}$$

Figure 7 shows exemplary realizations and phase plots of this fractionally integrated autoregressive process with fixed AR amplitude $\rho = 0.8$ and varying differencing parameter $d = -0.5,0,0.5$. Compared with the case in which the process is not long-range correlated but purely autoregressive [Fig. 7(a), middle, $\rho = 0.8, d = 0$], the combination of the AR stochastic oscillations with positive long-range correlations [Fig. 7(a), right, $\rho = 0.8$, $d = 0.5$] seems to slightly reduce the dependence of the present of the process on its past, while the opposite seems to occur when AR stochastic oscillations are combined with negative long-range correlations [Fig. 7(a), left, $\rho = 0.8$, $d = -0.5$]. The same effect, i.e., a decrease of the predictability of the present given the past for positive long-range correlations and an increase of this predictability for negative long-range correlations, is observed comparing the case of mixed AR and fractionally integrated processes [Fig. 7(a)] with the pure fractionally integrated process [Fig. 6(a)]. Thus, a process with both AR short-term dependencies and long-range correlations results less predictable than its pure autoregressive or pure fractionally integrated counterparts in the case of positive long-range correlations and more predictable in the case of anticorrelations.

III. RESULTS

This section provides the results for the application of the three entropy measures defined in Sec. II A (i.e., entropy, conditional entropy, and information storage) computed using the three entropy estimators presented in Sec. II B (i.e., linear, kernel, and knn) on the four types of stochastic processes discussed in Sec. II C (i.e., stationary AR processes, AR processes with different types of nonstationarity, power-law long-range correlated processes, and process with both AR structure and long-range correlation). For each type of process, we first theoretically obtain the true values of all three entropy measures through analytical derivations starting from the assigned model parameters. Then, using the same simulation

model and model parameters, we compute the estimated values of entropy measures using all three estimators for 100 realizations of the target process. Each realization typically lasts 300 points. All the analyzed processes have zero mean and are reduced to unit variance prior to the computation of entropy measures. All entropy estimations are performed using $m = 2$ lagged components to approximate the past of the process (i.e., $X_n^m = [X_{n-1}, X_{n-2}]$); this setting corresponds to choosing the true order of the simulated AR process, so as to make the interpretation of results free from issues related to an inappropriate selection of the embedding dimension.

A. Performance of entropy estimators and entropy measures for stationary AR processes

Figure 8 reports the characterization of entropy measures and entropy estimators for the case of a stationary AR process. For this process, the exact behavior of the entropy measures in response to changes in the analysis parameters can be studied by looking at the theoretical values (black solid lines), and can be compared with the distributions of values obtained applying the different estimators to multiple realizations of the process generated from setting specific values for the parameters (symbols and error bars).

The theoretical values of all entropy measures are obviously the same for different lengths of the generated realizations [Figs. 8(d), 8(e), and 8(f)]. Moreover, since this example deals with normalized Gaussian processes with zero mean and unit variance, the entropy of the process is constant at varying the AR parameters ρ and f [Figs. 8(g) and 8(j)]. On the other hand, when f is fixed and ρ increases, the theoretical value of conditional entropy decreases and that of information storage increases. When ρ is fixed and f increases, the theoretical value of conditional entropy increases for $f \in (0,0.25]$ and decreases for $f \in (0.25,0.5)$. The theoretical behavior of information storage is the opposite. The dependence of the measures of dynamical complexity on the AR amplitude is expected: a process with higher ρ exhibits a stronger dependence of the present on the past, and this better predictability is reflected by lower conditional entropy and higher information storage. On the other hand, the dependence of the entropy measures on the AR frequency, documented in Figs. 8(k) and 8(l) and in more cases in Fig. 10, is a less expected behavior, which indicates the existence of a complex relation between the statistical structure of a dynamic process and its information content.

Turning to the analysis of the entropy estimates first, we see that, as one may expect, the estimated values exhibit lower variability while increasing the time series length. This behavior is particularly evident for the kernel estimator, confirming the findings of previous studies [11,12]. The kernel estimator also shows a substantially higher variance compared to that of the linear and knn estimators [Figs. 8(e) and 8(f)]. In addition, we find that the kernel estimates of entropy and conditional entropy are strongly biased for all values of the analysis parameters [Figs. 8(d), 8(e), 8(g), 8(h), 8(j), and 8(k)]. The bias is less evident for the kernel estimates of information storage, and is generally low or negligible for the linear and knn estimates of all measures.

In Fig. 9 we investigate how the estimates of the different entropy measures are affected by the choice of the analysis parameters. The linear estimation approach has no free parameters and, for this case in which the amplitude distribution of the simulated process matches the assumption of Gaussianity made by the estimator, it returns very precise estimates for all measures [Figs. 9(a)–9(c)]. The kernel estimator turns out to be very sensitive to the choice of its free parameter, the threshold r. Specifically, as shown in Figs. 9(d) and 9(e), when r decreases from 0.5 to 0.2 and 0.1, we observe that the estimates of entropy and conditional entropy are higher and exhibit larger variability. Such sensitive dependence on the threshold r results from the partition rule of the state space used by the kernel estimator. The threshold r is the width of the Heaviside kernel function and determines the size of the cells used for probability estimation: when r decreases, less points are included in the cell used to estimate probabilities; as a result, the estimated probabilities are lower, leading to higher entropy estimates regardless of the true underlying value. On the contrary, when r increases, more points are included in the neighborhood of any reference point, increasing the estimated probability and thus leading to a lower entropy estimate. This holds regardless of the type of kernel function used for entropy estimation, and determines a substantial unreliability for the absolute values of entropy and conditional entropy estimated with the kernel method. The bias (but not the variance) is compensated for the estimates of information storage [Figs. 9(d)–9(f)]. On the contrary, results from the knn estimator are more accurate for the estimation of all entropy measures and much less dependent on the choice of its free parameter k denoting the number of neighboring points used for probability estimation.

Figure 10 provides a more detailed analysis of the dependence of entropy measures on the parameters of a stationary AR process. In this case, where both the AR amplitude ρ and the AR frequency f are varied, we see that the entropy measures reflect the signal properties observed in Fig. 5: increasing ρ with constant f, or moving f away from 0.25 with constant ρ, yields a decrease of conditional entropy and an increase of information storage that indicate lower complexity and higher regularity of the dynamics. Moreover, by comparing the theoretical and estimated values for the different estimators we found that—despite the bias in the kernel estimation of entropy and conditional entropy—all of the estimators can follow the changes in entropy measures when the internal dynamics of the stationary AR process changes. However, unlike the linear estimator which makes an accurate approximation of all entropy measures for all combinations of AR parameters, the kernel and knn estimators exhibit a bias when the AR amplitude is high ($\rho \geqslant 0.8$) and the AR frequency is very low ($f < 0.1$) or very high ($f > 0.4$).

In summary, the simulations reported in this section indicate that the assessment of entropy measures is not an easy task even for the simple case of stationary AR processes. Theoretically, the expected values of conditional entropy and information storage are dependent on the features of the process in a way that is not always straightforward. Moreover, the practical estimation of these measures is not an easy task: while for the linear estimator computation is accurate thanks to the close correspondence between model assumptions and properties

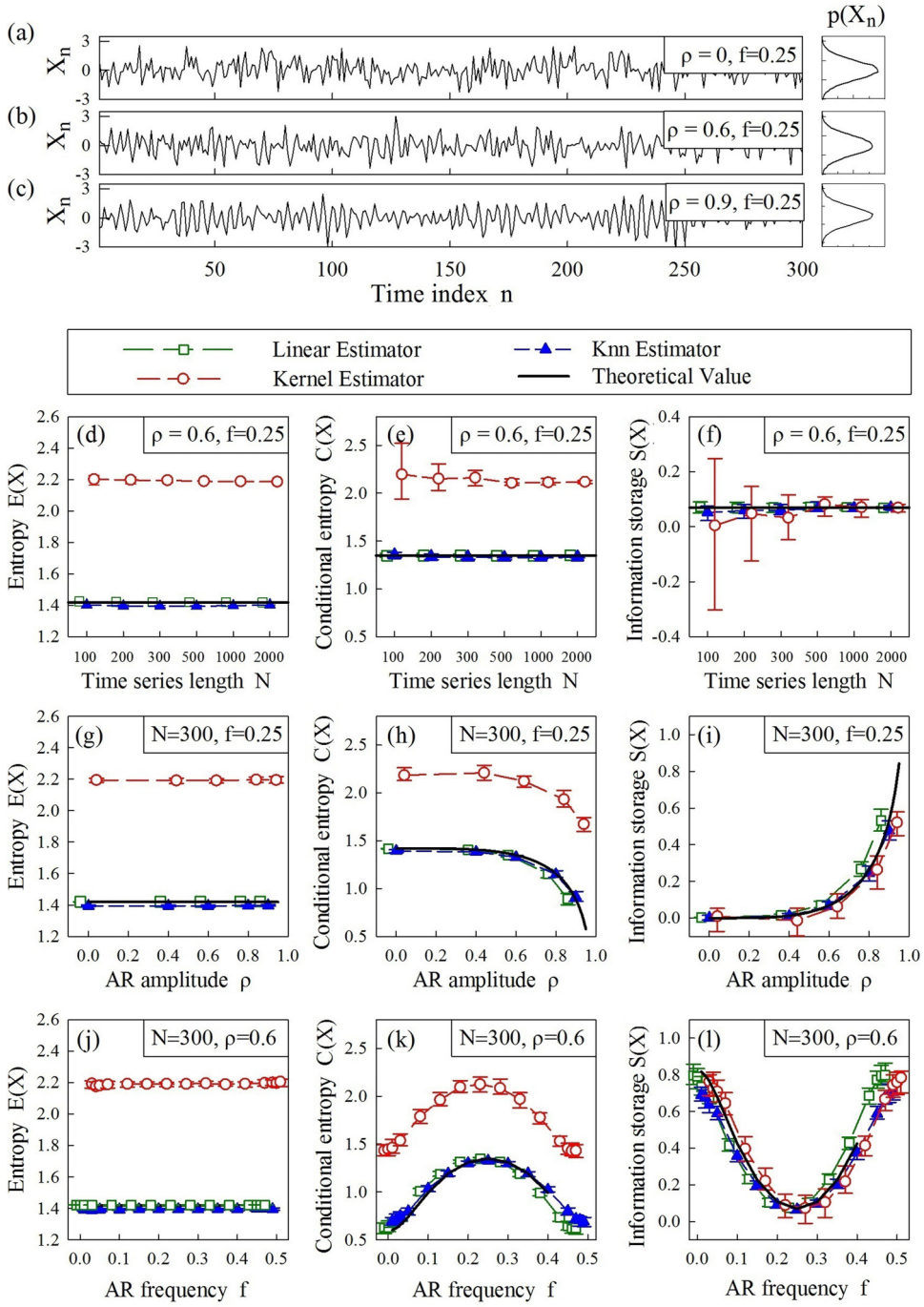

FIG. 8. Performance of entropy estimators and entropy measures for stationary AR processes. (a)–(c) Exemplary realizations of AR processes generated with fixed $f = 0.25$ and varying $\rho = 0,0.6,0.9$, with corresponding probability distributions reported on the right; note that if $\rho = 0$ the process is a white noise $\forall f$. (d)–(f) Dependence of entropy measures and entropy estimates, obtained for AR processes with fixed amplitude and frequency ($\rho = 0.6, f = 0.25$), on the length N of the time series generated as process realizations. (g)–(i) Dependence of entropy measures and entropy estimates on the AR amplitude ρ with fixed AR frequency ($f = 0.25$) and time series length ($N = 300$). (j)–(l) Dependence of entropy measures and entropy estimates on the AR frequency with fixed AR amplitude $\rho = 0.6$ and time series length ($N = 300$). Panels (d)–(l) report the theoretical values (black solid line) and the estimated distributions (mean and 25%–75% percentiles over 100 realizations) of entropy (d), (g), (j), conditional entropy (e), (h), (k), and information storage (f), (i), (l) obtained with the linear estimator (green squares), the kernel estimator implemented with threshold $r = 0.2$ (red circles), and the knn estimator implemented with $k = 10$ neighbors (blue triangles). Results: The expected values of all entropy measures do not change with the realization length N. Moreover, for these normalized time series the theoretical values of entropy are unaffected by the AR parameters. The conditional entropy decreases with the increase of ρ when f is fixed, and increases with increasing f for $f \in (0,0.25]$ and decreases for $f \in (0.25,0.5)$ when ρ is fixed. The theoretical behavior of information storage is the opposite of that of the conditional entropy. The estimates obtained with the linear and knn estimators are close to the theoretical values for all entropy measures, while the estimates of entropy and conditional entropy obtained with the kernel estimator are strongly biased and exhibit high variance for short time series.

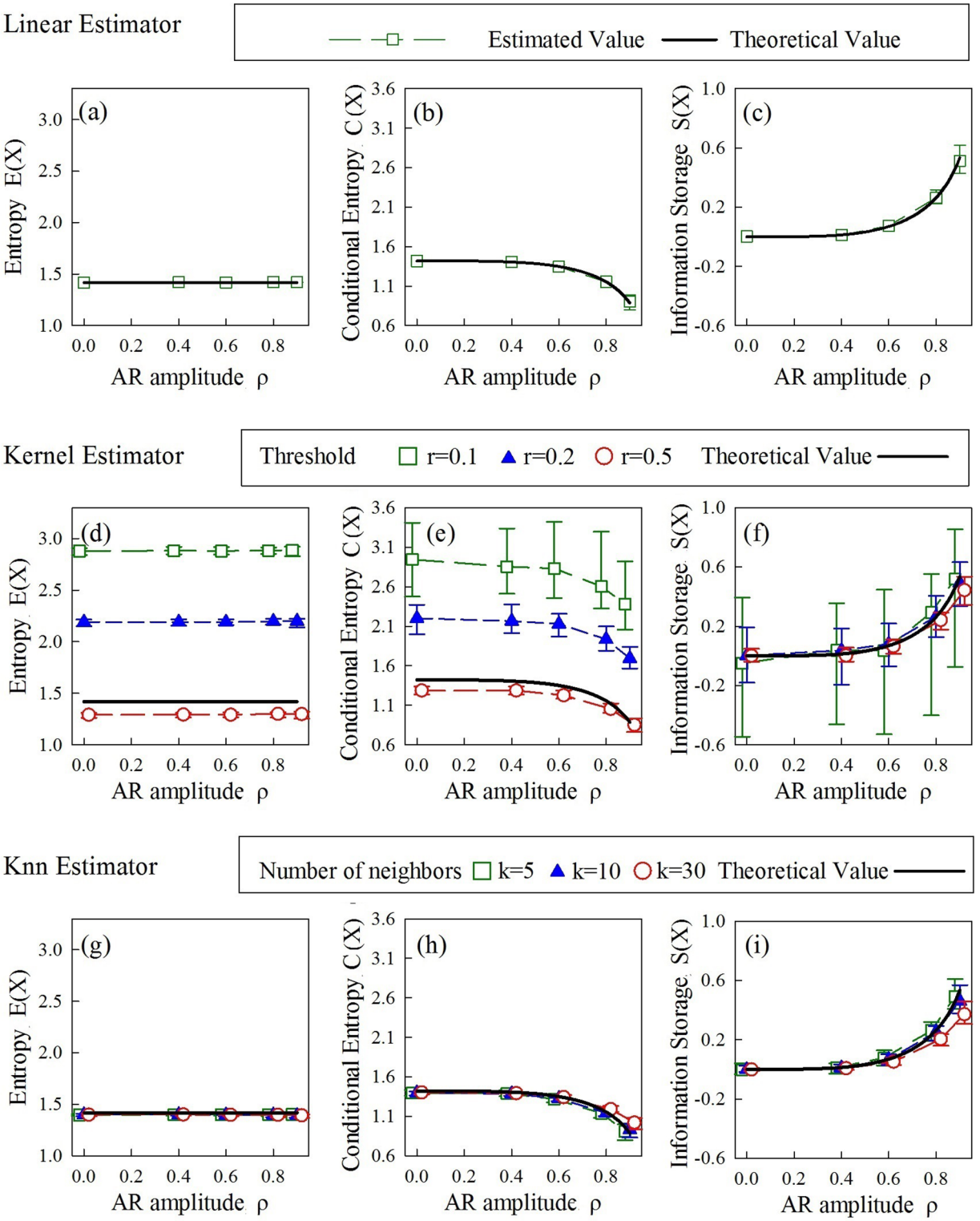

FIG. 9. Dependence of entropy estimators on estimator parameters for stationary AR processes. Plots depict the theoretical values (black solid line) and the estimated distributions (mean and 25%–75% percentiles over 100 realizations, colored symbols and error bars) of entropy (a), (d), (g), conditional entropy (b), (e), (h), and information storage (c), (f), (i) computed using the linear estimator (a)–(c), the kernel estimator (g)–(i) implemented with threshold $r = 0.1$ (green open squares), $r = 0.2$ (blue full triangles), and $r = 0.5$ (red open circles), and the knn estimator (g), (h) implemented with $k = 5$ (green open squares), $k = 10$ (blue full triangles), and $k = 30$ (red open circles) neighbors. Estimates are computed over realizations of length $N = 300$, generated with fixed AR frequency $f = 0.25$ and varying the AR amplitude in the range $\rho \in \{0,0.4,0.6,0.8,0.9\}$. Results: The kernel estimates of entropy and conditional entropy are strongly dependent on the parameter r setting the kernel threshold, whereas the knn estimates are much less sensitive to the parameter k setting the number of neighbors. Since the linear estimator assumes the form of the probability distributions, it has no free parameters.

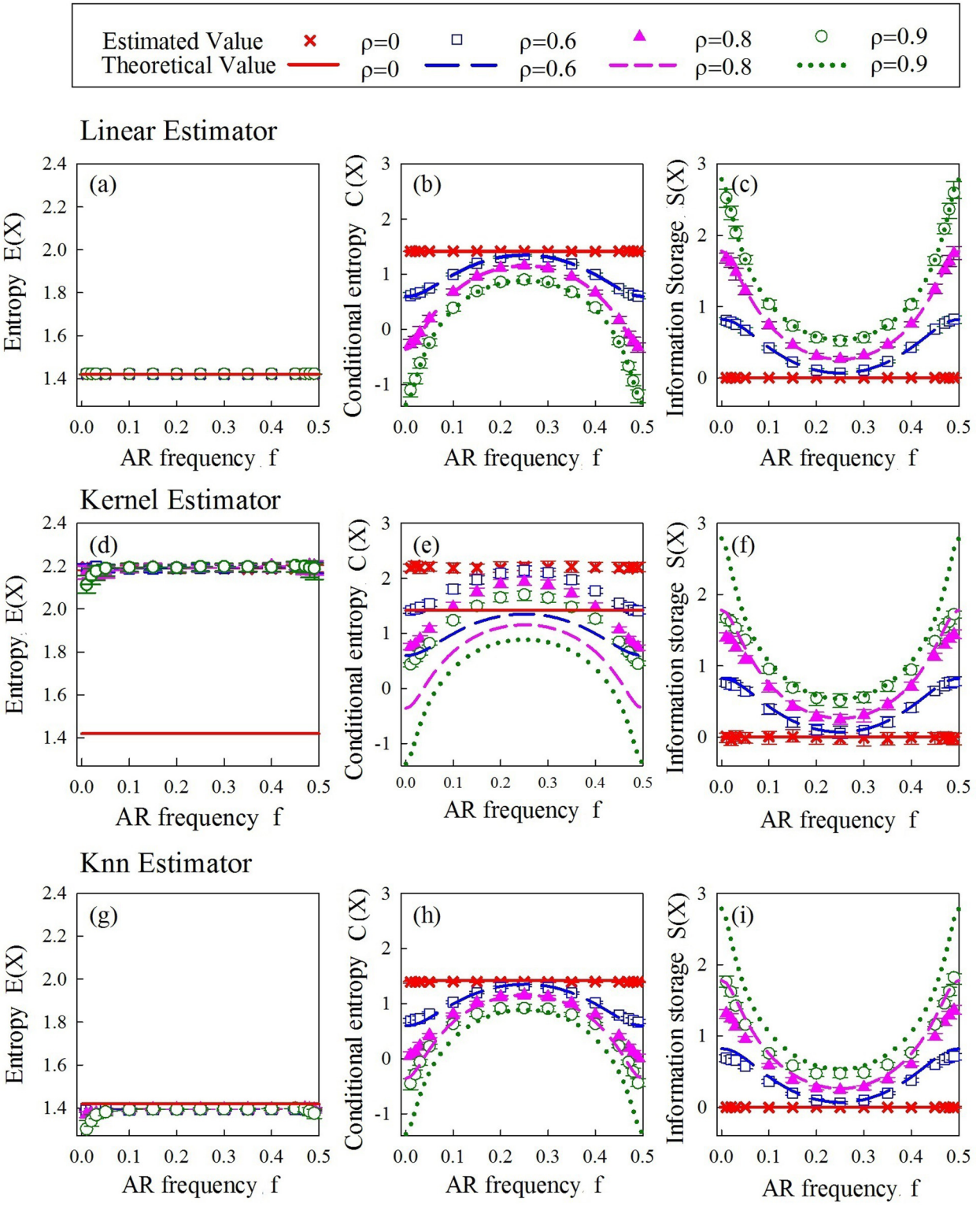

FIG. 10. Dependence of entropy measures and entropy estimators on AR process parameters. Plots depict the theoretical values (lines) and the estimated distributions (mean and 25%–75% percentiles over 100 realizations lasting $N = 300$ samples, colored symbols and error bars) of entropy (a), (d), (g), conditional entropy (b), (e), (h) and information storage (c), (f), (i) computed using the linear estimator (a)–(c), the kernel estimator implemented with threshold $r = 0.2$ (d)–(f), and the knn estimator implemented with $k = 10$ neighbors (g)–(i). Each measure is computed as a function of the AR frequency varying in the range $f \in (0–0.5)$ for different values of the AR amplitude ($\rho = 0$, red crosses and solid lines; $\rho = 0.6$, blue long-dashed lines and open squares; $\rho = 0.8$, pink short-dashed lines and full triangles; $\rho = 0.9$, green dotted lines and open circles). Results: The linear estimates of entropy measures are close to the theoretical values regardless of the values of the AR amplitude ρ and frequency f. The kernel and knn estimates exhibit a bias that is more evident for high values of the AR amplitude ($\rho \geqslant 0.8$), and for very low or very high values of the AR frequency ($f \leqslant 0.1, f \geqslant 0.4$).

of the simulated data (i.e., stationary Gaussian process), the model-free analysis is complicated by empirical factors, such as the data length, but also by the statistical properties of the underlying process. Specifically, we found that the estimates of conditional entropy and information storage are strongly biased for processes exhibiting very slow or very fast regular oscillations. Moreover, the kernel estimates of these measures, though being extremely popular when implemented in measures like approximate entropy and sample entropy, are highly biased with a bias strongly dependent on the estimation parameter. On the other hand, small bias and low estimation variance can be attained by computing the same measures through the knn method.

B. Performance of entropy estimators and entropy measures for AR processes with nonstationarities

Using the results for the stationary AR process as a benchmark, in this section we study the effect of nonstationary behaviors on the performance of the estimators of entropy, conditional entropy, and information storage. Starting from stationary AR processes, we induce three types of nonstationarity by superimposing sinusoidal trends, adding random spikes and inflating the amplitude of segments of the original time series. For each type of nonstationarity, we first compare the statistical properties and the entropy measures estimated for of individual realizations of AR processes with and without nonstationarity and then perform exhaustive analysis assessing how the estimation performance varies with the severity of the simulated nonstationary behavior.

1. Nonstationarity due to sinusoidal trends

The first type of nonstationarity we consider is the sinusoidal trend. As shown in Figs. 11(a) and 11(b), the presence of a sinusoidal trend changes the probability distribution of X_n, which departs from Gaussianity and becomes bimodal. Trends have also the effect of distorting the temporal relation between X_n and X_{n-2}, making it more evident but changing the sign of their correlation [see Figs. 11(c) and 11(d), where the cloud of points is less dispersed and the fitting line changes its slope]. In this case, the change of the distribution after superimposition of the trend is not reflected by alterations of the estimates of the entropy of the time series, while the higher predictability is reflected by a susbstantial decrease of the conditional entropy and a clear increase of the information storage [Figs. 11(e) and 11(f)]. These effects are evident regardless of the entropy estimator.

The effects described above are confirmed by the analysis of 100 process realizations with and without sinusoidal trends reported in Fig. 12. The analysis performed as a function of the AR amplitude shows that, regardless of the period or the amplitude of the trend, the presence of trends does not have big effects on the estimation of entropy but totally impairs the ability of all entropy estimators in following the variations of the regularity of the AR process. While such an ability was documented in Fig. 8 for the original stationary AR process, here we see that none of the estimators can correctly follow the theoretical behaviors of conditional entropy and information storage as a function of AR amplitude ρ, not even qualitatively [see the difference between the estimated values for signals with trends (colored lines with markers) and the theoretical value for original AR signals (black lines)]. Moreover, for all estimators we find that the estimation bias depends more on the trend amplitude A (represented by red line with cross and pink line with triangle) than on the trend period T (represented by blue line with square and green line with circle). With trend amplitude equal to the variance of the original process ($A = 1$), the conditional entropy is underestimated for $\rho < 0.7$ and overestimated for $\rho > 0.7$, while the opposite happens for the information storage; with trend amplitude $A = 5$ the conditional entropy is systematically underestimated and the information storage is systematically overestimated.

Overall, we find that trends have a big impact on the detection of the dynamical complexity of stochastic processes. In all cases, the negative impact of the presence of trends is documented by the flat response of the entropy measures to variations in the predictability of the underlying original process.

2. Nonstationarity due to spikes

Next, we consider the case in which the stationary AR process is corrupted by spikes with random temporal location and amplitude. Spikes are extremely common in real life signals [18,49,52,53,55,108] and may be manifested as artifacts originating from external conditions or from the intrinsic dynamics of the system. Here we simulate spikes with random temporal location and amplitude. As shown in Figs. 13(a) and 13(b), the presence of spikes concentrates the probability distribution of X_n in a way such that the largest part of the signal variance is due to the spikes, which are outliers of the distribution. As a result of the presence of random outliers, the points of the 2D phase plot of (X_{n-2}, X_n) are concentrated around the origin and the estimation of the temporal relation between X_n and X_{n-2} is strongly biased with respect to the clean case [see Fig. 13(d), where the linear fit follows the outliers rather than the noncorrupted points]. As documented in Figs. 13(e) and 13(f), the more concentrated probability distribution induced by the presence of spikes result in a decrease in the model-free estimate of entropy (kernel and knn estimators), while the linear model-based estimate is unchanged for these normalized time series. Moreover, the errors in the detection of the linear relation between time series samples result in a clear overestimation of conditional entropy and underestimation of information storage by the linear estimator. On the contrary, the kernel and knn estimators are still able to capture the predictability of the time series at least to some extent, as demonstrated by the detection of a significant amount of information storage resulting from the estimation of a decrease in the conditional entropy compared with the entropy. We ascribe the higher robustness to spikes of kernel and knn estimators of conditional entropy and information storage to the fact that these estimators explore locally the state space in the computation of probabilities, thus excluding from the estimate the points corrupted by spikes.

Figure 14 reports the results of the systematic analysis of the effects of spikes, performed studying the behavior of the entropy measures as a function of the AR amplitude of the uncorrupted AR process at varying the frequency of occurrence and the amplitude of the spikes. We found that

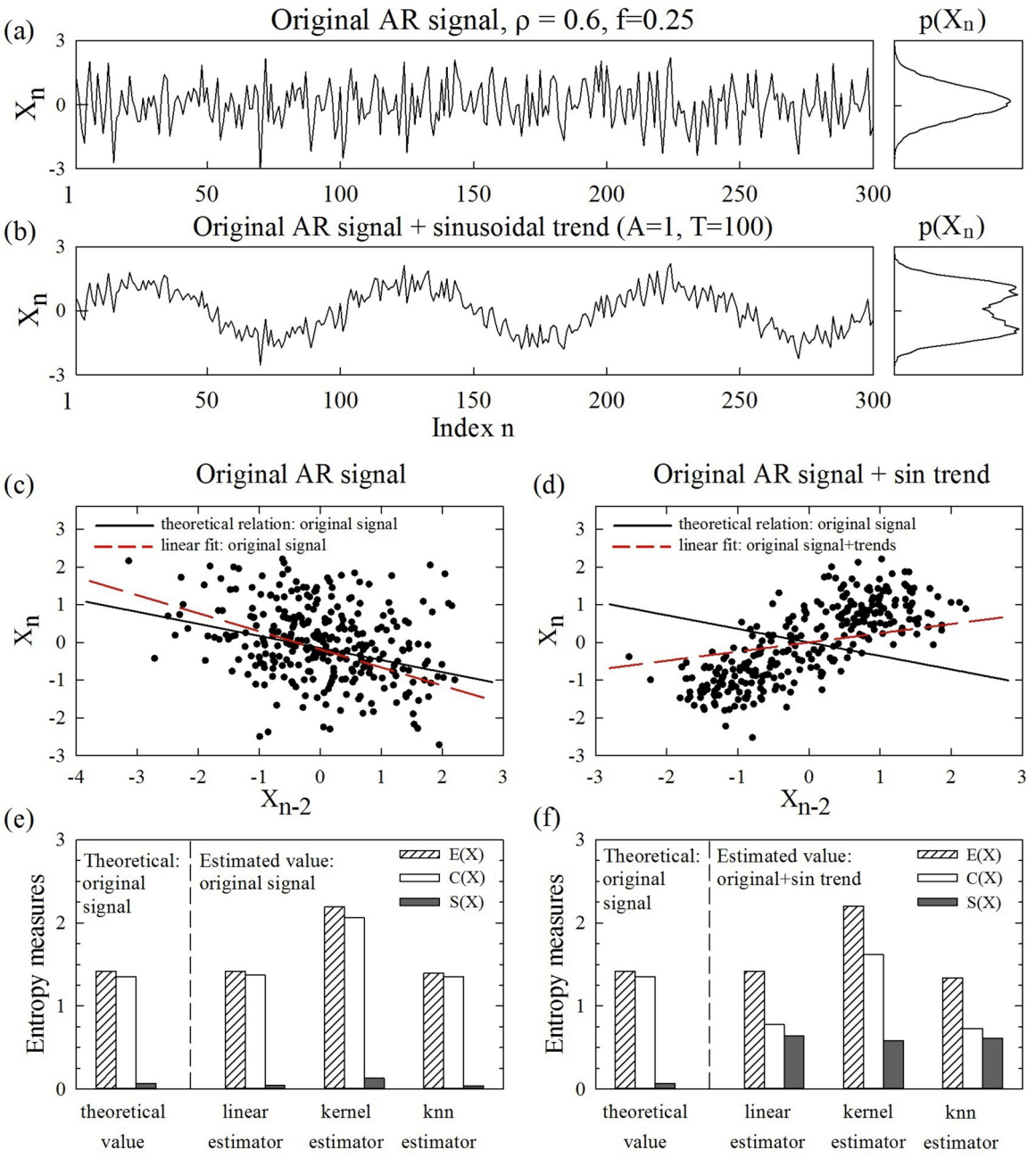

FIG. 11. Performance comparison for AR signals with and without sinusoidal trends: alteration of signal properties and entropy estimates. (a) and (b) show exemplary realizations of a stationary AR process with amplitude $\rho = 0.6$ and frequency $f = 0.25$ before and after superposition of a sinusoidal trend with amplitude $A = 1$ and period $T = 100$; the corresponding probability distributions are shown on the right. Signals are normalized to zero mean and unit variance. (c), (d) 2D phase plots of (X_n, X_{n-2}) derived from the time series in (a), (b). The generating equation of this AR process with $\rho = 0.6$ and $f = 0.25$ is $X_n = -0.36X_{n-2} + U_n$, which yields the theoretical temporal relation between X_n and X_{n-2} shown by the solid black line; the estimated temporal relation obtained through linear least-squares fit of the two clouds of points is shown by the red dashed lines. (e), (f) Entropy (shaded bars), conditional entropy (white bars) and information storage (gray bars) expressed as theoretical values computed for the stationary AR process without trends and estimated values computed for the time series in (a), (b). Estimations are performed using the linear estimator, the kernel estimator with threshold $r = 0.2$, and the knn estimator with $k = 10$ neighbors. Results: The presence of a sinusoidal trend superimposed to a realization of the AR process alters the probability distribution of X_n and distorts the temporal relation between X_n and X_{n-2}. This results in a significant decrease of the conditional entropy and in a significant increase of the information storage for all estimators.

the linear estimates of conditional entropy and information storage are highly affected by spikes, which blunt the capability of the measures to respond to changes in the AR amplitude [Figs. 14(b) and 14(c), except for the case of low spike amplitude ($A = 1$) and percentage ($P = 5\%$) in which a certain performance is preserved]. On the other hand, spikes were found to be less problematic for the kernel and knn estimates of the entropy measures. Figures 14(d)–14(i)

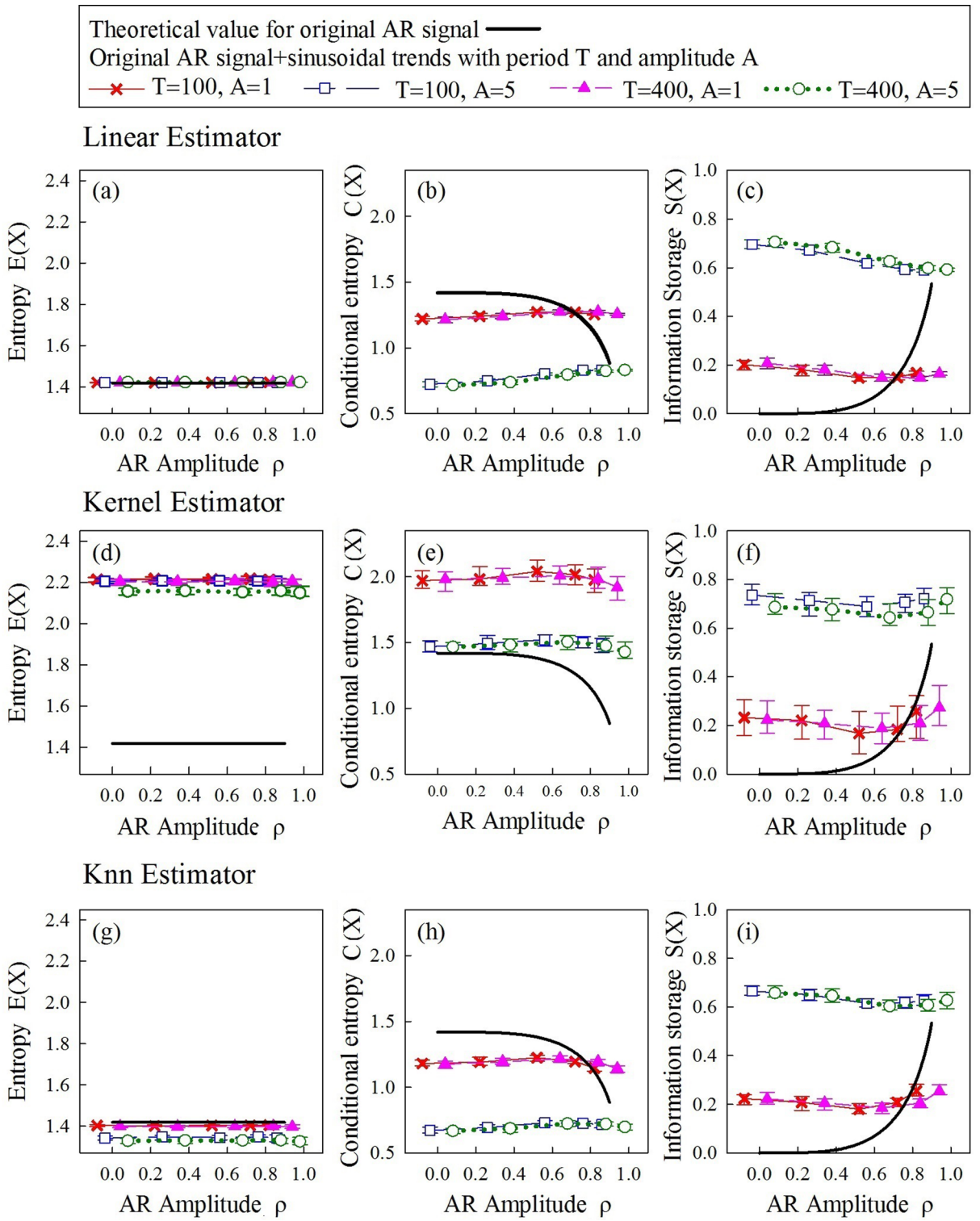

FIG. 12. Effects of nonstationarity due to sinusoidal trends on the estimation of entropy measures. Plots depict the behavior of entropy (a), (d), (g), conditional entropy (b), (e), (h) and information storage (c), (f), (i) computed as a function of the amplitude ρ of an AR process with fixed frequency $f = 0.25$, expressed as theoretical values computed for the original process without trends (black solid lines), and estimated distributions (mean and 25%–75% percentiles) computed over 100 realizations of $N = 300$ samples of the process, each corrupted with an additive sinusoidal trend of period T and amplitude A and normalized to zero mean and unit variance (colored symbols and error bars: $T = 100, A = 1$, red crosses and solid lines; $T = 100, A = 5$, blue open squares and long-dashed lines; $T = 400, A = 1$, pink full triangles and short-dashed lines; $T = 400, A = 5$, green open circles and dotted lines). Estimates are performed using the linear estimator (a)–(c), the kernel estimator implemented with threshold $r = 0.2$ (d)–(f), and the knn estimator implemented with $k = 10$ neighbors (g)–(i). Results: The presence of trends impairs the ability of all estimators to quantify the changes of conditional entropy and information storage induced by variations in the AR amplitude ρ. Moreover, trends induce an estimation bias bias proportional to the trend amplitude A.

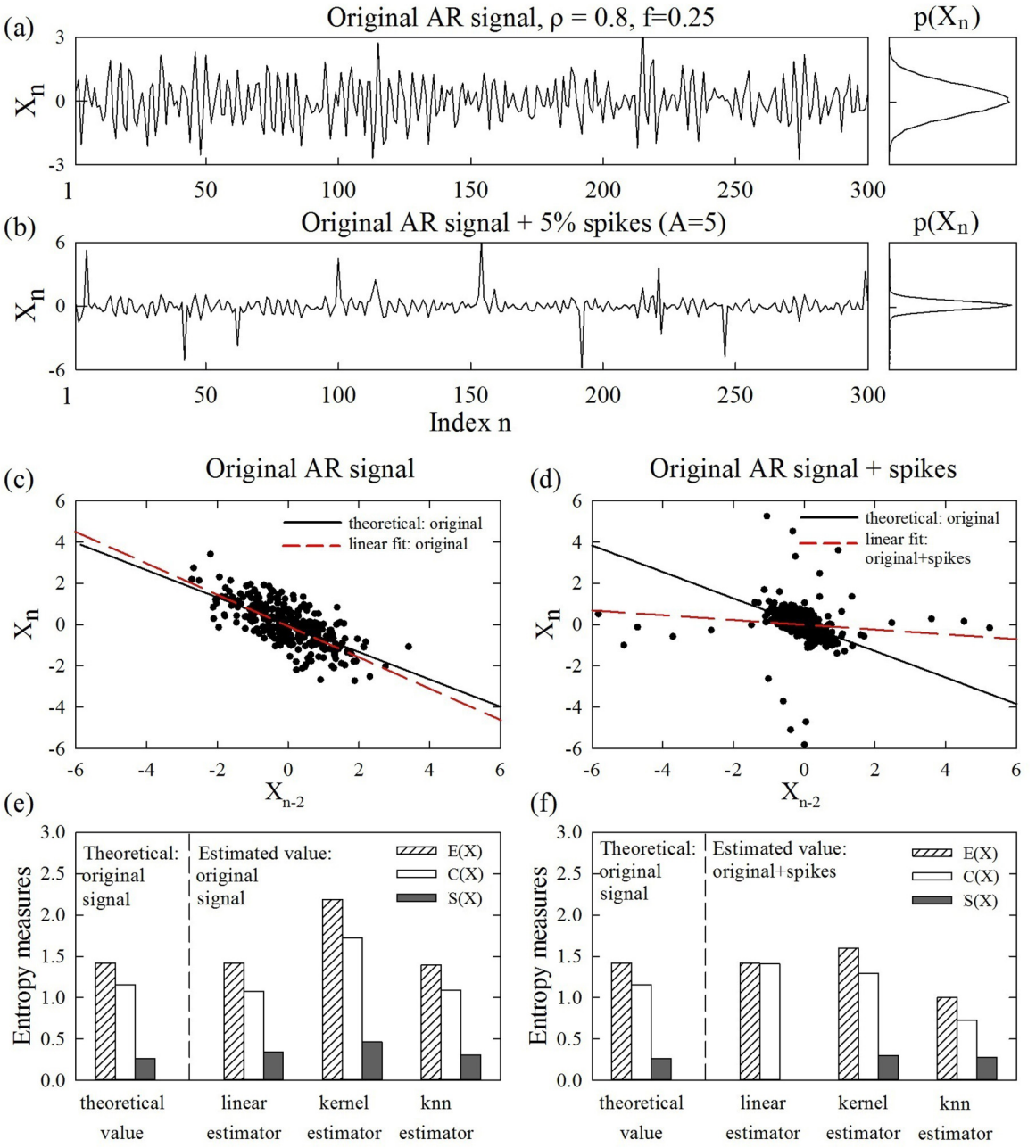

FIG. 13. Performance comparison for AR signals with and without random spikes: alteration of signal properties and entropy estimates. (a), (b) Exemplary realizations of a stationary AR process with amplitude $\rho = 0.8$ and frequency $f = 0.25$ before and after superposition of spikes with amplitude $A = 5$ to 5% of the time series points; the corresponding probability distributions are shown on the right. Signals are normalized to zero mean and unit variance. (c), (d) 2D phase plots of (X_n, X_{n-2}) derived from the time series in (a), (b). The generating equation of this AR process with $\rho = 0.8$ and $f = 0.25$ is $X_n = -0.64X_{n-2} + U_n$, which yields the theoretical temporal relation between X_n and X_{n-2} shown by the solid black line; the estimated temporal relation obtained through linear least-squares fit of the two clouds of points is shown by the red dashed lines. (e)–(f) Entropy (shaded bars), conditional entropy (white bars), and information storage (gray bars) expressed as theoretical values computed for the stationary AR process without spikes and estimated values computed for the time series in (a), (b). Estimations are performed using the linear estimator, the kernel estimator with threshold $r = 0.2$, and the knn estimator with $k = 10$ neighbors. Results: The presence of spikes superimposed to a realization of the AR process concentrates the probability distribution of the process and adds random outliers, which blurs the detection of the temporal relation between between X_n and X_{n-2}. This results in the inability of the linear estimator to detect the information storage in the process, while the kernel and knn estimators are less affected.

display that, apart from a negative bias in the estimation of entropy and conditional entropy, the estimated values of the entropy measures could correctly follow the variations in their theoretical values for the original process induced by changing the AR amplitude. The kernel and knn estimates of information storage exhibit a lower bias and a higher variance than the corresponding estimates of conditional entropy. The dependence of the estimation biases on the spike percentage P

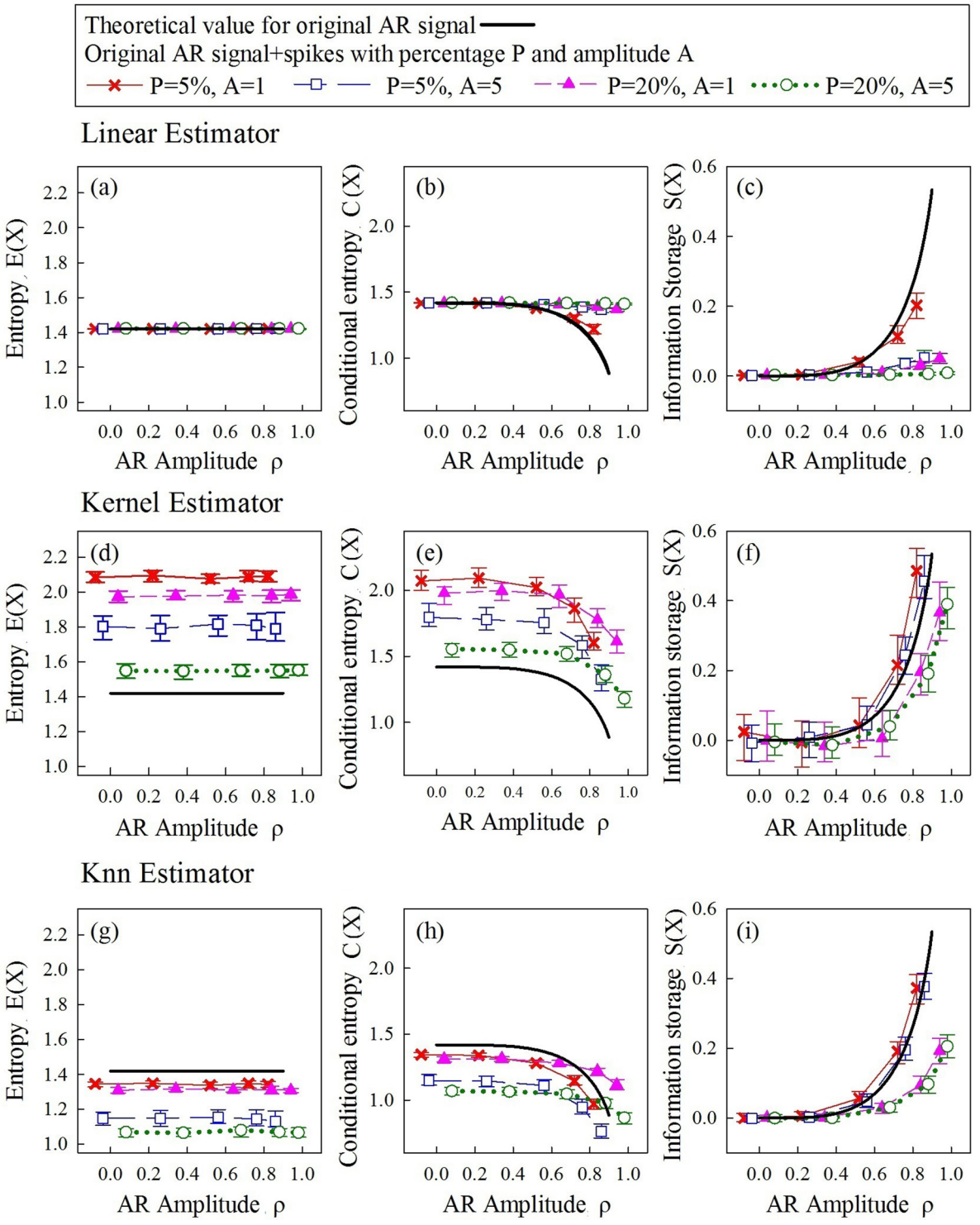

FIG. 14. Effects of nonstationarity due to random spikes on the estimation of entropy measures. Plots depict the behavior of entropy (a), (d), (g), conditional entropy (b), (e), (h), and information storage (c), (f), (i) computed as a function of the amplitude ρ of an AR process with fixed frequency $f = 0.25$, expressed as theoretical values computed for the original process without spikes (black solid lines), and estimated distributions (mean and 25%–75% percentiles) computed over 100 realizations of $N = 300$ samples of the process, each corrupted with additive random spikes of amplitude A occurring with probability P and normalized to zero mean and unit variance (colored symbols and error bars: $P = 5\%, A = 1$, red crosses and solid lines; $P = 5\%, A = 5$, blue open squares and long-dashed lines; $P = 20\%, A = 1$, pink full triangles and short-dashed lines; $P = 20\%, A = 5$, green open circles and dotted lines). Estimates are performed using the linear estimator (a)–(c), the kernel estimator implemented with threshold $r = 0.2$ (d)–(f), and the knn estimator implemented with $k = 10$ neighbors (g)–(i). Results: The presence of spikes partially impairs the ability to quantify the changes of conditional entropy and information storage induced by variations in the AR amplitude ρ; the impairment is more evident for the linear estimator and for high percentages of spikes. Moreover, spikes induce an estimation bias proportional to both the amplitude and the percentage of spikes.

and the spike amplitude A varies across estimators: the linear estimator fails when $P > 5\%$ or $A > 1$, the kernel estimator is equally affected by P and A and the knn estimator is more affected by P than A.

Overall, these results indicate that spikes have a deleterious impact on the model-based estimation of the measures of dynamical complexity. Since spikes are commonly encountered in a large variety of practical settings, we conclude that cautions should be used in adopting linear approaches to the computation of conditional entropy and information storage in the presence of these artifacts. On the contrary, model-free estimates are less affected by spikes and, in the presence of a moderate amount and amplitude of spikes, they are still sensitive to variations in the dynamical complexity of the clean time series.

3. Nonstationarity due to local changes in the signal variance

As a third nonstationary behavior, we consider the alteration in the amplitude of segments of the original AR process. As shown in Figs. 15(a) and 15(b), the local alteration of the signal variance has the effect of concentrating the probability distribution of X_n in a similar way than for the case of spikes. Similarly, the 2D phase plot of (X_{n-2}, X_n) exhibits a percentage of outliers that surround the cloud of points representing the non-corrupted portions of the original time series [Fig. 15(d)]. However, since this type of nonstationarity does not destroy the temporal relation between the time series samples, the linear fit in the 2D phase plot of (X_{n-2}, X_n) is still quite accurate. As a result, the linear estimation of conditional entropy and information storage is a bit degraded, but not fully impaired as in the case of random spikes [see Figs. 15(e) and 15(f)]. In this individual realization, the kernel and knn estimators provide slightly better performances in terms of estimation of conditional entropy and information storage. Note that, as in the case of random spikes, the concentration of the probability distribution is reflected by lower values of the entropy estimated using the kernel and knn estimators, while the linear estimates are again unaffected by the shape of the distribution.

Figure 16 reports the results of the complete analysis whereby the estimation of entropy measures is performed as a function of the AR amplitude for different values of the percentage and maximal amplitude of the segments of high variance imposed in the AR process. The main effect of the presence of segments of high variance is the introduction of a negative bias in the model-free estimates of entropy and conditional entropy, as well as of a positive bias in the model-free estimates of information storage [Figs. 16(d)–16(i)]; the bias in the information storage is higher for the kernel estimates than for the knn estimates. The linear estimates of the entropy measures are less affected by this bias [Figs. 16(a)–16(c)]. In spite of the bias we found that, in all conditions of local alteration of the signal variance, the values of the entropy measures computed using all estimators could follow the changes in their theoretical value imposed by varying the AR amplitude ρ.

These results suggest that the presence of nonstationarity due to segments of high variance is not as detrimental as other types of nonstationary behaviors, as it introduces a bias in the entropy measures but does not preclude the capability of these measures to detect changes in dynamical complexity induced by alterations of the predictable structure of the observed process.

C. Performance of entropy estimators and entropy measures for fractionally integrated white noise processes

In this section, we investigate the theoretical behavior of the entropy measures, as well as the performance of all entropy estimators in computing these measures, for processes with power-law long-range correlations. After setting the properties of fractionally integrated white noise processes as described in the methods (Sec. II C 3), the theoretical values of the entropy measures are computed as a function of the differencing parameter d, which controls the sign and the strength of long range correlations using the derivations described in the Appendix. We then compare these theoretical values with the distribution of the estimated values, in order to evaluate comparatively the efficacy of the various entropy estimators.

Results of this analysis are reported in Fig. 17. First, we find that both the theoretical and the estimated values of conditional entropy decrease, and the values of information storage increase, with the strength of long-range correlations modulated by the differencing parameter d. Additionally, the asymmetric behaviors of conditional entropy and information storage in response to positive or negative variations of the differencing parameter d [Figs. 17(b), 17(c), 17(e), 17(f), 17(h), and 17(i)] document that entropy measures are more sensitive to positive long-range correlations than to anti-correlations of the same strength. These results mirror the fact that signals with positive correlation are often associated with longer memory than signals with a negative correlation of the same strength.

Moreover we investigate the dependence of the entropy estimates on the time series length N, finding that not only the variance, but also the bias of of the estimates of conditional entropy and information storage, decrease for longer time series; this behavior is particularly evident for positive long-range correlations [Figs. 17(b), 17(c), 17(e), 17(f), 17(h), and 17(i)]. Similar discrepancies between numerical and mean estimated values of complexity measures were observed in [12] for $1/f$ noise time series, indicating that stationarity is an important prerequisite for the analysis of short time series, and trend-like behaviors may impair entropy estimation. A potential explanation for this finding lies in the the fact that, since signals with stronger positive correlation exhibit more trendlike behaviors [see Fig. 6(a) for a representative example], longer time series are needed to capture the similarity in the patterns that determines accurate estimates of conditional entropy and information storage. These results hold for all estimators, and also confirm the higher variance of the kernel estimator, compared to the linear and nearest-neighbor estimators, as is previously observed for the simulations of pure AR processes.

The findings reported in this section document that, in addition to traditionally used analytical tools for the quantification of long-range correlations such as the detrended fluctuation analysis (DFA), also the entropy measures studied in this work, which are commonly used to assess short-range dependencies, are able to quantify the the degree of long-range dependency

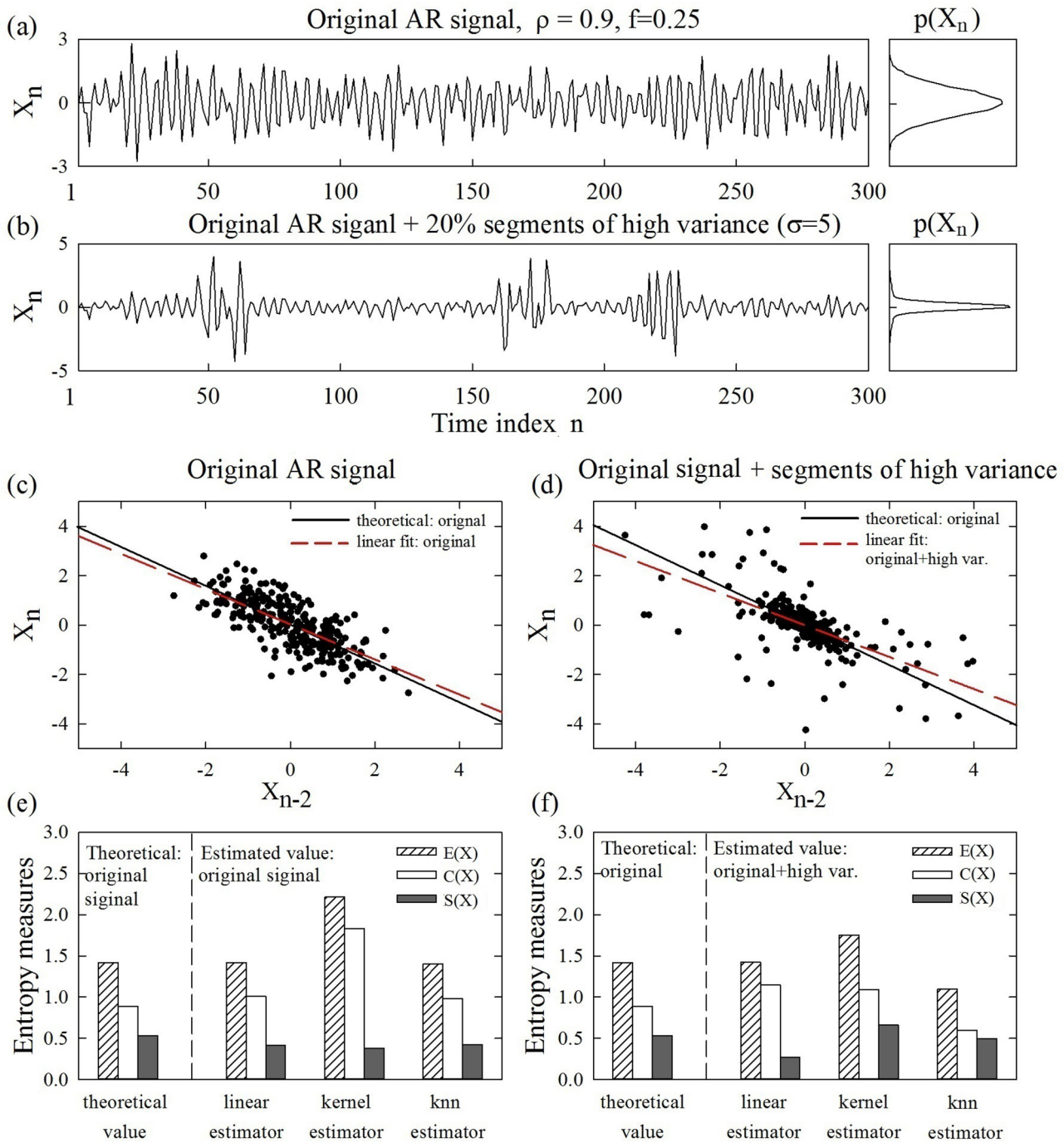

FIG. 15. Performance comparison for AR signals with and without segments of high variance: alteration of signal properties and entropy estimates in the presence of local changes in signal variance. (a), (b) Exemplary realizations of a stationary AR process with amplitude $\rho = 0.9$ and frequency $f = 0.25$ before and after inflating random segments by an amplification factor $\sigma = 5$ (a total of $P = 20\%$ of the time series points are inflated); the corresponding probability distributions are shown on the right. Signals are normalized to zero mean and unit variance. (c), (d) 2D phase plots of (X_n, X_{n-2}) derived from the time series in (a), (b). The generating equation of this AR process with $\rho = 0.9$ and $f = 0.25$ is $X_n = -0.81X_{n-2} + U_n$, which yields the theoretical temporal relation between X_n and X_{n-2} shown by the solid black line; the estimated temporal relation obtained through linear least-squares fit of the two clouds of points is shown by the red dashed lines. (e), (f) Entropy (shaded bars), conditional entropy (white bars), and information storage (gray bars) expressed as theoretical values computed for the stationary AR process without changes in variance and estimated values computed for the time series in (a), (b). Estimations are performed using the linear estimator, the kernel estimator with threshold $r = 0.2$, and the knn estimator with $k = 10$ neighbors. Results: The presence of segments with higher variance concentrates the probability distribution of the process and disperses a portion of the points without distorting their temporal relation. This results in a mild decrease of the conditional entropy and increase of the information storage for the kernel and knn estimators, while opposite changes are appreciated for the linear estimator.

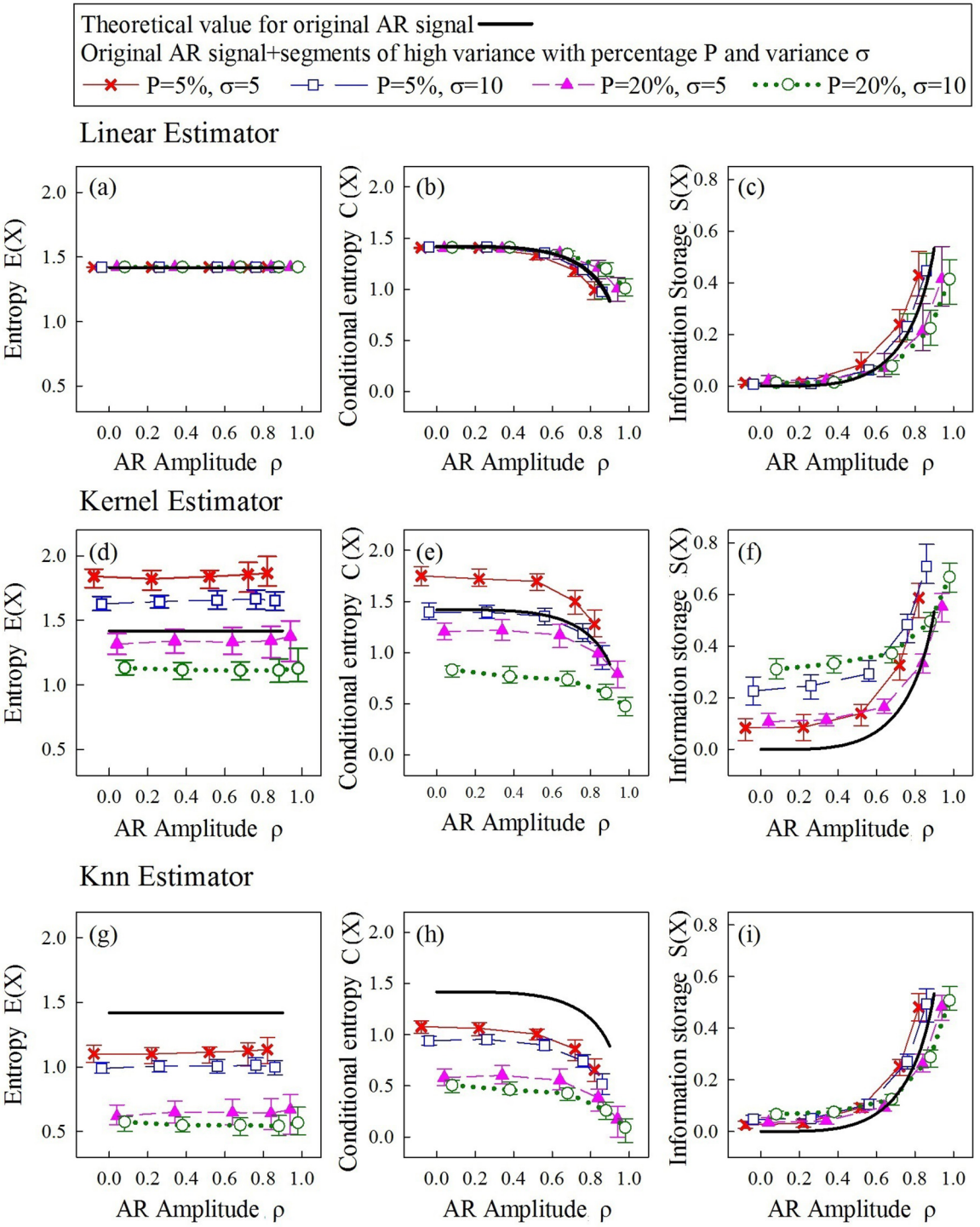

FIG. 16. Effects of nonstationarity due to local changes in the signal variance on the estimation of entropy measures. Plots depict the behavior of entropy (a), (d), (g), conditional entropy (b), (e), (h), and information storage (c), (f), (i) computed as a function of the amplitude ρ of an AR process with fixed frequency $f = 0.25$, expressed as theoretical values computed for the original process without changes in variance (black solid lines), and estimated distributions (mean and 25%–75% percentiles) computed over 100 realizations of $N = 300$ samples of the process, each corrupted by randomly distributed inflated segments and normalized to zero mean and unit variance. Each inflated segment lasts 20 points and is generated by magnifying original data points by a factor of σ, with the percentage of inflated points to the total signal length being $P\%$ (colored symbols and error bars: $P = 5\%, \sigma = 5$, red crosses and solid lines; $P = 5\%, \sigma = 10$, blue open squares and long-dashed lines; $P = 20\%, \sigma = 5$, pink full triangles and short-dashed lines; $P = 20\%, \sigma = 10$, green open circles and dotted lines). Estimates are performed using the linear estimator (a)–(c), the kernel estimator implemented with threshold $r = 0.2$ (d)–(f), and the knn estimator implemented with $k = 10$ neighbors (g)–(i). Results: The presence of segments with high variance does not impair significantly the ability to quantify the changes of conditional entropy and information storage induced by variations in the AR amplitude ρ. However, local changes in the signal variance induce an estimation bias proportional to the percentage of inflated points and—to a lower extent—to the amplitude of inflation.

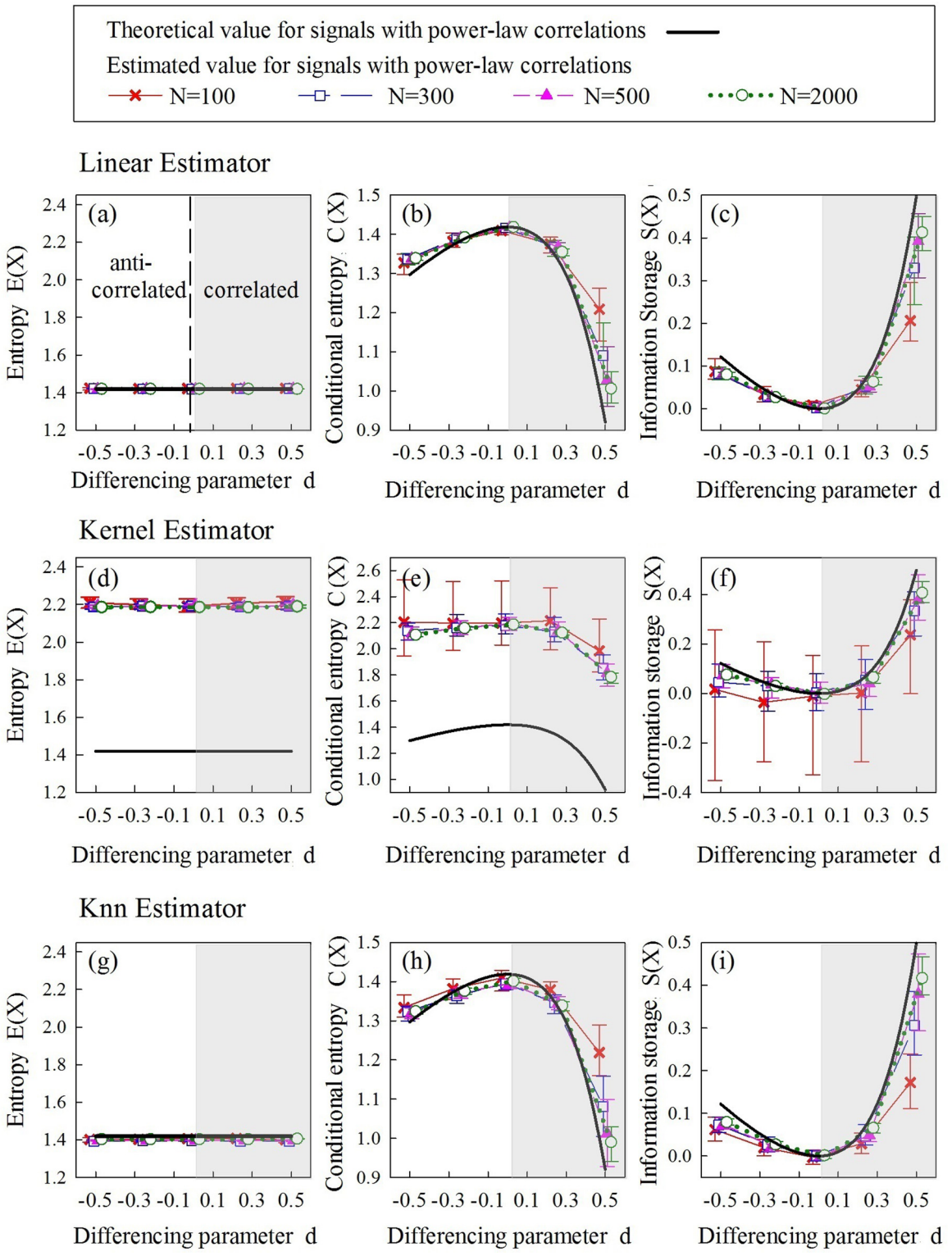

FIG. 17. Performance of entropy estimators and entropy measures for fractionally integrated processes with long-range power-law correlations. Plots depict the theoretical values (black solid lines) and the estimated distributions (mean and 25%–75% percentiles over 100 realizations of length $N = 100$ (red crosses and solid lines), $N = 300$ (blue open squares and long-dashed lines), $N = 500$ (pink solid triangles with short-dashed lines), and $N = 2000$ (green open circles with dotted lines)) of entropy (a), (d), (g), conditional entropy (b), (e), (h), and information storage (c), (f), (i) computed as a function of the differencing parameter d. Estimates are obtained using the linear estimator (a)–(c), the kernel estimator implemented with threshold $r = 0.2$ (d)–(f), and the knn estimator implemented with $k = 10$ neighbors (g)–(i). Results: Signals with positive correlations present lower conditional entropy and higher information storage than signals with anticorrelation of the same strength. All estimators reflect the changes in entropy measures with the type and strength of power-law correlation, with a higher accuracy for longer time series length, but exhibit a bias which increases with the correlation strength.

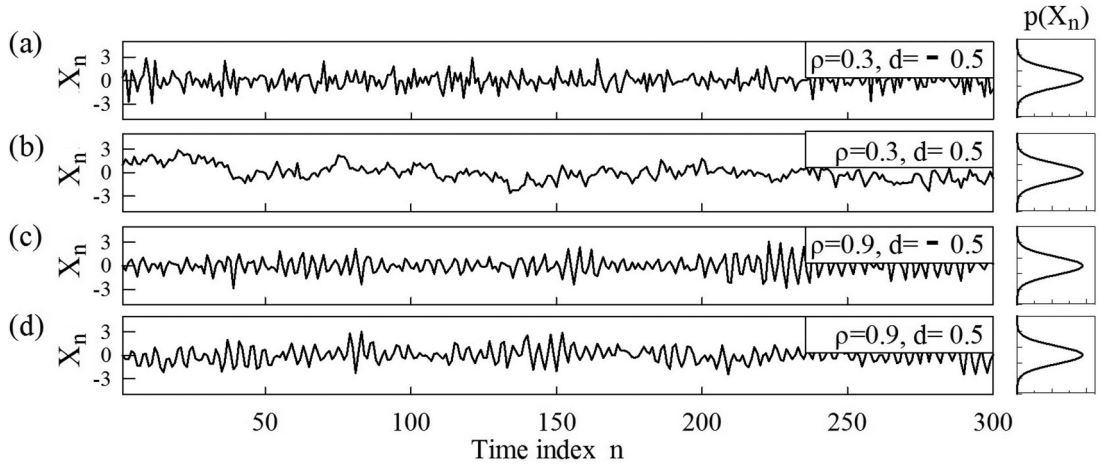

FIG. 18. Exemplary realizations of fractionally integrated autoregressive processes for varying AR amplitude ρ and differencing (correlation) parameter d. Plots depict signals with (a) weak autoregression ($\rho = 0.3$) and strong anti-correlation ($d = -0.5$), (b) weak autoregression ($\rho = 0.3$) and strong positive correlation ($d = 0.5$), (c) strong autoregression ($\rho = 0.9$) and strong anticorrelation ($d = -0.5$), (d) strong autoregression ($\rho = 0.9$) and strong positive correlation ($d = 0.5$). Slow trends in the signal are present for strong positive long-range correlation, and less evident for anticorrelated signals or in the presence of a strong AR component.

of the present of a process on its past values. However, the accuracy of the estimates is highly dependent on the time series length, indicating that—contrary to what happens for the estimation of short-range AR dependencies—very long realizations would be needed to yield accurate estimation of conditional entropy and information storage in the presence of strong positive long-range correlations.

D. Performance of entropy estimators and entropy measures for fractionally integrated autoregressive processes

The results reported in the previous sections describe the capability of entropy measures and entropy estimators to reflect changes in the temporal structure of both pure AR processes producing stochastic oscillations and pure fractionally integrated white noise processes exhibiting power-law long-range correlations. Here we extend the analysis by investigating whether and how the theoretical properties of the entropy measures and the performance of the entropy estimators change when the analyzed processes display both short-term AR dependencies and power-law long-range correlations. Representative examples of these processes are reported in Fig. 18, suggesting that their dynamical structure is altered in a different way depending on the strength of the stochastic oscillation and the sign of the long-range correlations. Specifically, we see that long-range correlations of the same strength ($|d| = 0.5$) determine different structure in the signals depending on their sign when the AR amplitude is low [$\rho = 0.3$, Figs. 18(a) and 18(b)], while they do not affect substantially or differently the dynamical structure when the AR amplitude is high [$\rho = 0.9$, Figs. 18(c) and 18(d)].

Results of the analysis performed at varying the AR amplitude ρ for different values of the differencing parameter d are reported in Fig. 19. First, the analysis confirms that, for these Gaussian normalized time series, the expected values of entropy are not dependent on the parameters ρ and d, and the estimates, apart from the bias of the kernel method known also before, are accurate for all approaches Figs. 19(a), 19(d), and 19(g). The theoretical values of conditional entropy and information storage deviate from their behavior for pure AR processes (pink short-dashed curves) in a way depending on the sign of long-range correlations: for anticorrelated processes ($d < 0$, red solid and blue long-dashed curves) the trend is similar to the case $d = 0$ apart from a shift of the curves toward lower conditional entropy and higher information storage; for positive long-range correlated processes (green dotted and gray dash-dot curves) conditional entropy and information storage are no more increasing monotonically with ρ, showing a nontrivial dependency especially for high values of the differencing parameter. As seen in Figs. 19(b), 19(c), 19(e), 19(f), 19(h), and 19(i), these theoretical trends were followed by the estimated values with a performance comparable to that observed for the various estimators applied to pure AR or fractionally integrated processes (i.e., with the strong bias typical of the kernel estimator and with a slightly better performance of the linear estimator compared with the knn estimator). The main difference is that in this case of combined AR and fractionally integrated processes all estimators (even the linear) produced biased estimates of the entropy measures. The bias was positive for conditional entropy estimates and negative for information storage estimates, increased with the differencing parameter d, and was more marked for positive d than for negative d.

Thus, the combined presence of stochastic oscillations arising from short-term interactions and of power-law long-range correlations, which is a very common situation of real-world time series, complicates both the theoretical behavior and the practical estimation of entropy measures. The interpretation of the values taken by these measures, as well as their accurate estimation, become problematic in the presence of very regular stochastic oscillations, when the increase of the conditional entropy (and the decrease of the information storage) with the strength of long-range correlations is more subtle, or in the presence of strong positive correlations, when it may happen that the conditional entropy does not decrease (and

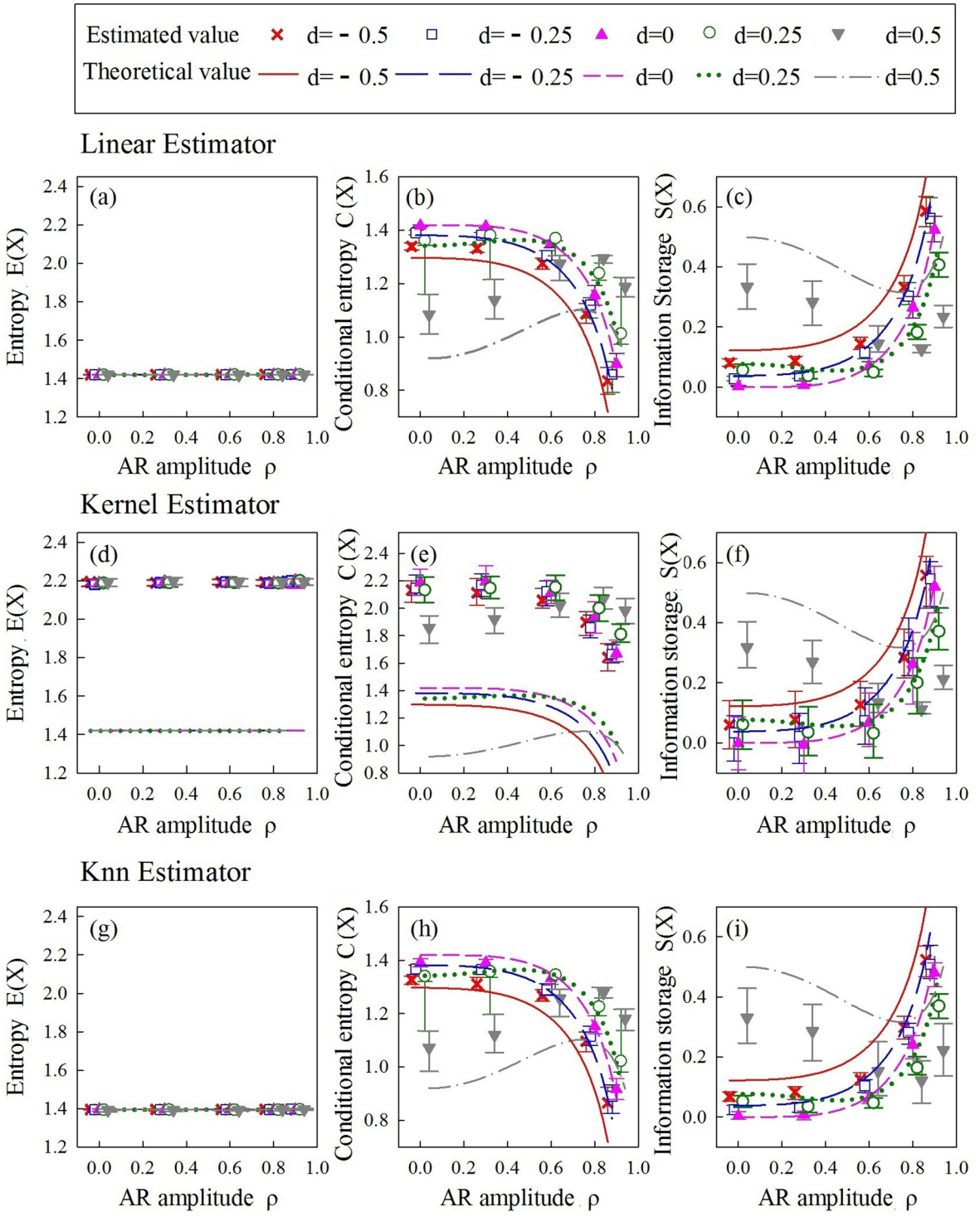

FIG. 19. Performance of entropy estimators and entropy measures for fractionally integrated autoregressive processes. Plots depict the theoretical values (lines) and the estimated distributions (mean and 25%–75% percentiles over 100 realizations lasting $N = 300$ samples, colored symbols and error bars) of entropy (a, d, g), conditional entropy (b), (e), (h), and information storage (c), (f), (i) computed using the linear estimator (a)–(c), the kernel estimator implemented with threshold $r = 0.2$ (d)–(f), and the knn estimator implemented with $k = 10$ neighbors (g)–(i). Each measure is computed as a function of the AR amplitude varying in the range $\rho \in (0,0.9)$ for fixed AR frequency ($f = 0.25$) and different values of the differencing parameter ($d = -0.5$, red crosses and solid lines; $\rho = -0.25$, blue long-dashed lines and open squares; $\rho = 0$, pink short-dashed lines and full triangles; $\rho = 0.25$, green dotted lines and open circles; $\rho = 0.5$, gray dash-dotted lines and open circles). Results: The presence of strong positive correlation alters markedly the dependence of conditional entropy and information storage on the AR amplitude, whereas anticorrelation induces only a shift in the measures without affecting substantially the dependence on ρ. All estimators are biased in approximating conditional entropy and information storage for fractionally correlated AR signals. The bias is more evident for positive than negative correlations and for stronger than weaker correlations.

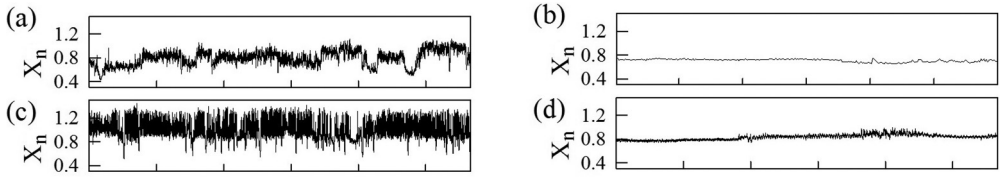

FIG. 20. Exemplary signals of consecutive heart beat intervals for one healthy subject (a), (c) and one CHF subject (b), (d), during day (a), (b) and night (c), (d). The original signals of healthy subjects typically exhibit larger variability and amplitude than those of CHF subjects.

the information storage does not increase) while increasing the regularity of the stochastic oscillations.

IV. APPLICATION TO HEART RATE VARIABILITY

Heart rate variability (HRV), the variation over time of the period between consecutive heartbeats, is a reliable reflection of the many physiological factors modulating the rhythm of the heart in healthy conditions, as well as of the alteration of these factors related to pathological states [114,115]. It is widely accepted that the assessment of HRV over temporal scales ranging from seconds to a few minutes allows the indirect investigation of the short-term mechanisms underlying cardiovascular control [116–118]. To investigate these short-term dynamics and their structural complexity, a viable and widely exploited approach is the use of entropy-based methods such as approximate entropy (ApEn), sample entropy (SampEn), corrected conditional entropy (CCE), and various refinements of these measures [2–4,19,20,30,119,120]. On the other hand, it is also known that heartbeat fluctuations exhibit long-range correlation properties manifested in $1/f$-like behavior, power-law correlations, multifractal spectrum and scaling behaviors that change with physiological state and disease [57,58,66,68,69,121,122]. Therefore, the assessment of the dynamical complexity of HRV remains a challenge because of the poorly investigated effects of long-range correlation properties on the patterns of short-term dynamics, and of the unclear role played by nonstandardized preprocessing steps and utilization of different entropy measures and estimators.

In order to test the ability of entropy measures in detecting changes in the static and dynamical properties of HRV signals in different physiological states and clinical conditions, as well as to assess the sensitivity of these measures to the adopted entropy estimator and preprocessing steps, here we study heartbeat dynamics measured in healthy subjects and congestive heart failure patients (CHF) during wake and sleep conditions [123]. Specifically, we considered a group of 18 healthy subjects (13 females and 5 males, with ages between 20 and 50, average 34.3 years) and a group of 12 patients suffering from CHF (3 females and 9 males, with ages between 22 and 71, average 60.8 years), in whom the time series of the consecutive heartbeat intervals were measured from the holter ECG recordings acquired continuously during six hours of wake (12 p.m. to 6 p.m.) and 6 h of sleep (12 a.m. to 6 a.m.); an example for one healthy subject and one CHF patient is reported in Fig. 20.

For each recording, entropy, conditional entropy, and information storage were computed over consecutive sequences of 300 interbeat intervals overlapped by half using the linear, kernel, and nearest-neighbor estimators. The analyses were performed under three types of preprocessing procedure: (1) the originally measured HRV time series in which the mean is removed within each 300-point window (local mean removal); note that removing the mean within each window will not affect the computation of entropy measures, but only serves as a prerequisite for the linear estimator because it was implemented without a constant term; (2) the same time series normalized to zero mean and unit variance within each 300-point window (local normalization); and (3) the same time series filtered by a linear high-pass filter (IIR with zero-phase, cutoff frequency at 3dB: 0.02 cycles/beat [124]) to remove slow trends and normalized to zero mean and unit variance within each 300-point window (slow-trend removal and local normalization). Exemplary signals for all three preprocessing procedures are given in Fig. 21.

Estimations were performed by setting standard commonly used values for the embedding and estimator-specific parameters, which also correspond to those used in the simulations: $m = 2$ points were chosen for representing the past of the processes; kernel entropy estimates were computed setting the threshold r equal to 0.2 times the standard deviation of the time series, and knn estimates were computed using $k = 10$ neighbors.

Then, the median value of the distribution of each entropy measure computed for each healthy subject or CHF patients during wake (W) and sleep (S) was retained for statistical analysis. A paired t-test was used to test the difference between measures derived during W and S inside the same group (Healthy or CHF), while an unpaired t-test was used to check differences between Healthy and CHF for a given analysis condition (W or S). A $p < 0.05$ was always considered as statistically significant.

Figure 22 collects the results of the analysis of the three entropy measures, computed using the three considered estimators applied to the HRV time series measured from the CHF patients and the healthy controls during wake and sleep conditions, as well as to the normalized and filtered versions of these time series. Results illustrate that different entropy measures can reflect different aspects of cardiac dynamics across physiological states and pathological conditions. They also provide evidence for the sensitivity of the measures to the estimator adopted and the preprocessing procedures applied. Our major findings are listed in the following.

A. Static measure of complexity in HRV signals: Entropy

Let us start by analyzing the entropy of the cardiac dynamics [Figs. 22(a), 22(d), and 22(g)]. Note that estimated values of Entropy measured by the linear and knn estimators depend on the units of measure of the time series amplitudes (in this case,

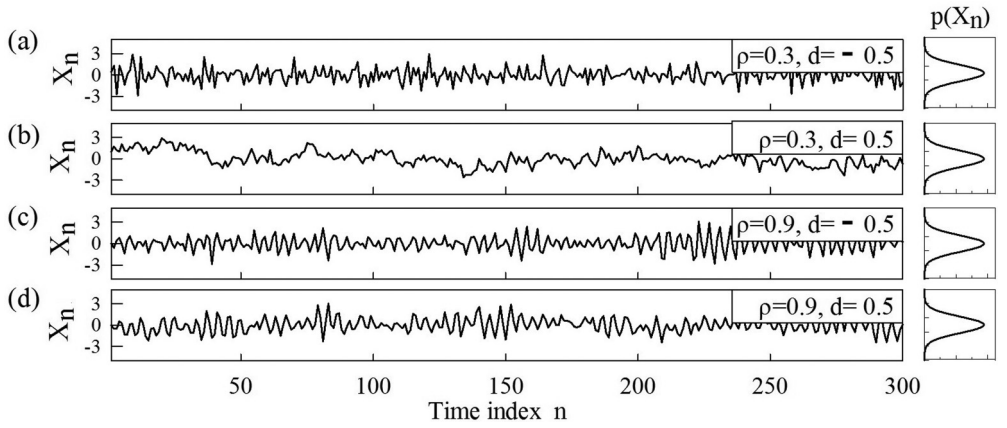

FIG. 21. Exemplary HRV time series of 300 points measured from the same CHF subject during day shown in Fig. 20(b) under different pre-processing procedures. Panels depict the original signal (a), the same signal in (a) normalized to zero mean (b), the same signal in (a) normalized to zero mean and unit variance (c), the same signal in (a) with slow trends removed by a linear high pass filter and normalized to zero mean and unit variance.

seconds), thus being little informative. As shown in Figs. 22(a) and 22(g), the values of entropy measured by the linear and knn estimators were negative for signals that are not normalized to unit variance and were positive for signals with unit variance, which is reasonable given that their estimations are dependent on signal variance [Eqs. (10) and (20)]. On the other hand, the kernel estimation of Entropy, being implemented by taking a percentage of the signal variance as similarity threshold when computing probabilities, is not sensitive to alterations in variance and thus yields very stable estimates for all preprocessing conditions [Fig. 22(d)].

The main finding about entropy is that it was markedly lower in CHF patients than in healthy subjects both during wake and sleep and was higher during sleep than during wake in both groups. These results, which were observed using the linear and knn estimators and hold only for the zero-mean time series without pre-processing, reflect, respectively, a depressed HRV in CHF patients [125,126] and a higher variance of the cardiac dynamics during sleep. Normalization of the time series to unit variance affects dramatically the values of Entropy as well as their variations across conditions: since the linear estimator relies only on variance to estimate entropy, after normalization it fails to detect changes in the overall signal variability [Fig. 22(a)]; the decrease of the knn estimates of entropy from wake to sleep in healthy subjects was statistically significant after normalization to unit variance [Fig. 22(g)], suggesting that the amplitude distribution of HRV is less skewed during sleep than wake.

B. Dynamic measures of complexity in HRV signals: Conditional entropy and information storage

Moving to the analysis of the measures of dynamical complexity of HRV, the first main finding is the significant increase of the conditional entropy and decrease of the information storage, observed moving from wake to sleep in healthy subjects. This behavior was consistently found for all three estimators and for both the original and the normalized time series (Figs. 22(b), 22(c), 22(e), 22(f), 22(h), and 22(i)] and confirms previous findings showing that HRV displays a higher short-term complexity during nighttime than during daytime, potentially due to the sympathetic withdrawal and parasympathetic enhancement commonly observed during sleep [21,120]. Interestingly, this increase of the HRV complexity during sleep was not observed anymore when conditional entropy and information storage were computed on the time series without slow trends obtained through the high-pass filtering. Thus, also according to our simulation results (see, e.g., Fig. 17), the higher conditional entropy and lower information storage during sleep than during wake for unfiltered signals with slow trends are likely to reflect a decrease of long-range correlations from wake to sleep, rather than alterations in the short-term dynamics. To support this hypothesis, we performed detrended fluctuation analysis on the whole 6-h time series measured for each subject in the wake and sleep conditions under the above mentioned preprocessing steps. Similar to previous work [52,53,66,67,69,127], we found that significant differences exist in the DFA exponents between wake and sleep as well as between healthy and CHF subjects for original time series, indicating stronger long range correlations during wake than during sleep, for CHF subjects than healthy subjects. Moreover, after normalizing and filtering the data, the DFA exponents decrease for all signals and the differences in DFA exponents across physiological states and clinical conditions are diminished, indicating the removal of long-range correlations due to the preprocessing procedure of normalization and filtering.

The second main result is the higher complexity in the short-term dynamics of HRV signals displayed by the CHF patients compared with the healthy subjects during both wake and sleep, which is consistent with previous findings [128]. This finding is documented by the significantly higher conditional entropy [Figs. 22(b), 22(e), and 22(h)] and lower information storage (Figs. 22(c), 22(f), and 22(i)] measured in CHF patients than healthy subjects in most cases of physiological condition and for most preprocessing procedures. This observation is consistent for all estimators applied to the preprocessed signals after normalization and detrending, suggesting that the short-term dynamics of HRV play a crucial role in determining the difference in the cardiac dynamics between

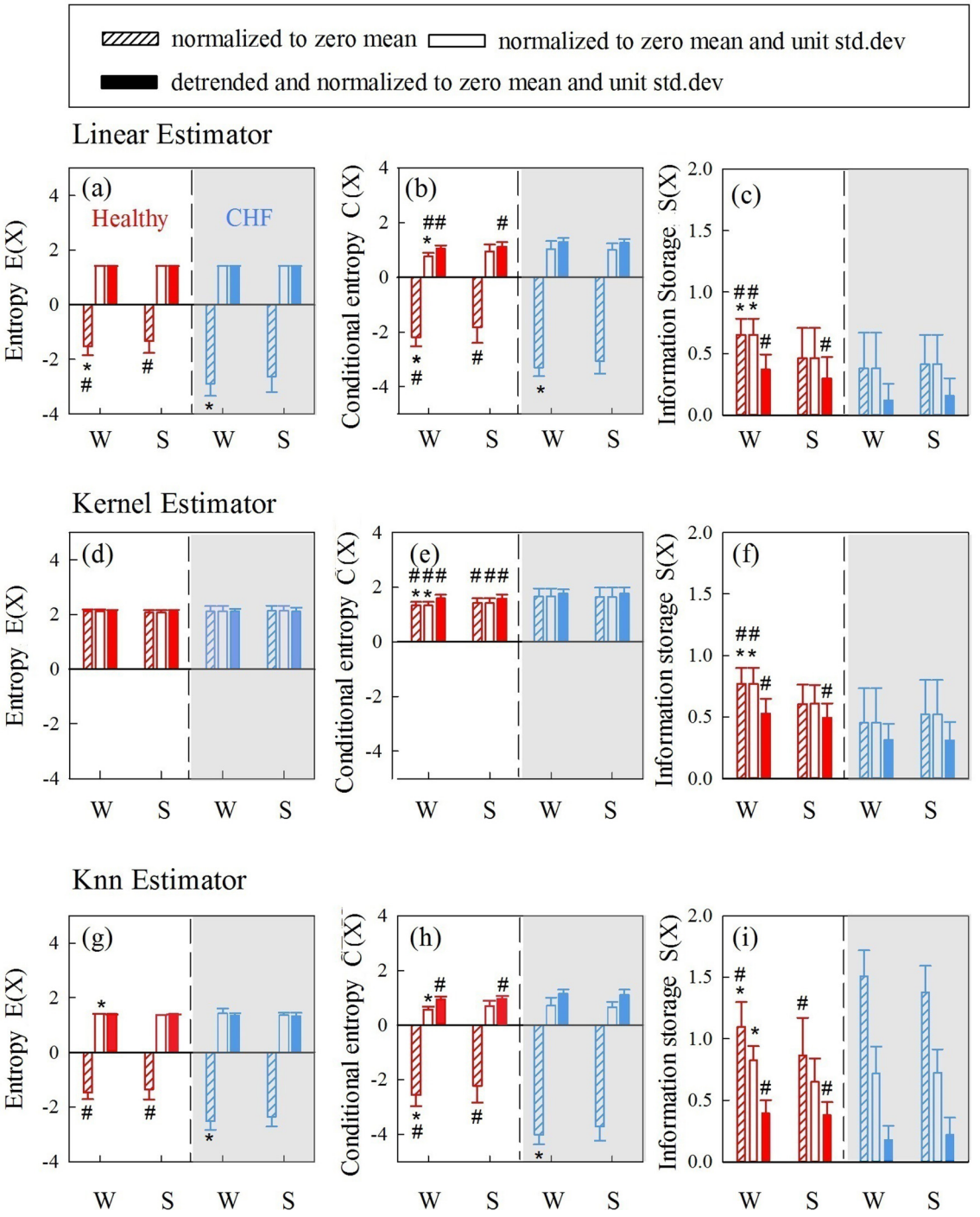

FIG. 22. Performance of entropy estimators and entropy measures for heart rate variability signals of healthy subjects and congestive heart failure (CHF) patients measured during wake (W) and sleep (S) states. Panels contain the behavior of entropy (a), (d), (g), conditional entropy (b), (e), (h), and information storage (c), (f), (i) computed for signals normalized to zero mean within each window (shaded bar), signals normalized to zero mean and unit standard deviation within each window (open bar), and signals detrended by a linear high-pass filter and normalized to zero mean and unit standard deviation within each window (solid bar) for healthy subjects (left side with white background and red bars) and CHF subjects (right side with gray background and blue bars) during wake (W) and sleep (S). Barplots depict the mean + standard deviation across subjects of the median value of entropy measures computed for consecutive windows of 300 data points with a 150 points overlap using $m = 2$ lagged components and implemented with the linear estimator (a)–(c), the kernel estimator with threshold $r = 0.2$ (d)–(f), and the knn estimator with $k = 10$ neighbors (g)–(i). Symbols denote statistical significance ($p < 0.05$) of the differences between W and S ($*$, paired t-test) or between healthy and CHF (#, unpaired t-test). Results: Compared to healthy subjects, CHF patients exhibit lower entropy values reflecting depressed HRV, and higher conditional entropy and lower information storage reflecting higher dynamical complexity. Healthy subjects display a day-to-night increase of the dynamical complexity (higher conditional entropy and lower information storage during S than during W) that is not observed in CHF patients.

CHF patients and healthy subjects, as was also reported in previous studies of the short-term dynamics of HRV time series [120].

On the other hand, the results for signals without slow trends removal evidence nonunivocal behaviors of the entropy measures and entropy estimators. This suggests the impact of normalization and the role of trends on the short-term analysis of HRV complexity, which complicates the interpretation of results. Using the linear estimator, the conditional entropy computed for the original zero-mean signals is lower in CHF patients than healthy subjects, while it is higher if computed on signals normalized to unit variance [Fig. 22(b)]. Nevertheless, such a discrepancy in the comparison between the healthy and CHF groups before and after normalization to unit variance was not observed for the linear estimation of information storage. This result documents the importance of normalization to unit variance, or of using the information storage as measure of dynamical complexity, in situations where the signal variance changes substantially across conditions. Using the kernel estimator, results for conditional entropy were independent of the pre-processing [Figs. 22(e) and 22(f)], confirming that this estimator is less sensitive to alterations in the signal variance across conditions. Using the knn estimator, results were highly dependent on the preprocessing, with CHF patients exhibiting lower conditional entropy and higher storage than healthy subjects for the original HRV series, no significant differences for the normalized series, and higher conditional entropy and lower storage for the detrended series (Figs. 22(h) and 22(i)]; besides the effects of changes in the signal variance, these different trends may be also ascribed to changes in the shape of the probability distribution of the HRV time series related to different impact of trends and long-range correlations in the CHF group. Indeed, the effect of stronger long-range correlations documented for CHF patients [67,126,129] may explain the lower dynamical complexity that was associated with heart failure in previous studies [128] and is here documented by the low conditional entropy and high storage measured for the original HRV series.

To summarize, in this section we reported a paradigmatic application to show on real signals the performance of entropy measures and entropy estimators in a field in which a big volume of work was performed using these approaches. The analysis of physiological time series reported in this Section indicates that entropy measures can only reveal specific types of dynamical features of real-world complex systems, which are also dependent on the choice of entropy estimators and preprocessing procedure. To achieve correct estimation and meaningful interpretations, careful assessment of entropy estimators and appropriate signal preprocessing have to be carried out. Specifically, our results document the usefulness of entropy measures to assess the overall signal variance, and also the importance of normalization in order to detect more cleanly alterations of the dynamical structure across states or conditions. They document also the big impact of long-range correlations on the values of entropy measures, which makes it important to remove slow trends in computing conditional entropy and information storage when the purpose is to use these measures to characterize the short-term dynamical properties of the observed system.

V. SUMMARY AND CONCLUSIONS

There is a large volume of studies in the literature where various entropy measures with different entropy estimators are applied to diverse dynamical systems across the fields of physics, biology, engineering, medicine, and economics. These studies consider a range of experimental conditions with different types of data artifacts and data limitations. It is a challenge to compare results obtained for different entropy measures applied to different systems, and to deduce information about the intrinsic complexity and underlying mechanisms. Thus, it is of paramount importance to assess the performance of entropy measures for different types of dynamics, often in the presence of nonstationarity and artifacts, and to be aware of how estimated values of these measures are affected by the choice of estimator-specific parameters. This paper provides a detailed recipe for the application of the most widely used entropy measures and entropy estimators on the most general dynamic processes encountered in physical and biological systems, and is a first account of the biases and limitations of entropy methods in presence of nonstationarities and data artifacts.

In this paper, we investigate the theoretical behavior of entropy measures, as well as the performance of entropy estimators, for various types of dynamical processes encountered in real-world systems. Specifically, we consider the measures of entropy, conditional entropy and information storage, computed by means of linear parametric estimators, and nonlinear nonparametric estimators, such as the kernel and nearest-neighbor methods.

We first define a set of models to generate stochastic processes. We obtain theoretically the true values of the entropy measures through analytical derivations where we incorporate the known parameter values of the models used to simulate stochastic processes. We next compare the true theoretical values of the entropy measures with the numerically estimated values of the same measures obtained from the generated time series.

Our investigations include the following dynamic processes: (i) stationary AR processes; (ii) nonstationary AR processes corrupted by sinusoidal trends, random spikes, and local changes in variance; (iii) fractionally integrated processes with long-range power-law correlations; (iv) AR processes combined with long-range correlations. We also apply linear, kernel, and nearest-neighbor estimators of entropy, conditional entropy, and information storage to physiological signals of consecutive heartbeat intervals recorded in different populations (healthy and congestive heart failure) and during different physiological states (wake and sleep).

Our major findings and observations are listed in the following points:

(i) We find that even for a process as simple as the stationary AR process, both the theoretical interpretation and the practical estimation of entropy measures are not always straightforward (Figs. 8 and 10). The dynamical complexity of the AR process, reflected by high values of conditional entropy and low values of information storage, varies not only with the AR amplitude ρ, a parameter which controls the predictability of the process, but also with the AR frequency parameter f, which defines the frequency of the stochastic

oscillations. Moreover, the accuracy in the estimation of entropy measures by all estimators is higher for longer time series [Figs. 8(d)–8(f)]. We find that the kernel and knn estimates of all entropy measures are biased and provide low accuracy for processes with high AR amplitude and very low or very high AR frequency. This bias is significantly reduced for AR processes with low and intermediate values of the AR amplitude in combination with intermediate values of AR frequency parameter. In contrast, we find that the accuracy of the linear estimator does not depend on the AR parameters [Figs. 8(g)–8(l)].

(ii) We also find that the estimation results of nonparametric estimators are affected by estimator-specific parameters (Fig. 9). For the knn estimator, changes in the number of neighbors k only have negligible impacts whereas the kernel estimates of entropy and conditional entropy can be strongly biased and highly variant when the threshold r (the width of Heaviside kernel function) varies. The sensitive dependence of the kernel estimator on its model parameter is not specific to any particular kernel function but results from the general approach adopted by the kernel methods to partition the state space. Despite such a defect, the kernel estimates of the conditional entropy are ubiquitously employed by the approximate entropy (ApEn) and sample entropy (SampEn) measures to assess the dynamical complexity of time series in a wide range of empirical studies. Although it is commonplace to consider a range of values for the threshold parameter r as appropriate for the computation of ApEn or SampEn (typically r is chosen between 0.1 and 0.3 times the standard deviation of the observed signal), it is noteworthy that kernel estimates of the conditional entropy can vary with the threshold r to an extent that can easily exceed any difference between the intrinsic complexity of the studied dynamics [Figs. 9(d)–9(f)].

(iii) The effects of nonstationarities due to various data artifacts on the estimation of entropy measures vary with the type of nonstationary behavior and with the type of the estimator that is used. Sinusoidal trends are detrimental to all estimators, reducing dramatically the capability of conditional entropy and information storage to reflect changes in the dynamical complexity of these processes (Figs. 11 and 12). Spikes impair the performance of the linear model-based estimates of dynamical complexity, while they affect less model-free methods, such as the kernel and nearest-neighbor estimators (Figs. 13 and 14). Local changes in signal variance appear to be less problematic, as they do not compromise the ability of all estimators to detect changes in dynamical complexity despite introducing a bias (Figs. 15 and 16). In addition to these systematical studies on the effects of nonstationarities, future studies are foreseeable, which assess comparatively the performance of different nonparametric entropy estimators in describing the complexity of signals exhibiting nonlinear and/or chaotic dynamics.

(iv) For processes with power-law long-range correlations (Fig. 17), we establish the theoretical dependence of the entropy measures on the correlation strength using the analytical derivations of the true values of entropy measures based on given parameters of the stochastic processes presented in the Appendix. We observe lower theoretical values of conditional entropy and higher values of information storage when increasing the strength of positive correlations or anticorrelations, while the theoretical values of entropy remain unchanged as it depends only the variance and is independent of the correlations in the signal. Moreover, we find that when the strength of long-range correlations is the same, the conditional entropy is higher and the information storage is lower for signals with anticorrelations than for signals with positive correlations. Such theoretical properties of conditional entropy and information storage are approximated fairly well by all estimators. The estimation bias is related to not only data length but also the sign and strength of the long-range correlations. In specific, the estimations by all estimators are less accurate when the data length is shorter and the absolute correlation strength is higher. In addition, for data with the same absolute correlation strength, the entropy estimates are more biased for signals with positive correlations than for signals with negative correlations.

(v) For processes with both AR dynamics and long-range power-law correlations (Fig. 19), we study both the theoretical behaviors of the entropy measures and the corresponding estimates by all three estimators. We show that the combined effect of autoregression and long-range correlations complicates the interpretation of conditional entropy and information storage already at the theoretical level. Specifically, for anticorrelated signals, we find that the response of conditional entropy and information storage to changes in the correlation strength or in the AR amplitude is preserved compared with the cases of pure correlated noise or pure AR process: the conditional entropy decreases and the information storage increases with stronger short-term dependence due to autoregression and with stronger long-range correlations (either positive or anticorrelations). On the contrary, for signals with positive long-range correlations, such a response may be inverted: when increasing the strength of short-term dependence due to autoregression in the presence of strong positive long-range correlations or when increasing the strength of positive long-range correlations in the presence of strong short-term dependence due to autoregression, the conditional entropy increases and the information storage decreases. As regards the estimation accuracy of entropy estimators we find that, with relatively short data length common of real-world time series, all estimators display a non-negligible bias, which is more pronounced when long-range correlations are positive and/or strong.

(vi) The practical analysis of the HRV series (Fig. 22) documents that, when properly applied, entropy measures are able to characterize changes of specific types in the cardiac system that are associated with different physiological and clinical states. However, a correct interpretation of the behavior of entropy measures for varying conditions requires clear understandings of the properties of the specific chosen measure and adopted estimator, and proper choice of preprocessing applied to the measured signals. In the reported application, (1) we observe that the linear estimate of entropy for data not normalized to unit variance successfully detected the wake-to-sleep increase of HRV and the depressed HRV of heart failure patients (smaller standard deviation). In contrast, such difference in the estimated values of entropy across physiological states and clinical conditions is not observed when other estimators are adopted or when the analyzed signals are normalized. This finding indicates that Entropy is a useful measure in characterizing the variability in the data

value, which is preferable to be computed on the original time series prior to any normalization; (2) for healthy subjects, we find that the conditional entropy is lower and the information storage is higher during wake than sleep (indicating higher dynamical complexity) when their estimated values are obtained by all three estimators both for the original and for the normalized signals. Such changes of these two entropy measures are lost when slow trends are removed from the analyzed signals through high-pass filtering. In addition, by comparing the results of detrended fluctuation analysis for the same data set before and after filtering as well as during wake and during sleep, we find that both the preprocessing procedure of detrending and the switch of physiological state from wake to sleep leads to decrease of long-range correlations, which indicates that the observed changes in conditional entropy and information storage may result from the variations of trending behaviors in the signals due to long-range correlations; (3) during both wake and sleep, we find that for normalized and detrended signals the estimates of conditional entropy are lower and the estimates of information storage are higher by all estimators for healthy subjects than CHF patients [black bars in Figs. 22(c), 22(f), 22(i)], indicating higher predictability in the short-term dynamics of HRV recordings in healthy subjects compared to CHF patients. For normalized signals without detrending, the results are inconsistent across entropy estimators and entropy measures when comparing clinical conditions and physiological states. When entropy methods are directly applied to original HRV signals, there are various factors present in the data, such as different signal variance, trends, or long-range correlations, which affect differently the entropy measures and estimators, and may thus lead to inconsistent results and impairs interpretation.

In conclusion, this paper provides a systematic overview of the entropy-based approaches to the quantification of the complexity of time series measured from dynamical systems. Entropy measures and estimators used in this paper represent or directly relate to very popular measures of complexity such as approximate entropy (ApEn), sample entropy (SampEn), multiscale entropy and permutation entropy, which are utilized in thousands of publications in all possible fields. We demonstrate that it is a challenging task to choose an entropy measure that adequately quantifies the target dynamical process and to provide a correct estimate of this measure from real-life time series.

Based on the summarized findings above, we give the following recommendations for the practical application of the discussed entropy methods:

(1) Entropy reflects the static properties of the investigated process, describing its amplitude distribution; it should be computed on original, nonnormalized time series, as this measure is related to the variance of the signal.

(2) Conditional entropy and information storage are complementary measures of the dynamical structure of the process, reflecting respectively its complexity and regularity intended in terms of predictability of the present given the past.

(3) Information storage should be preferred to conditional entropy, as it is less dependent on the signal variance and in general its estimated values are less biased for all entropy estimators.

(4) Linear estimates of the entropy measures are the most appropriate for Gaussian processes; this property is lost for nonstationary and nonlinear dynamics, which should be studied employing nonparametric and model-free estimators.

(5) Among model-free approaches, the knn estimator outperforms the kernel estimator in terms of bias and robustness for short time series. The kernel estimator, despite being extremely popular for the computations of approximate entropy and sample entropy computation, is highly biased with strong dependence on the threshold parameter r. Thus, we recommend to compute the conditional entropy using the knn estimator rather than the kernel estimator. If the kernel estimates of conditional entropy are computed (as in the extensive literature based on approximate entropy and sample entropy), we advise against the utilization of different values of the threshold r when comparing different experimental conditions.

(6) Entropy measures are affected in a different way by different types of artifacts and nonstationarities in the time series: slow trends are the most detrimental for all entropy estimators, and should be removed in preprocessing; spikes impair the linear estimation but have less impact on the kernel and the nearest neighbor estimators; in comparison, local changes in the variance of the time series lead to less bias for all three estimators considered in this study.

(7) In order to ensure that the variations in conditional entropy and information storage purely reflect changes in the dynamical properties of the underlying process (e.g., autoregression or long-range correlations), these measures should be computed after the normalization of the time series to zero mean and unit variance.

(8) In order to ensure that the variations in conditional entropy and information storage purely reflect short-term dynamical properties of the process (and not due to long-range correlations), conditional entropy and information storage should be computed after removing the slow trends in the time series through a high-pass filter with an appropriate cutoff frequency.

(9) The computation of conditional entropy and information storage on the original time series with intrinsic trends may reveal alterations of the long-range correlation properties across conditions, with a sensitivity that increases with the length of the analyzed time series.

The comprehensive evaluation of entropy measures and entropy estimators provided here can be used as a reference guide to compare and interpret results of existing studies. This systematic investigation of the performance of entropy measures and entropy estimators and their bias when applied to real-life time series from diverse systems with complex dynamics, nonstationarities, and artifacts can serve as a primer for researchers who apply entropy methods.

ACKNOWLEDGMENTS

We acknowledge support from the Fundamental Research Funds for the Central Universities (Grants No. 2014KJJCB29 and No. 2015KJJCA06), W. M. Keck Foundation, National Institutes of Health (NIH Grant No. 1R01-HL098437), and the Office of Naval Research (ONR Grant No. 000141010078).

APPENDIX: THEORETICAL COMPUTATION OF ENTROPY MEASURES FOR AUTOREGRESSIVE FRACTIONALLY INTEGRATED PROCESSES

The practical computation of the entropy measures presupposes to provide estimates of the entropy and conditional entropy for vector variables. In the most general case, and when nonlinear effects are relevant, nonparametric approaches are recommended to yield model-free estimates. In the case of Gaussian processes, exact computation can be performed according to the approach proposed in Ref. [13] for pure autoregressive processes, which is here extended to the more general case of fractionally integrated autoregressive processes with Gaussian distribution. In such a case, exact values of entropy, conditional entropy, and information storage are those obtained by Eqs. (10), (12), and (13), showing that these measures can be derived from the variance of the zero-mean process X, $\sigma_X^2 = \mathbb{E}[X_n^2]$, and from the partial variance of the process given its past, $\sigma_U^2 = \mathbb{E}[U_n^2]$, where U is the residual of a linear regression of the present of X on its past values [Eq. (11)]. A known result [96] is that for Gaussian variables the partial variance of X_n given $X_n^l = [X_{n-1} \cdots X_{n-l}]$ (l is the number of points used to approximate the past of the process) can be expressed in terms of covariance matrices as

$$\sigma\left(X_n|X_n^l\right) = \sigma_U^2 = \sigma_X^2 - \Sigma\left(X_n; X_n^l\right)\Sigma\left(X_n^l\right)^{-1}\Sigma\left(X_n; X_n^l\right)^T, \tag{A1}$$

with $\Sigma(\cdot)$ and $\Sigma(\cdot;\cdot)$ indicating, respectively, covariance and cross-covariance matrix. Thus, the computation of entropy measures amounts to calculate the terms in (A1) and use them in the definitions given by Eqs. (10), (12), and (13). In order to determine the subtrahend of Eq. (A1), we have to compute the autocovariance of the process X, which is defined as $R_k = \mathbb{E}[X_n X_{n-k}]$ for any time lag $k \geqslant 0$. In the following, we describe the procedure to derive the autocovariance of autoregressive fractionally integrated (ARFI) processes from the parametric representation of these processes.

The representation of an ARFI process is given by Eq. (30), from which the polynomial part can be rewritten as

$$A(L)(1-L)^d = \left(1 + \sum_{k=1}^{m} A_k L^k\right)\left(\sum_{k=0}^{\infty} G_k L^k\right), \tag{A2}$$

where $G_k = \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}$ (note that $G_0 = 1$). Thus, the ARFI process can be approximated as a finite order AR representation by truncating the fractional integration part at a given (arbitrarily high) lag q and solving the polynomial multiplication of Eq. (A2). This leads to representing the ARFI process as an AR process of order $p = m + q$:

$$X_n = \sum_{k=1}^{p} B_k X_{n-k} + U_n, \tag{A3}$$

where the coefficients B_k results from the polynomial multiplication. In the simulations treated in this paper where $m = 2$ [Eq. (27)], the coefficients become

$$B_1 = A_1 - G_1, \tag{A4}$$

$$B_k = A_2 G_{k-2} + A_1 G_{k-1} - G_k, \quad \forall k \geqslant 2, \tag{A5}$$

$$B_{q+1} = A_2 G_{q-1} + A_1 G_q, \tag{A6}$$

$$B_{q+2} = A_2 G_q. \tag{A7}$$

Then, we recall that the autocovariance of the process Eq. (A3) is related to the AR parameters B_k via the well-known Yule-Walker equations:

$$R_k = \sum_{l=1}^{p} B_l R_{k-l} + \delta_{k0}\sigma_U^2, \tag{A8}$$

where δ_{k0} is the Kronecker product. In order to solve Eq. (A8) for $R_k, k = 0, 1, \ldots, p-1$, we first express Eq. (A3) in a compact form as $\boldsymbol{\Phi}_n = \boldsymbol{A}\boldsymbol{\Phi}_{n-1} + \boldsymbol{E}_n$, where

$$\boldsymbol{\Phi}_n = [X_n X_{n-1} \cdots X_{n-p+1}]^T, \tag{A9}$$

$$\boldsymbol{A} = \begin{bmatrix} A_1 & \cdots & A_{p-1} & A_p \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 1 & \cdots & 0 & 0 \end{bmatrix}, \tag{A10}$$

$$\boldsymbol{E}_n = \left[\sigma_U^2 0 \cdots 0\right]^T. \tag{A11}$$

Then, the covariance matrix of $\boldsymbol{\Phi}_n$, $\boldsymbol{\Psi}$, takes the following form:

$$\boldsymbol{\Psi} = \mathbb{E}\left[\boldsymbol{\Phi}_n \boldsymbol{\Phi}_n^T\right] = \begin{bmatrix} R_0 & R_1 & \cdots & R_{p-1} \\ R_1 & R_0 & \cdots & R_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ R_{p-1} & R_{p-2} & \cdots & R_0 \end{bmatrix}. \tag{A12}$$

Since $\boldsymbol{\Psi}$ can be also expressed as a discrete-time Lyapunov equation, $\boldsymbol{\Psi} = \boldsymbol{A}\boldsymbol{\Psi}\boldsymbol{A}^T + \boldsymbol{\Xi}$, where $\boldsymbol{\Xi} = \mathbb{E}[\boldsymbol{E}_n \boldsymbol{E}_n^T]$ is the covariance of $\boldsymbol{E}_n$, we can solve for $\boldsymbol{\Psi}$ and obtain the autocovariance values $R_0 = \sigma_X^2$ and $R_1, \ldots, R_{p-1}$. Afterwards, by repeatedly applying Eq. (A8), the autocovariance R_k can be calculated recursively for any lag $k \geqslant 0$. This shows how the autocovariance sequence can be computed up to arbitrarily high lags starting from the parameters of ARFI representation of the observed Gaussian process. The autocovariances can then be used as elements in the covariance matrices in Eq. (A1) to obtain the partial variance of X_n given X_n^l.

The parameters determining the accuracy of the procedure are the number of lags q used to truncate the AR representation of the ARFI process, and the number of lags l used to approximate the past history of the process for the calculation of conditional entropy and information storage. In fact, considering the past up to lag l corresponds to calculating the autocovariance of the process Eq. (A3) up to the element R_l. As a rule of thumb, given that for a pure AR process the autocovariance decays exponentially with the lag, with a rate of decay depending on the modulus of the largest eigenvalue of $\boldsymbol{A}$, $\rho(\boldsymbol{A})$, it has been suggested to compute the autocovariance up to a lag l such that $\rho(\boldsymbol{A})^l$ is smaller than a predefined numerical tolerance [130]. This approximation should hold also for a stationary ARFI process when the lag q is chosen sufficiently high to detect the decay over time of the coefficients G_k in Eq. (A2). In this study, we set $q = 100$, observing that for this value both the coefficients G_k and the autocorrelation R_k (which was computed up to $k = p = q + m = 102$) decayed to very low values.

[1] S. M. Pincus, Proc. Natl. Acad. Sci. USA **88**, 2297 (1991).

[2] J. S. Richman and J. R. Moorman, Amer. J. Physiol.-Heart Circ. Physiol. **278**, H2039 (2000).

[3] A. Porta, G. Baselli, D. Liberati, N. Montano, C. Cogliati, T. Gnecchi-Ruscone, A. Malliani, and S. Cerutti, Biol. Cybern. **78**, 71 (1998).

[4] W. Chen, Z. Wang, H. Xie, and W. Yu, IEEE Trans. Neural Syst. Rehab. Eng. **15**, 266 (2007).

[5] S. Truebner, I. Cygankiewicz, R. Schroeder, M. Baumert, M. Vallverdu, P. Caminal, R. Vazquez, A. Bayés de Luna, and A. Voss, Biomedizinische Technik **51**, 77 (2006).

[6] C. Bandt and B. Pompe, Phys. Rev. Lett. **88**, 174102 (2002).

[7] M. Riedl, A. Müller, and N. Wessel, Eur. Phys. J. Spec. Topics **222**, 249 (2013).

[8] P. Li, C. Liu, K. Li, D. Zheng, C. Liu, and Y. Hou, Med. Biol. Eng. Comput. **53**, 77 (2015).

[9] M. Costa, A. L. Goldberger, and C.-K. Peng, Phys. Rev. Lett. **89**, 068102 (2002).

[10] M. Baumert, M. Javorka, A. Seeck, R. Faber, P. Sanders, and A. Voss, Comput. Biol. Med. **42**, 347 (2012).

[11] L. Angelini, R. Maestri, D. Marinazzo, L. Nitti, M. Pellicoro, G. D. Pinna, S. Stramaglia, and S. A. Tupputi, Artific. Intel. Med. **41**, 237 (2007).

[12] M. Costa, A. L. Goldberger, and C.-K. Peng, Phys. Rev. E **71**, 021906 (2005).

[13] L. Faes, A. Porta, and G. Nollo, Entropy **17**, 277 (2015).

[14] C. Gómez, J. T. Lizier, M. Schaum, P. Wollstadt, C. Grützner, P. Uhlhaas, C. M. Freitag, S. Sabine, S. Bölte, R. Hornero *et al.*, Front. Neuroinformat. **8**, 91 (2014).

[15] P. Grassberger and I. Procaccia, Physica D: Nonlin. Phenom. **13**, 34 (1984).

[16] K. Briggs, Phys. Lett. A **151**, 27 (1990).

[17] J. D. Farmer and J. J. Sidorowich, Phys. Rev. Lett. **59**, 845 (1987).

[18] M. Casdagli, Physica D: Nonlin. Phenom. **35**, 335 (1989).

[19] J. Kurths, A. Voss, P. Saparin, A. Witt, H. Kleiner, and N. Wessel, Chaos: Interdisc. J. Nonlin. Sci. **5**, 88 (1995).

[20] S. Vikman, T. H. Mäkikallio, S. Yli-Mäyry, S. Pikkujämsä, A.-M. Koivisto, P. Reinikainen, K. J. Airaksinen, and H. V. Huikuri, Circulation **100**, 2079 (1999).

[21] D. E. Vigo, J. Dominguez, S. M. Guinjoan, M. Scaramal, E. Ruffa, J. Solernó, L. N. Siri, and D. P. Cardinali, Autonom. Neurosci. **154**, 84 (2010).

[22] A. Voss, R. Schroeder, A. Peters, and S. Perz, PloS One **10**, e0118308 (2015).

[23] N. Wessel, A. Voss, J. Kurths, A. Schirdewan, K. Hnatkova, and M. Malik, Med. Biol. Eng. Comput. **38**, 680 (2000).

[24] L. Faes and A. Porta, in *Directed Information Measures in Neuroscience* (Springer, Berlin, 2014), pp. 61–86.

[25] A. Porta, B. D. Maria, V. Bari, A. Marchi, and L. Faes, IEEE Trans. Biomed. Eng. **PP**, 1 (2016).

[26] A. Porta, L. Faes, A. Marchi, V. Bari, B. De Maria, S. Guzzetti, R. Colombo, and F. Raimondi, Front. Physiol. **6**, 301 (2015).

[27] L. Faes, A. Porta, G. Rossato, A. Adami, D. Tonon, A. Corica, and G. Nollo, Autonom. Neurosci. **178**, 76 (2013).

[28] R. Hornero, M. Aboy, D. Abásolo, J. McNames, and B. Goldstein, IEEE Trans. Biomed. Eng. **52**, 1671 (2005).

[29] R. Alcaraz and J. J. Rieta, Biomed. Sign. Process. Control **5**, 1 (2010).

[30] S. Pincus, Economet. Rev. **27**, 329 (2008).

[31] D. Allen, M. McAleer, and A. Singh, *An Entropy Based Analysis of the Relationship Between the DOW JONES Index and the TRNA Sentiment Series*, Tech. Rep. EI2016-21 (2016).

[32] J. P. Kaipust, J. M. Huisinga, M. Filipi, and N. Stergiou, Motor Control **16**, 229 (2012).

[33] A. H. Khandoker, M. Palaniswami, and R. K. Begg, J. Neuroeng. Rehab. **5**, 4 (2008).

[34] J. J. Sosnoff, M. D. Goldman, and R. W. Motl, Multiple Sclerosis **16**, 868 (2010).

[35] J. E. Deffeyes, R. T. Harbourne, W. A. Stuberg, and N. Stergiou, Infant Behav. Dev. **34**, 81 (2011).

[36] A. D. Georgoulis, C. Moraiti, S. Ristanis, and N. Stergiou, J. Clin.l Monitor. Comput. **20**, 11 (2006).

[37] Z. Shuangcheng, Z. Qiaofu, W. Shaohong, and D. Erfu, Int. J. Climatol. **26**, 2131 (2006).

[38] G. Balasis, R. V. Donner, S. M. Potirakis, J. Runge, C. Papadimitriou, I. A. Daglis, K. Eftaxias, and J. Kurths, Entropy **15**, 4844 (2013).

[39] J. T. Lizier, M. Prokopenko, and A. Y. Zomaya, Informat. Sci. **208**, 39 (2012).

[40] S. Dasgupta, F. Wörgötter, and P. Manoonpong, Evolv. Syst. **4**, 235 (2013).

[41] H.-B. Xie, J.-Y. Guo, and Y.-P. Zheng, Ann. Biomed. Eng. **38**, 1483 (2010).

[42] D. Abásolo, R. Hornero, P. Espino, D. Alvarez, and J. Poza, Physiol. Meas. **27**, 241 (2006).

[43] R. Ferenets, A. Vanluchene, T. Lipping, B. Heyse, and M. M. Struys, J. Amer. Soc. Anesthesiol. **106**, 696 (2007).

[44] Y. Bai, Z. Liang, X. Li, L. J. Voss, and J. W. Sleigh, Physiol. Meas. **36**, 2483 (2015).

[45] C. Gómez, J. Poza, M. García, A. Fernández, and R. Hornero, in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (IEEE, Osaka, Japan), pp. 1765–1768.

[46] M. O. Sokunbi, Front. Neuroinformat. **8**, 1 (2014).

[47] Z. Wang, Y. Li, A. R. Childress, and J. A. Detre, Plos One **9**, e89948 (2014).

[48] A. Neiman, B. Shulgin, V. Anishchenko, W. Ebeling, L. Schimansky-Geier, and J. Freund, Phys. Rev. Lett. **76**, 4299 (1996).

[49] P. A. Varotsos, N. V. Sarlis, E. S. Skordas, and M. S. Lazaridou, Phys. Rev. E **70**, 011106 (2004).

[50] P. A. Varotsos, N. V. Sarlis, E. S. Skordas, H. K. Tanaka, and M. S. Lazaridou, Phys. Rev. E **73**, 031114 (2006).

[51] P. A. Varotsos, N. V. Sarlis, H. K. Tanaka, and E. S. Skordas, Phys. Rev. E **71**, 032102 (2005).

[52] P. Ch. Ivanov, A. Bunde, L. N. Amaral, S. Havlin, J. Fritsch-Yelle, R. M. Baevsky, H. E. Stanley, and A. L. Goldberger, Europhys. Lett. **48**, 594 (1999).

[53] Y. Ashkenazy, P. Ch. Ivanov, S. Havlin, C.-K. Peng, A. L. Goldberger, and H. E. Stanley, Phys. Rev. Lett. **86**, 1900 (2001).

[54] P. Ch. Ivanov, K. Hu, M. F. Hilton, S. A. Shea, and H. E. Stanley, Proc. Natl. Acad. Sci. USA **104**, 20702 (2007).

[55] P. Ch. Ivanov, A. Yuen, B. Podobnik, and Y. Lee, Phys. Rev. E **69**, 056107 (2004).

[56] P. Ch. Ivanov, A. Yuen, and P. Perakakis, PloS One **9**, e92885 (2014).

[57] P. Ch. Ivanov, M. G. Rosenblum, C. Peng, J. Mietus, S. Havlin, H. Stanley, A. L. Goldberger *et al.*, Nature **383**, 323 (1996).

[58] P. Ch. Ivanov, L. A. N. Amaral, A. L. Goldberger, S. Havlin, M. G. Rosenblum, Z. R. Struzik, and H. E. Stanley, Nature **399**, 461 (1999).

[59] D. T. Schmitt and P. Ch. Ivanov, Amer. J. Physiol.-Regul., Integr. Compar. Physiol. **293**, R1923 (2007).

[60] P. Ch. Ivanov, L. A. N. Amaral, A. L. Goldberger, S. Havlin, M. G. Rosenblum, H. E. Stanley, and Z. R. Struzik, Chaos: Interdisc. J. Nonlin. Sci. **11**, 641 (2001).

[61] P. Ch. Ivanov, Q. D. Y. Ma, R. P. Bartsch, J. M. Hausdorff, L. A. Nunes Amaral, V. Schulte-Frohlinde, H. E. Stanley, and M. Yoneyama, Phys. Rev. E **79**, 041920 (2009).

[62] E. Goldberg, *The Executive Brain: Frontal Lobes and the Civilized Mind* (Oxford University Press, New York, 2002).

[63] C. Shannon, Mathematical Reviews (MathSciNet): MR10, 133e (1948).

[64] S. M. Pincus and A. L. Goldberger, Amer. J. Physiol.y-Heart Circul. Physiol. **266**, H1643 (1994).

[65] K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, and J. Bhattacharya, Phys. Rep. **441**, 1 (2007).

[66] P. Bernaola-Galván, P. Ch. Ivanov, L. A. Nunes Amaral, and H. E. Stanley, Phys. Rev. Lett. **87**, 168105 (2001).

[67] C.-C. Lo, T. Chou, T. Penzel, T. E. Scammell, R. E. Strecker, H. E. Stanley, and P. Ch. Ivanov, Proc. Natl. Acad. Sci. USA **101**, 17545 (2004).

[68] A. L. Goldberger, L. A. Amaral, J. M. Hausdorff, P. Ch. Ivanov, C.-K. Peng, and H. E. Stanley, Proc. Natl. Acad. Sci. USA **99**, 2466 (2002).

[69] P. Ch. Ivanov, IEEE Eng. Med. Biol. Mag. **26**, 33 (2007).

[70] R. Karasik, N. Sapir, Y. Ashkenazy, P. Ch. Ivanov, I. Dvir, P. Lavie, and S. Havlin, Phys. Rev. E **66**, 062902 (2002).

[71] Z. Chen, K. Hu, P. Carpena, P. Bernaola-Galvan, H. E. Stanley, and P. Ch. Ivanov, Phys. Rev. E **71**, 011104 (2005).

[72] Z. Chen, P. Ch. Ivanov, K. Hu, and H. E. Stanley, Phys. Rev. E **65**, 041107 (2002).

[73] K. Hu, P. Ch. Ivanov, Z. Chen, P. Carpena, and H. E. Stanley, Phys. Rev. E **64**, 011114 (2001).

[74] L. Xu, P. Ch. Ivanov, K. Hu, Z. Chen, A. Carbone, and H. E. Stanley, Phys. Rev. E **71**, 051101 (2005).

[75] K. Hlavácková-Schindler, Appl. Math. Sci. **5**, 3637 (2011).

[76] M. U. Ahmed and D. P. Mandic, Phys. Rev. E **84**, 061918 (2011).

[77] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).

[78] D. Chicharro and A. Ledberg, Phys. Rev. E **86**, 041901 (2012).

[79] A. Rényi, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1 (University of California Press, 1961), pp. 547–561.

[80] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).

[81] O. Kafri and H. Kafri, *Entropy: God's Dice Game* (CreateSpace, Seatle, USA, 2013).

[82] K. Martinás, World Futures: J. Gen. Evol. **50**, 483 (1997).

[83] A. N. Kolmogorov, Dokl. Akad. Nauk SSSR **124**, 754 (1959).

[84] Y. G. Sinai, Dokl. Akad. Nauk. SSSR **124**, 768 (1959).

[85] W. Ebeling, World Futures: J. Gen. Evol. **50**, 467 (1997).

[86] J.-P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).

[87] K. Keller, A. M. Unakafov, and V. A. Unakafova, Entropy **16**, 6212 (2014).

[88] Y. Cao, L. Cai, J. Wang, R. Wang, H. Yu, Y. Cao, and J. Liu, Chaos: Interdisc. J. Nonlin. Sci. **25**, 083116 (2015).

[89] L. Faes, G. Nollo, F. Jurysta, and D. Marinazzo, New J. Phys. **16**, 105005 (2014).

[90] C. Gómez, J. T. Lizier, M. Schaum, P. Wollstadt, C. Grützner, P. Uhlhaas, C. M. Freitag, S. Schlitt, S. Bölte, R. Hornero *et al.*, Front. Neuroinformat. **8**, 89 (2014).

[91] M. Wibral, B. Rahm, M. Rieder, M. Lindner, R. Vicente, and J. Kaiser, Progr. Biophys. Mol. Biol. **105**, 80 (2011).

[92] X. R. Wang, J. M. Miller, J. T. Lizier, M. Prokopenko, and L. F. Rossi, PloS One **7**, e40084 (2012).

[93] M. Wibral, J. T. Lizier, and V. Priesemann, Front. Robot. AI **2**, 5 (2015).

[94] J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths, Phys. Rev. Lett. **108**, 258701 (2012).

[95] L. Faes, G. Nollo, and A. Porta, Phys. Rev. E **83**, 051112 (2011).

[96] L. Barnett, A. B. Barrett, and A. K. Seth, Phys. Rev. Lett. **103**, 238701 (2009).

[97] J. D. Victor, Phys. Rev. E **66**, 051903 (2002).

[98] S. Panzeri, R. Senatore, M. A. Montemurro, and R. S. Petersen, J. Neurophysiol. **98**, 1064 (2007).

[99] A. Kaiser and T. Schreiber, Physica D: Nonlin. Phenom. **166**, 43 (2002).

[100] L. Kozachenko and N. N. Leonenko, Prob. Peredachi Inf. **23**, 9 (1987).

[101] A. Kraskov, H. Stögbauer, and P. Grassberger, Phys. Rev. E **69**, 066138 (2004).

[102] A. Porta, T. Gnecchi-Ruscone, E. Tobaldini, S. Guzzetti, R. Furlan, and N. Montano, J. Appl. Physiol. **103**, 1143 (2007).

[103] C. K. Rhea, T. A. Silver, S. L. Hong, J. H. Ryu, B. E. Studenka, C. M. Hughes, and J. M. Haddad, PloS One **6**, e17696 (2011).

[104] D. Kugiumtzis, Phys. Rev. E **87**, 062918 (2013).

[105] J. Zhu, J.-J. Bellanger, H. Shu, and R. Le Bouquin Jeannès, Entropy **17**, 4173 (2015).

[106] L. Faes, D. Kugiumtzis, G. Nollo, F. Jurysta, and D. Marinazzo, Phys. Rev. E **91**, 032904 (2015).

[107] E. Koscielny-Bunde, A. Bunde, S. Havlin, H. E. Roman, Y. Goldreich, and H.-J. Schellnhuber, Phys. Rev. Lett. **81**, 729 (1998).

[108] Y. Ogata and K. Abe, Int. Stat. Rev. **59**, 139 (1991).

[109] W. Willinger, M. S. Taqqu, and V. Teverovsky, Finance Stochast. **3**, 1 (1999).

[110] A. Bunde, S. Havlin, J. W. Kantelhardt, T. Penzel, J.-H. Peter, and K. Voigt, Phys. Rev. Lett. **85**, 3736 (2000).

[111] R. T. Baillie, J. Econometr. **73**, 5 (1996).

[112] C. W. J. Granger and R. Joyeux, J. Time Ser. Anal. **1**, 15 (1980).

[113] B. Podobnik, D. Fu, H. E. Stanley, and P. Ch. Ivanov, Eur. Phys. J. B **56**, 47 (2007).

[114] R. E. Kleiger, P. K. Stein, and J. T. Bigger, Ann. Noninvas. Electrocardiol. **10**, 88 (2005).

[115] G. G. Berntson, Psychophysiology **34**, 623 (1997).

[116] S. Akselrod, D. Gordon, F. A. Ubel, D. C. Shannon, A. Berger, R. J. Cohen *et al.*, Science **213**, 220 (1981).

[117] A. Malliani, M. Pagani, F. Lombardi, and S. Cerutti, Circulation **84**, 482 (1991).

[118] M. A. Cohen and J. A. Taylor, J. Physiol. **542**, 669 (2002).

[119] U. R. Acharya, N. Kannathal, and S. Krishnan, Physiol. Meas. **25**, 1139 (2004).

[120] A. Porta, L. Faes, M. Masé, G. D'Addio, G. Pinna, R. Maestri, N. Montano, R. Furlan, S. Guzzetti, G. Nollo *et al.*, Chaos: Interdisc. J. Nonlin. Sci. **17**, 015117 (2007).

[121] M. Kobayashi and T. Musha, IEEE Trans. Biomed. Eng. **BME-29**, 456 (1982).

[122] C.-K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger, Chaos: Interdisc. J. Nonlin. Sci. **5**, 82 (1995).

[123] M. Heart Failure Database (Beth Israel Deaconess Medical Center, Boston, The database now includes 18 healthy subjects (13 females and 5 males, with ages between 20 and 50, average 34.3 years), and 12 congestive heart failure subjects (3 females and 9 males, with ages between 22 and 71, average 60.8 year) in sinus rhythm.

[124] G. Nollo, L. Faes, B. Pellegrini, A. Porta, and R. Antolini, in *Computers in Cardiology 2000* (IEEE, Massachusetts, USA, 2000), pp. 143–146.

[125] P. Ponikowski, S. D. Anker, T. P. Chua, R. Szelemej, M. Piepoli, S. Adamopoulos, K. Webb-Peploe, D. Harrington, W. Banasiak, K. Wrabec *et al.*, Amer. J. Cardiol. **79**, 1645 (1997).

[126] S. Guzzetti, S. Mezzetti, R. Magatelli, A. Porta, G. De Angelis, G. Rovelli, and A. Malliani, Autonom. Neurosci. **86**, 114 (2000).

[127] J. W. Kantelhardt, Y. Ashkenazy, P. Ch. Ivanov, A. Bunde, S. Havlin, T. Penzel, J.-H. Peter, and H. E. Stanley, Phys. Rev. E **65**, 051908 (2002).

[128] A. Bianchi, M. Mendez, M. Ferrario, L. Ferini-Strambi, and S. Cerutti, Methods Info. Med. **49**, 479 (2010).

[129] G. C. Butler, S.-I. Ando, and J. S. Floras, Clin. Sci. **92**, 543 (1997).

[130] L. Barnett and A. K. Seth, J. Neurosci. Methods **223**, 50 (2014).