Probing high order dependencies with information theory

Carlos Granero-Belinchon, Stéphane G. Roux, Patrice Abry, Nicolas B. Garnier

Probing high order dependencies with information theory

C. Granero-Belinchón*, S.G. Roux*, P. Abry (IEEE Fellow)*, N.B. Garnier*
*Univ Lyon, Ens de Lyon, Univ Claude Bernard, CNRS, Laboratoire de Physique, F-69342 Lyon, France.

***Abstract*—Information theoretic measures (entropies, entropy rates, mutual information) are nowadays commonly used in statistical signal processing for real-world data analysis. The present work proposes the use of Auto Mutual Information (Mutual Information between subsets of the same signal) and entropy rate as powerful tools to assess refined dependencies of any order in signal temporal dynamics. Notably, it is shown how two-point Auto Mutual Information and entropy rate unveil information conveyed by higher order statistic and thus capture details of temporal dynamics that are overlooked by the (two-point) correlation function. Statistical performance of relevant estimators for Auto Mutual Information and entropy rate are studied numerically, by means of Monte Carlo simulations, as functions of sample size, dependence structures and hyper parameters that enter their definition. Further, it is shown how Auto Mutual Information permits to discriminate between several different non Gaussian processes, having exactly the same marginal distribution and covariance function. Assessing higher order statistics via multipoint Auto Mutual Information is also shown to unveil the global dependence structure fo these processes, indicating that one of the non Gaussian actually has temporal dynamics that ressembles that of a Gaussian process with same covariance while the other does not.**

***Index Terms*—Information theory, Entropy rates, Mutual Information, higher-order temporal dependencies, Non Gaussian processes.**

I. Introduction

Context. To characterize a (stationary) process $X(t)$, evolving along a given dimension t, *e.g.*, time, the precise assessment of both its static properties (marginal distribution or one-point statistics) and temporal dynamics (joint distributions or multi-point statistics) is crucial.

Correlation and spectral estimates, providing practitioners with valuable insight into temporal dynamics are key tools in statistics and data analysis and have been widely used in many different fields. In neurosciences, the analysis of correlation between neuron firings [1], [2] or between stimulus and brain activity [3], [4] is essential. In fluid mechanics, Kolmogorov's empirical theory of turbulence [5] characterizes the multiscale distribution of energy with the power spectrum of the velocity field [6]. The analysis of correlations also supports useful descriptions in terms of complex networks for, e.g., ecology [7], climatology [8] or traffic [9].

It is however well known that, while the correlation function, or equivalently the power spectral density, exhaust the description of joint statistics and hence of temporal dynamics for jointly Gaussian processes, they provide characterizations of temporal dynamics for general processes, that are limited in at least two respects: not only they are restricted to two-point joint statistics but also to a specific measure of these two point-statistics. In general, the full analysis of temporal dynamics require the use of multi-point, or higher order, joint statistics. The goal of the present work is to address such an issue using information theoretic measures.

Related work. To assess dependencies beyond two-point correlations, multipoint correlation functions were defined. For instance, three-point correlation functions have been used in cosmology, where they were related to cosmological parameters [10], [11]. Alternatively, higher order-spectra, that combine Fourier transforms at several different frequencies and therefore takes into account higher order statistics, have also been used to generalize the power spectrum beyond second-order statistics [12], [13]. For instance, bi-spectrum and tri-spectrum were analyzed in depth in e.g., [14], [15] and used in applications as different as cosmology [16], [17], or medicine [18], [19] and other fields [20], [21]. While of popular use in applications, higher-order correlation functions and spectra suffers from major caveats, mostly related to estimation issues. Indeed, the assessment of order m-statistics with higher-order correlation functions and spectra requires the estimation of m-time point statistics, which, for a given fixed sample size, can only yield decreasing estimation performance when order m is increased.

Alternatively, Shannon's information theory, and the corresponding information theoretic measures, entropy, entropy rate and mutual information, can be considered as an efficient way to characterize high order statistics [22], [23], [24]. Applications of entropy rate were yet reported in a wide range of domains, such as biology [25], dynamical systems [26], fluid turbulence [27] or analysis of languages [28]. Mutual information is also very popular, especially in computer sciences, e.g., in machine learning [29], and other very different domains [30], [31], [32], [33], [34], [35]. Auto-Mutual Information (AMI) — sometimes called Information Storage [36] —, defined as the mutual information between the analyzed process and a delayed version of itself, has been used in different domains [37], [38], [39], [40]. Information theory additionally provides tools to explore causality relationships between processes [41], [42]. However, within this paper we are mainly interested in the characterization of their dependence structures.

Entropy rate and AMI have been envisaged as tools to describe temporal dynamics in time series beyond second-

This work was supported by the LABEX iMUST (ANR-10-LABX-0064) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

order statistics, or even at second-order statistics, in a richer way than correlation or Fourier spectrum, see [43], [37], [39], [27].

Goals, contributions and outline. The main goals of the article are i) to show the ability of AMI and entropy rate to describe high order dependences, even when considering only two-point interactions. Even though this property of information theory tools is well known, we are able to show it by analyzing processes with identical correlation and marginal distribution but different high order dependences. AMI and entropy rate appear as straightforward generalizations of correlation function and power spectrum when considering non-Gaussian processes analysis. ii) to show the effect of analysing interactions between more than two points, combining Takens embedding [44] with AMI. We characterize the complexity of the dependence structure of a process by measuring AMI for embedding dimension higher than 2. More precisely, we analyze the differences between power law decay and exponential decay dependence structures.

To demonstrate the ability of information theory to describe high order statistics, we illustrate the use of entropy rate and auto-mutual information to characterize non Gaussian processes across scales. As an example, we consider two synthetic log-normal processes with different dependence structures but with identical correlation function and identical marginal statistics. In section II, we introduce and define the theoretical information theory measures and our estimators. In section III, we interpret the information theory tools used along the article, we present the processes with identical correlation function and one-point statistics but different dependence structure that we study and finally we illustrate how AMI and entropy rate allow to distinguish between them. In section IV-A, we report the convergence and the performance of our estimators of AMI and entropy rate and their ability to measure high order dependencies. We present their evolution across scales and emphasized that, contrary to classical correlation analysis, they allow a fine characterization of non Gaussian processes.

II. Entropy, entropy rate and auto mutual information

A. Definitions

To analyze the temporal dynamics of univariate stationary processes $X = \{x_t\}_{t\in\mathbb{R}}$, it has been proposed in [44] to rely on the time delay-embedding procedure. It amounts to constructing the m-dimensional process $X^{(m,\tau)}$, whose elements are m-dimensional vectors:

$$\mathbf{x}_t^{(m,\tau)} = \left(x_t,\, x_{t-\tau},\, \cdots,\, x_{t-(m-1)\tau}\right)\,, \tag{1}$$

where the delay τ implicitly defines an analysis scale and where the embbeding dimension m controls the order of the statistics of X which are analyzed. The information theoretic measures defined below aim to characterize temporal dynamics via the analysis of the m-time point joint statistical distributions $p(\mathbf{x}_t^{(m,\tau)})$.

Within this article, we study ergodic stationary processes and then $p(\mathbf{x}_t^{(m,\tau)})$ depend on m and τ but not on t. Consequently $p(\mathbf{x}_t^{(m,\tau)}) = p(X^{(m,\tau)})$.

Shannon entropy, is defined as a functional of $p(\mathbf{x}_t^{(m,\tau)})$ [22]:

$$\begin{aligned} H(\mathbf{x}_t^{(m,\tau)}) &= -\int_{\mathbb{R}^m} p(\mathbf{x}_t^{(m,\tau)}) \ln p(\mathbf{x}_t^{(m,\tau)}) \mathrm{d}\mathbf{x}_t^{(m,\tau)} \\ &= H^{(m,\tau)}(X) \end{aligned}$$

By nature, it depends on the full joint distribution and therefore on statistics of any order. However, for stationary processes it does not depend on the mean value of the marginal distribution. For $m = 1$, Shannon entropy does not relate to temporal dynamics but to one-point statistics and thus obviously does not depend on τ and is hence denoted $H(X)$.

Entropy rate of order m, $h^{(m,\tau)}$, quantify variations of Shannon entropy between two successive time-embedded versions of a stationary process [22], [24]:

$$\begin{aligned} h^{(m,\tau)}(x_t) &= H(\mathbf{x}_t^{(m+1,\tau)}) - H(\mathbf{x}_{t-\tau}^{(m,\tau)}) \\ &= H^{(m+1,\tau)}(X) - H^{(m,\tau)}(X) \\ &= h^{(m,\tau)}(X). \end{aligned} \tag{2}$$

Entropy rate $h^{(m,\tau)}$, like $H^{(m,\tau)}$, probes joint statistics of any order, but in addition quantifies how much information is gained in the analysis of temporal dynamics by increasing by one the order of the embedding.

Mutual information $I(\mathbf{x}_t^{(m,\tau)}, \mathbf{y}_{t'}^{(p,\tau)})$ quantifies the amount of information shared by two processes X and Y, via their embbeded versions $\mathbf{x}_t^{(m,\tau)}$ and $\mathbf{y}_{t'}^{(p,\tau)}$ [22], [24]:

$$\begin{aligned} I(\mathbf{x}_t^{(m,\tau)},\mathbf{y}_{t'}^{(p,\tau)}) = \int_{\mathbb{R}^{m+p}} & p\left(\mathbf{x}_t^{(m,\tau)}, \mathbf{y}_{t'}^{(p,\tau)}\right) \\ & \ln\left(\frac{p\left(\mathbf{x}_t^{(m,\tau)}, \mathbf{y}_{t'}^{(p,\tau)}\right)}{p(\mathbf{x}_t^{(m,\tau)})p(\mathbf{y}_{t'}^{(p,\tau)})}\right) \mathrm{d}\mathbf{y}_{t'}^{(p,\tau)}\mathrm{d}\mathbf{x}_t^{(m,\tau)} \end{aligned} \tag{3}$$

When X and Y are jointly stationary processes, $I(\mathbf{x}_t^{(m,\tau)},\mathbf{y}_{t'}^{(p,\tau)}) = I(X_t^{(m,\tau)}, Y_{t'}^{(p,\tau)})$ obviously depends on $t' - t$ only.

Mutual information can be read as the Kullback-Leibler divergence between the joint probability $p(\mathbf{x}_t^{(m,\tau)}, \mathbf{y}_{t'}^{(p,\tau)})$ and its counterpart under independence hypothesis $p(\mathbf{x}_t^{(m,\tau)})p(\mathbf{y}_{t'}^{(p,\tau)})$.

To analyze the higher statistical order temporal dynamics (also referred to as nonlinear temporal dynamics) of a single stationary process X, one can use auto-mutual information [24], [36], defined as the mutual information between $\mathbf{x}_t^{(m,\tau)}$ and $\mathbf{x}_{t+p\tau}^{(p,\tau)}$:

$$I^{(m,p,\tau)}(X) = I(\mathbf{x}_t^{(m,\tau)}, \mathbf{x}_{t+p\tau}^{(p,\tau)})\,. \tag{4}$$

Because the concatenation of $\mathbf{x}_{t+p\tau}^{(p,\tau)}$ and $\mathbf{x}_t^{(m,\tau)}$ corresponds exactly to the $(m+p)$-dimensional time-embedded vector $\mathbf{x}_{t+p\tau}^{(m+p)}$:

$$\left(\mathbf{x}_t^{(m,\tau)}, \mathbf{x}_{t+p\tau}^{(p,\tau)}\right) = \mathbf{x}_{t+p\tau}^{(m+p,\tau)},$$

$I^{(m,p,\tau)}(X)$ quantifies the shared information between the past m-point dynamics and the future p-point dynamics. Auto mutual information quantifies the deviation from independence between the two arguments, beyond the simpler decorrelation.

These information theoretic quantities all depend on the chosen time lag τ. When $\tau \to 0$, dependences between two

consecutive coordinates of the embedding vector increase, therefore inducing an increase in $I^{(m,1,\tau)}(X)$ and a decrease in $h^{(m,\tau)}(X)$. Conversely, when $\tau \to +\infty$, dependencies decrease to 0, therefore inducing an increase of $h^{(m,\tau)}(X)$ up to the Shannon entropy $H(X)$ of the process, its largest value, and a decrease of $I^{(m,1,\tau)}(X) \to 0$.

Furthermore, entropy rate can then be expressed as [27]:

$$h^{(m,\tau)}(X) = H(X) - I^{(m,1,\tau)}(X)\,, \tag{5}$$

which shows that there are two different contributions to the entropy rate: First, $H(X)$, which only depends on the one-point distribution and hence is a static property ; second, $I^{(m,1,\tau)}(X)$ which gathers all information conveyed by linear and non linear temporal dynamics, irrespective of the variance of the distribution [27].

B. Gaussian process

When X is a stationary jointly Gaussian process, hence fully defined by its variance σ_x^2 and normalized correlation function $c_X(\tau)$, analytical expressions for entropy, entropy rate and auto mutual information are available [45]:

$$H(X) = \frac{1}{2}\ln(2\pi e\sigma_x^2) \tag{6}$$

$$I^{(m,p,\tau)}(X) = \frac{1}{2}\ln\left(\frac{|\Sigma^{(m)}||\Sigma^{(p)}|}{|\Sigma^{(m+p)}|}\right) \tag{7}$$

$$h^{(m,\tau)}(X) = \frac{1}{2}\ln(2\pi e\sigma_x^2) - \frac{1}{2}\ln\left(\frac{|\Sigma^{(m)}|}{|\Sigma^{(m+1)}|}\right) \tag{8}$$

with $\Sigma^{(m)}$ the m-dimensional correlation matrix, and $||$ denoting the determinant. By definition, $\Sigma^{(1)} = 1$.

C. Estimation procedures

To estimate these information theoretic quantities, two main categories of non-parametric procedures are available, based on either kernel density estimations (KDE) or nearest neighbor search (NNS). While the former essentially counts the number $k(\epsilon)$ of the observed samples in interval of a priori set size ϵ, the latter rather measure the size $\epsilon(k)$ of box that contain an a priori set number of k samples.

Because this second strategy has been documented as providing low bias estimates [46], [47], we chose k-nearest neighbors procedures to estimate the entropy $H(X)$ [48], [49], [50]:

$$\widehat{H}(X) = -\psi(k) + \psi(N) + \frac{d}{N}\sum_{i}^{N}\log(\epsilon_i(k)) \tag{9}$$

with ψ the Digamma function, d the dimension of X, N the number of samples of X, and using the Chebyshev distance (L_∞ metric).

To compute the mutual information, we use the k-nearest neighbors procedure presented by Kraskov, et al. [46]:

$$\widehat{I}(X,Y) = \psi(k) + \psi(N) - \langle\psi(n_x) + \psi(n_y)\rangle \tag{10}$$

The procedure chooses k for the joint distribution $Z(X,Y)$ and hence obtains the distance ϵ to the kth nearest neighbour. Then using ϵ, it calculates the number of neighbours $n_x + 1$ lying on a segment $\mathbf{x} = x \pm \frac{\epsilon}{2}$ of the X dimension, *i.e.* lying on $-\infty \leq Y \leq \infty$, $x - \epsilon/2 \leq X \leq x + \epsilon/2$. Consequently, $\epsilon/2$ is the distance to the $n_x + 1$st neighbour of x. The procedure uses the same development for Y and obtains that $\epsilon/2$ is the distance to the $n_y + 1$st neighbour of y.

Because benefits against a direct use of eq.(2) were documented in [51], $h^{(m,\tau)}(X)$ is estimated plugging the above estimates into eq.(5).

For an stationary process X, $p(\mathbf{x}_t^{(m,\tau)})$ does not depend on t, and consequently, we can exploit this property to estimate the information theoretic quantities of X by averaging on all the temporal samples of the process.

To perform an analysis across scales by varying the parameter τ we use an adapted Theiler prescription [52] to eliminate the contribution of dependencies from scales smaller than τ.

III. Temporal dynamics and dependence structure

The aim of this section is to try to characterize what part of the statistics of a given stationary process Y contribute to the values taken by information theoretic measures H, $h^{(m,\tau)}$ and $I^{(m,p,\tau)}$. To that end, we conduct a formal study of the specific yet very instructive case where processes Y with correlation function $c_Y(\tau)$ and marginal distributions p_Y are obtained as point bijective transformation F, $Y = F(X)$ of an underlying Gaussian process X with correlation function c_X, designed to reach the targeted covariance c_Y. Processes Y constructed in this way depend on F and $c_X(\tau)$ [53], [54].

A. Analytical calculations and interpretations

Given any bijective transformation F that maps a Gaussian signal X into a signal $Y = F(X)$, the information theoretic quantities of Y can be computed as:

$$\begin{aligned} H(Y) &= H(X) + \langle\ln F\rangle_X \\ &= \frac{1}{2}\ln(2\pi e\sigma_X^2) + \langle\ln F\rangle_X \end{aligned} \tag{11}$$

$$\begin{aligned} I^{(m,p,\tau)}(Y) &= I^{(m,p,\tau)}(X) \\ &= \frac{1}{2}\ln\left(\frac{|\Sigma_X^{(m)}||\Sigma_X^{(p)}|}{|\Sigma_X^{(m+p)}|}\right) \end{aligned} \tag{12}$$

$$h^{(m,\tau)}(Y) = \frac{1}{2}\ln(2\pi e\sigma_X^2) + \langle\ln F\rangle_X - \frac{1}{2}\ln\left(\frac{|\Sigma_X^{(m)}|}{|\Sigma_X^{(m+1)}|}\right) \tag{13}$$

with $\langle\ln F\rangle_X = \int p(\mathbf{x}_t^{(1)})\ln F(x)dx$, hence a constant, σ_X^2 and $\Sigma_X^{(m)}$ are the variance and the m-dimensional covariance matrix of the underlying Gaussian process X.

These analytical calculations permit to yield understanding of general validity regarding the nature of the information conveyed by each of these three information theoretic quantities:
i) Entropy H is only related to the marginal distribution of the process p_Y, hence is quantifying static properties only.
ii) Entropy rate $h^{(m,\tau)}$ mixes up marginal distribution features (variance and pointwise transform F), hence static properties with dynamical properties and dependencies of any order, as indeed $\Sigma_X^{(m)}$ depends jointly on c_Y and on F, hence on the

joint distribution of $\mathbf{y}_t^{(m,\tau)}$. The fact that $h^{(m,\tau)}$ gather both static and dynamics aspects of time series likely explains why it has been observed in many applications to be a often discriminant feature, compared to other features [40]. However, the entanglement of static and dynamic properties may also be considered a drawback for the analysis, characterization and understanding of data properties.

iii) Interestingly, auto mutual information $I^{(m,p,\tau)}$ performs a characterization mainly related to the dependence structure of Y, as indeed $\Sigma_X^{(m)}$ depends on the joint distribution of $\mathbf{y}_t^{(m,\tau)}$. It therefore quantifies temporal dynamics, irrespective of static properties. Thus, AMI is the sole measure that focuses exclusively on higher order dynamics and on the joint dependence structure, beyond correlation.

Often in the scientific literature, $h^{(m,\tau)}$ and $I^{(m,p,\tau)}$ are referred to as *nonlinear* features, to emphasize that they characterize temporal dynamics beyond the mere use of the correlation functions or power spectral densities. This is indeed correct. Even when two-point statistics are analyzed, i.e., $m = p = 1$, $I^{(1,1,\tau)}$ reflects dependencies, including and encompassing correlations as the whole joint two-point distribution function is involved. Nevertheless, $I^{(m,p,\tau)}$ remains mostly driven by the correlation function c_Y, and hence by *linear* effects. This is also the case for $h^{(m,\tau)}$, driven mostly both by the static variance σ_Y^2 and dynamical correlation function c_Y.

B. Process definition

1) Analytical derivations: To further illustrate that indeed $I^{(m,p,\tau)}$ and $h^{(m,\tau)}$ may capture differences in any-order dependencies beyond correlations, let us study the cases where two stationary processes Y_{F_1} and Y_{F_2} have exact same marginal distribution p_Y and exact same covariance c_Y, but are yet obtained from two different point transformations F_1 and F_2 — applied to Gaussian processes X_1 and X_2 — and hence have different joint distributions and hence different dependence structures.

Let denote F_Y the cumulative distribution function associated to the marginal distribution p_Y. A natural candidate to obtain a process with marginal p_Y is to apply the point transform

$$F_1(x) = F_Y^{-1} \circ \Phi(x), \tag{14}$$

to a Gaussian process X_1, where F_Y is the cumulative distribution function of the marginal distribution p_Y and $\Phi(x)$ the cumulative distribution function of a centered standard Gaussian variable, and where X_1 is a jointly Gaussian process with covariance function $c_{X_1}(\tau)$ to reach the targeted $c_Y(\tau)$ [53], [54].

However, many other transforms can yield processes Y with the exact same marginal distribution p_Y and the exact same covariance c_Y. In [54], it was for example proposed to use

$$F_2(x) = F_Y^{-1} \circ (2(\Phi(|x|) - 1/2)) \ , \tag{15}$$

on a jointly Gaussian process X_2. This thus yields a radically different joint dependence structure for Y_{F_2}, despite having the same prescribed correlation function as Y_{F_1}. As a consequence, Y_{F_1} and Y_{F_2} have different AMI functions $I^{(m,p,\tau)}(Y_{F_1}) \neq I^{(m,p,\tau)}(Y_{F_2})$, and entropy rates, and hence different joint dependence structure and thus temporal dynamics. Conversely, their entropies are identical $H(Y_{F_1}) = H(Y_{F_2})$. This is not surprising , as the entropy is completely defined by the one-point distribution.

2) The log-normal example: To quantify the qualitative analysis proposed in Section III-A , let us consider stationary processes Y_{F_1} and Y_{F_2} having the same log-normal marginal distribution p_Y, chosen as an example of distributions that may complicate estimation because of its being a *frontier* between distributions with all moments finite and heavy-tail distributions.

For process Y_{F_1} obtained with the natural transform (14), F_1 simplifies to $F_1(x) = \exp(x)$, which is a bijective transformation. One can therefore use eqs.(11) to (13) above with $\langle \ln F_1 \rangle_{X_1} = \mu_{X_1}$, σ_{X_1} and $\Sigma_{X_1}^{(m)}$ being the mean, variance and m-dimensional covariance matrix of the Gaussian variable X_1 [45]. The entropy $H(Y_{F_1})$ can be computed using the mean μ_Y and variance σ_Y^2 of Y_{F_1}:

$$\mu_{X_1} = \ln\left(\frac{\mu_Y^2}{\sqrt{\sigma_Y^2 + \mu_Y^2}}\right) \ , \ \sigma_{X_1} = 2\ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right). \tag{16}$$

Writing $I^{(m,p,\tau)}(Y_{F_1})$ thus only requires to express the correlation function of the underlying Gaussian process X_1 as a function of the targeted $c_Y(\tau)$:

$$c_{X_1}(\tau) = \frac{1}{\sigma_{X_1}^2} \ln((e^{\sigma_{X_1}^2} - 1)c_Y(\tau) + 1). \tag{17}$$

This permits to compute $h^{(m,\tau)}$ and $I^{(m,p,\tau)}$ analytically, using eqs.(12) and (13).

For stationary log-normal process Y_{F_2}, F_2 is not bijective. The entropy H can be computed analytically, but this is not the case for $I^{(m,p,\tau)}$ and $h^{(m,\tau)}$.

3) Correlation structures: short-range versus long range dependencies: To study the impact of the correlation structure, processes with short-range dependencies are compared against processes with long-range dependencies, the former being well-represented by an exponentially decaying covariance,

$$c_Y(\tau) = \sigma_Y^2 e^{-f_c|\tau|}, \ \ f_c > 0, \tag{18}$$

and the latter by an algebraically decaying covariance,

$$c_Y(\tau) = \frac{\sigma_Y^2}{2}\left[(\tau-1)^{2\mathcal{H}} - 2\tau^{2\mathcal{H}} + (\tau+1)^{2\mathcal{H}}\right] \ , \tag{19}$$

with $0.5 < \mathcal{H} < 1$.

C. Two-point dependencies: Auto mutual information versus correlation

For both short-range and long range correlations, we compare the two LN processes obtained with F_1 and F_2, denoted Y_{F_1}, Y_{F_2}, against a Gaussian process with same correlation function, denoted $Y_{\mathbb{I}}$, where $\mathbb{I}$ stands for the identity transform. First order statistics (probability density function) and second order statistics measured by the correlation function or the power spectral density as well as two point statistics (*i.e.*, $m = p = 1$) entropy, entropy rate and AMI estimates are

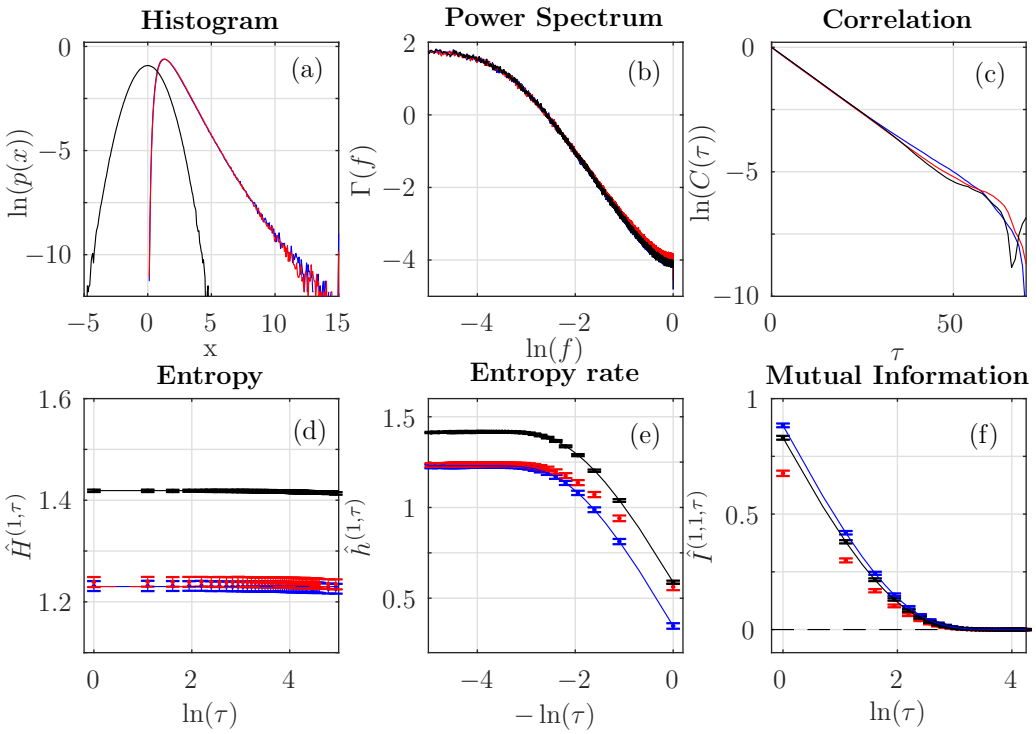

Fig. 1. **Short-range dependencies.** Classical analysis: histogram (a), power spectrum (b) and correlation function (c). Information theory analysis for $m = 1$, $p = 1$: entropy (d), entropy rate (e) and auto mutual information (f). The color indicates the process: blue for standard log-normal (Y_{F_1}), red for Y_{F_2} log-normal and black for Gaussian ($Y_{\mathbb{I}}$). The continuous lines in (e,d,f) correspond to analytical expressions. We used $N = 2^{16}$ and $k = 5$.

compared on Fig. 1 for $Y_{\mathbb{I}}$, Y_{F_1} and Y_{F_2} with short-range dependences and on Fig.2 for long-range dependences.

As expected, for each pair of short range and long range processes, the two log-normal processes, Y_{F_1} and Y_{F_2} cannot be distinguished one from the other using either PDF, correlation or power spectral estimation. Entropy being fully prescribed by the sole PDF, it is further not surprising that the two log-normal processes are also undistinguishable when using entropy estimates H (Fig (1,d)).

However, and interestingly, the estimates $I^{(1,1,\tau)}$ and $h^{(1,\tau)}$ of two-point AMI and entropy rate show differences as functions of τ between processes Y_{F_1} and Y_{F_2}, both for the short-range and long-range correlation cases. $I^{(1,1,\tau)}$ decreases and $h^{(1,\tau)}$ increases when $\tau \to \infty$, a predictable result as all dependencies vanish for large τ for stationary processes. Both functions (AMI and entropy rate) for processes Y_{F_1} and Y_{F_2} differ for small τ, thus quantifying the difference between the joint distributions of these two processes, despite identical correlation functions and marginal distributions.

Fig. 1 and Fig. 2 show that the two-point AMI estimates $\hat{I}^{(1,1,\tau)}$ for log-normal process Y_{F_1} ressembles far more to that of the Gaussian process $Y_{\mathbb{I}}$ than to that of the companion log-normal process Y_{F_2}, even they all three have the same correlation function. Very interestingly, it provides a deep insight into the temporal dynamics of the time series: processes Y_{F_1} and Y_{F_2} have same marginal distributions and same correlations functions, but the temporal dynamics of Y_{F_1} is close to that of the Gaussian process $Y_{\mathbb{I}}$ while that of Y_{F_2} is not. This also illustrates the benefits of AMI $I^{(1,1,\tau)}$ over entropy rate $h^{(1,\tau)}$ that mixes the impact of both marginal distribution and joint dynamical structure, thus making Y_{F_2} comparable to $Y_{\mathbb{I}}$ at small τ when dependencies matter and Y_{F_2} comparable to Y_{F_1} at large τ when only the marginal distribution contribute to $h^{(1,\tau)}$.

Figs 1,d,e,f and 2,d,e,f also report theoretical H, h and AMI for processes $Y_{\mathbb{I}}$ and Y_{F_1}, while theoretical h and AMI are not available for process Y_{F_2}. All estimates well match the predicted theoretical behaviors (solid lines in Figs 1, 2). The theoretical behaviors of AMI and entropy rate of process Y_{F_1} are obtained using eqs. (12) and (13) with the analytical expression (17).

These observations essentially hold both for the short-range (Fig. 1) and long-range (Fig. 2) correlations and yield the first key conclusion of this work: as opposed to correlation or spectral estimation, estimates of the two-points AMI ($I^{(1,1,\tau)}$) and entropy rate ($h^{(1,\tau)}$) are able to quantify differences in the temporal dynamics of processes even when they share the exact same marginal distributions and correlation functions, as long as they have different two-point joint statistics.

IV. Estimation performance and illustrations

The aim of this section is to assess, by means of Monte Carlo simulations, the performance of the estimators $\hat{H}$, $\hat{I}^{(m,p,\tau)}$ and $\hat{h}^{(m,\tau)}$ for processes with different statistical structures and to study the impacts of key parameters either entering the definition of the estimators (number of nearest neighbors k) or characterizing the dependence structures of the processes (strength of the correlation $c_Y(\tau)$) as well as with sample sizes N.

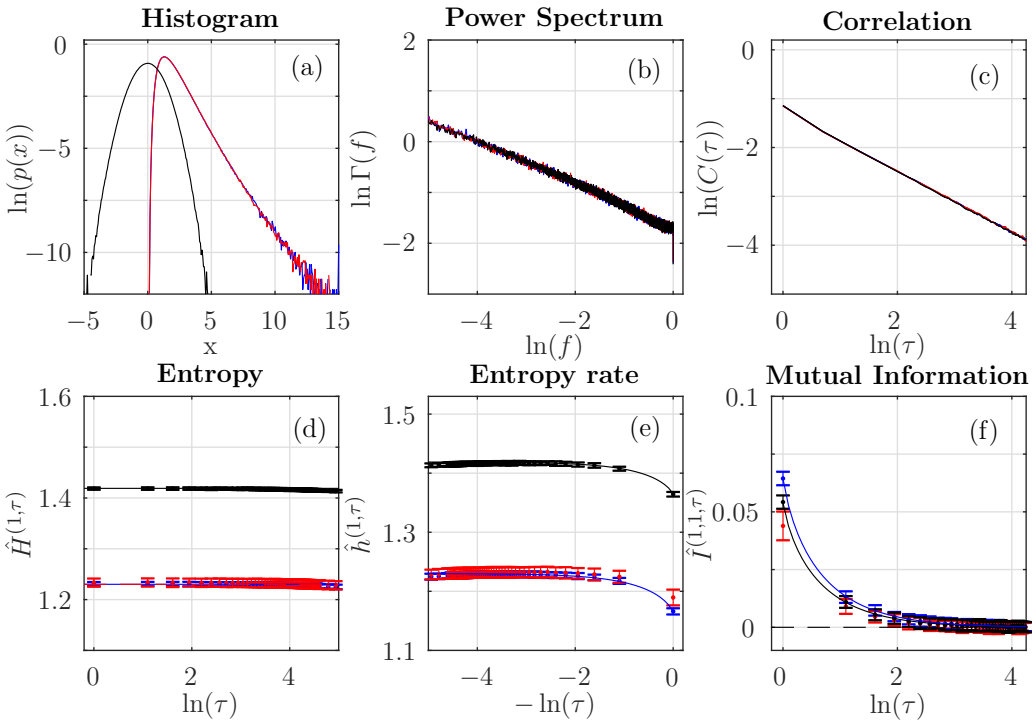

Fig. 2. **Long-range dependencies.** Classical analysis: histogram (a), power spectrum (b) and correlation function (c). Information theory analysis for $m = 1$, $p = 1$: entropy (d), entropy rate (e) and auto mutual information (f). The color indicates the process: blue for standard log-normal (Y_{F_1}), red for Y_{F_2} log-normal and black for Gaussian ($Y_{\mathbb{I}}$). The continuous lines in (e,d,f) correspond to analytical expressions. We used $N = 2^{16}$ and $k = 5$.

A. Estimation performance

1) Monte Carlo simulations Set-up: The processes are synthetized numerically using circulant matrix embedding strategies, with the toolbox designed in [54], [53] and made available at www.hermir.org. As we have access to analytical curves only for $Y_{\mathbb{I}}$ and Y_{F_1}, these are the processes used to characterize the performance of the estimators.

For short and long memory, parameters are set $f_c = 0.1$ and $\mathcal{H} = 0.7$ respectively. All variances are set to $\sigma_Y \equiv 1$. For each case, 100 realizations of processes are synthesized and performances are assessed as averages (and standard deviations) across realizations.

To study the impact of the marginal distribution on estimation performance, process Y_{F_1} is compared to process $Y_{\mathbb{I}}$.

2) Dependence with the sample size N: Fig. 3 reports performances as functions of the sample size N, for $2^8 \leq N \leq 2^{16}$, for a fixed $k = 5$. It clearly shows that the three estimators are asymptotically consistent (with vanishing biases and variances) irrespective of the marginal distribution and of the correlation being short-range or long-range. Unsurprisingly, variances are larger for the log-normal processes than they are for the Gaussian ones. Also, long-range dependencies imply much larger variances compared to short-range dependencies for a same $c_Y(\tau) = 0.32$ (which corresponds to $\tau = 1$ for the long range correlation process and $\tau = 11$ for the short range one), thus clearly showing that long memory does not induce biases but does imply larger variances. The proposed estimators are overall robust to marginal distributions and dependence structures [46], [47], and have satisfactory performance as soon as $N \geq 2^{10}$, hence $N = 2^{16}$ is used hereafter.

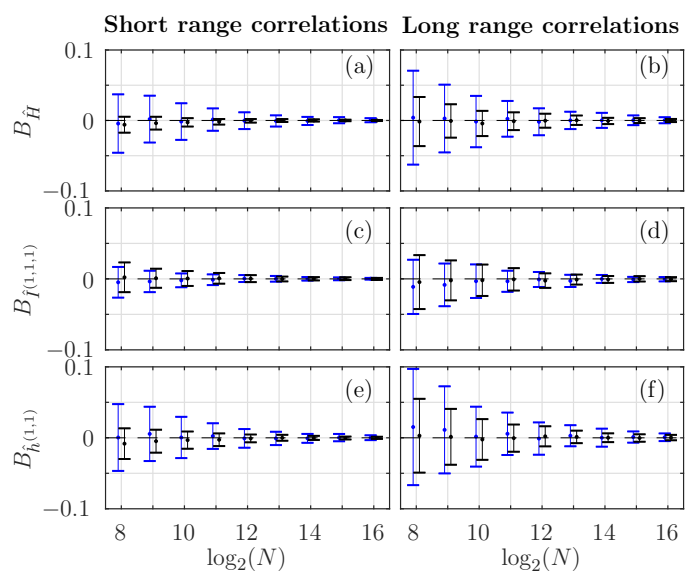

Fig. 3. Estimation performance for $\hat{H}$ (entropy), $\hat{I}^{(1,1,\tau)}$ (auto mutual information) and $\hat{h}^{(1,\tau)}$ (entropy rate), as functions of N, for log-normal (blue) and Gaussian (black) processes, with short-range (left) and long range (right) correlations. $f_c = 0.1$, $\mathcal{H} = 0.7$, $c(\tau) = 0.32$ and $k = 5$.

3) Dependence with the number of nearest neighbors k: Fig. 4 reports estimation performance as functions of k, for $N = 2^{16}$ and same correlation $c_Y = 0.32$ as above. Performances barely depend on k, but for a mild increase of the

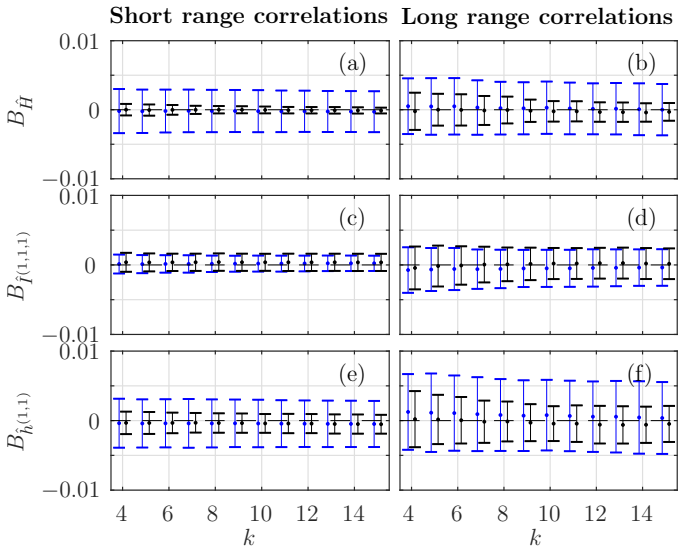

Fig. 4. Estimation performance for $\hat{H}$ (entropy), $\hat{I}^{(1,1,\tau)}$ (auto mutual information) and $\hat{h}^{(1,\tau)}$ (entropy rate), as functions of k, for log-normal (blue) and Gaussian (black) processes, with short-range (left) and long range (right) correlations. $f_c = 0.1$, $\mathcal{H} = 0.7$, $c(\tau) = 0.32$ and $N = 2^{16}$.

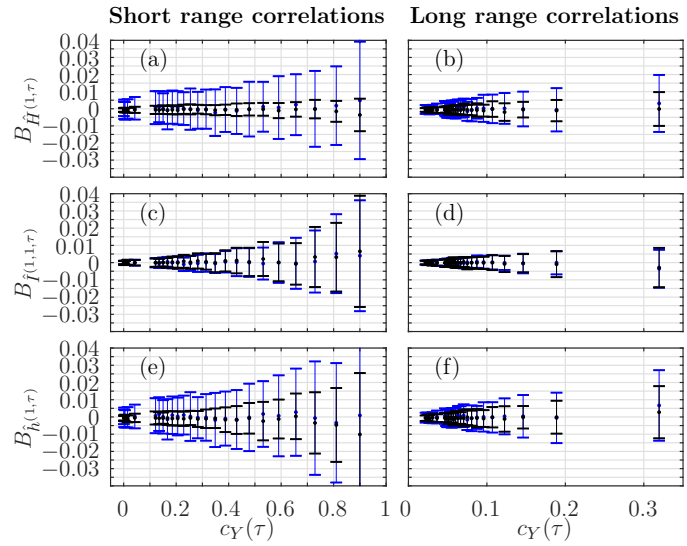

Fig. 5. Estimation performance for $\hat{H}$ (entropy), $\hat{I}^{(1,1,\tau)}$ (auto mutual information) and $\hat{h}^{(1,\tau)}$ (entropy rate), as functions of the correlation strength c_Y, for log-normal (blue) and Gaussian (black) processes, with short-range (left) and long range (right) correlations. $f_c = 0.1$, $\mathcal{H} = 0.7$, $k = 5$ and $N = 2^{12}$.

biases, and this irrespective of the marginal distributions and of short-range or long-range nature of the correlations. Because increasing k further implies increasing the computational load, $k = 5$ is used hereafter.

4) Influence of the correlation strength $c_Y(\tau)$: Fig. 5 reports estimation performance as a function of $c_Y(\tau)$ and shows that all biases slightly increase when $c_Y(\tau)$ increases. The variance appears to increase a little faster than linearly in c_Y. For both short-range and long-range correlations, the log-normal processes show larger standard deviations compared to the Gaussian ones, yet with similar biases. Further, for a given correlation level $c_Y(\tau)$ the long-range correlation structure shows larger standard deviations compared to the short-range correlation.

From this characterization, we can conclude that the entropy, AMI and entropy rate estimators show a low bias and standard deviation for the combination $k = 5$ and $N = 2^{16}$ of parameters used along the article. In addition, we can conclude that the standard deviation of all the three estimators is larger for strong correlations and non-Gaussian distributions, but their biases depend slightly on the marginal distribution and the correlation strength.

B. Process discrimination with auto mutual information

In Figs 1,e,f and 2,e,f, we show that both AMI and entropy rate across scales are able to discriminate between processes Y_{F_1} and Y_{F_2}. The discrimination is far more effective for the smaller scales for both correlation structures (Long and short range).

To quantify the differences between Y_{F_1} and Y_{F_2} that we found in Figs 1 and 2, we first present the histograms of the estimations of the entropy rate (Fig. 6a) and of the AMI (Fig. 6b) for fixed $N = 2^{16}$ and $\tau = 1$. We perform a statistical ranksum test using Mann-Whitney U-test (Wilcoxon rank-sum test) and compute the p-value, for varying τ and N. As seen in Fig. 6d, the p-value quickly decreases with increasing N, as expected. Fig. 6c quantifies the distinguishability between both log-normal processes across scales, showing that the discrimination is far better for smaller scales. This confirms the observations of Figs 1,e,f and 2,e,f.

C. Dependence structure beyond correlations

This sections further explores the ability of AMI $I^{(m,p,\tau)}$, with $m \geq 1$ and/or $p \geq 1$ to probe multiple-point dependences and hence fine details in temporal dynamics.

Fig. 7 reports higher-order AMI estimates $\hat{I}^{(m,p,\tau)}$ for several combination of m and p, as functions of the embedding dimension p for both short-range (left) and long-range (right) correlation functions, for processes $Y_{\mathbb{I}}$, Y_{F_1} and Y_{F_2}. From top to bottom the m value is equal to 1, 2, 3 and p. Consequently, the first three lines study the behaviour of AMI on p for fixed value of m, while the fourth line shows the behavior of AMI for $m = p$.

While two-point AMI yielded essentially identical conclusions for short-range and long-range correlation processes, higher order AMI reveals clear differences: For short-range correlation processes (left column), $Y_{\mathbb{I}}$, Y_{F_1} and Y_{F_2}, $\hat{I}^{(m,p,\tau)}$ and $I^{(m,p,\tau)}$ when it can be computed analytically, show a very mild dependence of AMI with m and p for a fixed τ. To the converse, for long-range correlation processes (right column), $Y_{\mathbb{I}}$, Y_{F_1} and Y_{F_2}, $\hat{I}^{(m,p,\tau)}$ and $I^{(m,p,\tau)}$ when it can be computed analytically, show constant increase of AMI when p or m increase, for a fixed τ. This fundamental observation clearly indicates that for short-range correlation, whatever τ, there is little or no gain in the analysis of temporal dynamics brought by the study of higher order AMI. Conversely, for long-range correlation processes, $I^{(m,p,\tau)}$ captures more and more information in the temporal dynamics as the embedding

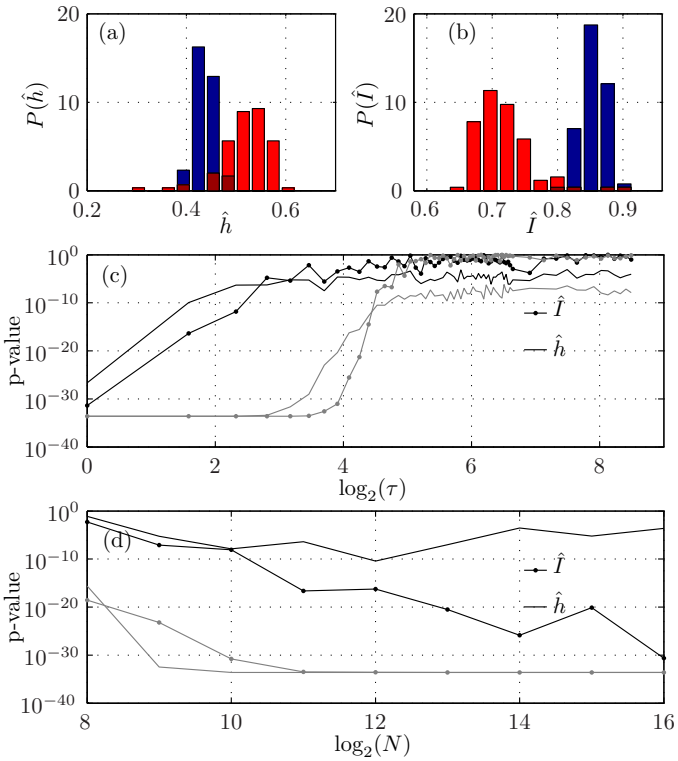

Fig. 6. **Statistical test.** Histogram of the estimation of $\hat{h}^{(1,1)}(1)$ (a) and $\hat{I}^{(1,1)}(1)$ (b) for the scale invariant processes (blue for Y_{F_1} and red for Y_{F_2} of size $N = 2^{16}$. (c) p-value of the Wilcoxon rank-sum test between the two log-nomal processes Y_{F_1} and Y_{F_2} as fonction of the scale for the statistics of $\hat{h}^{(1,1)}(\tau)$ (dotted line) and $\hat{I}^{(1,1)}(\tau)$ (full line) with $N = 2^{16}$. (d) same quantities as (c) for scale $\tau = 1$ as functions of the process size N. The statistics has been done on 100 realizations of both process. In (c) and (d) the gray curves correspond to the results obtained on the Exponential Decay processes. Mann-Whitney U-test ranksum.

dimensions are increased thus revealing a significant difference between short-range and long-range correlation structures. Analytical formulae for $I^{(m,p,\tau)}$ when $m, p > 1$, are only available for the identity $\mathbb{I}$ and the natural transform F_1, namely for $Y_{\mathbb{I}}$ and Y_{F_1}.

In order to quantify the evolution of AMI when the embedding increases, we compute the p-value of Wilcoxon test between the statistic of $\hat{I}^{(1,1,1)}$ and $\hat{I}^{(m,p,1)}$ for the same processes as in Fig. 7. Fig. 8 quantifies the increase of AMI when the embedding increases. This evolution of AMI is far more important in the long range correlations case, as expected from Fig. 7. Further, table I shows that, for fixed $m+p$, AMI evolves differently depending on the precise values of m and p separately. For a fixed $m+p$ AMI is larger when m and p are close, which can be understood as a closer similarity between the two embedded vectors appearing as the arguments of the AMI in eq.(4).

	$Y_{\mathbb{I}}$	Y_{F_1}	Y_{F_2}
Short range correlations	0.34	0.96	0.72
Long range correlations	$6e^{-15}$	$5e^{-20}$	$2e^{-3}$

TABLE I

Statistical test. p-VALUE OF WILCOXON TEST BETWEEN THE STATISTICS OF $\hat{I}^{(1,3)}(1)$ AND $\hat{I}^{(2,2)}(1)$.

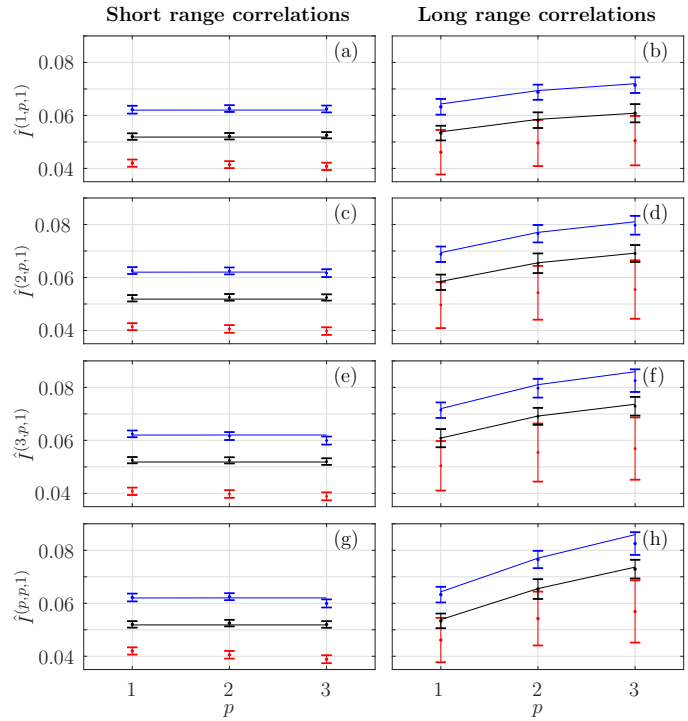

Fig. 7. **Evolution of AMI with embedding.** $\hat{I}^{(m,p)}(\tau = 1)$ for processes with exponentially decaying correlation (left column) and with power law decay (right column). Black is for $Y_{\mathbb{I}}$, blue is for Y_{F_1} and red for Y_{F_2}. First three lines are for fixed m and increasing p: $m = 1$ (a,b), $m = 2$ (c,d) and $m = 3$ (e,f). Last line is for $m = p$. Other parameters are $N = 2^{16}$, $k = 5$, $\tau = 1$.

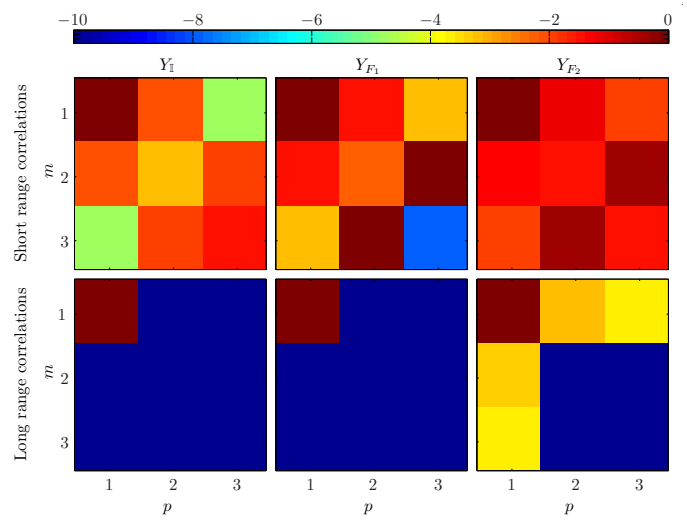

Fig. 8. **Statistical test.** p-value of Wilcoxon rank-sum test between the statistics of $\hat{I}^{(1,1)}(1)$ and $\hat{I}^{(m,p)}(1)$ for processes with exponentially decaying correlation (top line) or with a power law decay (bottom line). Left column is for $Y_{\mathbb{I}}$, middle column for Y_{F_1} and right column for Y_{F_2}.

V. CONCLUSIONS AND PERSPECTIVES

We explored the evolution of AMI ($I^{(m,p,\tau)}$) and entropy rate ($h^{(m,\tau)}$) on the scale parameter τ. Even when considering only two point interactions ($m = p = 1$) AMI and entropy rate probe statistics of any order, and hence non-linear dependences, and as such appear as unambiguous generalizations of

the correlation function and the power spectrum.

As an illustration, we analyzed two log-normal processes with identical 1-point distribution and identical correlation function. AMI ($I^{(1,1,\tau)}$) and entropy rate ($h^{(1,\tau)}$) are able to discriminate these two processes, which appear as identical for classical analysis, by enlightening differences in high order dependences. The behavior of these quantities along scales reveal the existence of stronger high-order dependences at smaller scale, which allows an easier separation of processes.

In addition, AMI generalizes easily to consider explicitly interactions between more than two points. We thus probe the complexity of the dependence structure above linear order, *i.e.*, the additional information measured by AMI when considering an extra point should indicate the existence and relevance of next-order interactions. As an example, comparing the effect of increasing the order of interactions for two dependence structures showed that while exponential decay — short range — dependences do not involve next order interactions, power law — long range — dependences do.

The same qualitative results were obtained when considering others one-point distribution, which supports the validity of the methodology for any distribution and dependence structure. This generality should make the methodology very interesting to perform non-Gaussian processes characterization in several different applications.

References

[1] G. P. Moore, J. P. Segundo, D. H. Perkel, and H. Levitan, "Statistical signs of synaptic interaction in neurons," *Biophysical Journal*, vol. 10, no. 9, pp. 876 – 900, 1970.

[2] M. Cohen and A. Kohn, "Measuring and interpreting neuronal correlations," *Nature Neuroscience*, vol. 14(7), pp. 811–819, 2011.

[3] C. Gray and D. McCormick, "Chattering cells: Superficial pyramidal neurons contributing to the generation of synchronous oscillations in the visual cortex." *Science*, vol. 274, pp. 109–113, 1996.

[4] K. J. Miller, S. Zanos, E. E. Fetz, M. den Nijs, and J. G. Ojemann, "Decoupling the cortical power spectrum reveals real-time representation of individual finger movements in humans," *Journal of Neuroscience*, vol. 29, no. 10, pp. 3132–3137, 2009.

[5] A. Kolmogorov, "The local structure of turbulence incompressible viscous fluid for very large reynolds number." *Dokl. Akad. Nauk SSSR*, vol. 30, pp. 299–303, 1941.

[6] F. Anselmet, Y. Gagne, E. Hopfinger, and R. Antonia, "High-order velocity structure functions in turbulent shear flow," *Journal of Fluid Mechanics*, vol. 140, pp. 63–89, 1984.

[7] W. Zhang, "Constructing ecological interaction networks by correlation analysis: hints from community sampling," *Network Biology*, vol. 1(2), pp. 81–98, 2011.

[8] K. Yamasaki, A. Gozolchiani, and S. Havlin, "Climate networks around the globe are significantly affected by el niño," *Phys. Rev. Lett.*, vol. 100, p. 228501, Jun 2008.

[9] Q. Ou, Y. Jin, T. Zhou, B. Wang, and B. Yin, "Power-law strength-degree correlation from resource-allocation dynamics on weighted networks," *Phys. Rev. E*, vol. 75, p. 021102, Feb 2007.

[10] M. Takada and B. Jain, "The three-point correlation function in cosmology," *Monthly Notices of the Royal Astronomical Society*, vol. 340(2), 2002.

[11] Z. Slepian and D. Eisenstein, "Computing the three-point correlation function of galaxies in $\mathcal{O}(\mathrm{n}^2)$," *Monthly Notices of the Royal Astronomical Society*, vol. 454(4), pp. 4142–4158, 2015.

[12] C. L. Nikias and J. M. Mendel, "Signal processing with higher-order spectra," *IEEE signal processing magazine*, vol. 10, no. 3, pp. 10–37, 1993.

[13] J.-L. Lacoume, P.-O. Amblard, and P. Comon, "Statistiques d'ordre supérieur pour le traitement du signal," p. 290, 1997, iSBN = 2-225-83118-1.

[14] W. Collis, P. White, and J. Hammond, "Higher-order spectra: The bispectrum and trispectrum," *Mechanical Systems and Signal Processing*, vol. 12, no. 3, pp. 375 – 394, 1998.

[15] V. Chandran, S. Elgar, and B. Vanhoff, "Statistics of tricoherence," *IEEE Transactions on Signal Processing*, vol. 42, pp. 3430–3440, 1995.

[16] T. Matsubara, "Analytic minkowski functionals of the cosmic microwave background: Second-order non-gaussianity with bispectrum and trispectrum," *Phys. Rev. D*, vol. 81, p. 083505, Apr 2010.

[17] P. Trivedi, T. R. Seshadri, and K. Subramanian, "Cosmic microwave background trispectrum and primordial magnetic field limits," *Phys. Rev. Lett.*, vol. 108, p. 231301, Jun 2012.

[18] T. Ning and J. Bronzino, "Autoregressive and bispectral analysis techniques: EEG applications," *IEEE Engineering in Medicine and Biology Magazine*, vol. 9, pp. 47–50, 1990.

[19] P. Husar and G. Henning, "Bispectrum analysis of visually evoked potentials," *IEEE Engineering in medicine and biology magazine*, vol. 16, pp. 57–63, 1997.

[20] J. T. Astola, K. O. Egiazarian, G. I. Khlopov, S. I. Khomenko, I. V. Kurbatov, V. Y. Morozov, and A. V. Totsky, "Application of bispectrum estimation for time-frequency analysis of ground surveillance doppler radar echo signals," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 9, pp. 1949–1957, Sept 2008.

[21] C. Courtney, S. Neild, P. Wilcox, and B. Drinkwater, "Application of the bispectrum for detection of small nonlinearities excited sinusoidally," *Journal of Sound and Vibration*, vol. 329, no. 20, pp. 4279 – 4293, 2010.

[22] C. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. XXVII, pp. 388–427, 1948.

[23] C. Shannon and W. Weaver, *The Mathematical Theory of Information*, Urbana, Ed. University of Illinois Press, 1949.

[24] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, Ed. John Wiley & Sons, Inc, 2006.

[25] H. Herzel, W. Ebeling, and A. O. Schmitt, "Entropies of biosequences: The role of repeats," *Phys. Rev. E*, vol. 50, pp. 5061–5071, Dec 1994.

[26] A. Lesne, J. Blanc, and L. Pezard, "Entropy estimation of very short symbolic sequences," *Phys. Rev. E*, vol. 79, p. 046208, Apr 2009.

[27] C. Granero-Belinchón, S. Roux, and N. Garnier, "Scaling of information in turbulence," *EPL*, vol. 115, p. 58003, 2016.

[28] R. Takahira, K. Tanaka-Ishii, and L. Debowski, "Entropy rate estimates for natural language —a new extrapolation of compressed large-scale corpora," *Entropy*, vol. 18, p. 364, 2016.

[29] G. Tourassi, E. Frederick, M. Markey, and C. F. Jr, "Application of the mutual information criterion for feature selection in computer-aided diagnosis," *Medical Physics*, vol. 28, pp. 2394–2402, 2001.

[30] C. Granger and J. Lin, "Using the mutual information coefficient to identify lags in nonlinear models," *Journal of time series analysis*, vol. 15, pp. 371–384, 1994.

[31] O. Michel and P. Flandrin, "Application of methods based on higher-order statistics for chaotic time series analysis," *Signal Processing*, vol. 53, pp. 133–148, 1996.

[32] G. A. Darbellay, "The mutual information as a measure of statistical dependence," in *IEEE International Symposium on Information theory*, 1997.

[33] D. T. Pham, "Fast algorithms for mutual information based independent component analysis," *IEEE Transactions on Signal Processing*, vol. 52, no. 10, pp. 2690–2700, 2004.

[34] T. E. Duncan, "Mutual information for stochastic signals and fractional Brownian motion," *IEEE Transactions on information theory*, vol. 54, no. 10, pp. 4432–4438, 2008.

[35] R. Malladi, D. H. Johnson, G. P. Kalamangalam, N. Tandon, and B. Aazhang, "Mutual information in frequency and its application to measure cross-frequency coupling epilepsy," *IEEE Transactions on signal processing*, vol. 66, no. 11, pp. 3008–3023, 2018.

[36] W. Xiong, L. Faes, and P. C. Ivanov, "Entropy measures, entropy estimators, and their performance in quantifying complex dynamics: Effects of artifacts, nonstationarity, and long-range correlations," *Phys. Rev. E*, vol. 95, p. 062114, Jun 2017.

[37] H. P. Bernhard, "A tight upper bound on the gain of linear and nonlinear predictors for stationary stochastic processes," *IEEE Transactions on signal processing*, vol. 46, no. 11, Nov. 1998.

[38] D. Albers and G. Hripcsak, "Estimation of time-delayed mutual information and bias for irregularly and sparsely sampled time-series," *Chaos, Solitons & Fractals*, vol. 45, no. 6, pp. 853 – 860, 2012.

[39] ——, "Using time-delayed mutual information to discover and interpret temporal correlation structure in complex populations," *Chaos*, vol. 22, p. 013111, 2012.

[40] C. Granero-Belinchon, S. Roux, P. Abry, M. Doret, and N. Garnier, "Information theory to probe intrapartum fetal heart rate dynamics," *Entropy*, vol. 19, no. 12, p. 640, 2017.

[41] P. O. Amblard and O. J. J. Michel, "On directed information theory and Granger causality graphs," *Journal of computational neuroscience*, vol. 30, pp. 7–16, 2011.

[42] ——, "The relation between Granger causality and directed information theory: a review," *Entropy*, vol. 15, no. 1, pp. 113–143, 2013.

[43] P. Gaspard and X. J. Wang, "Noise, chaos and (ϵ, τ)-entropy per unit time," *Physics reports*, vol. 235, no. 6, pp. 291–343, 1993.

[44] F. Takens, *Dynamical Systems and Turbulence, Warwick 1980: Proceedings of a Symposium Held at the University of Warwick 1979/80*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1981, ch. Detecting strange attractors in turbulence, pp. 366–381.

[45] K. Zografos and S. Nadarajah, "Expressions for Rényi and Shannon entropies for multivariate distributions," *Statistics and probability Letters*, vol. 71, pp. 71–84, 2005.

[46] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Physical Review E*, vol. 69, pp. 066 138–1/066 138–16, June 2004.

[47] W. Gao, S. Oh, and P. Viswanath, "Demystifying fixed k-nearest neighbor information estimators," *CoRR*, vol. abs/1604.03006, 2016.

[48] L. Kozachenko and N. Leonenko, "Sample estimate of entropy of a random vector," *Problems of Information Transmission*, vol. 23, pp. 95–100, 1987.

[49] H. Singh, N. Misra, V. Hnizdo, A. Fedorowicz, and E. Demchuk, "Nearest neighbor estimates of entropy," *American Journal of Mathematical and Management Sciences*, vol. 23, no. 3 and 4, pp. 301–321, 2003.

[50] P. O. Amblard, S. Zozor, O. J. J. Michel, and A. M. Cuculescu, "On the estimation of the entropy using k-th nearest neighbors," in *IMA conf Maths and signals Processing*, 2008.

[51] C. Granero-Belinchón, S. Roux, and N. Garnier, "Un estimateur du taux d'entropie basé sur l'information mutuelle," in *Colloque Gretsi*, 2017.

[52] J. Theiler, "Spurious dimension from correlation algorithms applied to limited time-series data," *Physical Review A*, vol. 34, pp. 2427–2432, 1986.

[53] H. Helgason, V. Pipiras, and P. Abry, "Fast and exact synthesis of stationary multivariate gaussian time series using circulant embedding," *Signal Process.*, vol. 91, no. 5, pp. 1123–1133, May 2011.

[54] ——, "Synthesis of multivariate stationary series with prescribed marginal distributions and covariance using circulant matrix embedding," *Signal Processing*, vol. 91, pp. 1741–1758, August 2011.