HAL
open science

Information theory for non-stationary processes with stationary increments

Carlos Granero-Belinchon, Stéphane G. Roux, Nicolas Garnier

To cite this version:

Carlos Granero-Belinchon, Stéphane G. Roux, Nicolas Garnier. Information theory for non-stationary processes with stationary increments. Entropy, 2019, Information Theoretic Measures and Their Applications, 21 (12), pp.1223. 10.3390/e21121223 . hal-02395022

HAL Id: hal-02395022

https://hal.science/hal-02395022v1

Submitted on 6 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Article

Information theory for non-stationary processes with stationary increments

Carlos Granero-Belinchon[(1,2)], **Stéphane G. Roux**[(1)], **Nicolas B. Garnier**[(1)]

[1] Univ Lyon, Ens de Lyon, Univ Claude Bernard, CNRS, Laboratoire de Physique, F-69342 Lyon, France;
[2] ONERA-DOTA, University of Toulouse, FR-31055 Toulouse, France;

Submitted to Entropy

Abstract: We describe how to analyze the wide class of non stationary processes with stationary centered increments using Shannon information theory. To do so, we use a practical viewpoint and define ersatz quantities from time-averaged probability distributions. These ersatz versions of entropy, mutual information and entropy rate can be estimated when only a single realization of the process is available. We abundantly illustrate our approach by analyzing Gaussian and non-Gaussian self-similar signals, as well as multi-fractal signals. Using Gaussian signals allow us to check that our approach is robust in the sense that all quantities behave as expected from analytical derivations. Using the stationarity (independence on the integration time) of the ersatz entropy rate, we show that this quantity is not only able to fine probe the self-similarity of the process but also offers a new way to quantify the multi-fractality.

1. Introduction

Many real world processes, like global weather data, water reservoir levels, biological or medical signals, economic time series, *etc*, are intrinsicaly non-stationary [1–5]: their probability density function (PDF) deforms when time evolves. Analyzing such processes requires a stationary hypothesis in order to apply classical analysis, like, *e.g.*, two-point correlations assessment [6]. The stationary hypothesis can be either strict or weak: while a strict stationarity requires all moments of the process — and hence its PDF — to be time-independent, a weak stationarity is achieved when the first moment and the covariance function are time-independent and the variance is finite at all time [7]. Even the weaker hypothesis is often very restrictive and not realistic over long time periods. When the signal has a drift or a linear trend, another approach is to focus on its time-increments or time-derivatives. Indeed, assuming that the increments or time derivative are stationary is then a more realistic hypothesis. For real world processes, the stationarity of the increments or even the stationarity of the signal is often argued to be valid when considering small chunks of data spanning short enough time range [8–10], so that slow evolutions of higher order moments can be neglected. The present article focuses on non-stationary processes with increments that are stationary and centered; this hypothesis ensures that the processes do not have any trend or drift.

Shannon information theory provides a very general framework to study stationary processes [11, 12], and some attempts to analyze non-stationary processes have been reported [13–15]. Contrary to most classical approaches, like, e.g., linear response theory in statistical physics or solid state physics, this framework is not restricted to the study of two-point correlations and linear relationships, and it allows to quantify higher order dependences [16] and nonlinear dynamics [12]. Information theory can be straightforwardly applied to any non-stationary time-process $X = \{x_t\}_{t\in\mathbb{R}}$: by carefully studying how the probability density and dependences of the process evolve in time, a time-evolving Shannon entropy $H_t(X)$ can be defined. The drawback of this approach is that it requires the knowledge of many realizations of the time-evolution of the process, as it relies on having enough statistics over the realizations [15].

Unfortunately, obtaining enough data is very difficult in real world systems where in the best case scenario a few realizations can be recorded experimentally, and usually only a single realization is accessible. In this paper, we develop a methodology that can be applied to a single realization, in order

to analyse a non-stationary signal with stationary centered increments. We describe a time-averaged framework that gathers all available data points in a time window representing a single realization, whether it is the full experimental time duration, or just a fraction of it [13].

The present paper is organized as follows. In section 2, we present the general framework of information theory for a non-stationary signal, and our new framework that exploit time averages. We then give a particular emphasis on self-similar processes. In section 3, we report a benchmarking of our framework in the special case of Gaussian self-similar signals, a model situation where it is possible to obtain analytical developments. In section 4, we explore the case of non-Gaussian self-similar processes. Finally, in section 5, we drop the hypothesis of self-similarity and we apply our framework to a multifractal process.

2. Information theory for non-stationary processes

2.1. Non-stationary processes with stationary increments

In this article, we consider non-stationary processes with stationary increments. Such a process can be written as a motion $M = \{m_t\}_{t\in\mathbb{R}}$ obtained by integrating a stationary noise $W = \{w_t\}_{t\in\mathbb{R}}$:

$$m_t = m_0 + \int_0^t w_{t'} \mathrm{d}t' \,. \tag{1}$$

where m_0 and w_0 are the values at time $t = 0$, both of which can be set to 0 without loss of generality.

Nowadays signals are recorded and stored on digital media, which amounts to consider in practice a set of data sampled at discrete times t_k where $k \in \mathbb{N}^+$. We further assume that the signals are equi-sampled, *i.e.*, dt is constant and we choose $dt = 1$. So we consider in this article discrete time processes and we express them as motions $M = \{m_t\}_{t\in\mathbb{N}^+}$ obtained by integrating a stationary noise $W = \{w_t\}_{t\in\mathbb{N}^+}$ according to:

$$m_t = m_0 + \sum_{k=1}^{t} w_k \,, \quad t > 0 \,, \tag{2}$$

where again $m_0 = w_0 = 0$. Eq.(2) can also be replaced by

$$m_t = m_{t-1} + w_t \,, \quad t > 0 \,. \tag{3}$$

If the noise W is not centered, *i.e.*, has a statistical mean $\mathbb{E}(W) = \beta \neq 0$, we introduce the centered noise $w'_t = w_t - \beta$. The equations for the motion M read:

$$m_t = m_0 + \beta t + \sum_{k=1}^{t} w'_k \,, \quad t > 0 \,, \tag{4}$$

$$= m_{t-1} + \beta + w'_t \,, \quad t > 0 \,. \tag{5}$$

The process M can be interpreted as a motion built on the stationary centered noise W' together with an additive deterministic drift, which is the linear trend βt.

In this article, we study motions without trend, so we impose that the noise W is centered, *i.e.*, that its statistical mean $\mathbb{E}(W) = \beta = 0$. Besides the simple centering of the increments $W \mapsto W - \mathbb{E}(W)$, any detrending method can be applied to M, e.g., using moving averages. As a consequence, the motion M is centered: its statistical mean is $\mathbb{E}\{m_t\} = m_0 = 0$ at all times $t > 0$. Nevertheless, its variance, and all its higher order moments, may depend on time: the motion M is a non-stationary process with stationary increments. Typical examples of such processes are Brownian motion and fractional Brownian motion [17], both of which have a variance that evolves with time.

2.2. General framework

For a generic non-stationary process $X_t = \{x_t\}_{t\in\mathbb{R}}$, the probability density function (PDF) $p_{x_t}(x_t)$ changes with time. The information theory framework can be applied to each random variable x_t, *i.e.*, at each time t. To do so, the PDF of x_t needs to be estimated at each time t, which in practice requires to have many realizations available [15].

To analyze the temporal dynamics of a random processes at a given time t, we consider the m-dimensional vector obtained with the Takens time-embedding procedure [18]:

$$\mathbf{x}_t^{(m,\tau)} = \left(x_t,\, x_{t-\tau},\, \cdots,\, x_{t-(m-1)\tau}\right) . \tag{6}$$

The embedding dimension m controls the order of the statistics that are considered, and the delay τ defines a time scale. We define below some information theory quantities that are functionals of the m-point joint-distributions $p_{\mathbf{x}_t^{(m,\tau)}}(\mathbf{x}_t^{(m,\tau)})$, in order to characterize linear and non-linear temporal dynamics.

Shannon entropy

The entropy of $\mathbf{x}_t^{(m,\tau)}$ is:

$$H(\mathbf{x}_t^{(m,\tau)}) = -\int_{\mathbb{R}^m} p_{\mathbf{x}_t^{(m,\tau)}}(\mathbf{x}) \log(p_{\mathbf{x}_t^{(m,\tau)}}(\mathbf{x}))\mathrm{d}\mathbf{x} . \tag{7}$$

This quantity depends on time t, as well as on embedding parameter m and delay τ. We further note it $H_t^{(m,\tau)}(X) = H(\mathbf{x}_t^{(m,\tau)})$, where the index t indicates the time and the parameters (m,τ) are indicated as upper indices. It measures the amount of information characterizing the m-dimensional PDF of the process X at time t sampled at scale τ. When $m = 1$, the entropy does not depend on τ and does not probe the dynamics of the process; we then note it $H_t(X)$, dropping the $(m = 1, \tau)$ upper indices. However, for embedding dimension $m > 1$ the entropy depends on the linear and non-linear dynamics of the process. Indeed, the entropy involves arbitrarily high order moments of the joint PDF $p_{\mathbf{x}_t^{(m,\tau)}}(\mathbf{x}_t^{(m,\tau)})$. As usual, the entropy does not depend on the first moment of the distribution.

Using the time-increments of size τ, $\delta_\tau x_t \equiv x_t - x_{t-\tau}$, it can be shown (see appendix A) that the amount of information measured by $H_t^{(m,\tau)}(X)$ is the same as the amount of information in the vector $\tilde{\mathbf{x}}_t^{(m,\tau)} \equiv (x_t, \delta_\tau x_t, \delta_\tau x_{t-\tau}, ..., \delta_\tau x_{t-(m-2)\tau})$, *i.e.*,

$$\begin{aligned} H_t^{(m,\tau)}(X) &= H(\tilde{\mathbf{x}}_t^{(m,\tau)}) \\ &= H\left(x_t, \delta_\tau x_t, \delta_\tau x_{t-\tau}, ..., \delta_\tau x_{t-(m-2)\tau}\right) . \end{aligned} \tag{8}$$

For processes with stationary increments, the marginal distribution of x_t may be strongly time-dependent, but the marginal distributions of any increment is time-independent. Eq.(8) thus suggests that the time-dependence of $H_t^{(m,\tau)}(X)$ originates mainly from x_t, the first component of the rewritten embedded vector $\tilde{\mathbf{x}}_t^{m,\tau}$. Nevertheless, it should be observed that although the $m-1$ increments, considered by themselves, have a stationary dependence structure, the covariance of x_t with any of the increments is *a priori* non-stationary.

Mutual information and auto-mutual information

The mutual information MI measures the amount of information shared by two processes. For two non-stationary time-embedded vectors $\mathbf{x}_{t_1}^{(m,\tau)}$ and $\mathbf{y}_{t_2}^{(n,\tau)}$, it is defined as :

$$MI(\mathbf{x}_{t_1}^{(m,\tau)}, \mathbf{y}_{t_2}^{(n,\tau)}) = H_{t_1}(\mathbf{x}_{t_1}^{(m,\tau)}) + H_{t_2}(\mathbf{y}_{t_2}^{(n,\tau)}) - H(\mathbf{x}_{t_1}^{(m,\tau)}, \mathbf{y}_{t_2}^{(n,\tau)}) \,. \tag{9}$$

In the following, we use auto-mutual information $I_t^{(m,n,\tau)}(X)$ to measure, for a single process X, the shared information between two successive time-embedded vectors of dimension m and n [19]:

$$I_t^{(m,n,\tau)}(X) = MI(\mathbf{x}_t^{(n,\tau)}, \mathbf{x}_{t-n\tau}^{(m,\tau)}) \,. \tag{10}$$

Auto-mutual information defined in (10) probes the dynamics of the process X_t at time t by measuring the dependencies between two consecutive chunks of m and n points sampled every τ.

Entropy rate

The entropy rate, or entropy gain [20], of order m at time t measures the increase of Shannon entropy when the embedding dimension is increased from m to $m+1$. It is defined as the variation of Shannon entropy between $\mathbf{x}_{t-\tau}^{(m,\tau)}$ and $\mathbf{x}_t^{(m+1,\tau)}$, two successive time-embedded versions of the process X:

$$h_t^{(m,\tau)}(X) = H_t^{(m+1,\tau)}(X) - H_{t-\tau}^{(m,\tau)}(X) \tag{11}$$

$$= H_t(X) - I_t^{(m,1,\tau)}(X) \,. \tag{12}$$

Within the general framework, the entropy, mutual information and entropy rate are well defined at any time t for a non-stationary process. Although this framework can formally be used to analyze non-stationary processes at any time t, in practice it is often impossible to assess statistics at a fixed time t, as the number of available realizations from real world datasets may be very small. To overcome this issue, we propose in the next section another framework that considers averages over a finite and possibly large time window, which represents for example the duration of an experimental measurement.

2.3. Practical time-averaged framework

We now focus on non-stationary processes with stationary increments. We develop in this section a pragmatic approach which can be applied when a single time trace of a non-stationary signal is available.

We first present a very formal perspective that defines a time-averaged PDF of a non-stationary process. We then propose a practical approach which uses a very simple estimation of such a time-average PDF. We finally use this practical approach to define all the information quantities that we are interested in.

2.3.1. Time-averaged framework

Using a formal perspective, we consider the global statistics of the dataset, when forgetting its time dynamics, and we formally consider the time-averaged probability density function in the time window $[t_0, t_0+T]$:

$$\bar{p}_{T,t_0,\mathbf{x}^{(m,\tau)}}(\mathbf{x}) = \frac{1}{T}\int_{t_0}^{t_0+T} p_{\mathbf{x}_t^{(m,\tau)}}(\mathbf{x})\mathrm{d}t \,. \tag{13}$$

Because of the time-average, this probability density function doesn't depend on a single time t but on the starting time t_0 and the duration T of the time window.

In the case of a stationary process, the PDF $p_{\mathbf{x}_t^{(m,\tau)}}(\mathbf{x})$ is independent of t, so the PDF $\bar{p}_{T,t_0,\mathbf{x}^{(m,\tau)}}(\mathbf{x})$ is independent of t_0 and T.

In the case of a non-stationary process with stationary centered increments, the dependence on t_0 only appears on the mean of the time-averaged PDF $\bar{p}_{T,t_0,\mathbf{x}^{(m,\tau)}}(\mathbf{x})$. As a consequence, since the Shannon entropy does not depend on the mean, none of the information theoretic quantities depends on t_0.

In the case of a non-stationary process with stationary but non-centered increments, there is a drift: the first moment of $p_{\mathbf{x}_t^{(m,\tau)}}(\mathbf{x})$ evolves linearly with time. When integrated in time in eq.(13), this induces a deformation of the time-averaged PDF $\bar{p}_{T,t_0,\mathbf{x}^{(m,\tau)}}(\mathbf{x})$, which affects *a priori* moments of any order. As a consequence, the Shannon entropy is then expected to depend on t_0.

In the following, we focus on non-stationary processes with stationary centered increments, described in section 2.1.

2.3.2. Practical framework

In practice, given a time series of length T, we propose to very roughly approximate the PDF $\bar{p}_{T,t_0,\mathbf{x}^{(m,\tau)}}$ defined in (13) with the normalized histogram $\hat{\bar{p}}_{T,t_0}$ of all data points $x_t^{(m,\tau)}$, $t \in [t_0, t_0+T]$ available in the time window. This is a very strong assumption, as $\bar{p}_{T,t_0,\mathbf{x}^{(m,\tau)}}$ is *a priori* very different from any $p_{x_t^{(m,\tau)}}$, and *a priori* very different from the histogram $\hat{\bar{p}}_{T,t_0}$ constructed after cumulating all the available data in the interval. This pragmatic approach comes down to treat the set of available data points in the time interval exactly in the same way as if it was a set of data points originating from a stationary, albeit unknown, process and then estimate its PDF.

In the following, we drop the hat in the notations, and consider only the ersatz probabilities $\hat{\bar{p}}$ in place of the time-averaged probabilities $\bar{p}$. As we discuss later in section 6, if several experimental realizations are available, it is of course possible to use them to enhance the estimation of the time-averaged PDF.

2.3.3. Information theory quantities in the practical framework

Given a time series of length T, and considering the ersatz PDFs $\bar{p}_{T,t_0,\mathbf{x}^{(m,\tau)}}$, we define $\bar{H}_T^{(m,\tau)}(X)$ the time-averaged Shannon entropy, $\bar{I}_T^{(m,n,\tau)}(X)$ the time-averaged auto-mutual information and $\bar{h}_T^{(m,\tau)(X)}$ the time-averaged entropy rate, as described below.

Ersatz Shannon entropy

We define the ersatz entropy of the time-embedded signal as the entropy of the time-averaged PDF $\bar{p}_{T,t_0,\mathbf{x}^{(m,\tau)}}$:

$$\bar{H}_T^{(m,\tau)}(X) = -\int_{\mathbb{R}^m} \bar{p}_{T,t_0,\mathbf{x}^{(m,\tau)}}(\mathbf{x}) \log(\bar{p}_{T,t_0,\mathbf{x}^{(m,\tau)}}(\mathbf{x}))\mathrm{d}\mathbf{x}\,. \tag{14}$$

$\bar{H}_T^{(m,\tau)}(X)$ gives the amount of information of the set of values of the signal $\mathbf{x}_t^{(m,\tau)}$ in the time interval $[t_0, t_0+T]$ and hence it can be interpreted as the total information characterizing the temporal trajectory $\{\mathbf{x}_t^{(m,\tau)}, t \in [t_0, t_0+T]\}$ of the process. If the process has stationary centered increments, the total amount of information in the trajectory depends only on its length T, and not on its starting time t_0. In that sense, the ersatz entropy $\bar{H}_T^{(m,\tau)}$ is not stationary.

Using the rewriting (8), we argue that this dependence in T originates from x_t — the first component of the vector $\tilde{\mathbf{x}}_t^{m,\tau}$ — which has a time-dependent marginal distribution. Because the $m-1$ other components of $\tilde{\mathbf{x}}_t^{m,\tau}$ are increments, they have by hypothesis a stationary dependence structure. So increasing the embedding dimension does not impact the dependence of the ersatz entropy on the window size T, but only its dependence on the increments size τ.

Auto-mutual information

We define the ersatz auto-mutual information as:

$$\bar{I}_T^{(m,n,\tau)}(X) = \bar{H}_T^{(m,\tau)}(X) + \bar{H}_T^{(n,\tau)}(X) - \bar{H}_T^{(m+n,\tau)}(X)\,. \tag{15}$$

Entropy rate

We define the ersatz entropy rate over a time interval of size T as:

$$\bar{h}_T^{(m,\tau)}(X) = \bar{H}_T^{(m+1,\tau)}(X) - \bar{H}_T^{(m,\tau)}(X) \tag{16}$$

$$= \bar{H}_T(X) - \bar{I}_T^{(m,1,\tau)}(X)\,. \tag{17}$$

From (16), we may expect a cancelation of the main dependence in T which is the same for $\bar{H}_T^{(m+1,\tau)}$ and for $\bar{H}_T^{(m,\tau)}$. As a consequence, the ersatz entropy rate $\bar{h}_T^{(m,\tau)}$ should be stationary, in the sense that it should not depend on the length T of the time interval that is considered.

If the available samples span a very large time window, one may consider using multiple non-overlapping time windows of size T starting at various times. Because of the stationarity and zero-mean of the increments and hence the independence of the ersatz quantities on t_0, it is possible to average the different estimations of the ersatz quantities obtained in each window. It is also possible to use all the non-overlapping windows to populate the histogram and thus enhance the estimation of the time-averaged PDF. Each of these two operations will increase the statistics and hence improve the estimation.

2.4. Self-similar processes

In this section, we focus on the special case of self similar processes, *i.e.*, signals which exhibit monofractal scale invariance [21]. Such processes have been used as a satisfying first approximation to model or describe very various phenomena, such as ionic transport [22], fluid turbulence [23], climate [24], river flows [25], cloud structure [26] or earthquakes [27], as well as neural signals [28], stock markets [29,30], texture patterns [31] or internet traffic [32]. A process X_t is monofractal scale-invariant if there exists a real number $\mathcal{H}$ such that for all $a \in \mathbb{R}^{+*}$, the probability density functions of x_{at} and $a^{\mathcal{H}} x_t$ are equivalent. $\mathcal{H}$ is called the Hurst exponent. If $\mathcal{H} < 0$, the process is stationary and called a fractional noise. If $0 \leq \mathcal{H} < 1$, the process is non-stationary with stationary increments. The case $\mathcal{H} = 1/2$ corresponds to the traditional Brownian motion.

Assuming $x_{t=0} = 0$, the scale invariance property can be expressed as [33]:

$$p_{x_{at}}(x) = \frac{1}{a^{\mathcal{H}}} p_{x_t}\left(\frac{x}{a^{\mathcal{H}}}\right)\,. \tag{18}$$

The scale invariance property of a process X_t transfers to its increments, as well as any of its time-embedded version:

$$p_{\mathbf{x}_{at}^{(m,a\tau)}}(\mathbf{x}) = \frac{1}{a^{m\mathcal{H}}} p_{\mathbf{x}_t^{(m,\tau)}}\left(\frac{\mathbf{x}}{a^{\mathcal{H}}}\right)\,. \tag{19}$$

This relation allows to express the non-stationary PDF of $\mathbf{x}_t^{(m,\tau)}$ at any time t as a function of the PDF at unit-time (t=1). This is done by using the factor $a = 1/t$ in eq.(19), *i.e.*, by rescaling each coordinate of the embedded vector by the factor $t^{\mathcal{H}}$.

Using eq.(13), it is straightforward to see that the scale invariant property of the form (19) is also valid for the time-averaged PDF $\bar{p}_{T,t_0,\mathbf{x}^{(m,\tau)}}(\mathbf{x})$.

Because of its definition (2) as a cumulative sum of a noise, a motion can be seen as accumulating the correlations between successive points of the noise. When performing a time-embedding, the particular case $\tau = 1$ is interesting: considering the relation (8), we may expect that the information

contained in the time-embedded motion $\mathbf{m}_t^{(m,\tau=1)}$ is closely related to the information contained in the time-embedded noise $\mathbf{w}_t^{(m,\tau=1)}$. This is not the case anymore when $\tau \geq 2$.

Fractional Brownian motion

The Fractional Brownian motion (fBm) was proposed by Mandelbrot and Van Ness [17] and quickly became a benchmark for self-similarity and long-range dependence. The fBm is the only Gaussian self-similar process with stationary increments. It is characterized by its Hurst exponent, $\mathcal{H}$.

The fBm is a motion, obtained by integrating according to (2) a fractional Gaussian noise (fGn), defined as a centered Gaussian process with the correlation structure

$$c_{\text{fGn}}(\tau) = \frac{\sigma_1^2}{2}\left[(\tau-1)^{2\mathcal{H}} - 2\tau^{2\mathcal{H}} + (\tau+1)^{2\mathcal{H}}\right]. \tag{20}$$

The fGn is a stationary noise with the standard deviation σ_1. It is scale-invariant with a Hurst exponent $\mathcal{H}-1$.

The non-stationary covariance structure of the fBm B reads

$$\mathbb{E}\{B_t B_{t-\tau}\} = \frac{\sigma_1^2}{2}\left[t^{2\mathcal{H}} + (t-\tau)^{2\mathcal{H}} - \tau^{2\mathcal{H}}\right], \tag{21}$$

where $\tau < t$.

2.4.1. General framework

We show below how the theoretical information quantities depend on time t and delay τ. We start from the relation (8) between the entropy of the time-embedded vector and the entropy of the increments and we normalize each component of the vector $\tilde{\mathbf{x}}_t^{(m,\tau)}$ by its standard deviation. The standard deviation σ_t of the motion x_t evolves with time as $\sigma_t = \sigma_1 t^{\mathcal{H}}$, while the standard deviation σ_τ of the increments $x_t - x_{t-\tau}$ is independent of t, thanks to the stationarity of the increments, and evolves with the size of the increment as $\sigma_\tau = \sigma_1 \tau^{\mathcal{H}}$. So we have:

$$\begin{aligned} H_t^{(m,\tau)}(X) &= H\left(x_t/\sigma_t, \delta_\tau x_t/\sigma_\tau, ..., \delta_\tau x_{t-(m-2)\tau}/\sigma_\tau\right) \\ &\quad + \ln\sigma_t + (m-1)\ln\sigma_\tau, \qquad (22) \\ &= H\left(x_t/t^{\mathcal{H}}, \delta_\tau x_t/\tau^{\mathcal{H}}, ..., \delta_\tau x_{t-(m-2)\tau}/\tau^{\mathcal{H}}\right) \\ &\quad + \mathcal{H}\ln t + (m-1)\mathcal{H}\ln\tau + m\ln\sigma_1. \qquad (23) \end{aligned}$$

We then use the scaling law (19) for $a = 1/t$ to relate the joint probability at a given time t to the joint probability at unit-time $t = 1$, which leads to:

$$\begin{aligned} H\left(x_t/t^{\mathcal{H}}, \delta_\tau x_t/\tau^{\mathcal{H}}, ..., \delta_\tau x_{t-(m-2)\tau}/\tau^{\mathcal{H}}\right) &= H\left(x_1, \delta_{\tau/t} x_1, ..., \delta_{\tau/t} x_{1-(m-2)\tau/t}\right) - m\ln\sigma_1 \qquad (24) \\ &= H\left(\tilde{\mathbf{x}}_{t=1}^{(m,\tau/t)}\right) - m\ln\sigma_1. \qquad (25) \end{aligned}$$

Using (8) again at time $t = 1$, we have $H\left(\tilde{\mathbf{x}}_{t=1}^{(m,\tau/t)}\right) = H_{t=1}^{(m,\tau/t)}(X)$, so we can express the time-dependent Shannon entropy (7) for self-similar processes as:

$$H_t^{(m,\tau)}(X) = H_{t=1}^{(m,\tau/t)}(X) + \mathcal{H}\ln t + (m-1)\mathcal{H}\ln\tau. \tag{26}$$

The entropy rate can be rewritten with (11) and (26) as:

$$h_t^{(m,\tau)}(X) = h_{t=1}^{(m,\tau/t)}(X) + \mathcal{H}\ln\tau, \tag{27}$$

where $h_1^{(m,\tau/t)}(X)$ is the entropy rate at time $t = 1$, using the rescaled time delay τ/t.

Although the two quantities $H_{t=1}^{(m,\tau/t)}(X)$ and $h_{t=1}^{(m,\tau/t)}(X)$ are considered at a fixed time $t = 1$, they still depend on t via the delay τ/t. Because τ/t is small as soon as $t \gg \tau$, we expect that the dependence of the entropy $H_t^{(m,\tau)}(X)$ on time t is mainly in $\mathcal{H} \ln t$, and that the entropy rate is almost time-independent.

Fractional Brownian motion

The PDF p_{B_t} of the fBm is Gaussian at any time t, so we can express its Shannon entropy and entropy rate at time t by using eq.(26) and the expression of the Shannon entropy of a Gaussian multivariate process [34]. We obtain the following approximated expressions:

$$H_t^{(m,\tau)}(B) \simeq m H_1^{\mathrm{FBM}} + \mathcal{H} \ln t + (m-1)\mathcal{H} \ln \tau \tag{28}$$

$$h_t^{(m,\tau)}(B) \simeq H_1^{\mathrm{FBM}} + \mathcal{H} \ln \tau \,, \tag{29}$$

where $H_1^{\mathrm{FBM}} \equiv \frac{1}{2} \ln \left(2\pi e \sigma_1^2\right)$ is the entropy of the fBm at unit-time. These formulae are exact for $m = 1, 2$, but for $m \geq 3$, constant terms as well as corrections in τ/t have been omitted for clarity.

2.4.2. Practical time-averaged framework

For a generic self-similar process, we are not able to derive any analytical results in the practical time-averaged framework. Nevertheless, the behaviors expected for a generic non-stationary process with stationary increments are holding: i) the ersatz entropy $\bar{H}_T^{(m,\tau)}$ is not stationary, in the sense that it depends on the length T of the time-interval, ii) the ersatz entropy rate $\bar{h}_T^{(m,\tau)}$ is stationary.

Fractional Brownian motion

The ersatz entropy of the fBm over a time window of size $T \gg \tau$ can be expressed by averaging its covariance structure on a time window of size T. We obtain [35]:

$$\bar{H}_T^{(1,\tau)}(B) = H_1^{\mathrm{FBM}} + \mathcal{H} \ln T \,. \tag{30}$$

The entropy of the fBm thus increases linearly with the logarithm of the window size T. The larger the time window, the more there is information in the trajectory.

The auto-mutual information of the fBm can be derived in the same way using (15) for $T \gg \tau$:

$$\bar{I}_T^{(1,1,\tau)}(B) = -\mathcal{H} \ln \left(\frac{\tau}{T}\right) + \mathcal{C}\left(\frac{\tau}{T}\right) \,, \tag{31}$$

where $\mathcal{C}\left(\frac{\tau}{T}\right)$ is a correction in τ/T that reads

$$\mathcal{C}\left(\frac{\tau}{T}\right) = \frac{1}{2} \ln \left(\frac{\frac{\tau}{2T} + \frac{1}{2\mathcal{H}+1}}{\left(\frac{\tau}{T} + \frac{1}{2\mathcal{H}+1}\right)} \right) \tag{32}$$

$$= -\frac{2\mathcal{H}+1}{4} \frac{\tau}{T} + \mathcal{O}\left(\left(\frac{\tau}{T}\right)^2\right) \,. \tag{33}$$

The ersatz mutual information depends logarithmically on the scale τ and the window size T. The larger the window-size T or the smaller the scale τ, the stronger the dependencies.

The ersatz entropy rate of order $m = 1$ is obtained by combining (30) and (31) according to (17):

$$\bar{h}_T^{(1,\tau)}(B) = H_1^{\mathrm{FBM}} + \mathcal{H} \ln \tau - \mathcal{C}\left(\frac{\tau}{T}\right) \tag{34}$$

which is independent of T up to corrections in τ/T, while being linear in $\ln(\tau)$ with a constant slope $\mathcal{H}$. The correction $-\mathcal{C}(\tau/T)$ in eq.(34) is positive, see eq.(33).

Comparing (28) with (30) shows that for the fBm, the ersatz entropy dependence on T is exactly the same as the entropy dependence on t. Comparing (29) with (34) shows that the entropy rate and the ersatz entropy rate do not depend on t or T up to corrective terms that are negligible if the scale τ is not too large. We also see explicitly that both quantities evolves with the scale τ in $\mathcal{H}\ln\tau$, again up to corrections of order τ/t and τ/T.

The example of the fBm suggests that for a scale-invariant process the evolution of any information theory quantity with the scale τ is the same within the practical time-averaged framework or the general framework. We push this analysis further in the next sections, by exploring if this property holds when the process is non-Gaussian.

3. Benchmarking the practical framework with the fBm

We focus in this section on the fractional Brownian motion, for which analytical expressions were derived in the previous sections. We use the fBm not only to benchmark our estimators of information theory quantities, but also to illustrate the use of the practical framework and the expected behavior of the ersatz quantities when used on a self-similar process of Hurst exponent $\mathcal{H}$.

3.1. Characterization of the estimates

3.1.1. Data

To obtain a fBm, we integrate a fractional Gaussian noise (fGn). We use circulant matrix method [36] to impose the correlation structure of the fGn (20) . Then, we center and normalize the noise such that the standard deviation, σ_{fGn}, is equal to one. We then take the cumulative sum to obtain the fBm. Through all this article, $\mathcal{H} = 0.7$ for all the processes used to illustrate our results, but we have checked that they hold for any other value $0 < \mathcal{H} < 1$.

3.1.2. Procedure

We estimate the Shannon entropy $\bar{H}_T^{(m,\tau)}$ with our own implementation of the k-nearest neighbors estimate from Kozachenko and Leonenko [37]. We estimate the auto-mutual information $\bar{I}_T^{(m,p,\tau)}$ with the algorithm provided by Kraskov, Stogbauer and Grassberger [38]. This estimator is also based on a nearest neighbors search and it provides — amongst several good properties — a build-in cancellation of the bias difference originating from each of the two arguments. In the following, we note k the number of neighbors, which is the only parameter of the estimators. The entropy rate $\bar{h}_T^{(m,\tau)}$ is then computed using eq.(17).

We generate for each motion a set of 100 independent realizations of fixed size T with a Hurst exponent $\mathcal{H} = 0.7$. We compute averages of the estimates on the realizations and use the standard deviation as error bars in the different graphs.

In subsections 3.1.3 and 3.1.4, we characterize respectively the bias and standard deviation (std) over realizations of our estimators of entropy, auto-mutual information and entropy rate.

3.1.3. Convergence / bias

We detail here how the ersatz entropy rate evolves with T and k. We report in Fig. 1a our results for all possible values of the couples $(\log_2(T), k) \in [9, ..., 17] \times [4, ..., 18]$, while τ is set to 1 here. According to eq.(34), the ersatz entropy rate of the fBm converges for large T to the value H_1^{fBm} (horizontal black line in Fig. 1a) thanks to the vanishing of the correction term $\mathcal{C}(\tau/T)$, according to (33).

Fig. 1a can be interpreted as describing the behavior of the bias of the estimator. This bias vanishes non-monotonically as $k/T^{\frac{1}{m+1}}$. When $k/T^{\frac{1}{m+1}}$ is reduced, first the bias is positive and diminishes toward negative values and then converges to zero. This behavior was previously reported for the

k-nn mutual information estimator applied for stationary processes [16,38,39], and we confirm it is valid for the fBm.

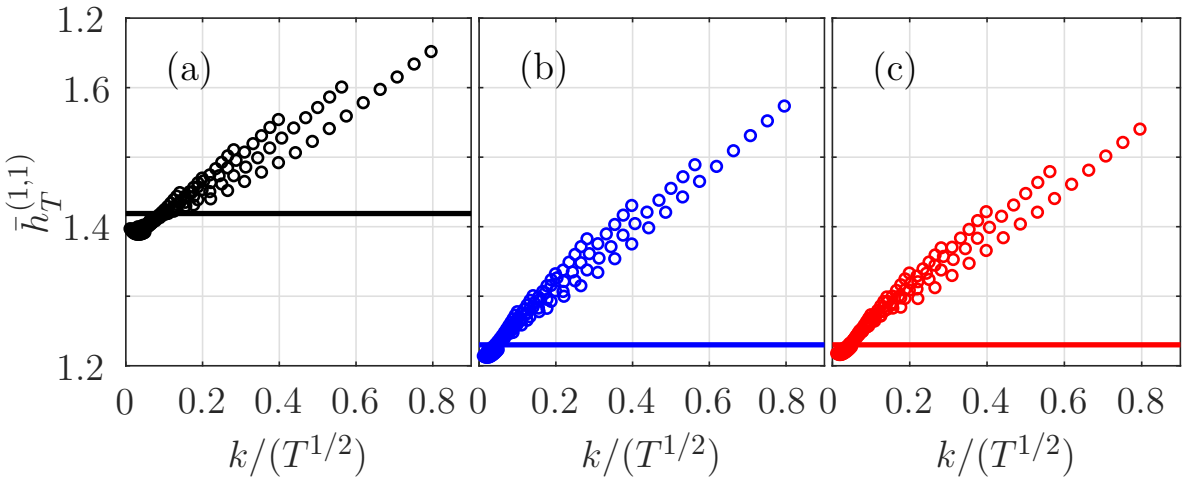

Figure 1. Dependence of $\bar{h}_T^{(1,\tau)}$ (for $\tau = 1$)) on $k/T^{\frac{1}{2}}$ for the fBm (a, in black) and for the Hermitian (b, in blue) and even-Hermitian (c, in red) log-normal processes.

We observed the same convergence for a large range of scales $\tau > 1$: the ersatz entropy rate then converges to $H^1_{\text{fBm}} + \mathcal{H} \ln \tau$ for large T with the same behavior of the bias.

3.1.4. Standard deviation of the estimates

We present in Fig. 2a the evolution of the standard deviation of the ersatz entropy, mutual information and entropy rate with T for $\tau = 1$. The standard deviation of both the entropy and mutual information is large, and does not decrease when T — and hence the number of samples — increases. On the contrary, the standard deviation of the entropy rate is much smaller and decreases when T increases. We attribute this feature to the dependence of the quantities on the observation time T, see eqs.(30) and (31) for the fBm. While $\bar{H}_T$ and $\bar{I}_T$ increase as $\ln T$, this is not the case for $\bar{h}_T$ which is independent on T (up to small corrections, negligible for smallish τ). Although it is difficult to explain why the standard deviation of the entropy and mutual information remain constant when T increases, it seems that this results from a balance between the non-stationarity (in $\ln T$) and the increased statistics. On the contrary, for the entropy rate which is stationary, the decrease of the std is as expected.

As a conclusion, both the bias and the standard deviation of the ersatz entropy rate increase when k increases or T decreases and can be made arbitrarily small by increasing the window size T. In the remainder of this article, we choose $k = 5$ and when studying the behavior of information theoretic quantities on the scale τ, we set $T = 2^{16}$.

3.2. Dependence on times T and τ

In this section, we present a detailed numerical study of the ersatz entropy, auto-mutual information and entropy rate of the fBm with $\mathcal{H} = 0.7$. In particular, we present a quantitative comparison with the analytical expressions (28, 29) in the general framework, as well as with analytical expressions (30, 31, 34) in the practical framework for the fBm. These comparisons allow: first, to validate the analytical expressions obtained for fBm in the practical framework, and second to show that the information theoretic quantities in the practical framework evolve in T and τ exactly as their counterparts evolve in the general framework in t and τ. To compare analytical and numerical results, we vary the window size T, the scale τ and the embedding dimension m.

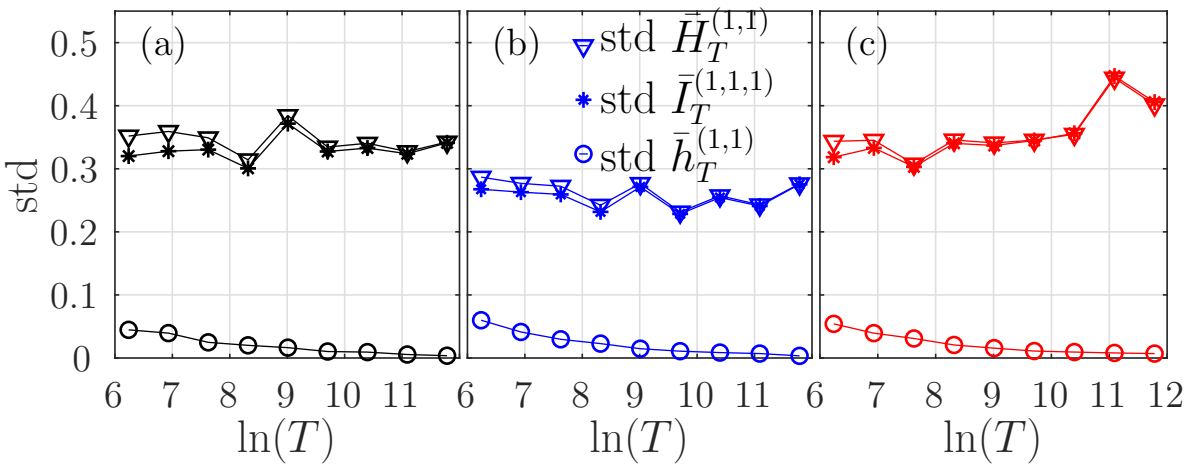

Figure 2. Standard deviations of $\bar{H}_T^{(1,\tau)}$ (triangles), $\bar{I}_T^{(1,1,\tau)}$ (circles) and $\bar{h}_T^{(1,\tau)}$ (stars), for $\tau = 1$, as functions of T, for the fBm (a, black) the Hermitian (b, in blue) and even-Hermitian (c, in red) log-normal processes.

3.2.1. Entropy and auto-mutual information

Dependence on T

The left column of Fig. 3 shows the ersatz Shannon entropy $\bar{H}_T^{(m,\tau)}$ (Fig. 3a) and auto-mutual information $\bar{I}_T^{(m,1,\tau)}$ (Fig. 3c) at a given scale $\tau = 1$, as a function of $\ln T$. The evolution of these two quantities for $m = 1$ is very close to $\mathcal{H} \ln T$, which is represented by a continuous black line. This is in agreement with eq.(30) and eq.(31). For $m > 1$, we obtain in the practical framework the behaviors predicted in the general framework, replacing t by T in the equations. We observe that the auto-mutual information does not depend on the embedding dimension m, while the entropy does, with an offset that seems to depend linearly on m. The dependence of the entropy and the auto-mutual information on the time window T is the signature of the non-stationarity of the signal.

Dependence on τ

The right column of Fig. 3 shows the ersatz Shannon entropy and auto-mutual information for a fixed window size $T = 2^{16}$ when varying the scale parameter τ. The ersatz Shannon entropy behaves as $(m-1)\mathcal{H} \ln(\tau)$, see Fig. 3b, in agreement with eq.(26) or eq.(28). The ersatz auto-mutual information behaves as $-\mathcal{H} \ln(\tau)$ for any embedding m, see Fig. 3d, in agreement with eq.(31), thus suggesting this formula is valid for any embedding dimension.

3.2.2. Stationarity of the entropy rate

Fig. 4a shows that the ersatz entropy rate $\bar{h}_T^{(m=1,\tau)}$ with embedding dimension $m = 1$ is almost constant when T is varied. For embedding dimensions $m > 1$, there is a small variation, of about 15%, much smaller than the 200% variation observed for either the entropy or the auto-mutual information (Fig. 3a,c) on the same range of T. This small dependence on T can be due to the correction in eq.(34), which may depend on m. We argue that it is mostly due to bias, which increases with the embedding dimension. Indeed, we observe that the entropy rate seems to converge for larger T to the same value close to H_1^{fBm} for all m. As a larger T corresponds to a larger sampling of the statistics, the bias is reduced, as reported in Fig. 1. Moreover, for $m = 1$, eq.(34) predicts a positive correction that vanishes when T is large: on the contrary we observe a convergence to a value lower than H_1^{fBm} which hints

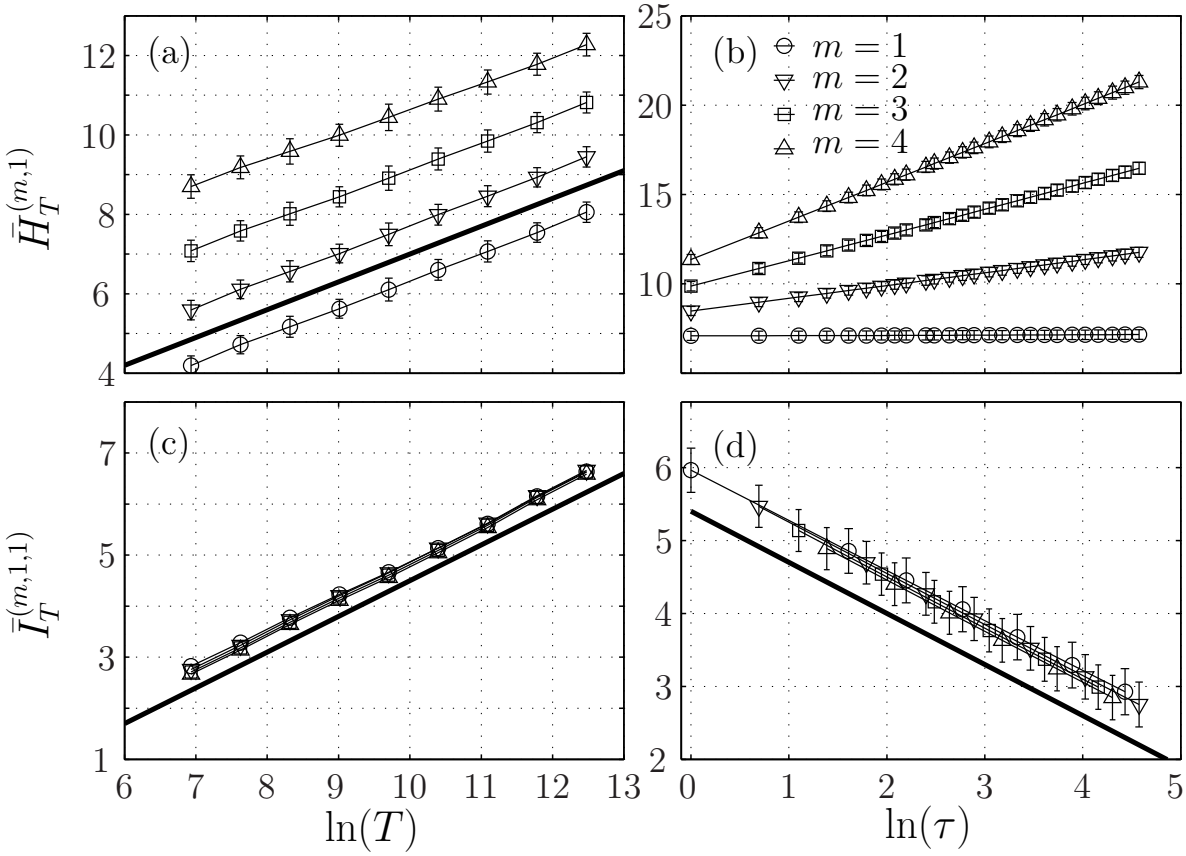

Figure 3. (a) Entropy $\bar{H}_T^{(m,\tau)}$ and (c) auto-mutual information $\bar{I}_T^{(m,1,\tau)}$ of the fBm in function of the logarithm of the window size $\ln(T)$ for a fixed scale $\tau = 1$. (b) Entropy and (d) auto-mutual information in function of the logarithm of the scale of analysis $\ln(\tau)$ for a fixed $T = 2^{16}$. Each symbol corresponds to a different embedding dimension m. In (a) and (c) the black line has a slope $\mathcal{H} = 0.7$, while in (d) its slope is $-\mathcal{H} = -0.7$.

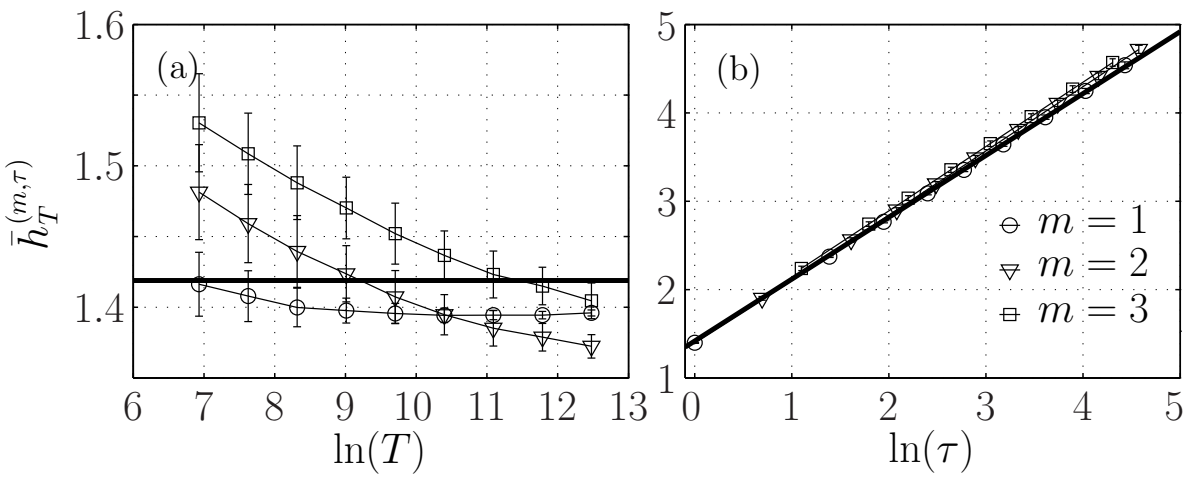

Figure 4. Ersatz entropy rate $\bar{h}_T^{(m,\tau)}$ of a fBm with $\mathcal{H} = 0.7$. (a): as a function of the window size T for fixed $\tau = 1$. (b) : as a function of the scale τ for $T = 2^{16}$. Each symbol corresponds to a different embedding dimension m. The horizontal black line in (a) indicates the theoretical value H_1^{fBm}. The black line in (b) represents the linear function $H_1^{\text{fBm}} + \mathcal{H} \ln \tau$ with $\mathcal{H} = 0.7$.

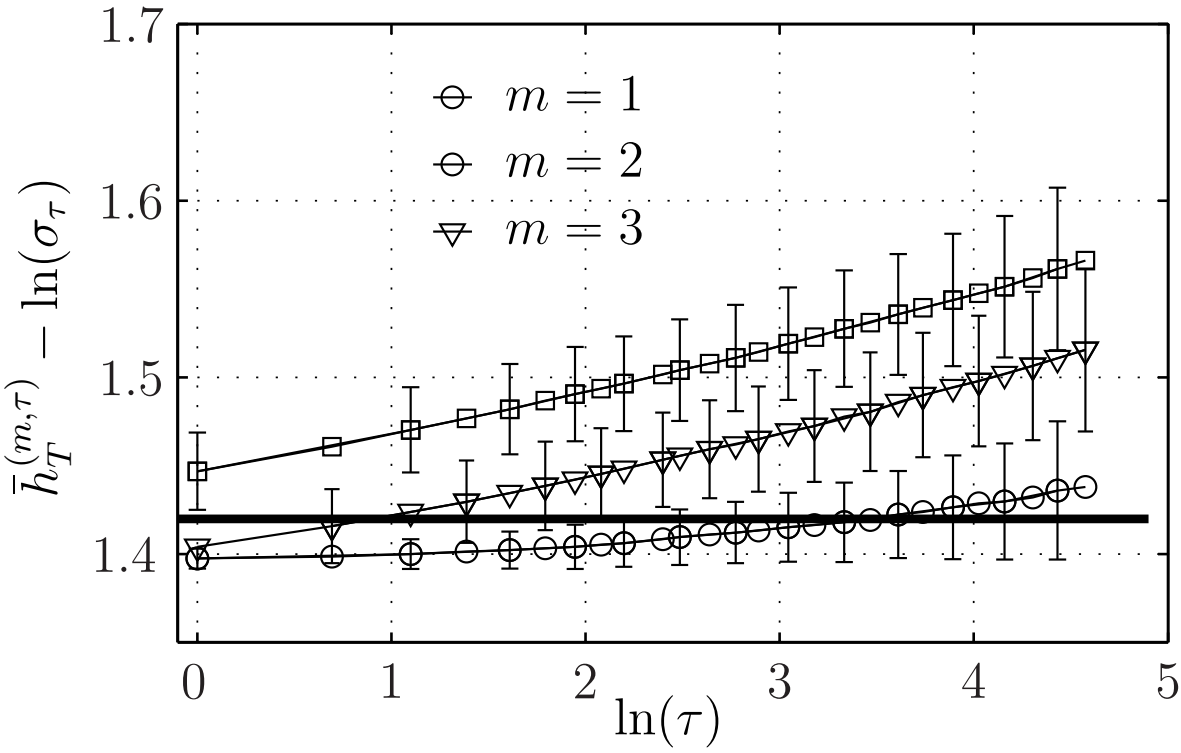

Figure 5. Ersatz entropy rate $\bar{h}_T^{(m,\tau)} - \ln(\sigma_\tau)$ of the fBm as a function of $\ln(\tau)$ for fixed $T = 2^{16}$ and varying embedding dimension m. The thick horizontal black line represents the constant value H_1^{fBm}.

that the bias is negative and larger than the theoretical correction. This suggests that the form of eq.(34) is still valid for embedding dimensions $m > 1$.

3.2.3. Entropy rate dependence on scale τ

Fig. 4b shows that for a fixed window size $T = 2^{16}$ the ersatz entropy rate is proportional to $\mathcal{H}\ln(\tau)$. We have added a black line defined by the linear function $H_1^{\mathrm{FBM}} + \mathcal{H}\ln\tau$, as suggested by eq.(34) without the corrective term. This black line perfectly describes the evolution of the entropy rate with the scale τ, which is independent on the embedding dimension m.

To observe the finer evolution of the entropy rate on the scale τ, we subtract the main contribution $\mathcal{H}\ln\tau$ to the entropy rate and we plot $\bar{h}_T^{(m,\tau)} - \mathcal{H}\ln\tau$ for different embedding dimensions in Fig. 5. We observe a slight increase, which is larger for larger embedding dimensions. For $m = 1$, the correction term can be evaluated from eq.(32), and is at most 2.10^{-3}, and does not account for the evolution reported here, which is probably due to the bias which increases when the number of points — which is proportional to T/τ — decreases and when the embedding m increases.

For a scale invariant self similar process, the standard deviation σ_τ of the increments of size τ behaves as $\sigma_\tau = \sigma_1\tau^{\mathcal{H}}$. Subtracting $\mathcal{H}\ln\tau$ amounts to subtracting $\ln\sigma_\tau$: for each scale τ, this corresponds to normalizing the down-sampled data (taking one point every τ points) by the standard deviation σ_τ of the increments of size τ. When the Hurst exponent is *a priori* unknown, σ_τ can be computed, and used to compute the main contribution $-\ln(\sigma_\tau)$; thus the fine evolution of the entropy rate with τ can be used as a tool to probe the deviation from the self similarity assumption, which is interesting for multifractal signals.

4. Application of the practical framework to non-Gaussian self-similar processes

In this section, we turn to non-Gaussian processes and describe the results obtained in the time-average framework generalized in this larger class of processes.

Procedure

We construct two different motions, in the very same way as we did for the fBm. We integrated two log-normal noises synthesized with the same log-normal marginal distribution and with the same correlation function (20) as the fGn, but different dependance structure. To generate these noises, we use the methodology proposed in [36] to obtain the log-normal marginal by applying two different transformations to the cumulative distribution function F_Z of a Gaussian white noise Z: the Hermitian transformation of rank 1 ($f^1(z) = F^{-1}(F_Z(z))$) and the even-Hermitian transformation of rank 2 : $f^1(z) = F^{-1}(2(F_Z(|z|) - \frac{1}{2}))$, where F is the cumulative distribution function of the targeted log-normal distribution. This synthesis is performed with the toolbox provided at www.hermir.org. Once the two log-normal noises have been generated, they are integrated using eq.(2) to obtain two non-stationary scale invariant processes with non-Gaussian statistics.

The dependence structures of the two log-normal noises were previously studied in detail [16]: while the correlation function is the same for the two noises — and identical to the targeted one of the fBm given by (20) — the complete dependence structure was shown to be different.

To study these two non-stationary and non-Gaussian motions, we use again realizations of $T = 2^{16}$ points, $k = 5$ neighbors and we focus on the case where embedding dimension $m = 1$ and Hurst exponent $\mathcal{H} = 0.7$.

Bias and standard deviation

We report in Fig. 1b and 1c the evolution of the ersatz entropy rate of the Hermitian and the even-Hermitian log-normal processes in function of $\frac{k}{T^{1/2}}$. We observe exactly the same behavior as for the fBm: the entropy rate converges to $H_1^{\ln}$, the entropy of the log-normal process at unit-time[1] (horizontal blue/red line in Fig. 1b,c), which then gives an estimation of the bias of our estimator, which appears to be the same as for the fBm.

We report in Fig. 2b and Fig. 2c the behavior of the standard deviation of the estimators. Again, exactly as for the fBm, the standard deviation is large for the ersatz entropy and the ersatz auto-mutual information, while it is much smaller for the ersatz entropy rate.

Again, both the bias and the standard deviation of the entropy rate increase when k increases or T decreases and can be made arbitrarily small by increasing T. These results do not depend on the marginal distribution: they have been obtained not only for the fBm with Gaussian statistics, but also for two motions built on log-normal noises.

Dependence on times T and τ

The evolution of $\bar{h}_T^{(1,\tau)}$ on the time window size T for the two motions is presented in Fig. 6a). As it was the case for the fBm, $\bar{h}_T^{(1,\tau)}$ depends only weakly on T, and seems to converge for larger T to the value $H_1^{\ln}$, up to a small corrective term.

---

[1] if X is a log-normal process of mean μ and standard deviation σ, then the process $\log X$ is Gaussian with the mean $\mu' = \log\left(\frac{\mu^2}{\sqrt{\mu^2+\sigma^2}}\right)$ and the standard deviation $\sigma' = 2\log\left(1 + \frac{\sigma^2}{\mu^2}\right)$ and the entropy of X can be expressed as [16]:

$$H_1^{\ln} = \frac{1}{2}\log(2\pi e\sigma') + \mu' .$$

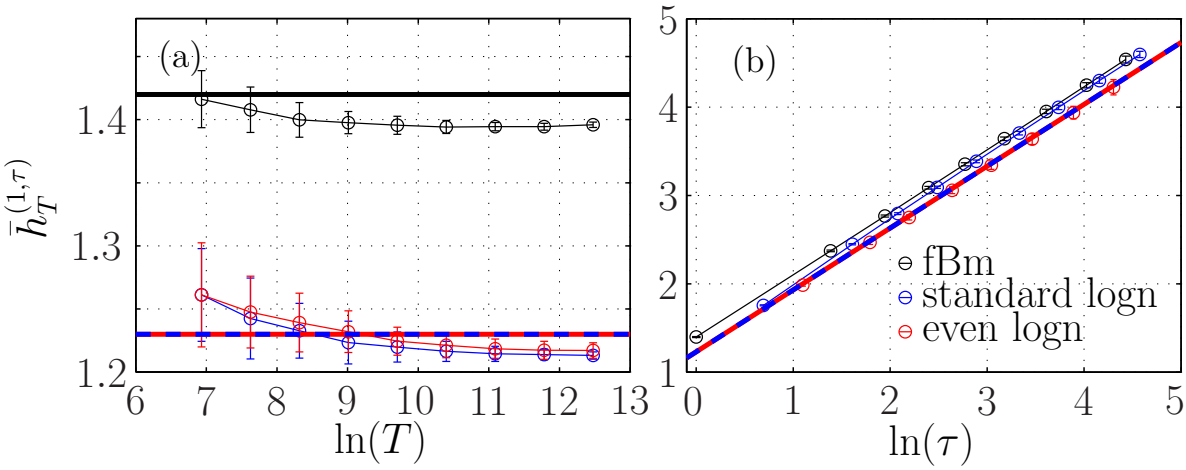

Figure 6. $\bar{h}_T^{(1,\tau)}$ for a motion built from a Hermitian (blue) or even-Hermitian (red) log-normal noise, as a function of a) the time window size T or b) the time scale τ. Results for the fBm (from Fig. 4 with $m = 1$) are reported in black for comparison. $T = 2^{16}$ and $k = 5$. The horizontal lines in (a) indicates the entropy H_1 of the noise (in black for the fBm and in red and blue for a log-normal process).

The evolution of $\bar{h}_T^{(1,\tau)}$ with the time scale τ is presented in Fig. 6b). In the same way as for the fBm, we again observe a large increase, almost proportional to $\ln \tau$. Because this strong tendency originates from the increase of the standard deviation σ_τ of the increments of size τ when τ increases, we again normalize the entropy rate by subtracting $\ln \sigma_\tau = \mathcal{H} \ln \tau$. Results are presented in Fig. 7, together with results for the fBm with $m = 1$ for comparison.

The normalized ersatz entropy rate of the motion built from the even-Hermitian log-normal noise appears as almost independent of τ. This behavior is identical to the one observed for the fBm, but the remaining constant value is different (H_1^{fBm} or $H_1^{\ln}$). The ersatz entropy rate of the fBm (in black) and the even-Hermitian motion (in red) both behaves exactly as $\mathcal{H} \ln \tau$, which is the expected behavior for a self-similar process, see eq.(27). On the contrary, the motion built with the Hermitian transformation of rank 1 exhibits an additional variation in τ: the normalized entropy rate $\bar{h}_T^{(m,\tau)} - \ln(\sigma_\tau)$ evolves from the value $H_1^{\ln}$ at $\tau = 1$ — expected for the motion built with a log-normal noise and obtained for the even-Hermitian process at any τ — up to the value H_1^{fBm} — expected for a Gaussian process, and obtained for the fBm at any τ.

As a conclusion, one can estimate the Hurst exponent of a perfectly self-similar process as the slope of the linear fit in $\ln \tau$ of the ersatz entropy rate. This is a valid approach for the fBm and the motion built from the noise constructed with the even-Hermitian transformation, because the ersatz entropy rate then behaves linearly in $\ln \tau$. On the contrary, the motion built using an hermitian transformation of rank 1 does not appear as perfectly self-similar. This can be indeed verified by plotting the normalized PDFs (setting the standard deviation to unity) of the increments $m_t - m_{t-\tau}$ of the motions for various values of τ. As can be seen in Fig. 8 the PDFs of the increments of the "standard log-normal process" varies with the scale τ, while these of the "even-Hermitian motion" remain identical. For $\tau = 1$, the increments are nothing but the log-normal noises, which are log-normal, as prescribed. For large τ, the increments of the "even-Hermitian motion" remain log-normal, while the increments of the standard log-normal motion" deforms and seems to become more Gaussian. The ersatz entropy rate catches this fine evolution perfectly.

5. Application of the practical framework to a multifractal process

We now explore the proposed time-averaged framework on the Multifractal Random Walk, to illustrate how it performs on a multifractal process. The multifractal random walk (MRW) [40,41]

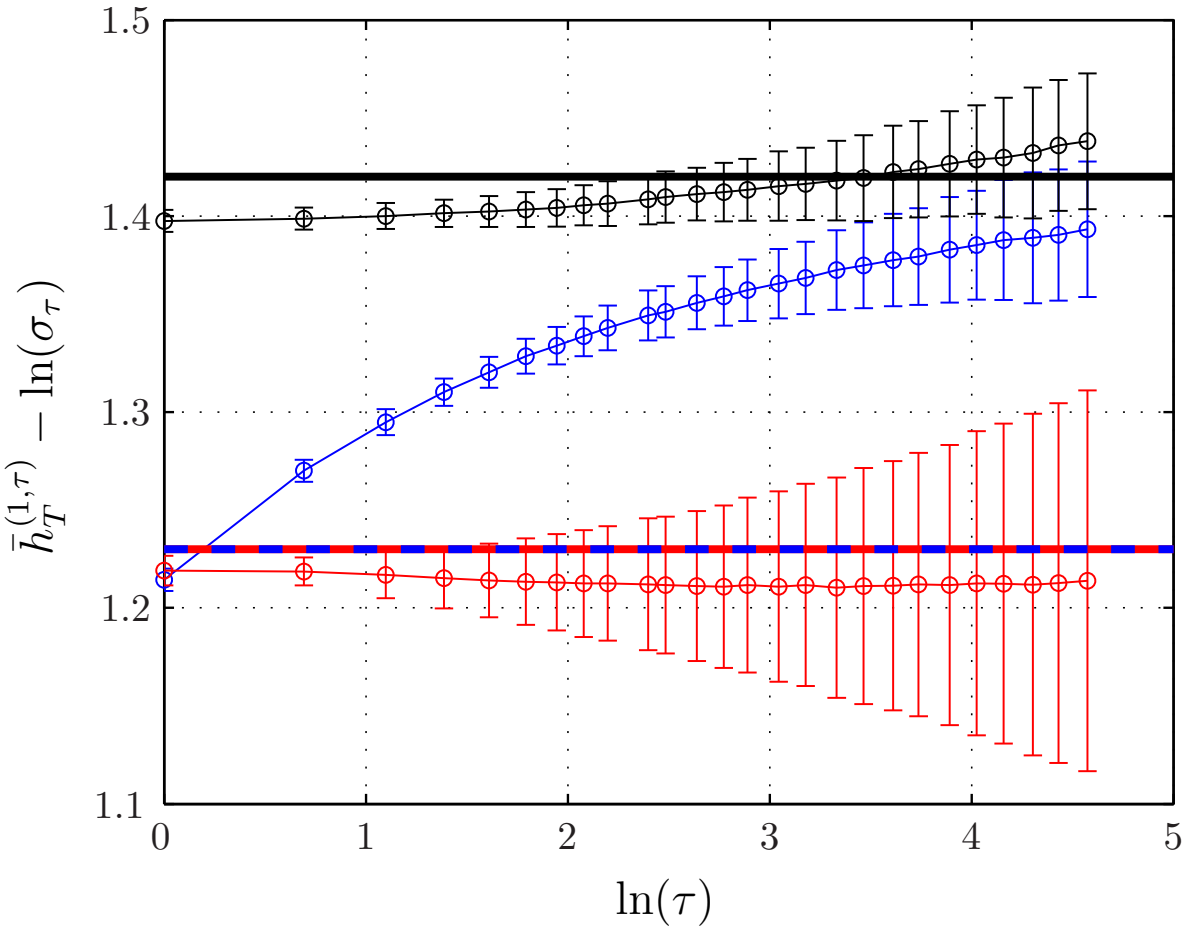

Figure 7. Ersatz entropy rate $\bar{h}_T^{(m=1,\tau)} - \ln(\sigma_\tau)$ for motions built on Hermitian (blue) or even-Hermitian (red) log-normal noise, together with results for the fBm (black) as a function of $\ln \tau$. $T = 2^{16}$ and $k = 5$. The straight lines in (b) indicate the theoretical values of the entropy of the processes.

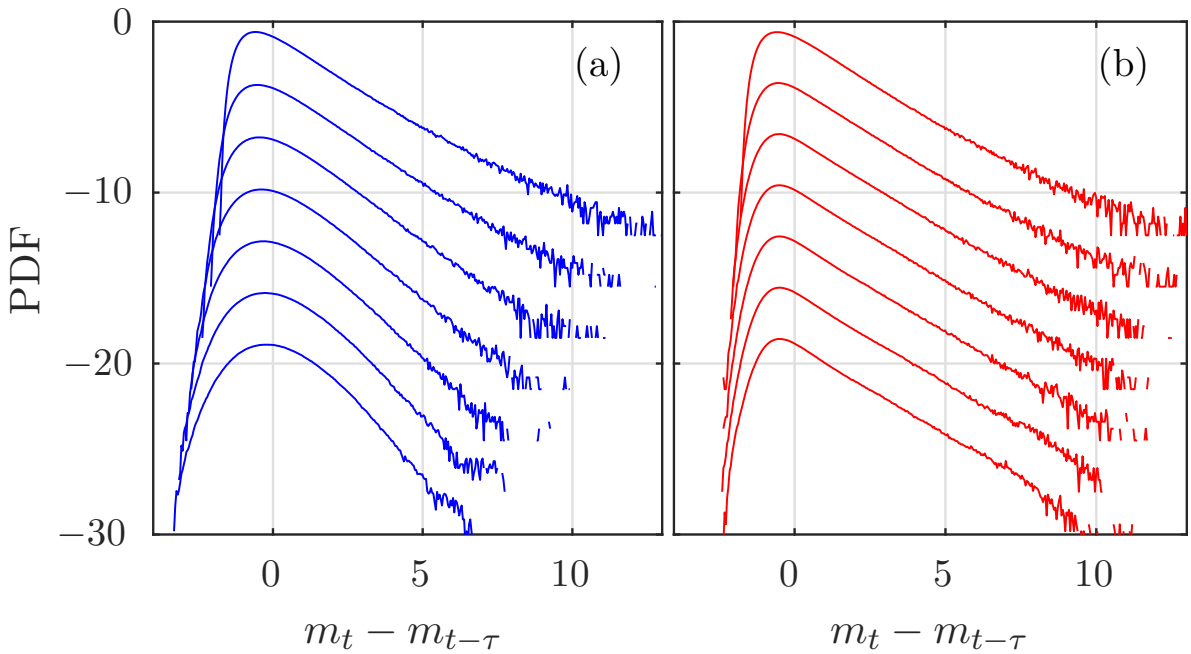

Figure 8. PDF of the increments of the (a) Hermitian and (b) even-Hermitian log-normal motions of size $\tau = 2^j$, from $j = 0$ (bottom) up to $j = 6$ (up). Curves have been arbitrarily shifted on the Y-axis for clarity.

is a popular multiplicative cascade process widely used to model systems that exhibit multifractal properties [42]. Like the fBm, the MRW is a motion obtained by integrating — again with eq. (2) — a stationary noise $W^{\mathrm{MRW}} = \{w_t^{\mathrm{MRW}}\}_{t\in\mathbb{R}}$ such that

$$w_t^{\mathrm{MRW}} = w_t^{\mathrm{fGn}} e^{\omega_t} \tag{35}$$

where $W^{\mathrm{fGn}} = \{w_t^{\mathrm{fGn}}\}_{t\in\mathbb{R}}$ is a fGn with parameter $\mathcal{H}^{\mathrm{fGn}}$ and $\Omega = \{\omega_t\}_{t\in\mathbb{R}}$ is a Gaussian random process, independent of X^{fGn} with a correlation function

$$\begin{aligned} c_\omega(\tau) &= -c_2 \log\left(\frac{L}{|\tau|+1}\right) && \text{if} \quad |\tau| < L \\ &= 0 && \text{otherwise} \end{aligned} \tag{36}$$

where L is the integral scale, set here to $L = T$.

The MRW is a scale invariant process: the power spectrum of its time-derivative W^{MRW} behaves as a power law with an exponent $2(\mathcal{H}^{\mathrm{fGn}} - c_2) + 1$, which would be the Hurst exponent obtained for a fGn with parameter $\mathcal{H}^{\mathrm{fGn}} - c_2$. Any moment of order q of the increments of size τ behaves as a power law of τ with the exponent $\zeta(q)$. Contrary to the fBm, the MRW is not exactly self similar and exhibits intermittency: $\zeta(q) = \mathcal{H}^{\mathrm{fGn}} q - \frac{c_2}{2} q^2$ is not a linear function of q, as expected for a self-similar process. As a consequence, the shape of the PDF of the increments depends on the scale.

We choose the parameter $\mathcal{H}^{\mathrm{fGn}}$ such that the power spectrum of the noise W^{MRW} is identical to the one of the fBm used in the former sections, *i.e.*, $\mathcal{H}^{\mathrm{fGn}} = 0.7 + c_2$. We set the parameter $c_2 = 0.025$, a value widely used to model the intermittency of Eulerian turbulent velocity field [23].

Figure 9 compares the evolution of the PDF of the increments of the fBm and the MRW. As expected, no change is observed for the fBm, while the PDF of the MRW has wider tails for smaller τ. The fBm is perfectly self-similar, while the MRW exhibits intermittency [43]: the PDF of its increments

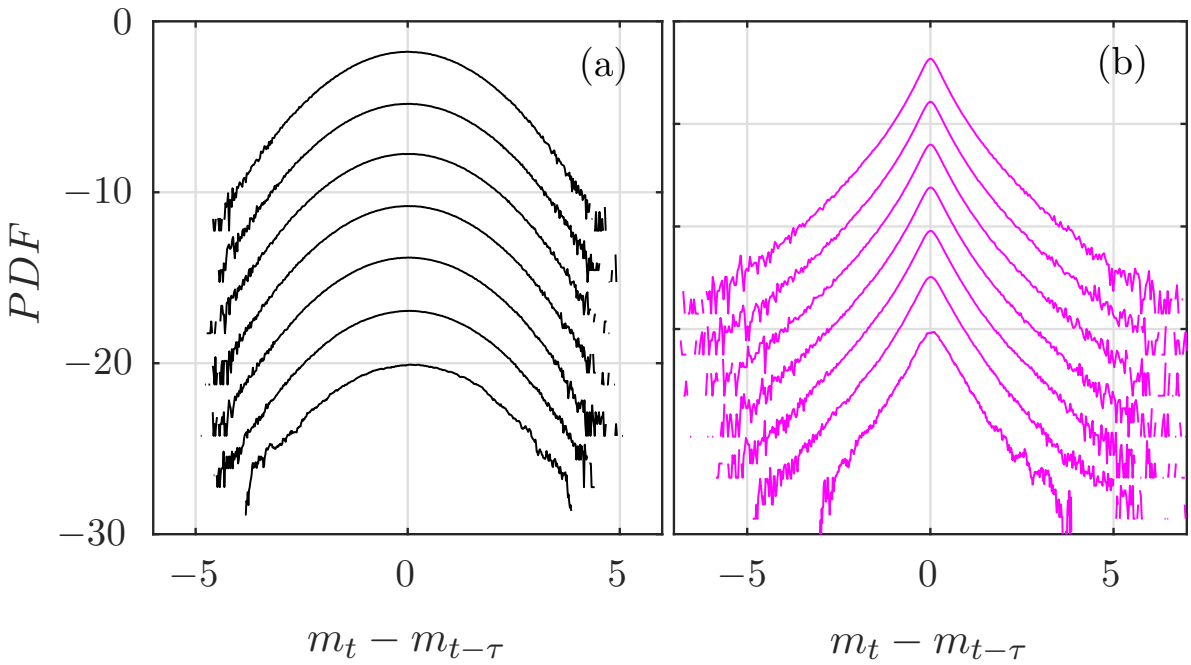

Figure 9. PDF of the increments of the (a) fBm and (b) of a MRW of size $\tau = 2^j$, from $j = 0$ (bottom) up to $j = 6$ (up). Curves have been arbitrarily shifted on the Y-axis for clarity.

is deformed when the scale τ of the increments is varied, although no analytical expression of the PDF is available.

We apply our practical framework and plot in Fig. 10a) the evolution of the ersatz entropy rate of the MRW with T. Again, the entropy rate seems to be independent of T. We nevertheless observe a small tendency to increase towards the value H_1^{MRW} the entropy of the MRW at unit-time. Here, because there is no analytical expression of the PDF, H_1^{MRW} cannot be derived analytically and we numericaly estimate its value.

The dependence in τ is plotted in Fig. 10b). We again observe a strong linear evolution of the ersatz entropy rate in $\mathcal{H} \ln \tau$. After subtracting this strong tendency (Fig. 11), we still observe an evolution with τ, but this evolution appears much weaker than for the Hermitian log-normal (blue curve in Fig. 7). Indeed, the deformation of the PDFs of the increments when varying τ is much slower for the MRW (Fig. 9b) than for the Hermitian log-normal (Fig. 8a).

6. Discussion and Conclusions

We proposed a new framework in information theory to analyze a non-stationary process by considering it as resulting from a gedanken stationary process and estimating the PDF by cumulating all available samples in a time interval of size T. This framework hence considers a PDF obtained by time-averaging over a time window $[t_0 : t_0 + T]$, and then proceeds to compute the associated information theory quantities. In particular, the ersatz entropy $\bar{H}_T(X)$ that is then defined can be interpreted as the amount of information characterizing the complete trajectory $\{X_t, t \in [t_0, t_0 + T]\}$ of the process X. If we assume that the increments of X are stationary and centered, then $\bar{H}_T(X)$ and all other information ersatz theoretical quantities depend only on the duration T and not of the first time t_0.

We illustrated our approach by focusing first on a model system: the fractional Brownian motion. We derived in this context the analytical expressions of the ersatz entropy, ersatz auto-mutual information and ersatz entropy rate, which allowed a pedagogical description of our new information theory quantities. We also reported how the ersatz quantities behave when the time-interval size

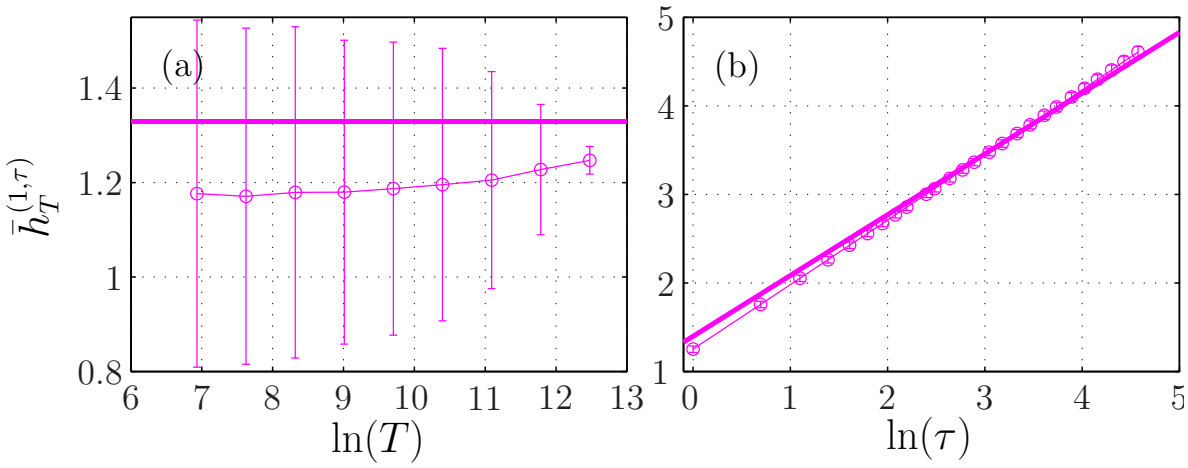

Figure 10. Ersatz entropy rate $\bar{h}_T^{(m=1,\tau)}$ of a MRW with $\mathcal{H} = 0.7$. (a): as a function of the window size T for fixed $\tau = 1$. (b): as a function of the scale τ for $T = 2^{16}$. The horizontal line in (a) indicates the numerical value H_1^{MRW} of the noise. The straight line in (b) has a slope $\mathcal{H} = 0.7$.

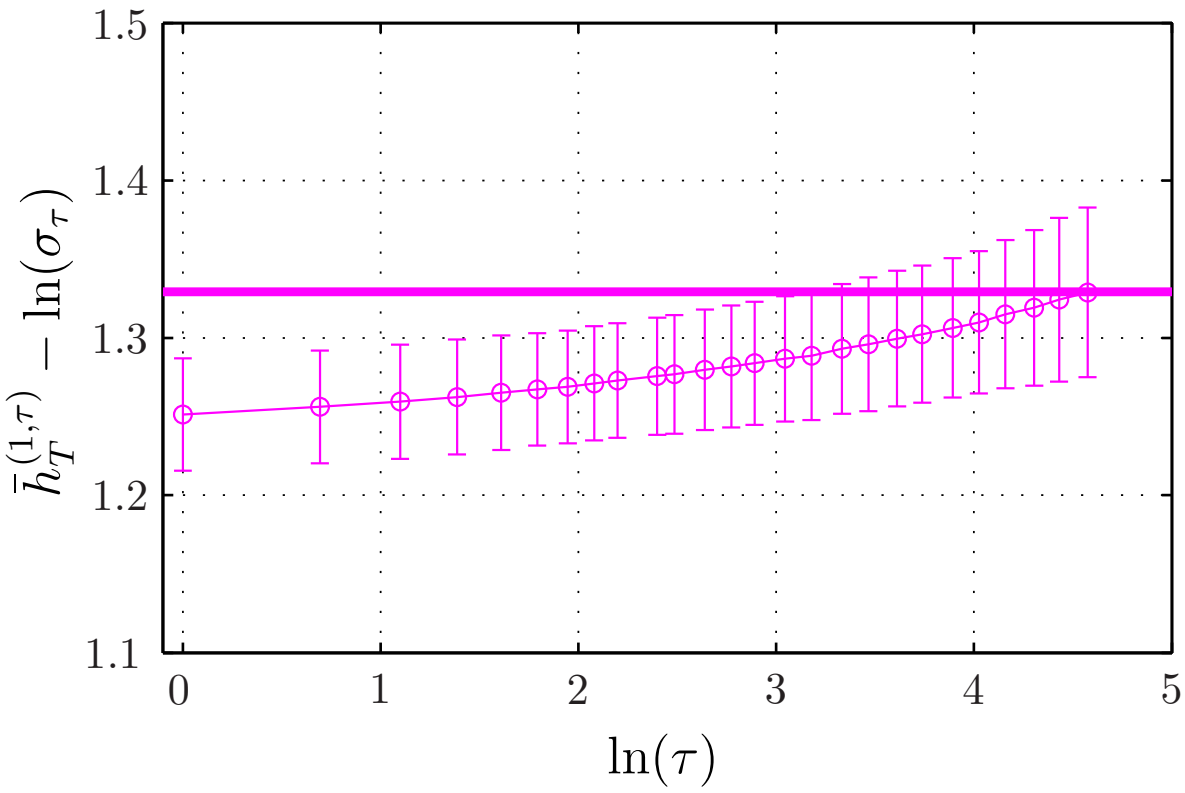

Figure 11. Ersatz entropy rate $\bar{h}_T^{(m=1,\tau)} - \ln(\sigma_\tau)$ of the MRW as a function of $\ln(\tau)$ for fixed $T = 2^{16}$.

T and the embedding time scale τ are varied: we obtained analytical expressions for embedding dimension $m = 1$, and confirmed them numerically for $m \geq 1$. Besides the fBm, we reported numerical observations for various self similar or multifractal processes. The ersatz entropy $\bar{H}_T^{(m,\tau)}$ always diverges logarithmically in T while the ersatz entropy rate $\bar{h}_T^{(m,\tau)}$ always behaves as almost independent of T. The examination of how the ersatz entropy rate $\bar{h}_T^{(m,\tau)}$ depends on the scale τ provides a fine exploration of either the self-similarity or the multifractality of the process.

This exploration of the multifractality of a non-stationary process with stationary increments using the ersatz entropy rate $\bar{h}_T^{(m,\tau)}(M)$ gives a viewpoint very similar to the one reported when analyzing the increments of the process with the regular Shanon entropy, as reported in [43]. We are currently investigating how to relate quantitatively the two approaches.

In the same vein, the ersatz entropy rate allowed us to discriminate two different non-stationary processes, and obtain fine differences in their self similarity properties (figure 7), in close relation to a method using the entropy rate of the increments of the signal, as exposed in [16]. A possible connection is also under investigation.

Through all this article, we have estimated the ersatz quantities of a process on a single trajectory $[t_0; t_0 + T]$ of this process; this situation corresponds to the worst case scenario where only a single realization of the process is know. If enough experimental data are available, one can improve the estimation of the ersatz quantities in two ways. First, if the same experiment has been conducted multiple times, and thus multiple realizations are available over the time interval $[t_0; t_0 + T]$, one can use all these independent realizations to enhance the estimation of the time-averaged PDF. Second, if a single but long enough realization of size $\mathcal{T}$ is available, one can split it into multiple time intervals $[kT; (k+1)T]$, $k \in [0..\lfloor \mathcal{T}/T \rfloor]$ and the use these intervals as independent realizations as in the first case. This later situation is made possible by the assumption that the increments of the signal are not only stationary, but also centered.

Appendix Entropy of a time-embedded signal

The time-embedded vector $\mathbf{x}_t^{(m,\tau)}$ (eq. 6) can be mapped into the vector $\tilde{\mathbf{x}}_t^{(m,\tau)} \equiv (x_t, \delta_\tau x_t, \delta_\tau x_{t-\tau}, ..., \delta_\tau x_{t-(m-2)\tau})$ by the linear transformation Q^m:

$$\mathbf{x}_t^{(m,\tau)} \mapsto \tilde{\mathbf{x}}_t^{(m,\tau)} = Q^m.\mathbf{x}_t^{(m,\tau)}, \tag{A1}$$

where Q^m is the band matrix defined as:

$$Q^m \equiv \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 0 \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{pmatrix}. \tag{A2}$$

which determinant satisfies $|\det(Q_m)| = 1$. As a consequence, $H(\tilde{\mathbf{x}}_t^{(m,\tau)}) = H(\mathbf{x}_t^{(m,\tau)})$, which proves (8).

Acknowledgments: The authors wish to thank L. Chevillard for stimulating discussions. 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).

References

1. Andreas, E.; Geiger, C.; Treviño, G.; Claffey, K. Identifying nonstationarity in turbulence series. *Boundary-Layer Meteorology* **2008**, *127*, 37–56.

2. Nerini, D.; Besic, N.; Sideris, I.; Germann, U.; Foresti, L. A non-stationary stochastic ensemble generator for radar rainfall fields based on the short-space Fourier transform. *Hydrology and Earth System Sciences* **2017**, *21*, 2777–2797.
3. Boashash, B.; Azemi, G.; O'Toole, J. Time-frequence processing of nonstationary signals. *IEEE Signal Processing Magazine* **2013**, *November*, 108–119.
4. Couts, D.; Grether, D.; Nerlove, M. Forecasting non-stationary economic time series. *Management Science* **1966**, *18*, 1–151.
5. Young, P. Time-variable parameter and trend estimation in non-stationary economic time series. *Journal of Forecasting* **1994**, *13*, 179–210.
6. Yang, K.; Shahabi, C. On the stationarity of multivariate time series for correlation-based data analysis. Fifth IEEE International Conference on Data Mining (ICDM'05), 2005.
7. Dębowski, L. On processes with summable partial autocorrelations. *Statistics and Probability Letters* **2007**, *77*, 752–759.
8. Yaglom, A. Correlation theory of processes with random stationary nth increments. *Mat. Sb. (N.S.)* **1955**, *37(79)*, 141–196.
9. Ibe, O., 11-Levy processes. In *Markov processes for stochastic modeling*; Elsevier, 2013; pp. 329–347.
10. Frisch, U. *Turbulence: the legacy of A.N. Kolmogorov*; Cambridge University Press, 1995.
11. Shannon, C. A Mathematical Theory of Communication. *The Bell System Technical Journal* **1948**, *XXVII*, 388–427.
12. Kantz, H.; Schreiber, T. *Nonlinear time series analysis*; Cambridge University Press, 2003.
13. Vu, V.Q.; Yu, B.; Kass, R.E. Information In The Non-Stationary Case. *Neural Computation* **2009**, *21*, 688–703.
14. Ray, A.; Chowdhury, A.R. On the characterization of non-stationary chaotic systems: Autonomous and non-autonomous cases. *Physica A* **2010**, *389*, 5077–5083.
15. Gómez Herrero, G.; Wu, W.; Rutanen, K.; Soriano, M.C.; Pipa, G.; Vicente, R. Assessing coupling dynamics from an ensemble of time series. *Entropy* **2015**, *17*, 1958–1970.
16. Granero-Belinchón, C.; Roux, S.; Abry, P.; Garnier, N.B. Probing high-order dependencies with information theory. *IEEE Transactions on Signal Processing* **2019**, *67*, 3796–3805.
17. Mandelbrot, B.; Van Ness, J. Fractional brownian motions fractional noises and applications. *SIAM Review* **1968**, *10*, 422–437.
18. Takens, F., Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick: proceedings of a symposium held at the University of Warwick 1979/80*; Springer Berlin Heidelberg, 1981; pp. 366–381.
19. Granero-Belinchon, C.; Roux, S.; Abry, P.; Doret, M.; Garnier, N. Information Theory to Probe Intrapartum Fetal Heart Rate Dynamics. *Entropy* **2017**, *19*, 640.
20. Crutchfield, J.; Feldman, D. Regularities unseen, randomness observed: The entropy convergence hierarchy. *Chaos* **2003**, *15*, 25–54.
21. Mandelbrot, B. *The fractal geometry of nature*; W.H. Freeman and Co., 1982.
22. Mauritz, K. Dielectric relaxation studies of ion motions in electrolyte-containing perfluorosulfonate ionomers: 4. long-range ion transport. *Macromolecules* **1989**, *22*, 4483–4488.
23. Chevillard, L.; Castaing, B.; Arneodo, A.; Lévêque, E.; Pinton, J.; Roux, S. A phenomenological theory of Eulerian and Lagrangian velocity fluctuations in turbulent flows. *Comptes Rendus Physique* **2012**, *13*, 899 – 928.
24. Kavvas, M.; Govindaraju, R.; Lall, U. Introduction to the focus issue: physics of scaling and self-similarity in hydrologic dynamics, hydrodynamics and climate. *Chaos: an interdisciplinary journal of nonlinear science* **2015**, *25*, 075201.
25. Rigon, R.; Rodriguez-Iturbe, I.; Maritan, A.; Giacometti, A.; Tarboton, D.; Rinaldo, A. On Hack's law. *Water Resources Research* **1996**, *32*, 3367–3374.
26. Gotoh, K.; Fujii, Y. A fractal dimensional analysis on the cloud shape parameters of cumulus over land. *Journal of Applied Meteorology* **1998**, *37*, 1283–1292.
27. Console, R.; Lombardi, A.; Murru, M.; Rhoades, D. Bath's law and the self-similarity of earthquakes. *Journal of Geophysical Research: Solid Earth* **2003**, *108*, 2128.
28. Ivanov, P.C.; Ma, Q.D.Y.; Bartsch, R.P.; Hausdorff, J.M.; A. Nunes, A.L.; Schulte-Frohlinde, V.; Stanley, H.; Yoneyama, M. Levels of complexity in scale-invariant neural signals. *Physical Review E* **2009**, *79*, 041920.

29. Drozdz, S.; Ruf, F.; Speth, J.; Wojcik, M. Imprints of log-periodic self-similarity in the stock market. *The European Physical Journal B– Condensed Matter and Complex Systems* **1999**, *10*, 589–593.
30. Cont, R.; Potters, M.; Bouchaud, J.P. Scaling in stock market data: stable laws and beyond. Scale Invariance and Beyond, 1997, Vol. 7, pp. 75–85.
31. Uhl, A.; Wimmer, G. A systematic evaluation of the scale invariance of texture recognition methods. *Pattern Analysis and Applications* **2015**, *18*, 945–969.
32. Chakraborty, D.; Ashir, A.; Suganuma, T.; Mansfield-Keeni, G.; Roy, T.; Shiratori, N. Self-similar and fractal nature of internet traffic. *Network Management* **2004**, *14*, 119–129.
33. Flandrin, P. Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Transactions on Information Theory* **1992**, *38*, 910–917.
34. Zografos, K.; Nadarajah, S. Expressions for Rényi and Shannon entropies for multivariate distributions. *Statistics and Probability Letters* **2005**, *71*, 71–84.
35. Granero-Belinchon, C.; Roux, S.G.; Garnier, N.B. Scaling of information in turbulence. *Epl* **2016**, *115*, 58003.
36. Helgason, H.; Pipiras, V.; Abry, P. Synthesis of multivariate stationary series with prescribed marginal distributions and covariance using circulant matrix embedding. *Signal Processing* **2011**, *91*, 1741–1758.
37. Kozachenko, L.; Leonenko, N. Sample estimate of entropy of a random vector. *Problems of Information Transmission* **1987**, *23*, 95–100.
38. Kraskov, A.; Stöbauer, H.; Grassberger, P. Estimating mutual information. *Physical Review E* **2004**, *69*, 066138.
39. Gao, W.; Oh, S.; Viswanath, P. Demystifying Fixed k-Nearest Neighbor Information Estimators. *IEEE Transactions on Information Theory* **2018**, *64*, 5629–5661.
40. Bacry, E.; Delour, J.; Muzy, J.F. Multifractal random walk. *Physical Review E* **2001**, *64*, 026103.
41. Bacry, E.; Muzy, J. Multifractal stationary random measures and multifractal random walk with log-infinitely divisible scaling laws. *Physical Review E* **2002**, *66*, 056121.
42. Delour, J.; Muzy, J.; Arnéodo, A. Intermittency of 1D velocity spatial profiles in turbulence: a magnitude cumulant analysis. *The European Physical Journal B* **2001**, *23*, 243–248.
43. Granero-Belinchón, C.; Roux, S.G.; Garnier, N.B. Kullback-Leibler divergence measure of intermittency: Application to turbulence. *Physical Review E* **2018**, *97*, 013107.