

## Probabilistic analysis of recurrence plots generated by fractional Gaussian noise

Sofiane Ramdani, Frédéric Bouchara, and Annick Lesne

Citation: *Chaos* **28**, 085721 (2018); doi: 10.1063/1.5030522

View online: <https://doi.org/10.1063/1.5030522>

View Table of Contents: <http://aip.scitation.org/toc/cha/28/8>

Published by the [American Institute of Physics](#)

---

---



**Chaos**  
An Interdisciplinary Journal of Nonlinear Science

**Fast Track Your Research. *Submit Today!***

# Probabilistic analysis of recurrence plots generated by fractional Gaussian noise

Sofiane Ramdani,<sup>1,a)</sup> Frédéric Bouchara,<sup>2</sup> and Annick Lesne<sup>3,4</sup>

<sup>1</sup>LIRMM, University of Montpellier, CNRS, Montpellier, France

<sup>2</sup>Université de Toulon, Aix Marseille University, CNRS, LIS, Toulon, France

<sup>3</sup>Sorbonne Université, CNRS, Laboratoire de Physique Théorique de la Matière Condensée, LPTMC, F-75252 Paris, France

<sup>4</sup>Institut de Génétique Moléculaire de Montpellier, University of Montpellier, CNRS, Montpellier, France

(Received 22 March 2018; accepted 10 August 2018; published online 30 August 2018)

Recurrence plots of time series generated by discrete fractional Gaussian noise (fGn) processes are analyzed. We compute the probabilities of occurrence of consecutive recurrence points forming diagonals and verticals in the recurrence plot constructed without embedding. We focus on two recurrence quantification analysis measures related to these lines, respectively, the percent determinism and the laminarity (*LAM*). The behavior of these two measures as a function of the fGn's Hurst exponent  $H$  is investigated. We show that the dependence of the laminarity with respect to  $H$  is monotonic in contrast to the percent determinism. We also show that the length of the diagonal and vertical lines involved in the computation of percent determinism and laminarity has an influence on their dependence on  $H$ . Statistical tests performed on the *LAM* measure support its utility to discriminate fGn processes with respect to their  $H$  values. These results demonstrate that recurrence plots are suitable for the extraction of quantitative information on the correlation structure of these widespread stochastic processes. *Published by AIP Publishing.* <https://doi.org/10.1063/1.5030522>

**Discrete fractional Gaussian noises are ubiquitous stochastic processes depending on a single parameter, the Hurst exponent  $H$ , which entirely describes their time correlations. White noise is recovered for  $H=0.5$ . The variation with  $H$  of two remarkable features of their recurrence plots (the determinism and the laminarity) is analyzed by confronting exact probabilistic results in infinite-size and finite-size simulation. This investigation provides new results about recurrence quantification analysis from diagonal and vertical lines, presents a methodology yielding their analytical derivation for infinite-size recurrence plots, and overall opens a novel research direction for the estimation of the exponent  $H$ .**

## I. INTRODUCTION

In 1987, Eckmann, Kamphorst, and Ruelle<sup>1</sup> introduced the concept of recurrence plots (RPs) to analyze data produced by nonlinear dynamical systems. Few years later, recurrence quantification analysis (RQA) was proposed by Zbilut and Webber<sup>2,3</sup> to extract quantitative information from recurrence plots (RPs). Since then, many works have shown that several measures can be derived from RPs<sup>4,5</sup> which provide information on the predictability, the stationarity, the cyclicities, or the laminar nature of the underlying dynamics or process generating the considered data.

After the recurrence rate (*REC*), the measures introduced next were related to diagonal structures that are parallel to the main diagonal of the RP. Among these measures, the percent determinism (*DET*) is certainly the best acknowledged.

*DET* is related to the predictability of the underlying process generating the analyzed data.<sup>4</sup>

Some years later, Marwan *et al.*<sup>6</sup> proposed new measures derived from the vertical structures of RPs. In particular, a measure inspired by *DET* was defined and called laminarity *LAM*. The recurrence points forming these vertical lines are related to the concept of *sojourn* points introduced by Gao.<sup>7,8</sup> It has been shown that such vertical structures can be related to intermittency phenomena and useful for the detection of chaos-chaos or chaos-order transitions.<sup>4,6</sup>

Although RPs and RQA were first intended to investigate deterministic chaotic dynamics, studies have shown that they can also be used to explore data generated by stochastic processes. More specifically, the measures derived from the diagonal structures of the RPs have been shown to be related to the time correlations of the considered random process (see Refs. 4 and 9 and references therein).

In the present work, the statistical properties of RPs of data generated by discrete fractional Gaussian noise (fGn) are investigated. This stochastic process was introduced in the seminal work of Mandelbrot and van Ness as the increment process of fractional Brownian motion.<sup>10</sup> Such processes have been widely used to model random time series in several fields such as biology, physics, or finance.<sup>11</sup> They are characterized by the so-called Hurst exponent  $H$ , which is the parameter specifying the autocovariance of the process.

Extending a methodology introduced in a previous paper,<sup>9</sup> we compute the measures *REC*, *DET*, and *LAM* of fGn processes RPs constructed without embedding (i.e., with an embedding dimension equal to 1). The effect on these measures of varying the  $H$  exponent is explored. The results show that, unlike the *DET* measure, the laminarity *LAM* is

<sup>a)</sup>sofiane.ramdani@umontpellier.fr

a monotonic increasing function of  $H$ . These findings suggest that the vertical structures of RPs of processes that can be modeled by fGn can be used to characterize their time correlations.

## II. DEFINITIONS OF DET AND LAM MEASURES

In this section, the RP construction and the definitions of *DET* and *LAM* indices are briefly recalled.

Let  $(x_i)$  be a time series. The first step of the RP construction is to perform the time delay embedding procedure<sup>12–14</sup> to define the time delay vector  $\mathbf{x}_i$  of dimension  $d$  and using a time delay  $\tau$ :

$$\mathbf{x}_i = [x_i, x_{i+\tau}, x_{i+2\tau}, \dots, x_{i+(d-1)\tau}]^T. \quad (1)$$

Then using some norm (Euclidean or maximum norm in general), all the distances between  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are compared to a given threshold (or radius)  $\varepsilon$ . The RP is simply the representation of the so-called recurrence points corresponding to locations  $(i, j)$  for which the distance between  $\mathbf{x}_i$  and  $\mathbf{x}_j$  is less than  $\varepsilon$ . No dot is represented otherwise.

The first measure currently computed is the recurrence rate *REC*, which is simply the rate of recurrence points. The second measure is the rate of recurrence points forming lines of length at least  $n$  that are parallel to the main diagonal (defined by  $i = j$ ). This so-called percent determinism *DET* is given by<sup>4</sup>

$$DET(\varepsilon, n) = \frac{\sum_{k=n}^{N_v} kJ_k(\varepsilon)}{\sum_{k=1}^{N_v} kJ_k(\varepsilon)}, \quad (2)$$

where  $J_k(\varepsilon)$  is the number of diagonal lines of length exactly  $k$  and  $N_v$  is the size of the RP.

Similarly, one can define the rate of recurrence points forming vertical lines of length at least  $n$ , namely, the laminarity *LAM* given by<sup>4,6</sup>

$$LAM(\varepsilon, n) = \frac{\sum_{k=n}^{N_v} kV_k(\varepsilon)}{\sum_{k=1}^{N_v} kV_k(\varepsilon)}, \quad (3)$$

where  $V_k(\varepsilon)$  is the number of vertical lines of length exactly  $k$ .

## III. FRACTIONAL GAUSSIAN NOISE

Fractional Gaussian noise (fGn) is defined as the increment process of fractional Brownian motion (fBm) (see Refs. 10 and 11). It is a centered, stationary, and Gaussian process. In the discrete-time case, the autocovariance function  $\gamma$  is given by

$$\gamma(k) = \frac{\sigma^2}{2} (|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}) \quad (4)$$

for  $k \in \mathbb{Z}$  and where  $\sigma^2$  is the variance of the process (we consider here  $\sigma^2 = 1$ ) and  $H$  is the Hurst exponent ( $0 < H < 1$ ,  $H = 0.5$  corresponding to white noise). Note that  $\gamma(-k) = \gamma(k)$ .

A classical approach for the simulation of fGn's sample paths is based on the Cholesky decomposition of the covariance matrix  $\mathbf{\Gamma}$ , whose  $(i, j)$  entry can be written  $\Gamma_{ij} = \gamma(i - j)$  due to covariance stationarity. This procedure is quite straightforward. Let  $\mathbf{L}$  be the lower triangular matrix related to the symmetric positive definite matrix  $\mathbf{\Gamma}$  according to the equality  $\mathbf{\Gamma} = \mathbf{L}\mathbf{L}^T$ . Denoting  $\boldsymbol{\eta}$  a  $N$ -dimensional column vector of white Gaussian noise samples, a time series with the desired fGn properties is obtained by computing  $\mathbf{x} = \mathbf{L}\boldsymbol{\eta}$ .<sup>11</sup>

Figure 1 shows the RPs of two sample paths of fGn processes for two different values of the parameter  $H$ .

## IV. THEORETICAL ANALYSIS

In this section, the methodology used to compute the theoretical values of the percent determinism and laminarity of time series generated by discrete fGns is described. These theoretical counterparts are denoted  $DET_{th}$  and  $LAM_{th}$ . We should emphasize at this point that the term theoretical in this context indicates that the RQA measures correspond to theoretical infinite RPs. Thus, one can refer to these theoretical quantities as asymptotic. To achieve these computations, the derivation of the probabilities of the occurrence of a diagonal or a vertical line is necessary.

The methodology is based on a previous work for which only diagonal lines were considered.<sup>9</sup> It should be underlined that this approach is not specific to fGns but is generic for the analysis of the RP measures of any discrete-time stationary Gaussian stochastic process.

Thus, in this section, we consider that  $\mathbf{x}$  is a real-valued, discrete time, wide-sense stationary, centered Gaussian stochastic process, with unit variance. For the sake of simplicity, the random variables and their realizations are denoted with the same symbols.

For the rest of the analysis, we will consider the case of RPs constructed without embedding. In this case, the vector  $\mathbf{x}_i$  reduces to the scalar  $x_i$  [see Eq. (1)].

In Secs. IV A–IV G, we will present the computations of the theoretical probabilities:

- $P_{i,j}$ : the probability of the occurrence of a recurrence point at location  $(i, j)$  of the RP.
- $P_{i,j}^n$ : the probability of the occurrence of a diagonal of length  $n$ , starting from point  $(i, j)$ .
- $Q_{i,j}^n$ : the probability of the occurrence of a diagonal of length exactly  $n$ , starting from point  $(i, j)$ .
- $T_{i,j}^n$ : the probability of the occurrence of a vertical of length  $n$ , starting from point  $(i, j)$ .
- $U_{i,j}^n$ : the probability of the occurrence of a vertical of length exactly  $n$ , starting from point  $(i, j)$ .

The use of an embedding dimension equal to 1 is motivated by technical aspects of the computation of probabilities  $P_{i,j}^n$  and  $T_{i,j}^n$ . As it can be seen in Secs. IV B and IV E below, considering an embedding dimension  $d > 1$  introduces an overlapping of the time delay vectors involved in the computation of these probabilities, which will result in a much more complex formalization of the inequalities defining diagonal or vertical lines. In addition, taking  $d = 1$  automatically discard

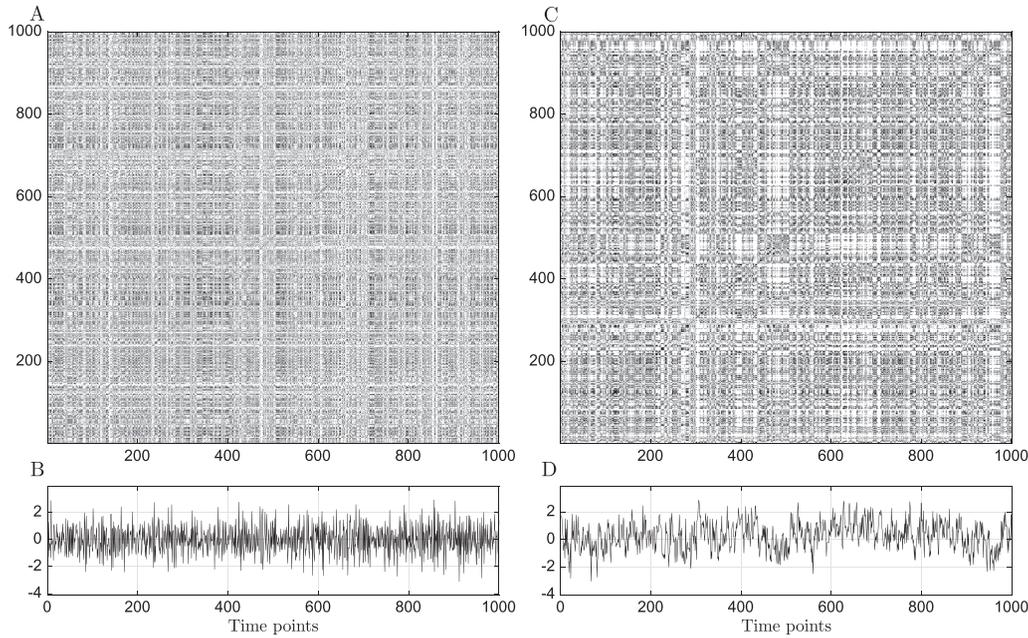


FIG. 1. Recurrence plots [(a),(c)] of unit-variance fGn sample paths [(b),(d)] with  $H = 0.20$  and  $H = 0.80$ , respectively. The RPs were constructed without embedding and using a threshold  $\varepsilon = 0.5$ .

the issue of the selection of a time delay, which is not straightforward for data generated by fGn processes. This point will be discussed in Sec. VI.

By definition,  $DET_{th}$  and  $LAM_{th}$  can be expressed as functions of  $Q_{ij}^n$  and  $U_{ij}^n$ . In the following, we will show that the probabilities  $Q_{ij}^n$  and  $U_{ij}^n$  can be derived from the probabilities  $P_{ij}^n$  and  $T_{ij}^n$ . Note also that in order to simplify the notations of these probabilities, their dependence on the threshold  $\varepsilon$  is not explicitly mentioned.

### A. Computation of the probability $P_{i,j}$

In the case of an embedding dimension  $d = 1$ , the computation of the probability of the occurrence of a recurrence point at location  $(i, j)$  of the RP was addressed in a previous work.<sup>9</sup> For a stochastic process  $\mathbf{x}$ , with variance  $\sigma^2$  and autocovariance function  $\gamma$ , satisfying the above mentioned conditions, it can be written

$$P_{i,j} = \operatorname{erf}\left(\frac{\varepsilon}{\sqrt{2\alpha_{i,j}}}\right), \quad (5)$$

with  $\alpha_{i,j} = 2[\sigma^2 - \gamma(i-j)]$  for  $i \neq j$ . In the case  $i = j$ , we simply get  $P_{i,i} = 1$ . For  $|i-j|$  large enough, the probability  $P_{i,j}$  is independent of  $(i, j)$  and it provides an approximation of the theoretical recurrence rate  $REC_{th}$ .<sup>9</sup> Thus, for  $|i-j|$  large enough,  $P_{i,j}$  can be simply denoted  $P$ . The evolution of  $P_{i,j}$  as a function of  $|i-j|$ , shown in Sec. V, provides a quantitative support of this assumption.

### B. Computation of the probability $P_{i,j}^n$

To compute  $P_{i,j}^n$ , a method developed in Ref. 9 for the general case of a stationary Gaussian random process  $\mathbf{x}$  is applied. On a recurrence plot constructed without embedding,  $P_{i,j}^n$  is the probability of the occurrence of  $n$  joint events given by  $|x_i - x_j| \leq \varepsilon$ ,  $|x_{i+1} - x_{j+1}| \leq \varepsilon, \dots, |x_{i+n-1} - x_{j+n-1}| \leq \varepsilon$ .

If  $\mathbf{y}_{ij}^n$  is an  $n$ -dimensional random vector given by

$$\mathbf{y}_{ij}^n = (x_i - x_j, x_{i+1} - x_{j+1}, \dots, x_{i+n-1} - x_{j+n-1})^T. \quad (6)$$

Thus,  $P_{i,j}^n(\varepsilon)$  is the probability to have  $\|\mathbf{y}_{ij}^n\|_\infty \leq \varepsilon$ , where  $\|\cdot\|_\infty$  is the maximum norm.

Note that  $\mathbf{y}_{ij}^n$  is the difference of two Gaussian random vectors of dimension  $n$ , respectively, constituted of the random variables  $x_i$  et  $x_j$  and their  $(n-1)$  following variables. Consequently, the random vector  $\mathbf{y}_{ij}^n$  is also Gaussian as a vector composed of differences of joint normal components of the process  $\mathbf{x}$ .<sup>15,16</sup>

Let  $\Omega_{k,l}^{i,j}$  be the covariance matrix of the random vector  $\mathbf{y}_{ij}^n$ . The components of this matrix are defined by

$$\Omega_{k,l}^{i,j} = \langle (x_{i+k-1} - x_{j+k-1})(x_{i+l-1} - x_{j+l-1}) \rangle, \quad (7)$$

for  $(k, l) \in \{1, 2, \dots, n\}^2$ , where the brackets  $\langle \cdot \rangle$  denote expectation.

Consequently, the probability density function (pdf) associated with  $\mathbf{y}_{ij}^n$  is given by the following multivariate Gaussian function:

$$f_{ij}^n(\mathbf{y}) = \frac{1}{(2\pi)^{n/2} |\Omega^{i,j}|^{1/2}} \exp\left[-\frac{1}{2} \mathbf{y}^T (\Omega^{i,j})^{-1} \mathbf{y}\right], \quad (8)$$

where  $|\Omega^{i,j}|$  is the determinant of the matrix  $\Omega^{i,j}$ .

The probability  $P_{i,j}^n$  to have a diagonal of  $n$  consecutive recurrence points starting from a point  $(i, j)$  on the RP reads

$$P_{i,j}^n = \int_{\mathbb{M}(\varepsilon)} f_{ij}^n(\mathbf{y}) d\mathbf{y}, \quad (9)$$

with  $\mathbb{M}(\varepsilon) = \{\mathbf{y} : \mathbf{y} \in \mathbb{R}^n, \|\mathbf{y}\|_\infty \leq \varepsilon\}$ .

In the generic case, this probability can not be calculated analytically, but it can be estimated numerically. To achieve this, the method proposed by Genz<sup>17</sup> is used. A description of this method can be found in the Appendix of Ref. 9.

It can be numerically shown that  $P_{ij}^n$  is approximately independent of  $(i, j)$  for  $|i - j| \gg 1$ , i.e., for points  $(i, j)$  that are not close to the main diagonal of the RP. This can be justified by the stationarity of the analyzed random process.<sup>9</sup> This will be illustrated in Sec. V and the dependence of  $P_{ij}^n$  with respect to  $|i - j|$  will be shown for different values of  $H$ .

### C. Computation of the probability $Q_{ij}^n$

The computation of  $Q_{ij}^n$  can be achieved in the general case and without any assumption on the nature of the pdf of the stochastic process. It can be expressed as follows:

$$Q_{ij}^n = (P_{ij}^n - P_{ij}^{n+1}) - (P_{i-1,j-1}^{n+1} - P_{i-1,j-1}^{n+2}), \quad (10)$$

where  $P_{ij}^n$  is given by Eq. (9).

In order to compute the theoretical value  $DET_{th}$ , we also use the fact that  $P_{ij}^n$  is approximately independent of  $(i, j)$  for  $|i - j| \gg 1$ . Thus, the notations  $P_{ij}^n$  and  $Q_{ij}^n$  can be, respectively, replaced with  $P^n$  and  $Q^n$ . Equation (10) then leads to

$$Q^n \simeq P^n - 2P^{n+1} + P^{n+2}, \quad (11)$$

where  $P^n$ ,  $P^{n+1}$ , and  $P^{n+2}$  are estimated using Eq. (9) for  $|i - j| \gg 1$ .

### D. Computation of the theoretical percent determinism $DET_{th}$

After a normalization of the numerator and denominator in the definition, Eq. (2), of the empirical  $DET$ , the following expression for its theoretical (i.e., asymptotic) counterpart is obtained:

$$DET_{th}(\varepsilon, n) = \frac{\sum_{k \geq n} kQ^k}{\sum_{k \geq 1} kQ^k}. \quad (12)$$

For the numerical computation of  $DET_{th}$ , a reasonable approximation can be obtained by considering only the first terms of the involved sums. This is a consequence of the fast vanishing of probability  $Q^k$  with  $k$ , especially for small  $\varepsilon$  values. A different and refined approximation can be derived by replacing  $Q^k$  with the expression given by Eq. (11). According to this result, the numerator of  $DET_{th}$  can be written

$$\sum_{k \geq n} kQ^k \simeq \sum_{k \geq n} k(P^k - 2P^{k+1} + P^{k+2}). \quad (13)$$

If we detail the first terms of the involved sum, we get

$$\begin{aligned} \sum_{k \geq n} kQ^k &\simeq n(P^n - 2P^{n+1} + P^{n+2}) \\ &+ (n+1)(P^{n+1} - 2P^{n+2} + P^{n+3}) \\ &+ (n+2)(P^{n+2} - 2P^{n+3} + P^{n+4}) \\ &+ (n+3)(P^{n+3} - 2P^{n+4} + P^{n+5}) \\ &\dots, \end{aligned} \quad (14)$$

which leads to

$$\begin{aligned} \sum_{k \geq n} kQ^k &\simeq nP^n - 2nP^{n+1} + nP^{n+2} \\ &+ (n+1)P^{n+1} - (2n+2)P^{n+2} + (n+1)P^{n+3} \\ &+ (n+2)P^{n+2} - (2n+4)P^{n+3} + (n+2)P^{n+4} \\ &+ (n+3)P^{n+3} - (2n+6)P^{n+4} + (n+3)P^{n+5} \\ &\dots. \end{aligned} \quad (15)$$

Through this last expression, one can observe that the intermediary terms  $P^{n+2}$ ,  $P^{n+3}$ ,  $P^{n+4}$ , ... cancel out. Thus, an approximation of the numerator of  $DET_{th}$  can be written

$$\sum_{k \geq n} kQ^k \simeq nP^n - (n-1)P^{n+1}. \quad (16)$$

Consequently, for the denominator of  $DET_{th}$  [see Eq. (12)], we get

$$\sum_{k \geq 1} kQ^k \simeq P^1. \quad (17)$$

Note that  $P^1$  is, in fact, equal to the probability  $P$ , that is, the theoretical recurrence rate [see Eq. (5)]. Finally, an approximation of  $DET_{th}$  is given by

$$DET_{th}(\varepsilon, n) \simeq \frac{nP^n - (n-1)P^{n+1}}{P^1}. \quad (18)$$

### E. Computation of the probability $T_{ij}^n$

The probability of the occurrence of a vertical of length  $n$ , starting from point  $(i, j)$  corresponds to the joint events defined by  $|x_i - x_j| \leq \varepsilon$ ,  $|x_i - x_{j+1}| \leq \varepsilon, \dots, |x_i - x_{j+n-1}| \leq \varepsilon$ . Similar to the case of diagonal lines, we introduce a new  $n$ -dimensional random vector  $\mathbf{z}_{ij}^n$  defined by

$$\mathbf{z}_{ij}^n = (x_i - x_j, x_i - x_{j+1}, \dots, x_i - x_{j+n-1})^T. \quad (19)$$

Consequently,  $T_{ij}^n(\varepsilon)$  is the probability to have  $\|\mathbf{z}_{ij}^n\|_\infty \leq \varepsilon$ .

Denoting  $\Phi^{ij}$  the covariance matrix of the Gaussian random vector  $\mathbf{z}_{ij}^n$ , we have

$$\Phi_{k,l}^{ij} = \langle (x_i - x_{j+k-1})(x_i - x_{j+l-1}) \rangle, \quad (20)$$

for  $(k, l) \in \{1, 2, \dots, n\}^2$ . The pdf associated with  $\mathbf{z}_{ij}^n$  is also given by a multivariate Gaussian function

$$g_{ij}^n(\mathbf{z}) = \frac{1}{(2\pi)^{n/2} |\Phi^{ij}|^{1/2}} \exp \left[ -\frac{1}{2} \mathbf{z}^T (\Phi^{ij})^{-1} \mathbf{z} \right]. \quad (21)$$

Finally, the probability  $T_{ij}^n$  to observe a vertical line of  $n$  recurrence points starting from a point  $(i, j)$  on the RP can be written

$$T_{ij}^n = \int_{\mathbb{D}(\varepsilon)} g_{ij}^n(\mathbf{z}) d\mathbf{z}, \quad (22)$$

with  $\mathbb{D}(\varepsilon) = \{\mathbf{z} : \mathbf{z} \in \mathbb{R}^n, \|\mathbf{z}\|_\infty \leq \varepsilon\}$ .

As done for  $P_{ij}^n$ , we exploit the fact that  $T_{ij}^n$  is approximately independent of  $(i, j)$  for  $|i - j| \gg 1$ . This will be numerically validated in Sec. V where the dependence of  $T_{ij}^n$  as function of  $|i - j|$  is depicted for different values of  $H$ .

### F. Computation of the probability $U_{ij}^n$

The computation of the probability  $U_{ij}^n$  of the occurrence of a vertical of length exactly  $n$ , starting from point  $(i, j)$ , can be performed thanks to the reasoning used for the case of diagonals. Thus, we have

$$U_{ij}^n = (T_{ij}^n - T_{ij}^{n+1}) - (T_{i-1, j-1}^{n+1} - T_{i-1, j-1}^{n+2}). \quad (23)$$

Once again, the fact that  $T_{ij}^n$  is approximately independent of  $(i, j)$ , for  $|i - j| \gg 1$ , is used. Replacing the notation  $T_{ij}^n$  with  $T^n$  and  $U_{ij}^n$  with  $U^n$ , we then get from Eq. (23)

$$U^n \simeq T^n - 2T^{n+1} + T^{n+2}, \quad (24)$$

where  $T^n$ ,  $T^{n+1}$ , and  $T^{n+2}$  are computed using Eq. (22) for  $|i - j| \gg 1$ .

### G. Computation of the theoretical laminarity $LAM_{th}$

From Eq. (3) defining the empirical  $LAM$ , its theoretical (i.e., asymptotic) counterpart can be defined by

$$LAM_{th}(\varepsilon, n) = \frac{\sum_{k \geq n} kU^k}{\sum_{k \geq 1} kU^k}. \quad (25)$$

As for the theoretical percent determinism, this expression can be simplified. Indeed, by replacing  $U^n$  with the expression given by Eq. (24) involving  $T^n$ , the following approximation for the denominator of  $LAM_{th}$  is obtained:

$$\sum_{k \geq n} kU^k \simeq \sum_{k \geq n} k(T^k - 2T^{k+1} + T^{k+2}). \quad (26)$$

After some algebra detailed in Eqs. (14) and (15) for the case of diagonal lines and for the computation of  $DET_{th}$ , we get

$$\sum_{k \geq n} kU^k \simeq nT^n - (n - 1)T^{n+1}. \quad (27)$$

Similarly, the denominator of  $LAM_{th}$  is given by

$$\sum_{k \geq 1} kU^k \simeq T^1. \quad (28)$$

The approximation of  $LAM_{th}$  then can be written

$$LAM_{th}(\varepsilon, n) \simeq \frac{nT^n - (n - 1)T^{n+1}}{T^1}. \quad (29)$$

## V. NUMERICAL RESULTS

In this section, we present the results of the computation of the theoretical probabilities  $P_{ij}$ ,  $P_{ij}^n$ , and  $T_{ij}^n$  as a function of  $|i - j|$  for different values of the parameter  $H$  and for  $n = 2$  and  $n = 4$ . Then, we show the results of the numerical computations of the theoretical values  $REC_{th}$ ,  $DET_{th}$ , and  $LAM_{th}$  for a range of  $H$  values (from 0.05 to 0.95) and compare them to the empirical  $REC$ ,  $DET$ , and  $LAM$  values obtained for simulated paths of fGn processes. In the case of  $DET_{th}$  and  $LAM_{th}$ , we also explore the effect of the diagonal and vertical minimal length  $n$  on the obtained values.

The simulation of discrete fGn time series can be performed through different methods.<sup>11,18</sup> A classical approach relies on the Cholesky decomposition of the covariance

matrix of the process. Alternatives are based on the so-called Lowen's circulant method,<sup>19</sup> truncated symmetric moving average filters,<sup>20</sup> or wavelet-based synthesis.<sup>21</sup> For our study, we performed simulations using these three approaches and obtained very similar qualitative results. Thus, we only present here the outcomes of the simulations based on the Cholesky decomposition of the covariance matrix.

According to Eqs. (18) and (29), the computation of  $DET_{th}$  and  $LAM_{th}$  relies, respectively, on the computation of the probabilities  $P^n$  and  $T^n$ . These probabilities are estimated thanks to Eqs. (9) and (22) for  $|i - j| \gg 1$  and are obtained through the integration of the multivariate Gaussian functions defined by Eqs. (11) and (29). This can be achieved using the approach proposed by Genz<sup>17</sup> and detailed in the Appendix of Ref. 9.

The covariance matrices involved in the functions defined by Eqs. (8) and (21) were, respectively, denoted by  $\Omega^{ij}$  and  $\Phi^{ij}$ . According to Eqs. (7) and (20), these matrices are composed of terms that can be expressed through the autocovariance function  $\gamma$  (defined in Sec. III) as follows:

$$\Omega_{k,l}^{ij} = 2\gamma(k - l) - \gamma(i - j + k - l) - \gamma(j - i + k - l), \quad (30)$$

$$\Phi_{k,l}^{ij} = 1 - \gamma[i - (j + l - 1)] - \gamma[(j + k - 1) - i] + \gamma(k - l), \quad (31)$$

for  $(k, l) \in \{1, 2, \dots, n\}^2$ .

For the computations of  $REC_{th}$ ,  $DET_{th}$ , and  $LAM_{th}$ , we set  $|i - j| = 500$  and  $\varepsilon = 0.5$ . This value of the threshold radius  $\varepsilon$  ensured large enough probabilities of occurrence of diagonal and vertical lines in almost all numerically explored cases (values of  $H$  and  $n$ ). Note that we consider zero-mean and unit variance processes. This implies that the selected relative radius corresponds to 50% of the standard deviation of the theoretical process. This value was also selected according to our previously published results regarding a sensitivity analysis with respect to the threshold value.<sup>9</sup> Indeed, for 1000-point sample paths, the findings support that such value ensures reliable estimations of RQA measures, i.e., consistent with their theoretical values and with low variability.

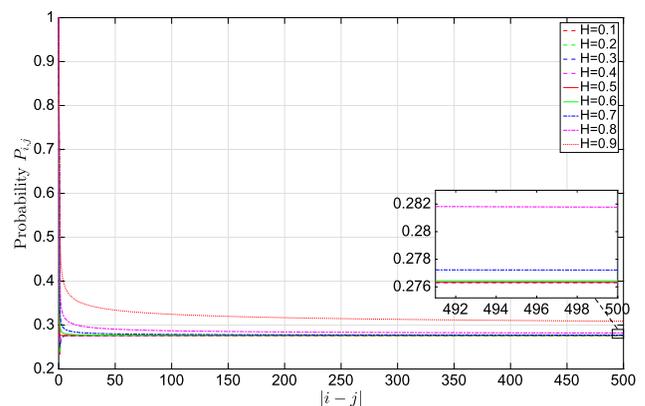


FIG. 2. Probability  $P_{ij}$  as a function of  $|i - j|$  obtained for  $H$  values ranging from 0.1 to 0.9, with a step of 0.1. The curves for  $H \leq 0.5$  are almost identical for largest values of  $|i - j|$ .

Figure 2 shows the values of the probability  $P_{ij}$  (which provides an estimation of  $REC_{th}$ ) for  $|i - j|$  ranging from 0 to 500 for different values of the parameter  $H$ . Figures 3 and 4 show similar results for the probabilities  $P_{ij}^n$  and  $T_{ij}^n$ , respectively, involved in the computation of  $DET_{th}$  and  $LAM_{th}$ , for  $n = 2$  and  $n = 4$ . The results depicted in these figures confirm the numerically observed asymptotic independence of the theoretical probabilities of  $(i, j)$  for  $|i - j| \geq 500$ . Note that as expected and unlike  $T_{ij}^n$ , the probability  $P_{ij}^n$  is equal to 1 for  $|i - j| = 0$ , which corresponds to the main diagonal of the RP.

For the estimation of the empirical values of  $REC$ ,  $DET$ , and  $LAM$ , we set the embedding dimension  $d = 1$  and the same threshold  $\varepsilon = 0.5$ . We excluded the main diagonal of the RPs for these estimations (for Figs. 1 and 6, displaying RPs, the main diagonal was not discarded). For each value of  $H$ , these computations were performed for simulated time series of 1000-point length. Thirty realizations of each process were simulated in order to get statistics (mean and standard deviation) of the  $REC$ ,  $DET$ , and  $LAM$  values. We performed the comparison between the theoretical and empirical RQA ( $DET$  and  $LAM$ ) measures for three values of the diagonal and vertical lines length, namely,  $n = 2, 3, 4$ .

The empirical RQA measures were obtained by means of the Cross Recurrence Plot Toolbox developed by N. Marwan<sup>6</sup> (available at <http://www.agnld.uni-potsdam.de/~marwan/toolbox.php>).

In Fig. 4, the empirical and theoretical values of  $REC$ ,  $DET$ , and  $LAM$  are depicted for  $H$  values ranging from 0.05 to 0.95 with a step of 0.05. For the empirical values, the mean

value of each measure (over thirty simulated sample paths) is shown and the error bars correspond to the standard deviations. This figure also shows these values for different values of  $n$  (2, 3 and 4), namely, the minimal length of the diagonal and vertical lines, respectively, considered for the computation of  $DET$  and  $LAM$  measures. As expected, the  $DET$  and  $LAM$  values decrease with  $n$ . Note that, in Fig. 4, the scale for the  $REC$  values is reduced in comparison to the scales used to depict  $DET$  and  $LAM$  values.

In order to get a deeper insight into the potential discriminative power of empirical  $LAM$  measure, which is monotonic with respect to  $H$ , we also performed statistical tests. We used a pairwise two-tailed Student's test for independent samples to compare, for all possible pairs of  $H$  values (19 values ranging from 0.05 to 0.95), the sample means of  $LAM$  values obtained from 30 simulated time series. The results for three different values of  $n$ , namely, 2, 3, and 4, are presented in panels (a), (b), and (c) of Fig. 5, in a graphical matrix form using a 2-D binary plot where the white squares represent the cases for which the null hypothesis of equality of means was rejected. A black square is plotted when the null hypothesis was not rejected. The significance level  $\alpha$  value was set to 0.05 for these tests. The use of the Student test was supported by checking the normality of each sample of 30 empirical  $LAM$  values using the Kolmogorov-Smirnov test: for all  $H$  and  $n$  values, we found that the null hypothesis of a normal underlying distribution was not rejected. In panels (d), (e), and (f) of Fig. 5, we present the power of these statistical tests using heatmap plots, for each value of  $n$ .

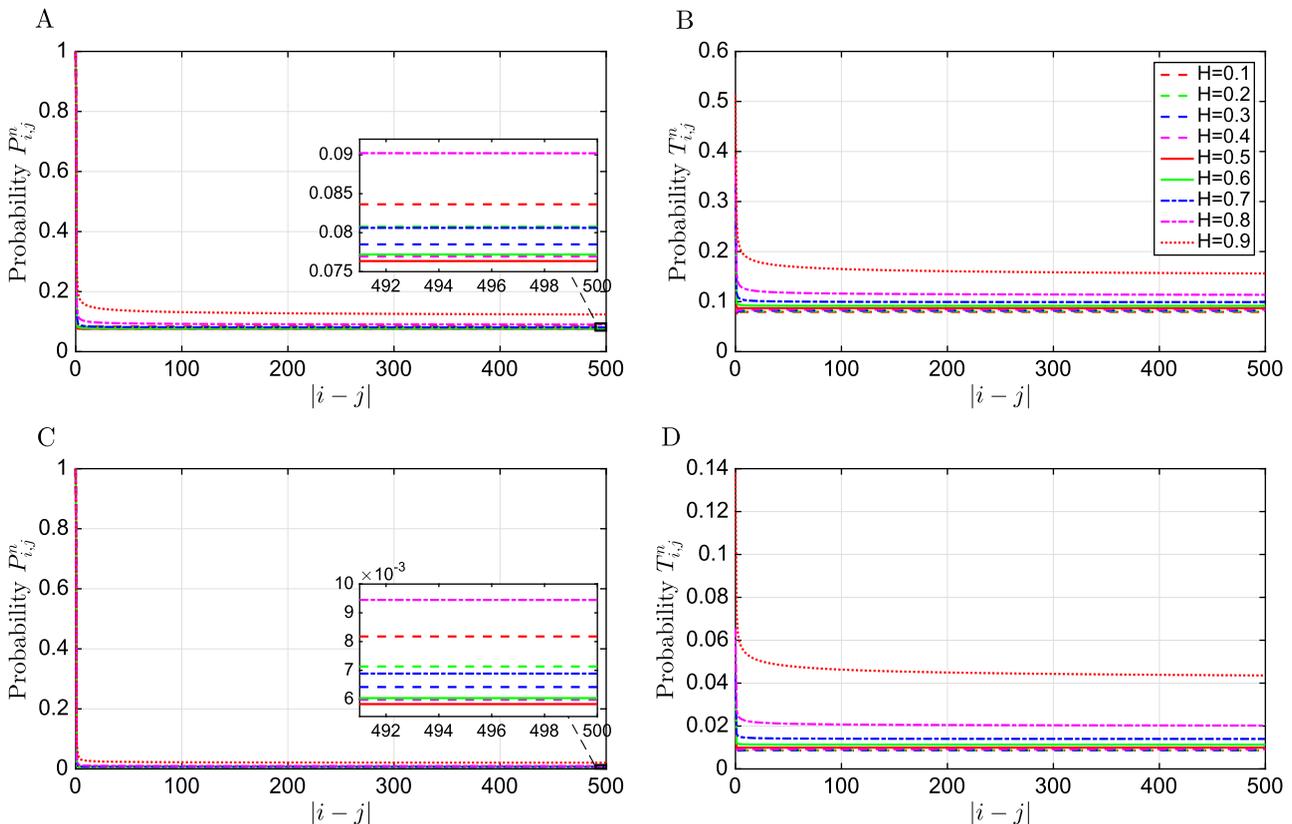


FIG. 3. Probabilities  $P_{ij}^n$  [panels (a) and (c)] and  $T_{ij}^n$  [panels (b) and (d)] as functions of  $|i - j|$  for  $H$  values ranging from 0.1 to 0.9, with  $n = 2$  [(a) and (b)] and  $n = 4$  [(c) and (d)]. The value of probability  $T_{ij}^n$  shown in panels (b) and (d) increases with  $H$ .

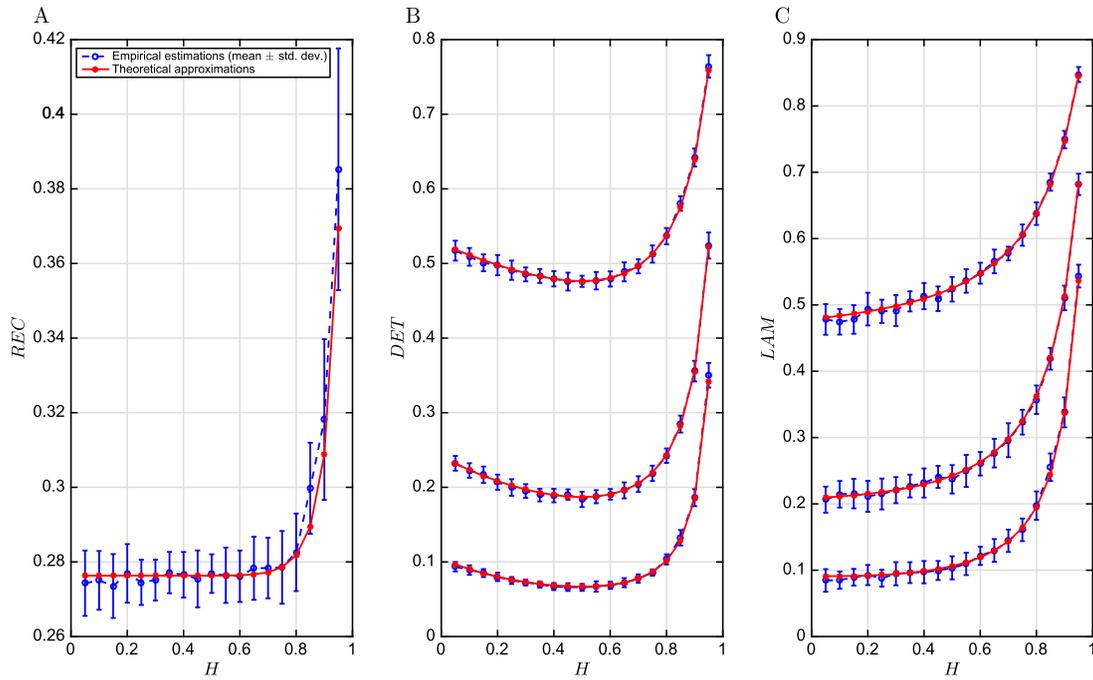


FIG. 4. Empirical and theoretical  $REC$  (a),  $DET$  (b), and  $LAM$  (c) measures obtained for  $H$  values ranging from 0.05 to 0.95. In the case of  $DET$  and  $LAM$ , the values obtained for  $n = 2$ ,  $n = 3$ , and  $n = 4$  are reported. The top-most curve corresponds to  $n = 2$  and the bottom-most to  $n = 4$ . Note that the scale used to depict  $REC$  values is reduced in comparison to the scales of  $DET$  and  $LAM$ .

**VI. DISCUSSION**

The computation of the theoretical RQA measures  $REC_{th}$ ,  $DET_{th}$ , and  $LAM_{th}$  is directly related to the computation of the probabilities  $P_{i,j}$ ,  $P^n_{i,j}$ , and  $T^n_{i,j}$  [see Eqs. (5), (18), and (29)].

The results shown in Figs. 2 and 3 indicate that the setting of  $|i - j| = 500$  for the computation of  $REC_{th}$ ,  $DET_{th}$ , and  $LAM_{th}$  ensures a very good approximation in almost all cases.

We also observe from Fig. 3 that the probability related to the vertical lines  $T^n_{i,j}$  [shown in panels (b) and (d)] is more

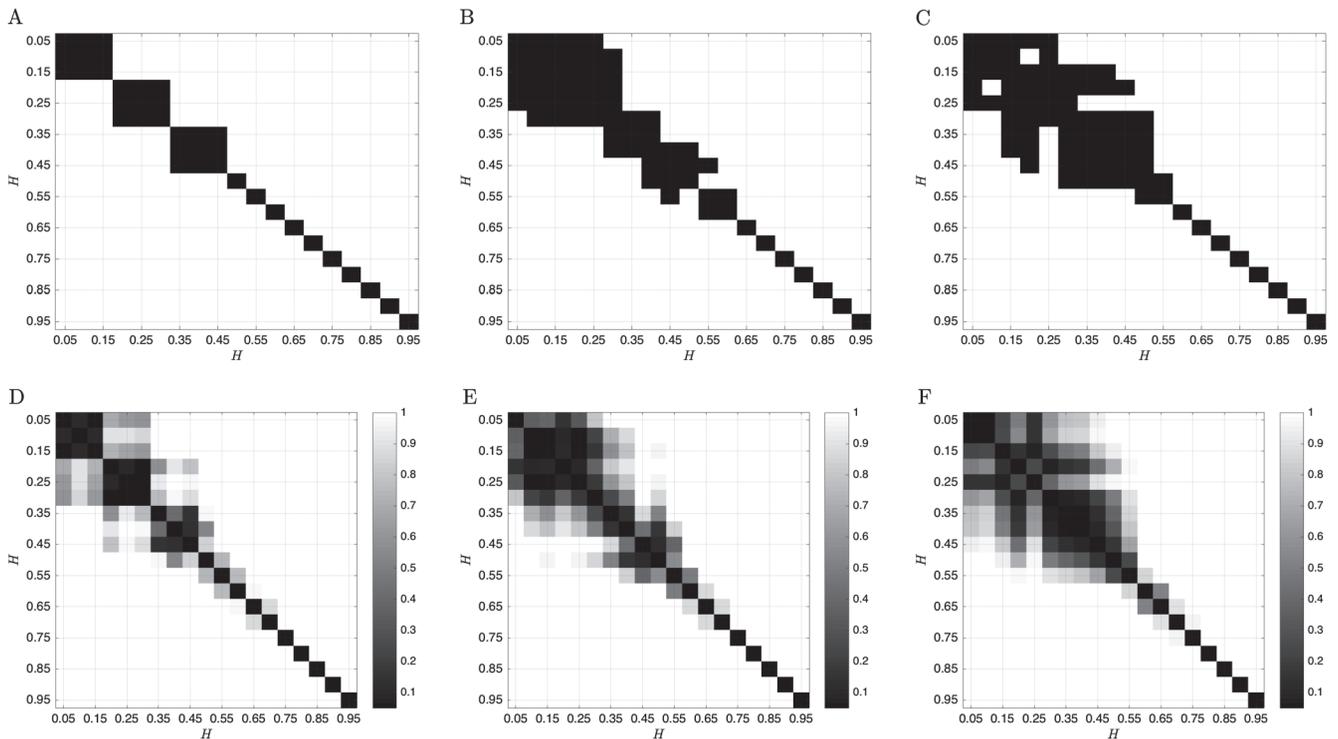


FIG. 5. 2-D binary plots [(a)–(c)] and heatmap plots [(d)–(f)], respectively, displaying t-test results and their corresponding power for  $n = 2, 3$ , and  $4$  (from left to right). These plots were obtained by comparing empirical  $LAM$  means for all possible pairs of  $H$  values ranging from 0.05 to 0.95. Note that the diagonal black squares result from the comparison of two identical samples of  $LAM$  values and thus they do not convey a statistical meaning.

relevant than  $P_{ij}^n$  [depicted in panels (a) and (c), in which a zoom is added] to distinguish  $H$  values lower than 0.5 to those larger than 0.5 ( $H = 0.5$  corresponds to the case of white noise). This can be confirmed after an adjustment of the scale of  $P_{ij}^n$  to match the  $T_{ij}^n$  scale.

The results displayed in Fig. 4 show that the theoretical and empirical values of the three considered RQA measures are very coherent. For the  $REC$  measure [Fig. 4, panel (a)], we observe a moderate overestimation for  $H \geq 0.85$ . We also observe a very stable value of the theoretical value  $REC_{th}$  for  $H \leq 0.80$ . For larger values of  $H$ , low-frequency components (local trends) appear in the time series, inducing modifications on the expected statistical properties, e.g., mean and standard deviation, of the simulated sample paths. This could explain the overestimation of the  $REC$  values of the simulated signals with respect to their theoretical counterparts. This probable explanation should be confirmed in the future by a dedicated study. Note that the local trends induce white bands and less homogeneous RPs, as apparent in the two RPs displayed in Fig. 6, generated by two fGn sample paths, with  $H = 0.7$  and  $H = 0.9$  (see also Ref. 4 and references therein regarding non-stationarity issues).

We observe that the variability of the empirical  $DET$  (quantified by the standard deviation) is, in general, lower than the variability obtained for the  $LAM$  measure [see Fig. 4, panels (b) and (c)]. We also note that the variability of  $DET$  is decreased when the number of the minimal length  $n$  of the considered diagonals is increased.

Figure 4 shows that unlike  $DET_{th}$ ,  $LAM_{th}$  is monotonic. The  $DET$  measure reaches a minimal value for  $H = 0.5$ , which corresponds to white Gaussian noise. This is expected as  $DET$  is positively correlated to the predictability of the process. The statistical results displayed in Fig. 5 suggest that the  $LAM$  measure computed for  $n = 2$  can potentially

be used to distinguish different values of  $H$  (with a 0.05 step resolution) when  $H \geq 0.5$  [see Figs. 5(a) and 5(d)]. For  $n = 3$  and  $n = 4$ , this would require higher values of  $H$ , namely,  $H > 0.6$  [see Figs. 5(b), 5(c), 5(e), and 5(f)]. As for a possible limitation coming from data length, and specifically for the results of Fig. 4, simulations performed with 200-point fGn sample paths provided consistent findings (not shown) with those presented here using 1000-point length time series. The main differences were a moderately larger overestimation of the  $REC$  measure for values of  $H$  larger than 0.80 and an expected increase of the variability of the empirical estimations. Concerning the results of the statistical tests, as expected, the discriminative power of the  $LAM$  measure was weakened so that the significant statistical differences were only found for  $H$  values larger than 0.75.

In order to confirm our results regarding the potential effect of the quality of the simulated fGn paths, we also investigated the RQA measures as functions of the *a posteriori* estimated  $H$  values instead of using the preset theoretical  $H$  values. To estimate these values of  $H$ , we used two classical methods, namely, detrended fluctuation analysis (DFA) and an approach using a wavelet-based discrete second-order derivative estimator (see, for instance, Refs. 22–26). We thus obtained  $H$  for each 1000-point simulated path and then averaged over the 30 realizations to get a unique statistic. The dependence of the three considered RQA measures on these estimated  $H$  values was similar and consistent with the one based on the predefined  $H$  values. Actually, the estimated  $H$  values are only slightly shifted with respect to the predefined  $H$  values. The results of the second estimation method are displayed in Fig. 7.

Few studies have addressed the theoretical derivation of RQA measures for large classes of stochastic processes. These studies had different objectives and were based on

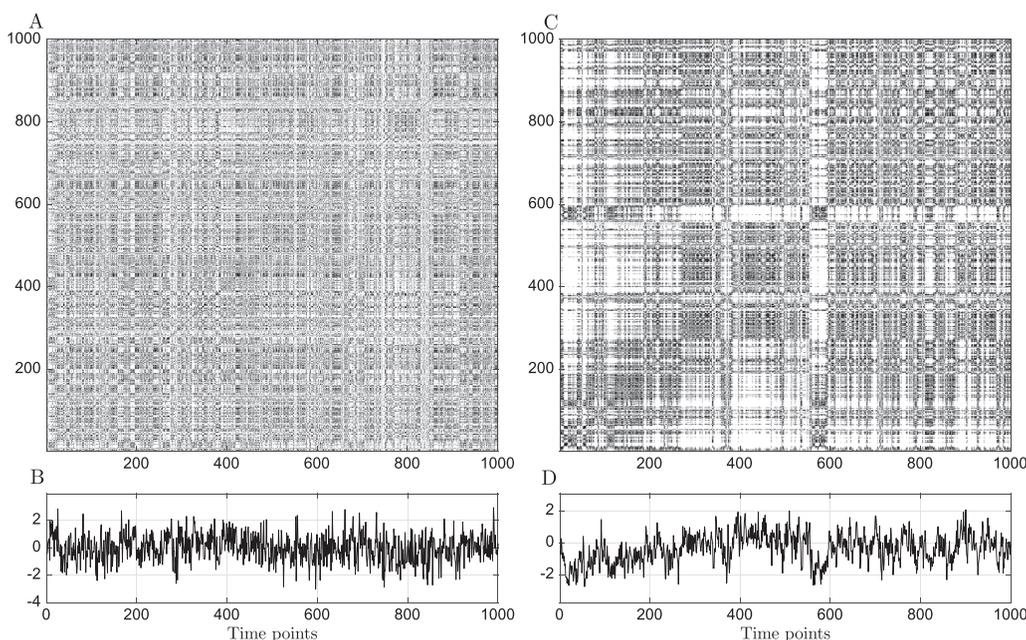


FIG. 6. Recurrence plots [(a),(c)] of unit variance fGn sample paths [(b),(d)] with  $H = 0.70$  and  $H = 0.90$ , respectively. The RPs were constructed without embedding and using a threshold  $\varepsilon = 0.5$ .

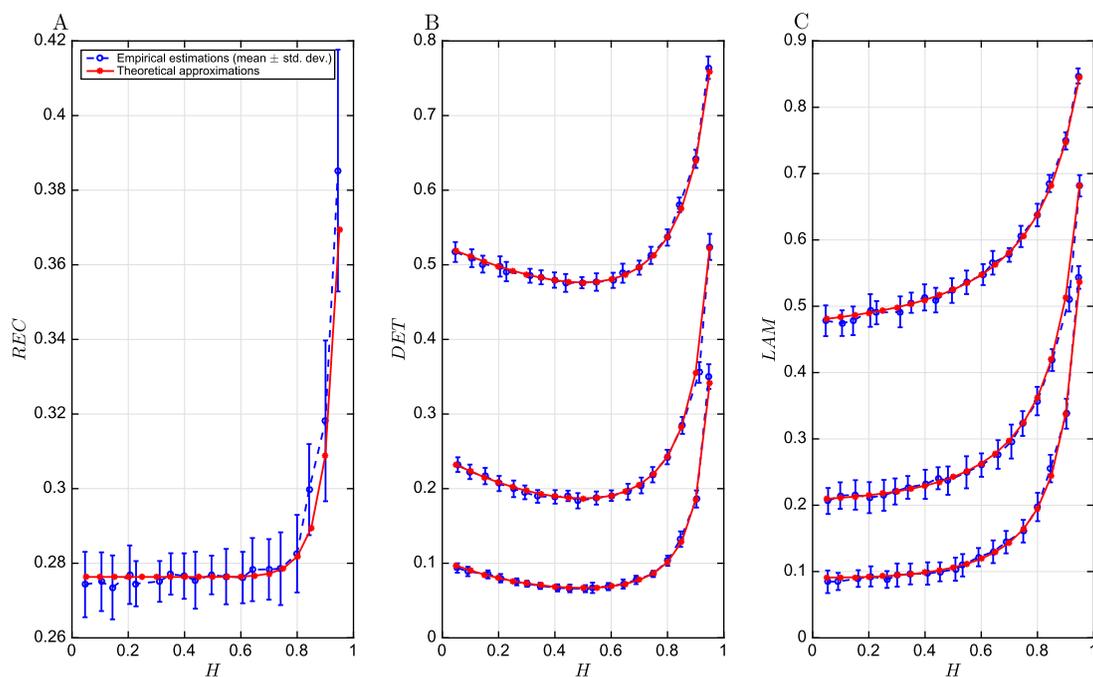


FIG. 7. Empirical and theoretical *REC* (a), *DET* (b), and *LAM* (c) measures obtained for  $H$  values ranging from 0.05 to 0.95. Values obtained for  $n = 2, 3$ , and 4 are reported for *DET* and *LAM* measures. The top-most curve corresponds to  $n = 2$  and the bottom-most to  $n = 4$ . For the empirical estimations of the three measures, the corresponding  $H$  values shown on  $x$ -axis were computed *a posteriori* using a wavelet-based discrete second-order derivative method, explaining the slight horizontal shift between the points of the blue and red curves.

completely different methodological approaches than ours. They were not specific to fGn processes. In Ref. 27, a mathematical and statistical analysis of some specific RQA measures (the  $k$ -recurrence rate, the percent determinism, and the average length of diagonal lines) was performed. The authors exploited correlation sums to analytically express the asymptotic values of these measures and applied them to i.i.d. processes, Markov chains, and autoregressive processes. In another work, Schultz *et al.*<sup>28</sup> derived approximations of diagonal line based RQA-measures including the percent determinism. The results were expressed in terms of pairwise proximity measures, which basically count the number of pairs of neighboring states in the reconstructed phase space. In the same line, Spiegel *et al.*<sup>29</sup> extended this result to approximate the laminarity using a measure of stationary states of the embedded trajectory in a phase space with distances defined by maximum norm. In Ref. 30, fGn processes were investigated through recurrence networks. The authors empirically showed that these networks can be useful to analyze such processes provided that the embedding dimension and delay are properly selected. They also emphasized that selecting the embedding dimension is not straightforward for fGn processes. One advantage of our approach is that the embedding dimension is set to 1 and thus no time delay is required to construct the RPs.

Overall, our results establish an analytical relationship between the main RQA measures and the covariance structure of a wide class of stationary Gaussian processes. These findings demonstrate that measures extracted from RPs can be exploited to identify and distinguish fGn processes. The theoretical values of specific RQA measures can be used to control the quality of the estimation of the  $H$  exponent.

## ACKNOWLEDGMENTS

S.R. would like to thank V. Kleptsyn for useful discussions concerning probabilistic computations. The authors would like to thank two anonymous referees for their constructive criticisms and insightful comments.

- <sup>1</sup>J. Eckmann, S. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *Europhys. Lett.* **4**, 973–977 (1987).
- <sup>2</sup>J. Zbilut and C. Webber, "Embeddings and delays as derived from quantification of recurrence plots," *Phys. Lett. A* **171**, 199–203 (1992).
- <sup>3</sup>C. Webber and J. Zbilut, "Dynamical assessment of physiological systems and states using recurrence plot strategies," *J. Appl. Physiol.* **76**, 965–973 (1994).
- <sup>4</sup>N. Marwan, M. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Phys. Rep.* **438**, 237–329 (2007).
- <sup>5</sup>C. Webber and N. Marwan, *Recurrence Quantification Analysis* (Springer International Publishing, 2015).
- <sup>6</sup>N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, and J. Kurths, "Recurrence plot based measures of complexity and its application to heart rate variability data," *Phys. Rev. E* **66**, 026702 (2002).
- <sup>7</sup>J. B. Gao, "Recurrence time statistics for chaotic systems and their applications," *Phys. Rev. Lett.* **83**, 3178–3181 (1999).
- <sup>8</sup>J. Gao and H. Cai, "On the structures and quantification of recurrence plots," *Phys. Lett. A* **270**, 75–87 (2000).
- <sup>9</sup>S. Ramdani, F. Bouchara, J. Lagarde, and A. Lesne, "Recurrence plots of discrete-time gaussian stochastic processes," *Physica D* **330**, 17–31 (2016).
- <sup>10</sup>B. Mandelbrot and J. van Ness, "Fractional brownian motions fractional noises and applications," *SIAM Rev.* **10**, 422–437 (1968).
- <sup>11</sup>J. Beran, *Statistics for Long-Memory Processes* (Chapman & Hall, 1994).
- <sup>12</sup>N. Packard, J. Crutchfield, J. Farmer, and R. Shaw, "Geometry from a time series," *Phys. Rev. Lett.* **45**, 712–716 (1980).
- <sup>13</sup>F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980* (Springer, 1981), pp. 712–716.
- <sup>14</sup>H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, 2nd ed. (Cambridge University Press, 2004).
- <sup>15</sup>A. Papoulis and S. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. (McGraw-Hill Higher Education, 2002).

- <sup>16</sup>C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning* (MIT Press, 2006).
- <sup>17</sup>A. Genz, "Numerical computation of multivariate normal probabilities," *J. Comput. Graph. Stat.* **1**, 141–149 (1992).
- <sup>18</sup>J.-M. Bardet, G. Lang, G. Oppenheim, A. Philippe, S. Stoev, and M. Taqqu, "Generators of long-range dependence processes: A survey," in *Theory and Applications of Long-Range Dependence* (Birkhäuser, 2003), pp. 579–623.
- <sup>19</sup>S. Lowen, "Efficient generation of fractional brownian motion for simulation of infrared focal-plane array calibration drift," *Methodol. Comput. Appl.* **1**, 445–456 (1999).
- <sup>20</sup>S. Stoev and M. Taqqu, "Simulation methods for linear fractional stable motion and FARIMA using the fast Fourier transform," *Fractals* **12**, 95–121 (2003).
- <sup>21</sup>P. Abry and F. Sellan, "The wavelet-based synthesis for the fractional brownian motion proposed by F. Sellan and Y. Meyer: Remarks and fast implementation," *Appl. Comput. Harmon. Anal.* **3**, 377–383 (1996).
- <sup>22</sup>C. Peng, J. Mietus, J. Hausdorff, S. Havlin, H. Stanley, and A. Goldberger, "Long-range anti-correlations and non-Gaussian behavior of the heartbeat," *Phys. Rev. Lett.* **70**, 1343–1346 (1993).
- <sup>23</sup>J. Moreira, J. Kamphorst, L. da Silva, and S. Kamphorst, "On the fractal dimension of self-affine profiles," *J. Phys. A* **27**, 8079–8089 (1994).
- <sup>24</sup>D. Delignieres, S. Ramdani, L. Lemoine, K. Torre, M. Fortes, and G. Ninot, "Fractal analyses for 'short' time series: A re-assessment of classical methods," *J. Math. Psychol.* **50**, 525–544 (2006).
- <sup>25</sup>P. Abry, P. Flandrin, M. Taqqu, and D. Veitch, "t," in *Theory and Applications of Long-Range Dependence* (Birkhäuser, 2003), pp. 527–556.
- <sup>26</sup>J.-M. Bardet, G. Lang, G. Oppenheim, A. Philippe, S. Stoev, and M. Taqqu, "Semi-parametric estimation of the long-range dependence parameter: A survey," in *Theory and Applications of Long-Range Dependence* (Birkhäuser, 2003), pp. 557–577.
- <sup>27</sup>M. Grendár, J. Majerová, and V. Špitalský, "Strong laws for recurrence quantification analysis," *Int. J. Bifurcat. Chaos* **23**, 1350147 (2013).
- <sup>28</sup>D. Schultz, S. Spiegel, N. Marwan, and S. Albayrak, "Approximation of diagonal line based measures in recurrence quantification analysis," *Phys. Lett. A* **379**, 997–1011 (2015).
- <sup>29</sup>S. Spiegel, D. Schultz, and N. Marwan, "Approximate recurrence quantification analysis (arqa) in code of best practice," in *Recurrence Plots and Their Quantifications: Expanding Horizons*, Springer Proceedings in Physics Vol. 180, edited by C. Webber, C. Ioana, and N. Marwan (Springer, Cham, 2016), pp. 113–136.
- <sup>30</sup>Y. Zou, R. V. Donner, and J. Kurths, "Analyzing long-term correlated stochastic processes by means of recurrence networks: Potentials and pitfalls," *Phys. Rev. E* **91**, 022926 (2015).