Theoretical Study of Multiscale Permutation Entropy on Finite-Length Fractional Gaussian Noise

Antonio Dávalos, Meryem Jabloun, Philippe Ravier, Olivier Buttelli
PRISME Laboratory, University of Orleans, France
Email: antonio.davalos-trevino@etu.univ-orleans.fr

***Abstract*—Permutation Entropy has been used as a robust and fast approach to calculate complexity of time series. There have been extensive studies on the properties and behavior of Permutation Entropy on known signals. Similarly, Multiscale Permutation Entropy has been used to analyze the structures at different time scales. Nevertheless, the Permutation Entropy is constrained by signal length, a problem which is accentuated with Multiscaling. We have analyzed the fractional Gaussian noise under a Multiscale Permutation Entropy analysis, taking into account the effect of finite-length signals across all scales. We found the Permutation Entropy value of Fractional Gaussian noise to be invariant to time scale. Nonetheless, a finite-length linear approximation for scale dependency is found as a result solely from the finite-length constrains of the method.**

***Index terms*—** Multiscale Permutation Entropy, Fractional Brownian Motion, Fractional Gaussian Noise, Finite-Length Time Series

I. Introduction

Entropy measurements have helped researchers explore the degree of complexity on data series in a wide range of fields. Particularly in the context of biomedical signals, Goldberger proposed that pathologies and aging can be distinguished by measuring physiological complexity [1]. These propositions has been applied to distinguish Alzheimer states [2], classify signals from Parkinson's disease patients [3], among many other applications.

Since the original proposition by Shannon [4], many different definitions for Entropy measures have been proposed for this purpose. In particular, the Permutation Entropy (PE) [5] has been used because of its robustness to noise and simple implementation. Also, to capture complexities at different time scales, and distinguish randomness from complexity, Costa [6] proposed a Multiscale Entropy concept. Since any Entropy measurement is compatible with the multiscale approach, Aziz and Arif [7] introduced the Multiscale Permutation Entropy (MPE), which extends the properties of PE across different scales.

There have been significant efforts to understand the underlying behavior of the Permutation Entropy under known structures, most notably by Bandt and Shiha [8] and Zunino [9] using fractional Brownian motion and fractional Gaussian noise. Nonetheless, the PE's behavior of these processes has not been analyzed under the scope of multiscaling. There is previous work in this regard, where Little [10] analyzed the approximate behavior of white noise under PE finite-length constrains. One of the limitations of PE is the requirement of a sufficiently large number of data points to yield an accurate result. There several proposed ideas to mitigate this limitation [11]. Nevertheless, since the multiscaling process reduces the amount of available data, the treatment of short signals is an unavoidable problem.

Hence, the purpose of this document is to study the dependency of time scale with a random signal with build-in correlations, by studying the Multiscale Permutation Entropy on fractional Gaussian noise. We expect to capture the added complexity of long-range correlation with the time scale. Also, we will address the problem of data length, which is accentuated for high scales, to quantify its added effect in practical measurements.

The article is organized as follows: Section II describes the theoretical background of Permutation Entropy, Multiscale Permutation Entropy, and fractional Gaussian noise. Section III explores the PE of fractional Gaussian noise in the context of multiscaling, and the scale dependency introduced by the coarse-graining procedure. Section IV and conclusions discuss the results and the implications of the previous analysis.

II. Theorethical Basis

In this section, we review the key concepts of Permutation Entropy, Multiscale Permutation Entropy and fractional Gaussian noise, which are necessary for our propositions.

A. Permutation Entropy

Bandt and Pompe [5] proposed an ordinal-based method to measure the entropy of a signal. Given an arbitrary time series with weak stationary assumption, an embedding dimension d is selected. For each possible segment of size d, a particular order (pattern) is obtained. In general, $d!$ different patterns are possible. For $d = 2$, only two possible patterns exist; $x_n < x_{n+1}$ or $x_n > x_{n+1}$, therefore labelled as π_{12} and π_{21}. For $d = 3$, six different patterns are possible (ex. π_{123}). The cases where $x_n = x_{n+1}$ are excluded, since real data applications have at least a minimum amount of noise, and this case will not occur almost surely.

Given an time series x_n, where $n = 1, \ldots, N$ (being N the size of the series), we compute the relative frequency for the $d!$ possible patterns π,

$$\widehat{p}(\pi) = \frac{\sharp\{n | n \leq N-(d-1)\lambda, (x_{n+1}, \ldots, x_{n+d})\ type\ \pi\}}{N-(d-1)\lambda} \tag{1}$$

where λ is a downsampling parameter, and the empirical probability of observing pattern π is defined as $\widehat{p}(\pi)$. For our purposes $\lambda = 1$, (see section II-B). Given this empirical probability distribution, the PE is then defined as,

$$\widehat{\mathcal{H}} = \frac{-1}{\ln(d!)} \sum_{i=1}^{d!} \widehat{p}(\pi_i) \ln \widehat{p}(\pi_i) \tag{2}$$

The maximum entropy is obtained when all patterns have the exact same probability (i.e. the mass probability function is uniform).

Note that the PE is just the widely known Shannon Entropy, using the relative frequencies for all the patterns present in the time series x_n. The PE is robust, simple to implement, and invariant with respect to nonlinear monotonous transformations [9]. Since the method only works with the ordinal patterns of the series, it does not need further assumptions over the distribution of the signal. For the feasibility of the method, sufficiently long data series are needed, with the general condition that $N \gg d!$.

B. Multiscale Coarse-grain procedure

Costa and Goldberg [6] introduced the concept of coarse-graining procedure to Entropy measurements to analyze the behavior of a time series at different time scales, thus defining the Multiscale Entropy (MSE). For a deterministic signal, the associated information Entropy will be low, and will approach its maximum value with random noise. However, pure uncorrelated noise is simple to characterize. Therefore, high Entropy implies randomness but not necessarily complexity [6]. For higher time scales, random noise tends to cancel out, and thus, gives a low Entropy measurement, where complex signals maintain high Entropy.

For a one-dimensional process, a coarse-grained time series of scale m is computed. Given a time series $\{x_1, \ldots, x_N\}$,

$$x_j^{(m)} = \frac{1}{m} \sum_{i=m(j-1)+1}^{jm} x_i \tag{3}$$

for $1 \leq j \leq N/m$. By doing this, each new element of the coarse-grained signal $x^{(m)}$ is the average of a non-overlapping segment (size m) of the original signal. Costa and Goldberg [6] then proceed to calculate the Sample Entropy on the coarse-grained time series for different scales.

Any Entropy measurement can be implemented with the multiscale approach, including the PE. Aziz and Arif [7] proposed the MPE, using the definition of PE in Section II-A to build the pattern distribution for different time scales. This approach preserves the ordinal advantages given by the PE, while taking in account the information discarded with the downsampling at $\lambda > 1$. The same length concern remains, and the coarse-graining scale must satisfy the condition $N/m \gg d!$.

C. Fractional Gaussian Noise

Fractional Brownian motion (fBm) and fractional Gaussian noise (fGn) were originally proposed by Mandelbrot and Van Ness [12] as a generalization of the Gaussian random walk phenomenon, where different measurements of autocorrelation are introduced by the Hurst parameter $0 < H < 1$. $H = 1/2$ corresponds to white Gaussian noise. For $H < 1/2$ the series present short memory (the sum of the autocorrelations tend to zero), and long memory for $H > 1/2$ (the sum or autocorrelations tend to infinity). The fractional Brownian motion $B_H(n)$ refers to the incremental process of correlated Gaussian variables, where the fractional Gaussian noise $G_H(n)$ refers to the individual steps. fGn is the only process which is Gaussian, self-similar, and have stationary increments [9]. They are related as follows:

$$\begin{aligned} B_H(n) &= \sum_{i=0}^{n} G_H(i) \\ G_H(n) &= B_H(n) - B_H(n-1) \end{aligned} \tag{4}$$

being n the discrete time step of the series. For our purposes, we will work with discrete fGn.

As Gaussian processes, fGn and fBm are completely represented by their mean and their autocovariance function [13].

$$E[G_H(n)] = 0, E[B_H(n)] = 0 \tag{5}$$

$$\begin{aligned} &cov(G_H(n), G_H(n+k)) = \\ &\qquad \frac{\sigma^2}{2}(|k+1|^{2H} + |k-1|^{2H} - 2|k|^{2H}) \end{aligned} \tag{6}$$

$$\begin{aligned} &cov(B_H(n), B_H(n+k)) = \\ &\qquad \frac{\sigma^2}{2}(|n|^{2H} + |n+k|^{2H} - |k|^{2H}) \end{aligned} \tag{7}$$

$$\rho_G(k) = \frac{1}{2}(|k+1|^{2H} + |k-1|^{2H} - 2|k|^{2H}) \tag{8}$$

Being $n = 0, \ldots, N$, $k \geq 0$ the difference in position between variables, $\rho_G(k) = cov(G_H(n), G_H(n+k))/var(G_H(n))$ the autocorrelation function of fGn (independent of n), and σ^2 the variance of any individual Gaussian step $G_H(n)$.

III. MPE applied to fractional Gaussian noise

In this section we expand previous results of PE on fGn using the Multiscale Approach. In section III-A we define the properties of a coarse-grained fGn. Section III-B reviews the symmetry properties of fGn, which we apply to the coarse-grained fGn. Lastly, in Section III-C we define a scale dependency from the finite-length constrain of the MPE, and we compare the results with simulated fGn.

A. Coarse-Graining on Fractional Gaussian Noise

To test the behavior of MPE on fGn, first it is necessary to know the distribution of noise with complex correlations within itself. By using the definition in (3) on fGn and the relation between fGn and fBm in (4), we define the

ISBN 978-90-827970-1-5

coarse-grained fractional Gaussian noise (cfGn) as,

$$\begin{aligned} G_H^{(m)}(j) &= \frac{1}{m}\sum_{i=1}^{m} G_H(m(j-1)+i) \\ &= \frac{1}{m}(B_H(mj) - B_H(m(j-1))) \end{aligned} \tag{9}$$

for $j = 1, \ldots, N/m$. The Coarse-Graining procedure, by definition, is a linear combination of the elements in the segment m. Since the sum of multiple Gaussian variables is also Gaussian, the cfGn can be completely determined by the expected value and covariance function.

First, we establish the expected value $E[G_H^{(m)}(j)] = 0$, as the expected value of a sum of independent random variables is the sum of the expected value of each variable. By using eq. (7) we obtain the variance of $G_H^{(m)}(j)$,

$$\begin{aligned} var(G_H^{(m)}(j)) &= E[(G_H^{(m)}(j))^2] \\ &= \frac{1}{m^2}E[(B_H(mj) - B_H(m(j-1)))^2] \\ &= \sigma^2 m^{2(H-1)} \end{aligned} \tag{10}$$

we perform a similar analysis for the autocovariance function of $G_H^{(m)}(j)$

$$\begin{aligned} &cov(G_H^{(m)}(j), G_H^{(m)}(j+k)) \\ &=\sigma^2 m^{2(H-1)}(|(k+1)|^{2H} + |(k-1)|^{2H} - 2|k|^{2H}) \\ &=\sigma^2 m^{2(H-1)}\rho_G(k) \end{aligned} \tag{11}$$

Notice that the structure of the autocovariance function is the same as the original fGn, but with the added information of the scale size m. This also implies

$$\rho_{G,(m)}(k) = \frac{\sigma^2 m^{2(H-1)}\rho_{G_H}(k)}{\sigma^2 m^{2(H-1)}} = \rho_G(k) \tag{12}$$

That is, we conclude that there is no difference between the correlation function of $\rho_G(k)$ and $\rho_{G,(m)}(k)$. Therefore, the autocovariance function of the original fGn signal is invariant to the Coarse-Graining procedure.

B. Permutation Entropy on Fractional Gaussian Noise

Bandt and Shiha [8] found that, for a process with stationary increments, the following symetries apply to patterns of dimension $d = 3$

$$\begin{aligned} &p(\pi_{123}) = p(\pi_{321}) = p \\ &p(\pi_{132}) = p(\pi_{213}) = p(\pi_{231}) = p(\pi_{312}) = \frac{1-2p}{4} \end{aligned} \tag{13}$$

where p is a single probability. From (13), they obtained the expected mass probability distribution of all patterns by finding only the value of $p(\pi_{123})$ for any Gaussian process with stationary increments. Both $G_H(n)$ and $G_H^{(m)}(j)$ satisfy these conditions. In these cases, p depends on the Hurst parameter H.

For $d = 2$, $p(\pi_{12}) = 1/2$ regardless of H. For $d = 4$, some of the probabilities obtained have complex values, and thus, the results are difficult to interpret. For $d \geq 5$, the probabilities have no closed-form expression [8].

Therefore, for $d = 3$, only $p(\pi_{123})$ is needed to get the PE of fGn. First the auxiliary variable $\Delta G_H^{(m)}(j) = G_H^{(m)}(j+1) - G_H^{(m)}(j)$ is defined, so that

$$\begin{aligned} &p_G(\pi_{123}) \\ &\quad= p(G_H^{(m)}(j) < G_H^{(m)}(j+1), G_H^{(m)}(j+1) < G_H^{(m)}(j+2)) \\ &\quad= p(\Delta G_H^{(m)}(j) > 0, \Delta G_H^{(m)}(j+1) > 0) \end{aligned} \tag{14}$$

By following Bandt and Shiha's [8] derivation, and the identity $\arcsin(\rho) = 2\arcsin(\sqrt{(1+\rho)/2}) - \pi/2$,

$$\rho_{\Delta G}(1) = \frac{2\rho_G(1) - 1 - \rho_G(2)}{2(1-\rho_G(1))} \tag{15}$$

$$\begin{aligned} p_G(\pi_{123}) &= \frac{1}{4} + \frac{1}{2\pi}\arcsin(\rho_{\Delta G}(1)) \\ &= \frac{1}{\pi}\arcsin\sqrt{\frac{1+\rho_{\Delta G}(1)}{2}} \\ &= \frac{1}{\pi}\arcsin\left(\frac{1}{2}\sqrt{\frac{1-\rho_G(2)}{1-\rho_G(1)}}\right) \end{aligned} \tag{16}$$

Where $\rho_{\Delta G}(k)$ is the correlation between $\Delta G_H^{(m)}(j)$ and $\Delta G_H^{(m)}(j+k)$. In other words, the probability $p_G(\pi_{123})$ of a Gaussian process with stationary increments depends solely on the autocorrelation function $\rho_{\Delta G}(k)$, evaluated in $k = 1$. With eq. (15) we obtained an explicit form for $p_G(\pi_{123})$ depending only on the autocorrelation of the original series.

With these results, and using equation (8), we can obtain an explicit formula for MPE on fGn,

$$p_G(\pi_{123}) = \frac{1}{\pi}\arcsin\frac{1}{4}\sqrt{\frac{1+2^{2H+1}-3^{2H}}{1-2^{2(H-1)}}} \tag{17}$$

As we proved in (12), $\rho_{G,(m)}(k) = \rho_G(k)$. This leads to the conclusion that $p_{G,(m)}(\pi_{123}) = p_G(\pi_{123})$, and thus, invariant to scale.

By using the definition of PE (2) and the symmetry properties in (13), we arrive to the expression of the MPE on cfGn

$$\begin{aligned} &\mathcal{H}(p_{G,(m)}, d=3) = \\ &\qquad - p\ln(2p) - (1-2p)\ln(\frac{1-2p}{4}) \end{aligned} \tag{18}$$

Since $p_{G,(m)}(\pi_{123}) = p_G(\pi_{123})$ is invariant to scale, the MPE of fGn also remains invariant. This implies that any time-scale dependency will come only from the finite-length constrains of the MPE, and not from the properties of fGn. We will explain this phenomenon in section III-C.

C. MPE dependency of Finite-Length cfGn

For all the empirical calculations of PE, N is assumed to be large. For a small N, the estimations of the pattern probabilities diverge from the true values from an hypothetical infinite length signal. Little [10] derived an approximation

ISBN 978-90-827970-1-5 © EURASIP 2018

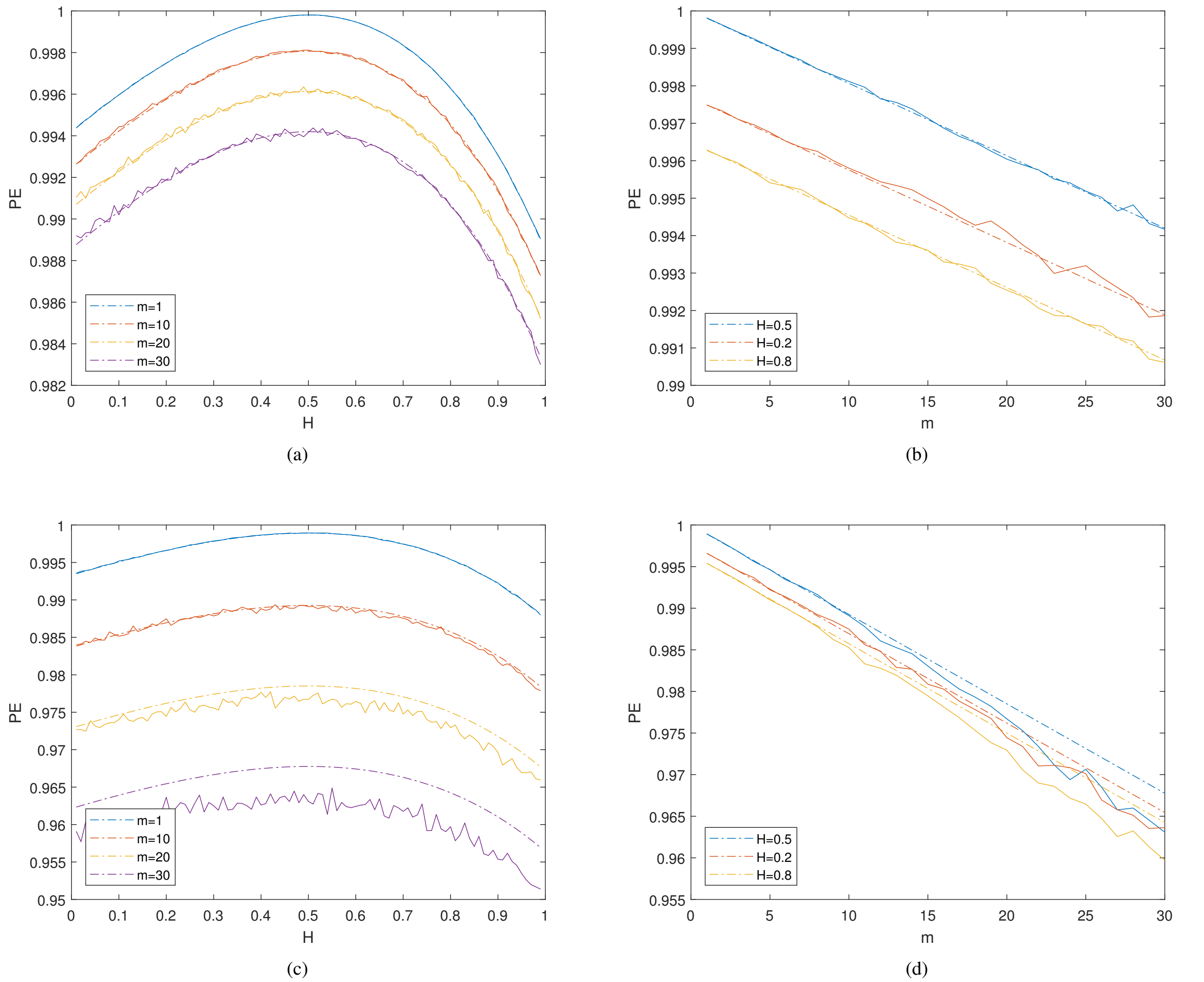

Fig. 1. Simulations of PE of 1500 signals at scales $m = 1, 10, 20, 30$ (lines), compared to their theoretical predictions (dotted lines), for signal length $N = 5000$ (figures a and b) and $N = 900$ (figures c and d)

of the PE taking into account the dependency of the signal length. The random variable Y_i is defined, with multinomial distribution, which corresponds with the count of patterns π_i in the original series $\{x_1, \ldots, x_N\}$. In the case of white noise, all patterns have equal probability $p_i = 1/d!$.

$$\widehat{Y}_i = \frac{N}{d!} + \Delta Y_i, \{Y_1, \ldots, Y_{d!}\} \sim Mu\left(N, p_1, \ldots, p_{d!}\right) \quad (19)$$

Using Taylor series expansion,

$$\begin{aligned} \widehat{\mathcal{H}} &= 1 - \frac{1}{\ln(d!)} \sum_{l=2}^{\infty} \frac{(-1)^l}{l(l+1)} \frac{d!^{l-1}}{N^l} \sum_{i=1}^{d!} (\Delta Y_i)^l \\ E[\widehat{\mathcal{H}}] &\approx 1 - \frac{d!-1}{2N\ln(d!)} \end{aligned} \quad (20)$$

being ΔY_i the deviation from the parameter $N/d!$.

We performed similar analysis taking into account the multiscale coarse-graining procedure. Here, we assumed an arbitrary pattern probability mass function $\widehat{Y}_i$ for any scale m, which deviates from the uniform discrete distribution.

$$\widehat{Y}_i = \frac{N}{m} p_i + \Delta Y_i, \{Y_1, \ldots, Y_{d!}\} \sim Mu\left(\frac{N}{m}, p_1, \ldots, p_{d!}\right) \quad (21)$$

$$\begin{aligned} E[\widehat{\mathcal{H}}] &= \mathcal{H} - \\ &\frac{m}{N\ln(d!)} \sum_{l=1}^{\infty} \frac{(-1)^{l+1}}{l(l+1)} \left(\frac{m}{N}\right)^l \sum_{i=1}^{d!} \frac{E[\Delta {Y_i}^{l+1}]}{p_i^l} \quad (22) \\ E[\widehat{\mathcal{H}}] &\approx \mathcal{H} - \frac{d!-1}{2N\ln(d!)} m \end{aligned}$$

where $\mathcal{H}$ is the theoretical PE for an infinite-length series. We note that the approximation of first order is always independent of the probability mass distribution, which is convenient when

ISBN 978-90-827970-1-5 © EURASIP 2018

analyzing arbitrary signals. Also, the function introduces a linear dependency on scale m which comes only form the length constrains. In general, the subsequent terms of the expansion depend of the higher moments of ΔY_i, which require some additional information of the probabilities of each pattern.

Figures 1a and 1b show a downward linear shift in the simulations of PE of cfGn, for $N = 5000$ data points. For $m = 30$, the length of the resulting coarse-grained signal is $N/m = 166$, which still follows a linear behavior, regardless of H.

In figures 1c and 1d, we performed the same analysis for $N = 900$. The maximum scale has a length of $N/m = 30$, which deviates significantly from the linear model. Here, the deviation suggests the need of a quadratic term approximation which, in general, depends on H.

IV. Discusion

The time scale dependency of the fGn has been explored. Given that MPE for fGn (for $d = 3$) has a closed expression dependent only of one pattern probability, it is possible to state the invariant behavior with respect scale m. This proves that infinite-length MPE in fGn does not show any relevant structures, even though the long and short memory correlations produce very complex behavior. General signals would not have this property, but any behavior that arises from the MPE analysis will differentiate the underlying phenomenon from noise, correlated or otherwise. More work is needed to assess the effect of other noise models.

It is necessary, nonetheless, to decouple the true underlying structure in respect to scale, from the scale dependency introduced by the finite-length restrictions, which is unavoidable in the coarse-graining for MPE analysis.

The linear model approximation (22) is convenient by its simplicity and by its independency respect to the pattern probability mass distribution. By adding a factor of $+\frac{d!-1}{2N\ln d!}m$ to the MPE measure (22), the resulting estimator comes closer to an infinite-length estimation of PE, thus mitigating the length problem.

It is important to note that the linear model does not hold for extreme cases (short length and high scale), and higher order terms are needed to explain the scale dependency introduced by the coarse-graining. It is still necessary to propose a theoretical quadratic term to the model that is independent of the pattern distribution.

V. Conclusion

We performed a theoretical analysis of the behavior of fractional Gaussian noise under the Multiscale Permutation Entropy analysis. We proved its behavior is invariant to the coarse-graining decomposition procedure, regardless of the correlation introduced by the Hurst parameter. Thus, the intrinsic properties and complex behavior of fGn are not captured by the MPE.

Simulations show a clear dependency to scale m, which comes from the finite-length restrictions of a time series, and thus, from the coarse-graining procedure itself.

Although the full behavior is complex to model, it is possible to correct the MPE measurement by a linear term that is independent to the signal distribution. This mitigates the length limitations of the MPE, which the coarse-graining makes difficult to ignore.

Further corrections are possible for short signals or high scale, but they depend on the distribution of the pattern length, which requires further knowledge of the signal. A deeper analysis is needed to propose a method that does not compromise the robustness of the Permutation Entropy, and ensure higher precision on short signals.

Acknowledgment

The authors would like to thank the Science and Technology Council of Mexico (CONACyT) for providing the funding for this research.

References

[1] A. L. Goldberger, C.-K. Peng, and L. A. Lipsitz, "What is physiologic complexity and how does it change with aging and disease?" *Neurobiology of Aging*, vol. 23, no. 1, pp. 23–26, 2002.

[2] F. C. Morabito, D. Labate, F. La Foresta, A. Bramanti, G. Morabito, and I. Palamara, "Multivariate multi-scale permutation entropy for complexity analysis of alzheimers disease EEG," *Entropy*, vol. 14, no. 7, pp. 1186–1202, 2012.

[3] Y. Wu, P. Chen, X. Luo, M. Wu, L. Liao, S. Yang, and R. M. Rangayyan, "Measuring signal fluctuations in gait rhythm time series of patients with parkinson's disease using entropy parameters," *Biomedical Signal Processing and Control*, vol. 31, pp. 265–271, 2017.

[4] C. E. Shannon, "A mathematical theory of communication," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 2001.

[5] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Physical Review Letters*, vol. 88, no. 17 174172, pp. 1–4, 2002.

[6] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical Review Letters*, vol. 89, no. 6 068102, pp. 1–4, 2002.

[7] W. Aziz and M. Arif, "Multiscale permutation entropy of physiological time series," in *2005 Pakistan Section Multitopic Conference*, 2005, pp. 1–6.

[8] C. Bandt and B. Shiha, "Order patterns in time series," *Journal of Time Series Analysis - Wiley Online Library*, pp. 646–665, 2007.

[9] L. Zunino, D. G. Prez, M. T. Martn, M. Garavaglia, A. Plastino, and O. A. Rosso, "Permutation entropy of fractional brownian motion and fractional gaussian noise," *Physics Letters A*, vol. 372, no. 27, pp. 4768–4774, 2008.

[10] D. J. Little and D. M. Kane, "Permutation entropy of finite-length white-noise time series," *Physical Review. E*, vol. 94, no. 2 022118, pp. 1–6, 2016.

[11] H. Azami and J. Escudero, "Improved multiscale permutation entropy for biomedical signal analysis: Interpretation and application to electroencephalogram recordings," *Biomedical Signal Processing and Control*, no. 23, pp. 28–41, 2015.

[12] B. Mandelbrot and J. Van Ness, "Fractional brownian motions, fractional noises and applications," *SIAM Review*, vol. 10, no. 4, pp. 422–437, 1968.

[13] D. Delignires, "Correlation properties of (discrete) fractional gaussian noise and fractional brownian motion," *Mathematical Problems in Engineering*, no. 485623, pp. 1–7, 2015.