Multiscale Permutation Entropy: Statistical Characterization on Autoregressive and Moving Average Processes

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

***Abstract*—Multiscale Permutation Entropy (MPE), an extension of Permutation Entropy (PE), was proposed to better capture the information content in long range trends. This technique has been extensively used in biomedical applications for diagnosis purposes. Although PE theory is well established and explored, there is still a lack of theoretical development for MPE. In the present paper, we expand the theory by formulating an explicit MPE model of first order Autoregressive (AR) and Moving Average (MA) processes, which are well known and used in signal modeling. We first build the autocorrelation function of coarse-grained AR and MA models, which are a prerequisite for MPE calculation. Next, we use the resulting autocorrelation functions to establish the theoretical value of MPE as a function of time scale and AR or MA parameters. The theoretical result is tested against MPE measurements from simulations. We found the MPE of the 1° order AR model to converge to the maximum entropy with increasing time scale. Nonetheless, the convergence is not always monotonic. For AR parameter values greater than the Golden Ratio, the MPE curve presents a local minimum at a time scale different than one, which implies a more regular structure than the one measured with PE. The MPE of the 1° order MA model converges rapidly to the maximum entropy with increasing time scales, regardless of the MA parameter value, which is in accordance to our expectations.**

***Index terms*—** Autoregressive model, Moving Average model, Multiscale Permutation Entropy, Coarse-graining procedure.

I. Introduction

Since its introduction, Information Entropy [1] has been used in a wide range of fields to study the complexity of data series. Different versions of Entropy measurements have been successfully used in medical fields to analyze physiological signals for diagnostic purposes. Goldberger, in particular, [2] proposed that pathologies can be distinguished by measuring statistical complexity.

Different approaches have been used in this field. Approximate Entropy [3], Sample Entropy [4] and Fuzzy Entropy [5] have been used to compare the similarity between different sections of a signal. A more symbolic approach is that of the Permutation Entropy (PE) [6], where the ordinal patterns of the signal are analyzed. This method has the advantage of being fast and easy to compute, and of being resistant to outliers. A more recent consideration is the analysis of different time scales using Multiscale Entropy [7], and Multiscale Permutation Entropy (MPE) [8], to capture the hidden similarities that could arise at lower frequencies, often lost in the analysis of their traditional counterparts.

Previous work has been done in the characterization of PE [9] [10]. Although several variants and improvements to MPE exist [11] [12] [13], there is not an comprehensive theory regarding the characterization of MPE. In our previous paper [14], we have already studied the expected value of MPE applied on fractional Gaussian noise, where we found a bias which is only dependent on the time scale. In the present paper, we will further expand the theory regarding the characterization of MPE by applying it to the 1° order Autoregressive (AR) and the 1° order Moving Average (MA) processes. These models are widely used in biomedical signal modeling, and offer an excellent starting point to better understand the MPE with respect to parameter variation. Our contribution is two-fold: we will develop an explicit formulation for the autocorrelation function for the coarse-grained signals of 1° order AR and MA models, which is a necessary step for the MPE calculation. We will also provide and explicit formulation of MPE as a function of time scale and AR or MA parameters.

The article is organized as follows, Section II introduces the theoretical background needed for PE, MPE, AR and MA models. Section III develops the mathematical expressions for coarse-grained 1° order AR and MA models and their respective MPE's. Finally, in Section IV we will apply MPE computation on AR and MA simulations, to provide a benchmark to compare the predictions of our models from Section III.

II. Theorethical Background

In this section we are going to lay the necessary background and tools. In subsection II-A we will establish the formulation of PE. In subsesection II-B we will explain the coarse-graining procedure to calculate the MPE. Finally, in subsection II-C, we will explain briefly the 1° order AR and MA models, necessary for later development.

A. Permutation Entropy

PE [6] measures the level of information contained in the ordinal patterns of the signal. For a fixed dimension d, an ordinal pattern arises in any d consecutive data points. Being N the signal length, n the time index, and dimension $d = 2$, there are two possible patterns in the signal $x_1, \ldots, x_n, \ldots, x_N$: $x_n < x_{n+1}$, and $x_n > x_{n+1}$. For dimension $d = 3$, there

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

exists six possible ordinal patterns. In general, for each chosen dimension d, there are d factorial ($d!$) possible patterns. By counting the number of instances of each pattern, we can construct an estimator of probabilities for each event in the sample space. This estimator is given by:

$$\widehat{p}_i = \frac{\sharp\{n | n \leq N-(d-1), (x_{n+1}, \ldots, x_{n+d}) \ type \ i\}}{N-(d-1)} \quad (1)$$

where $\widehat{p}_i$ is the estimated probability of finding the ordinal pattern i, and $i = [1, \ldots, d!]$. We compute the estimated pattern probability by counting the number patterns i, divided by the total pattern count. Typically, a delay term is included in (1) to deal with oversampled signals. For the remainder of this paper, we will assume the signal is properly sampled, and the patterns are directly obtained from consecutive data points. With these probability estimators established, we can calculate the PE, denoted by $\widehat{\mathcal{H}}$, using information Entropy definition [1]:

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

for all possible $d!$ patterns. The entropy is normalized, so that $\widehat{\mathcal{H}}_{max} = 1$.

PE has the advantages of being simple and fast to compute, and invariant to nonlinear monotonous transformations [10]. Since the method only works with ordinal values, there are no assumptions regarding the probability distribution of the raw data, which makes it robust to a wide variety of applications. The limitations of the method involve the signal length, where the condition $N \gg d!$ must be satisfied to ensure a good probability estimation [10].

B. Coarse-Grained procedure and Multiscale Permutation Entropy

To obtain the MPE, we must first apply a coarse-graining procedure to the original signal to analyze [8]. We divide the signal into adjacent non-overlapping segments of size m, which is a parameter that represents the time scale of interest. We then compute the average of all the data points in each segment, as shown in (3). The new coarse-grained signal of size N/m is composed of the averages of each segment.

$$x_k^{(m)} = \frac{1}{m} \sum_{j=m(k-1)+1}^{km} x_j \quad (3)$$

The MPE technique consists on applying the PE (2) on the resulting coarse-graining signals (3). There is a practical limit to the upper value of the time scale paramenter m, where the condition $N/m \gg d!$ is eventually not satisfied.

C. AR and MA Models

The Autoregressive (AR) and Moving-Average (MA) models are extensively used in the characterization and simulation of time series, particularty in the context of biomedical signals. These models are thus well suited for testing signal processing tools. ARMA models combine the two stochastic processes (AR and MA) in a single formulation. The Autoregressive part weights the influence of a data point in terms of p previous terms, plus a new error element. The Moving-Average, on the other hand, consist of a linear combination of the q previous stochastic error terms. More formally, the ARMA(p,q) model can be written as,

$$x_n = c + \varepsilon_n + \sum_{i=1}^{p} \phi_i x_{n-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{n-j} \quad (4)$$

The ε terms are assumed to be normal, independent, identically distributed (iid) random variables with zero-mean, and variance σ^2. The parameter p is the number of autoregressive terms ϕ_i, and q is the number of moving average terms θ_i. For the simplest case, the 1° degree MA ($p = 0$ and $q = 1$) is explicitly written as:

$$x_n = c + \varepsilon_n + \theta \varepsilon_{n-1} \quad (5)$$

where the expected value and variance are

$$E[X_n] = c \quad (6)$$

$$var(X_n) = \sigma^2(1 + \theta^2). \quad (7)$$

The MA normalized autocorrelation function is given by

$$\rho(\lambda) = \begin{cases} 1, & \text{if } \lambda = 0 \\ \theta/(1+\theta^2), & \text{if } |\lambda| = 1 \\ 0, & \text{otherwise,} \end{cases} \quad (8)$$

where λ represents the distance between data points x_n and $x_{n\pm\lambda}$. Similarly, the 1° degree AR ($p = 1$ and $q = 0$), can be written as,

$$x_n = c + \varepsilon_n + \phi x_{n-1} \quad (9)$$

where

$$E[x_n] = c/(1-\phi) \quad (10)$$

$$var(x_n) = \sigma^2/(1-\phi^2) \quad (11)$$

$$\rho(\lambda) = \phi^{|\lambda|}. \quad (12)$$

As we will study in section III-C, the MPE of AR and MA signals can be expressed as a function of their autocorrelation function. Therefore, it is necessary to first obtain the autocorrelation function of the coarse-grained AR and MA signals, based on (8) and (12).

III. COARSE-GRAINED ARMA MODEL AND MPE

In this section we will explicitly formulate the autocorrelation functions of 1° order MA and AR models, which are necessary to obtain the MPE of these models. As a first approach to the problem, we will limit our analysis to first order models.

A. Coarse-Grained First Order MA

By applying the coarse grained procedure expression (3) into the 1° order MA process definition (5), we get,

$$x_{k,MA}^{(m)} = c + \frac{\theta}{m}\varepsilon_{m(k-1)} + \frac{1}{m}\varepsilon_{mk} + \frac{1+\theta}{m} \sum_{j=m(k-1)+1}^{mk-1} \varepsilon_j, \quad (13)$$

k being the index variable of the new coarse grained signal, and m the scale (hence, the index mk refers to the term m times k from the original signal). From this expression, we can derive the autocovariance function,

$$Cov\big(x^{(m)}_{k,MA}, x^{(m)}_{k+\lambda,MA}\big) \tag{14}$$

$$= \begin{cases} \frac{\sigma^2}{m}\Big(1+\theta^2 \quad +2\Big(\frac{m-1}{m}\Big)\theta\Big), & \text{if } \lambda = 0 \\ \frac{\theta}{m}\sigma^2, & \text{if } |\lambda| = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

(16)

We also obtain the autocorrelation function from (13),

$$\rho^{(m)}_{MA}(\lambda) = \begin{cases} 1, & \text{if } \lambda = 0 \\ \frac{m\theta}{1+\theta^2+2(\frac{m-1}{m})\theta}, & \text{if } |\lambda| = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

B. Coarse-Grained First Order AR

Similarly, if we apply the coarse grained procedure expression in (3) into the 1° order AR process (9), we get,

$$x^{(m)}_{k,AR} = \frac{\phi}{m}\Big(\frac{1-\phi^m}{1-\phi}\Big)x_{m(k-1),AR} + \sum_{j=1}^{m}\Big(\frac{1-\phi^j}{1-\phi}\Big)\varepsilon_{mk+1-j} \tag{18}$$

The variance of (18) is

$$var(x^{(m)}_{k,AR}) = \frac{\sigma^2}{m^2(1-\phi^2)}\Bigg[m\left(\frac{1+\phi}{1-\phi}\right) - \frac{2\phi}{1-\phi}\left(\frac{1-\phi^m}{1-\phi}\right)\Bigg]. \tag{19}$$

The autocovariance function of (18) is given by

$$cov\big(x^{(m)}_{k,AR}, x^{(m)}_{k+\lambda,AR}\big) = \frac{\phi^{m\lambda+2}}{m^2}\Big(\frac{1-\phi^m}{1-\phi}\Big)^2 var(x_{m(k-1),AR}) + \frac{\sigma^2}{m^2}\phi^{m(\lambda+1)}\frac{1-\phi^m}{1-\phi}\sum_{j=1}^{m}\frac{1-\phi^j}{1-\phi} \tag{20}$$

and the autocorrelation is,

$$\rho^{(m)}_{AR}(\lambda) = \phi^{m(\lambda-1)+1}\left[\frac{(1-\phi^m)^2}{m(1-\phi^2)-2\phi(1-\phi^m)}\right] \tag{21}$$

for $|\lambda| > 0$, and $\rho^{(m)}_{AR}(\lambda = 0) = 1$.

It is important to point out that the autocovariance (20) is not equal to the variance (19) for $\lambda = 0$. The variance of each coarse-grained segment (3) consists of the sum of all the individual variances of each element, plus the sum of the covariance of every possible pair of elements within the segment. To obtain the covariance of (18), we compare the elements of two different segments, with no elements in common. This means no variance terms are present, and the calculation of (20) consists only on the sum of the covariance of each possible matching pair of elements between segments. Since no variance term appears, the resulting (20) is not valid when $\lambda = 0$.

Both correlations functions (17) and (21) are necessary for the proper theoretical calculation of their respective MPE's in the next subsection.

C. MPE of First Order MA and AR

For a Gaussian process with stationary increments [9], the MPE is a function only of the autocorrelation of the said process. Furthermore, by exploiting the symmetries of this kind of signals, we only need to calculate the probability of one pattern [9]. For dimension $d = 2$, this yields to the conclusion that both possible patterns have equal probabilities of appearing $(1/2)$, and the resulting PE is always maximum. For dimension $d = 3$ we have the relation

$$p_1 = \frac{1}{\pi}arcsin\left(\frac{1}{2}\sqrt{\frac{1-\rho(2)}{1-\rho(1)}}\right). \tag{22}$$

where $\pi = 3.1415$ and p_i is the pattern probability defined in (2). This is the probability of obtaining the first pattern ($x^{(m)}_k < x^{(m)}_{k+1} < x^m_{k+2}$) of a coarse-grained signal with these conditions. By exploiting the symmetries explained in [9], we have

$$\begin{aligned} p_1 &= p_6 \\ p_2 &= p_3 = p_4 = p_5 = \frac{1-p_1}{4}. \end{aligned} \tag{23}$$

By applying the autocorrelation function (17) of a coarse-grained 1° order MA process into the pattern probability calculation for p_1 (22), we get,

$$p^{MA}_1 = \frac{1}{\pi}arcsin\left(\frac{1}{2}\sqrt{\frac{1+\theta^2+2(\frac{m-1}{m})\theta}{1+\theta^2-(\frac{m^2-2m+2}{m})\theta}}\right) \tag{24}$$

Similarly, if we introduce the autocorrelation of the coarse-grained 1° order AR process (21), we get,

$$p^{AR}_1 = \frac{1}{\pi}arcsin\left(\frac{1}{2}\sqrt{\frac{m(1-\phi^2)-\phi(2-\phi^m)(1-\phi^{2m})}{m(1-\phi^2)-\phi(1-\phi^m)(3-\phi^m)}}\right) \tag{25}$$

Both expressions (24) and (25) are now dependent on the time scale m and the model parameters θ or ϕ, depending on the model. Consequently, by using these probabilities directly into the original Entropy calculation (2), the MPE for $d = 3$ can be described as follows,

$$\mathcal{H} = \frac{-1}{\ln(d!)}\left(2p_1\ln\big(p_1\big) + \big(1-p_1\big)\ln\Big(\frac{1-p_1}{4}\Big)\right) \tag{26}$$

These are the theoretical values for the MPE of coarse-grained MA (24) and AR (25) processes. Since the autocorrelation function depends on the scale m, so does the MPE. We can now compare, in the next section, the models' MPE based on parameters, with 1° order MA and AR simulations.

We limited our analysis for dimension $d = 3$. For dimension

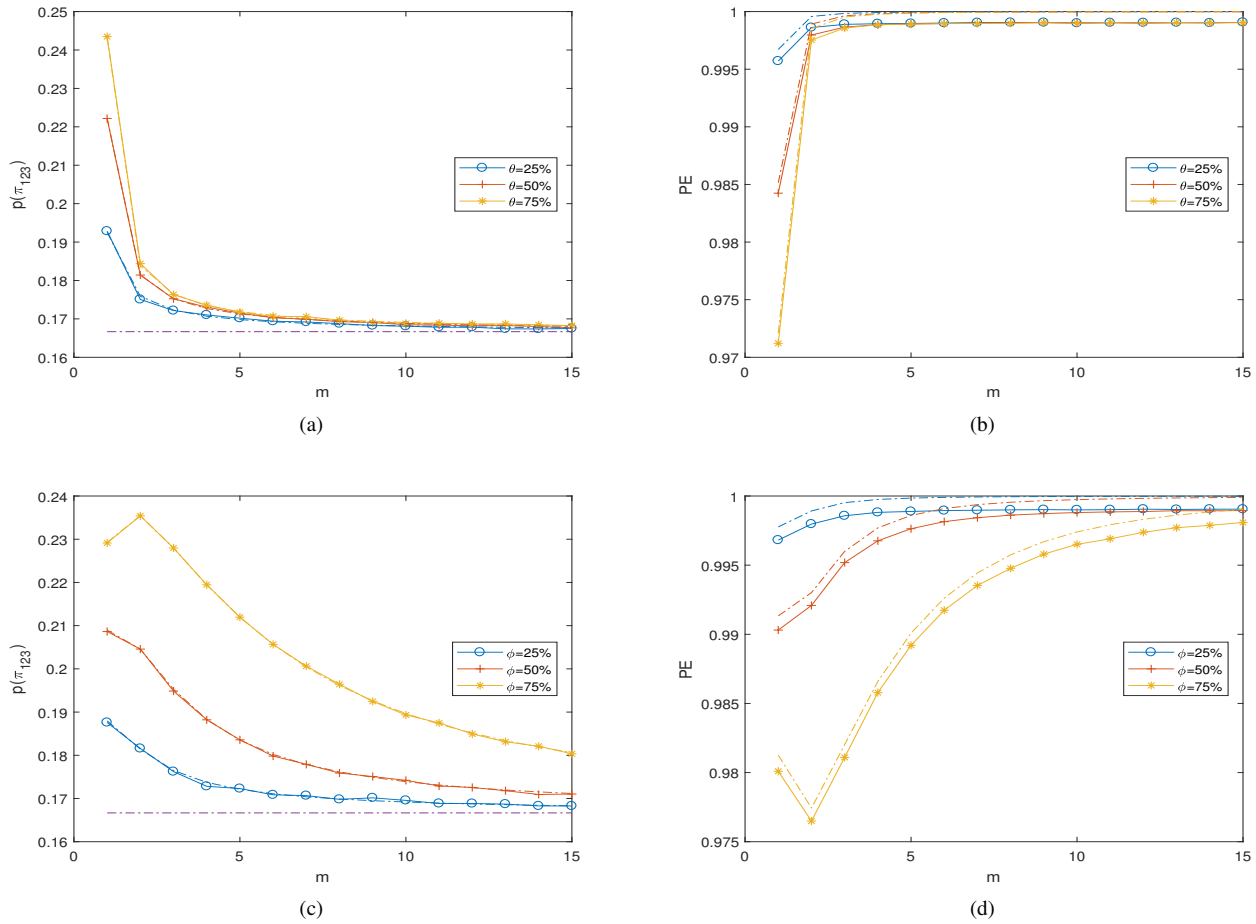

Fig. 1. Theoretical results (dotted lines) vs. average results for p_1 and MPE simulated signals for 1° order MA (top) and 1° order AR (bottom). Each parameter value has 1500 simulated signals of $N = 1000$ data points each. The bottom dotted line in (a) and (c) represent theoretical p_1 of uncorrelated white noise, which corresponds to maximum entropy of one.

$d = 4$, some of the resulting pattern probabilities are complex, so the interpretation of the results become difficult. For dimension $d \geq 5$, the pattern probabilities have no closed expression [9].

IV. RESULTS

To test the theoretical values of the MPE of the 1° order AR and MA processes (26), we performed simulations using increasing values of the model parameter from zero to one to ensure stability. For each parameter test value, we performed 1500 simulations to generate signals of length $N = 1000$ each. The resulting processes were subject to the coarse-graining procedure (3) for time scales $m = 1, \ldots, 15$. Finally, we obtained the MPE of these coarse-grained signals using (2). To avoid the finite-length bias [14], we increased the N at each time scale, to ensure that the coarse-grained signal length remains constant $N/m = 1000$. Albeit this cannot be done on real signals, this assumption allows us to observe the MPE of the models without any artifact effects due to the MPE algorithm itself, or any of it's variants.

Fig. 1a shows the probability p_1 (24) versus time scale in a MA process, for different values of θ. Fig. 1b shows of the corresponding MPE for a MA process (26). The dotted lines represents the theoretical results from the (24) while the solid lines are the result of the simulations explained above.

The differences in probabilities and MPE are heavily dependent on the value of θ in the first time scale. When m increases, the differences become less pronounced, asymptotically approaching to the maximum entropy value, characteristic of an uncorrelated noise. This agrees with our expectations, as the average values of coarse-graining blocks of increasing length become less correlated between them.

Similarly for the 1° order AR process, Fig. 1c presents p_1 (25) respect to time scale, for different values of ϕ, and Fig. 1d shows the respective MPE (26). The values in dotted lines come from the theoretical calculations, where the solid lines come directly from simulations.

The MPE curve for AR also presents an asymptotical approximation to the maximum entropy, which implies that for increasing values of m, the process becomes similar to uncorrelated noise. Nonetheless, the convergence is slower than the MA. This is also in accordance to our expectations,

as the AR model presents longer range correlations between data points.

An interesting MPE curve arises for high values of ϕ, where the curve does not present a monotonical increase. We can observe a minimum MPE for a scale different than $m = 1$. This is an interesting and unexpected result, as it is not evident from the structure of (25). To obtain the critical ϕ where this behavior begins to manifest, we take the argument of the arcsin function (22), and choose a ϕ such that the first two scales have the same probability and MPE ($p_1^{AR}|_{m=1} = p_1^{AR}|_{m=2}$)

$$\begin{aligned} \frac{1-\rho_{AR}^{(1)}(2)}{1-\rho_{AR}^{(1)}(1)} &= \frac{1-\rho_{AR}^{(2)}(2)}{1-\rho_{AR}^{(2)}(1)} \\ \phi(\phi-1)(\phi^2+\phi-1) &= 0 \\ \phi = -1/2 + \sqrt{5}/2 &\approx 0.618, \end{aligned} \tag{27}$$

in which the only root in the range of stability for ϕ corresponds to the inverse of the Golden Ratio. This represents the limit point for ϕ. Below this value, the MPE increases monotonically, as expected. On the other hand, if $\phi > 0.618$ we find a minimum value of MPE at $m > 1$. This suggests that highly correlated AR signals have more regular structures at higher scales, a fact that is not obvious from (18).

We can observe that even a relatively simple model like AR can present less information content at time scales greater than one, and thus, have more regularity and structure. This implies MPE has the potential to extract long-range regularities for real biomedical signals, even when they can be modelled and approximated to simple, well known time series models.

Lastly, we must address the comparison between the predicted MPE values and the simulations. The results closely resemble the predictions in both models, which supports the propositions made in section III-C. It is worth noting that the simulated MPE results, albeit close, are consistently below the predicted values. This can be explained by the small variations in the probability p_1 (23). When the distribution of probabilities does not match the precise symmetries used in section III-C. These variations are translated as a loss in MPE, as we measure some patterns as sightly more probable than others, thus reducing the overall MPE.

V. CONCLUSION

In the present paper, we have built an explicit formulation of the MPE for the 1° order AR and MA processes. This was achieved by means of calculating the autocorrelation function of coarse-grained signals. This yields to an equation that depends both on the process parameters, and the time scale. The theoretical MPE predictions were found to closely resemble the simulations, which add evidence to its validity.

The most outstanding discovery was the non-monotonous increase in MPE curve for high values of the 1° order AR parameter. For an AR parameter value greater than the inverse of the Golden Ratio, the MPE curve presents a minimum at higher scales than the first one, which suggest a hidden, complex relation in the signal trends. This proves that MPE is a powerful tool to detect long-term correlations and information content. Both MA and AR models converge asymptotically to the maximum entropy for high time scales, which is in accordance to our expectations. The MPE of simulated signals, albeit close, present a sightly lower value than the predictions, mainly because they present small variations from the symmetries in the patterns. This is to be expected from simulated random processes.

Since AR and MA models are heavily used in the modeling of biomedical signals, they are well suited as a starting point for testing different signal processing tools and methods. In the case of MPE, it is necessary to know what to expect from these models, to better understand and interpret the results when applying this method to real 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] C. E. Shannon, "A mathematical theory of communication," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 2001.

[2] 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.

[3] S. M. Pincus, "Approximate entropy as a measure of system complexity." *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, 1991.

[4] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.

[5] C. Liu, K. Li, L. Zhao, F. Liu, D. Zheng, C. Liu, and S. Liu, "Analysis of heart rate variability using fuzzy measure entropy," *Computers in Biology and Medicine*, vol. 43, no. 2, pp. 100–108, 2013.

[6] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Physical Review Letters*, vol. 88, no. 17, p. 174102, 2002.

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

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

[9] C. Bandt and B. Shiha. Order patterns in time series. journal of time series analysis. vol 28, no. 5, pp. 646-665, 2007.

[10] 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.

[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 Vol. 23 P28-41. 2016.

[12] A. Humeau-Heurtier, C.-W. Wu, and S.-D. Wu. Refined composite multiscale permutation entropy to overcome multiscale permutation entropy length dependence - IEEE signal processing letters. vol. 22 , issue: 12 , dec. 2015.

[13] J. Zheng, H. Pan, S. Yang, and J. Cheng, "Generalized composite multiscale permutation entropy and laplacian score based rolling bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 99, pp. 229–243, 2018.

[14] A. Dávalos, M. Jabloun, P. Ravier, and O. Buttelli, "Theoretical study of multiscale permutation entropy on FiniteLength fractional gaussian noise," *26th European Signal Processing Conference (EUSIPCO)*, pp. 1092–1096, 2018.