On the Deterministic Estimaton of Multiscale Permutation Entropy of High-Order Autoregressive-Moving-Average Processes as a Function of ARMA Parameters

Antonio Dávalos
PRISME Laboratory
Université d'Orléans
Orléans, France
antonio.davalos-trevino@etu.univ-orleans.fr

Meryem Jabloun
PRISME Laboratory
Université d'Orléans
Orléans, France
meryem.jabloun@univ-orleans.fr

Philippe Ravier
PRISME Laboratory
Université d'Orléans
Orléans, France
philippe.ravier@univ-orleans.fr

Olivier Buttelli
PRISME Laboratory
Université d'Orléans
Orléans, France
olivier.buttelli@univ-orleans.fr

***Abstract*—Multiscale Permutation Entropy (MPE) is one of the most common techniques to assess the ordinal information content within a time series. In the present paper we propose an explicit, deterministic function of the MPE of a general ARMA process, as a function only of its parameters and time scale. We compare our theoretical results with the MPE of corresponding simulated signals, which further support our formulation. We also present an exploration of the effects of the ARMA parameters on the MPE curve, where we found a monotonic decrease of entropy for long-term time scales, and highly non-linear effects on short scales. With these results, we aim to provide a benchmark for the MPE of any real time series modelled as an ARMA process.**

***Index Terms*—Multiscale Permutation Entropy, ARMA Process, Coarse-graining.**

I. Introduction

Permutation Entropy (PE) [1] and Multiscale Permutation Entropy (MPE) [2] are one of the main ordinal techniques in the assessment of information content inside time series. This type of analysis allows the researchers to explore the complexity of signals in widely different contexts, including finance [3] [4] and biomedical data [5] [6] [2] [7]. The PE and MPE are particularly fast and easy to compute, as well as being robust to non-monotonous transformations [1] [2].

One of the first approaches in understanding a stationary signal in time series is to adjust a model based on Gaussian innovations. On top of the well-known properties of Gaussian models, the particular symmetry they present allows the explicit PE calculation [8]. In previous articles, we have extended the mathematical framework of these results for the MPE, by obtaining the explicit MPE values for Fractional Gaussian Noise [9] and first order Autoregresive (AR) and Moving Average (MA) models [10], as well as characterizing the MPE expected value, bias, and variance [11]. For this later case we have confirmed that the MPE can be completely characterized by knowing the AR and MA parameters.This naturally leaves us with a follow-up question: Is the MPE on ARMA(p,q) process of arbitrary order fully determined by its parameters? During the present work, we will extend the results presented in [10] to the general ARMA cases, by exploiting the Gaussian ordinal symmetries in the model under the coarse-graining procedure, which is a necessary step to obtain the MPE.

The paper is organized as follows: Section II summarizes the necessary theoretical background for our analysis. We will present the Permutation Entropy, the Coarse-graining procedure for the Multiscale PE, and the formulation of ARMA models. Section III will develop the MPE directly from the Coarse-Grained ARMA model formulation. We will test the results of our proposition against simulated ARMA signals in Section IV. Finally, in Section V we will discuss the main findings.

II. THEORETICAL BACKGROUND

A. Multiscale Permutation Entropy

For a given time series $\boldsymbol{x} = [x_1, \ldots, x_N]^T$ of size N, the PE [1] measures the information content from the ordinal patterns present. If we select any d consecutive data points $[x_i, \ldots, x_{i+d-1}] \in \mathbb{R}^d$ from $\boldsymbol{x}$, we can extract a unique rank order from the cardinal values. For any embedded dimension d, there are $d!$ possible patterns in the event set. Using the time series $\boldsymbol{x}$, it is possible provide an estimation $\widehat{\boldsymbol{p}} = [\widehat{p}_1, \ldots, \widehat{p}_{d!}]$

Thanks to CONACyT for funding.

of the pattern probabilities $\boldsymbol{p} = [p_1, \ldots, p_{d!}]$ by counting the number of patterns [12],

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

where $i = 1, \ldots, d!$ is the pattern type for d, and τ is a positive integer downsampling parameter for $\boldsymbol{x}$ [13]. For the purposes of this article, we will set $\tau = 1$ without loss of generality.
With $\widehat{\boldsymbol{p}}$, we compute the PE estimate using the Shannon's formulation [1],

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

being $\widehat{\mathcal{H}}$ the estimated PE. This statistic has the advantage of being invariant to nonlinear monotonous transformations [1]. The main disadvantage consists in the necessity of a large N for the computation to be useful. In general, the condition $N \gg d!$ must be satisfied [14].
In [2] the MPE is computed by applying a coarse-grained procedure [15] on the original signal $\boldsymbol{x}$. In order to build a coarse-grained signal $\boldsymbol{x}^{(m)}$, where $m = 1, 2, \ldots$ represents the scale, we must define the coarse-graining procedure as,

$$x_j^{(m)} = \frac{1}{m} \sum_{i=m(j-1)+1}^{jm} x_i = \frac{1}{m} \mathbf{1}^T \left[x_{(j-1)m+1}, \ldots, x_{jm}\right]^T$$
$$= \frac{1}{m} \mathbf{1}^T \boldsymbol{x}_j^{(m)}. \quad (3)$$

This equation divides $\boldsymbol{x}$ into non-overlapping segments of size m. For the vectorial notation, the vertical vector $\boldsymbol{x}_j^{(m)}$ (not to be confused with $x_j^{(m)}$, the coarse grained point on $\boldsymbol{x}^{(m)}$) is composed of all the elements from the segment, and $\mathbf{1}^T = [1, \ldots, 1]$. The resulting coarse signal $\boldsymbol{x}^{(m)}$ is the result of computing the average value of all the data points within these segments. Finally, we obtain the MPE by applying the Entropy definition (2) on $\boldsymbol{x}^{(m)}$.
The MPE is designed to explore the regularities under longer time scale trends, which are not always captured at the original time resolution. Nonetheless, this approach accentuates the need of a large N, since the length of each coarse signal $\boldsymbol{x}^{(m)}$ is reduced by a factor of $1/m$.

B. ARMA Process

The ARMA(p,q) model is one of the most common stationary models for simulation purposes. The model is defined as,

$$X_t = c + \varepsilon_n + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{i=1}^{q} \theta_j \varepsilon_{t-j} \quad (4)$$

where t is a discrete time unit and ε_t is the present innovation, which is an independent Gaussian variable with mean zero and variance σ^2. The AR(p) part has the parameters $\phi_1, \ldots, \phi_p$, while the MA(q) has $\theta_1, \ldots, \theta_q$.

The Autocorrelation function for a pure MA(q) model (which corresponds to the model ARMA(0,q)) is given by [16],

$$\rho_{MA}(\lambda, q) = \begin{cases} \frac{\sum_{i=0}^{q-|\lambda|} \theta_j \theta_{j+|\lambda|}}{\sum_{i=0}^{q} \theta_j^2}, & \text{if } |\lambda| \leq q \\ 0, & \text{if } |\lambda| > q \end{cases} \quad (5)$$

Nontheless, in the case of a general ARMA(p,q), we need to solve the generalized Yule-Walker equations [17] using the AR and MA parameters as the system coefficients. This allows us to obtain the general ARMA autocorrelation function $\rho_{ARMA}(\lambda, p, q)$, which is directly linked to the MPE, as we will explain below.

III. MPE AND COARSE-GRAINED ARMA PROCESS

A. Stationary Gaussian Process Properties

Bandt and Shiha [8] noted that, for a stationary Gaussian process, the pattern probabilities are completely characterized by the signal's autocorrelation function. Moreover, for dimension $d = 3$, they provided the reader with the explicit expression,

$$p_1 = \frac{1}{\pi} \arcsin\left(\frac{1}{2}\sqrt{\frac{1-\rho(2)}{1-\rho(1)}}\right) \quad (6)$$

where ρ is the autocorrelation function, and p_1 is the probability of finding the pattern $x_t < x_{t+1} < x_{t+2}$ (note that this is not an estimator, but the true probability). Thus, if we know the random process beforehand, we do not need to estimate the pattern probabilities from equation (1). The stationary Gaussian signals also satisfy the following symmetries [8],

$$p_1 = p_6$$
$$p_2 = p_3 = p_4 = p_5 = \frac{1-2p_1}{4}. \quad (7)$$

The pattern probability p_6 corresponds to the inverse pattern of p_1, where $x_t > x_{t+1} > x_{t+2}$. The probabilities $p_2, \ldots, p_5$ correspond to all the other possible pattern probabilities in the signal. Therefore, the pattern probability distribution is completely determined by the autocorrelation function. We will limit our analysis to the embedded dimension $d = 3$, since higher dimensions present complex probabilities and, in general, do not have a closed form [8].
The ARMA(p,q) process is indeed a stationary Gaussian process, so the symmetries in (7) apply. Coarse-grained ARMA signals have not the same formulation as the original signal, but are nonetheless stationary Gaussian, since the Coarse-Graining procedure is built with a linear combination of the elements of $\boldsymbol{x}$. We have previously provided an explicit formulation for the Coarse-Grained MA(1) and AR(1) signal in [10]. In the following section, we will provide a MPE expression for the general ARMA(p,q), as a function of its parameters.

B. MPE for General ARMA Process

To obtain the MPE of an ARMA process, it is necessary to find the autocorrelation function for the coarse-grained signal, so we can compute the pattern probabilities using (6) and (7). Let $\boldsymbol{X}$ be a ARMA(p,q) process of length N, with parameters $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$. The coarse-grained signal $\boldsymbol{X}^{(m)} = [X_1^{(m)}, \ldots, X_{N/m}^{(m)}]$ is obtained by applying the Coarse-Grained procedure (3) to the process $\boldsymbol{X}$ (We use capital $\boldsymbol{X}$ here to denote the signal as a random process, instead of a real measured signal $\boldsymbol{x}$ in Section II-A). If we write the autocovariance function explicitly in terms of the segment vectors (3), we have,

$$\begin{aligned}
& cov\left(X_j^{(m)}, X_{j+|\lambda|}^{(m)}\right) \\
&= E\left[X_j^{(m)}\, X_{j+|\lambda|}^{(m)}\right] - E\left[X_j^{(m)}\right] E\left[X_{j+|\lambda|}^{(m)}\right] \\
&= E\left[\mathbf{1}^T \boldsymbol{X}_j^{(m)}\, \boldsymbol{X}_{j+|\lambda|}^{(m)} \mathbf{1}\right] - E\left[\mathbf{1}^T \boldsymbol{X}_j^{(m)}\right] E\left[\mathbf{1}^T \boldsymbol{X}_{j+|\lambda|}^{(m)}\right] \\
&= \tfrac{1}{m^2} \mathbf{1}^T \left(E\left[\boldsymbol{X}_j^{(m)}\, \boldsymbol{X}_{j+|\lambda|}^{(m)}\right] - E\left[\boldsymbol{X}_j^{(m)}\right] E\left[\boldsymbol{X}_{j+|\lambda|}^{(m)}\right]\right) \mathbf{1} \\
&= \tfrac{1}{m^2} \mathbf{1}^T \boldsymbol{K}\left(\boldsymbol{X}_j^{(m)}, \boldsymbol{X}_{j+|\lambda|}^{(m)}\right) \mathbf{1} \\
&= \tfrac{\sigma^2}{m^2} \mathbf{1}^T \boldsymbol{R}(m\lambda) \mathbf{1} \qquad (8)
\end{aligned}$$

where the matrix $\boldsymbol{K}$ is the crosscovariance matrix between the vectors $\boldsymbol{X}_j^{(m)}$ and $\boldsymbol{X}_{j+|\lambda|}^{(m)}$, which represent different segments on the coarse-grained signal $\boldsymbol{X}^{(m)}$. We obtain the crosscorrelation matrix $\boldsymbol{R}(m\lambda)$ by decoupling the constant innovation variance σ^2 from $\boldsymbol{K}$. Explicitly, $\boldsymbol{R}(m\lambda)$ is,

$$\boldsymbol{R}(m\lambda) = \begin{bmatrix} \rho(m\lambda) & \rho(m\lambda+1) & \ldots & \rho(m\lambda+m-1) \\ \rho(m\lambda-1) & \rho(m\lambda) & \ldots & \rho(m\lambda+m-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(m\lambda-m+1) & \rho(m\lambda-m+2) & \ldots & \rho(m\lambda) \end{bmatrix} \qquad (9)$$

which is a Toeplitz matrix where the diagonal contains the autocorrelation function from the original $\boldsymbol{X}$ with lag $m\lambda$. The quadratic form in equation (8) is a compact form to express the summation of all the elements in the matrix $\boldsymbol{R}(m\lambda)$.

If we divide equation (8) by the variance of $X_j^{(m)}$, we obtain the autocorrelation function $\rho^{(m)}(\lambda, m)$ of the Coarse-Grained ARMA signal,

$$\rho^{(m)}(\lambda, m) = \frac{\mathbf{1}^T \boldsymbol{R}(m\lambda) \mathbf{1}}{\mathbf{1}^T \boldsymbol{R}(0) \mathbf{1}}, \qquad (10)$$

where $\boldsymbol{R}(0)$ is

$$\boldsymbol{R}(0) = \begin{bmatrix} 1 & \rho(1) & \ldots & \rho(m-1) \\ \rho(1) & 1 & \ldots & \rho(m-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(m-1) & \rho(m-2) & \ldots & 1 \end{bmatrix} \qquad (11)$$

which is symmetric. Since the coarse grained ARMA signal autocorrelation function is completely characterized, we can obtain the pattern probability for dimension $d = 3$ using equation (6), which also automatically determines the theoretical MPE of the system by means of equation (2). With this model, it is now possible to obtain a complete MPE prediction, as a function of the time scale, for any ARMA model solely as a function of its parameters and the time scale.

IV. RESULTS

To test the validity of the obtained theoretical MPE expression from (10), we proposed a series of experiments involving simulated signals for several ARMA(p,q) model variations. We selected the following models:

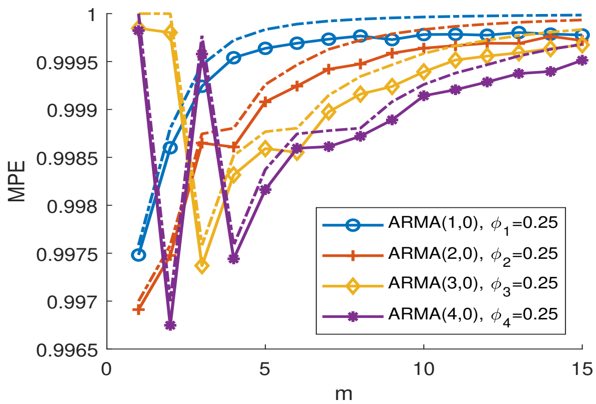
1) ARMA(p,0) with a single parameter $\phi_p = 0.25$, with increasing order p, with all lower order parameters set to zero.
2) ARMA(0,q) with a single parameter $\theta_q = 0.25$ with increasing q, also with lower order parameters equal to zero.
3) ARMA(1,q), with fixed AR parameter $\phi_1 = 0.5$, and adding a new MA terms $\theta_1 = \cdots = \theta_q = 0.1$ with increasing order.
4) ARMA(p,1), with fixed MA parameter $\theta_1 = 0.5$, and adding a new MA terms $\phi_1 = \cdots = \phi_p = 0.1$, also with increasing order.

Each dotted curve is composed of theoretical values using equations (6) and (2), using the coarse-grained ARMA autocorrelation function (10). Each solid curve is composed of the average MPE values of 100 simulated signals from the corresponding models. The signal length was fixed at $N = 5000$. As we did in previous work regarding ARMA models [10], we increased the signal length at each time scale, using $m * N$ data points, to avoid the decreasing linear bias due to the coarse-graining effect [9]. This is not possible for real signals, but will suffice as a benchmark to test our results, as shown in Figure 1.

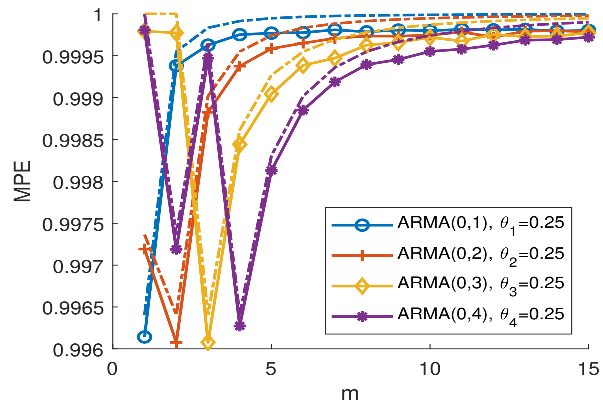
V. DISCUSSION

We first observe that both theoretical curves and and the simulations behave almost identically. This is observed on all curves tested. The simulation MPE results appear consistently below the predicted values, which agrees with our previous findings [9] where we found the MPE statistic to be biased. Since the experiments keep a constant signal length of $N * m = 5000$, the corresponding MPE bias of remains constant along the curve.

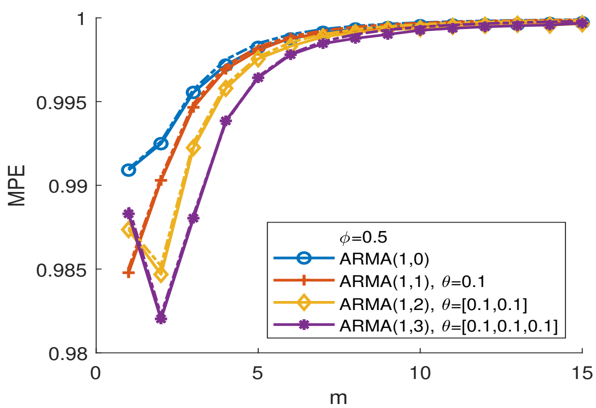
Figures 1a and 1b show the effect of increasing the order of the ARMA(p,0) and ARMA(0,q) models, respectively. In both cases, we found the curves to monotonically decrease their MPE value with increasing order, but only for $m \geq p$ or $m \geq q$. For time scales below the models' order, the MPE shows a highly nonlinear behavior. We must note this is not a statistical effect from the random process, as the theoretical deterministic results match. This effect is not clearly evident in Figures 1c and 1d, albeit we can still claim a monotonical decrease for large m. We need to perform further tests to fully characterize these interactions, as well as its effect on



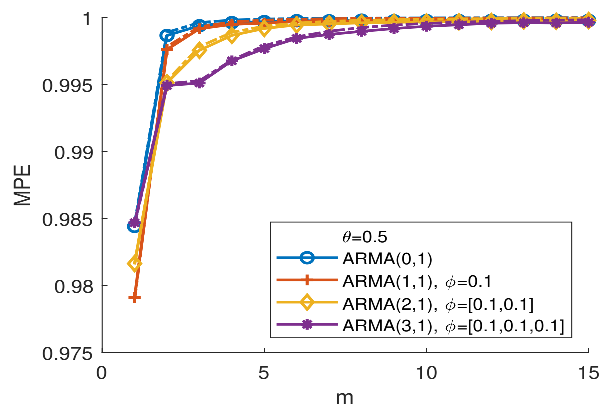
(a)



(b)



(c)



(d)

Fig. 1: MPE vs. time scale for: (a) ARMA(p,0) and (b) ARMA(0,q) with a single parameter of increasing order, c) ARMA(1,q) and d) ARMA(p,1) with increasing number of parameters. The dotted lines represent the MPE results from our model, while the solid lines follow the average MPE from simulations.

the measured MPE.

Since a disminution of the MPE value implies a disminution of information content [1], we can conclude that a high order ARMA model contains more information within its parameters, and less information in the random signal itself, but only if we consider long term effects. The short time scales seem to be ruled by non-linear interactions, which need further analysis. As a general remark, we should also note that the curves for the ARMA(1,0) and ARMA(0,1) model match the results of our previous work [10], where the coarse-grained autocorrelation, and thus, the MPE, are explicitly presented.

VI. Conclusion

In the present paper we have developed the mathematical framework to obtain the theoretical Multiscale Permutation Entropy of a general ARMA model, as a deterministic function of its parameters. We tested the MPE predictions against Entropy measurements from simulated signals, which closely follow the predicted curve, supporting the validity of our formulation.

This work extend the previous results we obtained for the first order Autoregressive and Moving Average processes [10]. Since the symmetry properties of Gaussian models allows the calculation of the exact pattern probability distribution for embedded dimension equal to three, we are able to find the theoretical MPE exactly, based on the signal autocorrelation function. This is true even after the Coarse-Graining procedure, which allows us to obtain the MPE at any desired time scale. Our proposed function allowed us to explore the parameters' effect on the MPE curve, finding a monotonical decrease in Entropy for long-term time scales, and a strongly non-linear variation in short time scales.

With this work, we can conclude that the Multiscale Permutation Entropy of any ARMA model is a deterministic function which only depends on the model's parameters. This allows us to assess the information content linked to the model itself, for short fluctuations as well as for trends. By knowing the theoretical MPE benchmark, we can compare real-signal ARMA model estimations by means of Permutation Entropy.

References

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

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

[3] Y. Yin and P. Shang, "Weighted multiscale permutation entropy of financial time series," *Nonlinear Dynamics*, vol. 78, no. 4, pp. 2921–2939, 2014.

[4] L. Zunino, M. Zanin, B. Tabak, D. Pérez, and O. Rosso, "Forbidden patterns, permutation entropy and stock market inefficiency," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 14, pp. 2854–2864, 2009.

[5] M. Zanin, L. Zunino, O. Rosso, and D. Papo, "Permutation entropy and its main biomedical and econophysics applications: A review," *Entropy*, vol. 14, no. 8, pp. 1553–1577, 2012.

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

[7] H. Azami and J. Escudero, "Improved multiscale permutation entropy for biomedical signal analysis: Interpretation and application to electroencephalogram recordings," *Biomedical Signal Processing and Control 23 P28-41*, 2015.

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

[9] A. Dávalos, M. Jabloun, P. Ravier, and O. Buttelli, "Theoretical study of multiscale permutation entropy on finite-length fractional gaussian noise," *26th European Signal Processing Conference (EUSIPCO). Rome, Italy. IEEE Xplore*, pp. 1092–1096, 2018.

[10] ——, "Multiscale permutation entropy: Statistical characterization on autoregressive and moving average processes," *27th European Signal Processing Conference (EUSIPCO). A Coruña, Spain. IEEE Xplore*, pp. 1092–1096, 2019.

[11] ——, "On the statistical properties of multiscale permutation entropy: Characterization of the estimator's variance," *Entropy*, vol. 21, no. 5, p. 450, 2019.

[12] L. Zunino, D. G. Pérez, M. T. Martín, 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.

[13] L. Zunino, D. G. Pérez, A. Kowalski, M. T. Martín, M. Garavaglia, A. Plastino, and O. A. Rosso, "Fractional brownian motion, fractional gaussian noise, and tsallis permutation entropy," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 24, pp. 6057–6068, 2008.

[14] A. M. Kowalski, M. T. Martín, A. Plastino, and O. A. Rosso, "Bandt–pompe approach to the classical-quantum transition," *Physica D: Nonlinear Phenomena*, vol. 233, no. 1, pp. 21–31, 2007.

[15] M. Costa, A. Goldberg, and C. Peng, "Multiscale entropy analysis of biological signals," *Phys. Rev. E 71, 021906*, 2005.

[16] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 2nd ed., ser. Springer Texts in Statistics. Springer-Verlag, 2002. [Online]. Available:

[17] ——, *Time Series Analysis, Forecasting and Control. Revised Edition*. Holden-Day, 1976.