engineering proceedings

MDPI

Proceeding Paper

Assessing the Preprocessing Benefits of Data-Driven Decomposition Methods for Phase Permutation Entropy—Application to Econometric Time Series †

Erwan Pierron ‡ and Meryem Jabloun *,‡

Prisme Laboratory, Université d'Orléans, 12 Rue de Blois, 45067 Orléans, France; erwan.pierron@etu.univ-orleans.fr

* Correspondence: meryem.jabloun@univ-orleans.fr

† Presented at the 10th International Conference on Time Series and Forecasting, Gran Canaria, Spain, 15–17 July 2024.

‡ These authors contributed equally to this work.

Abstract: This paper investigates the efficacy of various data-driven decomposition methods combined with Phase Permutation Entropy (PPE) to form a promising complexity metric for analyzing time series. PPE is a variant of classical permutation entropy (PE), while the examined data-driven decomposition methods include Empirical Mode Decomposition (EMD), Variational Mode Decomposition (VMD), Empirical Wavelet Transform (EWT), Seasonal and Trend decomposition using Loess (STL), and Singular Spectrum Analysis-based decomposition (SSA). To our knowledge, this combination has not been explored yet. Our primary aim is to assess how these preprocessing methods affect PPE's ability to capture temporal structural complexities within time series. This evaluation encompasses the analysis of both simulated and econometric time series. Our results reveal that combining SSA with PPE produces superior advantages for measuring the complexity of seasonal time series. Conversely, VMD combined with PPE proves to be the less advantageous strategy. Overall, our study illustrates that combining data-driven preprocessing methods with PPE offers greater benefits compared to combining them with traditional PE in quantifying time series complexity.

Keywords: time series; permutation entropy; phase permutation entropy; data-driven signal decomposition; EMD; VMD; EWT; STL; SSA

check for updates

Citation: Pierron, E.; Jabloun, M. Assessing the Preprocessing Benefits of Data-Driven Decomposition Methods for Phase Permutation Entropy—Application to Econometric Time Series. *Eng. Proc.* **2024**, *68*, 28. https://doi.org/10.3390/engproc2024068028

Academic Editors: Olga Valenzuela, Fernando Rojas, Luis Javier Herrera, Hector Pomares and Ignacio Rojas

Published: 9 July 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Time series analysis plays an important role in various scientific fields, spanning from economics to biology and physics. Understanding temporal structures, emerging trends, and fluctuations in time series can be addressed through reliable complexity measures. For instance, permutation entropy (PE) and its variant Phase PE (PPE) have emerged as significant tools in this domain [1,2]. PE, introduced in [1], serves as a natural complexity measure exploring ordinal patterns extracted from raw signals, whereas PPE, defined in [2], applies PE to instantaneous signal phases obtained through the Hilbert transform.

Many attempts have combined data-driven decomposition methods and PE to obtain a multiscale effect to unveil hidden signal structures. However, our investigation in [3] has shown limited efficacy of such a combination in enhancing PE interpretation. Conversely, to our knowledge, the combination of such methods with PPE has not been addressed yet, although data-driven decomposition methods like Empirical Mode Decomposition (EMD), Variational Mode Decomposition (VMD), Empirical Wavelet Transform (EWT), Seasonal and Trend decomposition using Loess (STL), and Singular Spectrum Analysis-based decomposition (SSA) have emerged as effective methods that provide oscillating components that are well suited for the Hilbert transform, facilitating the extraction of phase information.

In the present paper, we aim to explore the effectiveness of using data-driven signal decomposition methods as preprocessing prior to the PPE measure in order to capture dynamic complexities. Specifically, we aim to identify a preprocessing method capable of enhancing the interpretation of the PPE measure.

The paper is organized as follows: Section 2 provides a brief overview of the complexity measures, namely PE and PPE, as well as the considered signal decomposition methods. Section 3 presents our findings, offering critical reflections on the implications of signal decomposition methods for improving the interpretation of the PPE measure. Finally, Section 4 concludes the paper.

2. Permutation Entropy and Phase Permutation Entropy

PE is a complexity measure derived from transforming a time series into a sequence of fixed-length ordinal patterns (OPs) and counting the number of unique permutations among these OPs. An OP of length d is obtained by arranging each d consecutive sample values in ascending order. Consider N consecutive samples of the time series $X = x_1, x_2, \ldots, x_N$; normalized PE is given by:

$$PE = -\frac{1}{\log(d!)} \sum_{i=1}^{d!} p_i \log(p_i), \tag{1}$$

where d is the OP length and p_i is the probability of occurrence of the i-th OP Π_i, defined by:

$$p_i = -\frac{\{\#t, \{x_t, x_{t+1}, \ldots, x_{t+d-1}\} \text{ of type } \Pi_i\}}{N - d + 1}. \tag{2}$$

Here, # denotes cardinality and OP Π_i is obtained through a permutation of numbers from 1 to d. PE (1) enables the capture of the time series structure without imposing strong assumptions on the data distribution or the presence of specific trends. Specifically, a high PE is associated with chaotic or unpredictable time series, while a lower PE signifies highly regular or predictable time series.

PPE [2] is an extension of PE that incorporates phase information in evaluating the complexity of time series by substituting Φ_t for x_t in (2), with Φ_t defined as:

$$\Phi_t = \arctan\left(\frac{\tilde{x}_t}{x_t}\right), \text{ and } \tilde{x}_t = \text{Hilbert transform of } x_t. \tag{3}$$

Unlike PE, PPE preserves information about the instantaneous phase and proves to be particularly useful in the analysis of time series characterized by oscillatory dynamics, such as biological signals or periodic phenomena.

Preprocessing steps, such as linear filtering or nonlinear signal decomposition, have been proposed in the literature as a preliminary step before calculating PE. This step aims to explore signal structures across different scales, enabling the consideration of both local and global OPs. However, in [3], it was shown that the influence of data-driven signal decomposition on the PE measure is closely linked to the shift in the mean frequency of obtained components, especially for OPs of length $d = 3$ and 4. This type of processing does not provide new insights into the complexity measure or time series.

However, considering the construction of PPE, we are confident that employing data-driven signal decomposition as preprocessing will enhance the interpretation of PPE. To our knowledge, this aspect has not been addressed yet. We will explore several data-driven decomposition methods: Empirical Mode Decomposition (EMD) [4], Variational Mode Decomposition (VMD) [5], Empirical Wavelet Transform (EWT) [6], Seasonal-Trend Decomposition using LOESS (STL) [7], and Singular Spectrum Analysis-based decomposition (SSA) [8]. Each method has its advantages and drawbacks, and is suitable for specific types of time series and analytical applications.

EMD is a widely used method for decomposing nonlinear and non-stationary time series. It iteratively extracts oscillatory components, known as Intrinsic Mode Functions (IMFs), which are well suited for the Hilbert transform and capture various scales of signal variations [4].

Unlike EMD, VMD employs a total energy minimization approach with specific constraints to theoretically achieve an optimal signal decomposition into a sum of IMFs while offering increased flexibility in separating different scales of variation. Compared to EMD, VMD is known to be time-consuming [5].

EWT is a wavelet-based data decomposition method that aims to separate different temporal components using adaptive filters. One of its key advantages is its ability to automatically adjust its filters based on the local characteristics of the signal, allowing for efficient decomposition even in signals with rapid or non-stationary variations [6].

STL, on the other hand, is specifically designed for separating seasonal and trend components of time series using a local smoothing approach known as LOESS. This method is particularly valuable in fields where seasonal variations play a significant role, such as climate or economic time series analysis [7].

SSA has a theoretical foundation that aligns closely with principal component analysis. SSA is helpful in identifying principal processes and enhancing hidden periodicities within the data. Extracted components hold physical significance, representing either orthogonal oscillatory components with a narrow spectral band [8].

In the following section, we demonstrate the benefits of data-driven signal decomposition-based preprocessing for PPE interpretation when applied to econometric time series.

3. Results and Discussion

To attest to the benefit of data-driven decomposition methods to PPE compared to PE, we consider both simulated data and real econometric data.

3.1. Simulated Data

We propose to study four signal-generating models:

- Zero mean white Gaussian noise WGN with unitary variance,
- An autoregressive (AR) model of order 16 whose coefficients are randomly chosen as

$$\begin{aligned}[a_0, a_1, \ldots, a_{16}] = [&-0.0021, -0.0108, -0.0274, -0.0410, -0.0265, 0.0374,\\ &0.14350.2465, 0.2895, 0.2465, 0.1435, 0.0374, -0.0265,\\ &-0.0410, -0.0274, -0.0108, -0.0021]\end{aligned}$$

- Two random Fourier series with distinct reduced fundamental frequencies, namely, $\nu_0 = 0.013$ and 0.0013, defined as

$$x_t = \sum_{n=0}^{+\infty} \epsilon_n X_n \sin(2\pi n \nu_0 t + \varphi_n). \quad (4)$$

The random Fourier series is a stochastic process that extends the concept of a traditional periodic Fourier series by introducing randomness into its coefficients. The randomization of these coefficients has a smoothing effect (regularity) [9,10]. Indeed, the singularities have no more specific locations but are "spread everywhere". The independent Rademacher random variables, ϵ_n, take values ± 1 with equal probabilities, each having a probability of $\frac{1}{2}$. The deterministic coefficients X_n are assumed to satisfy $\sum_{n=0}^{\infty} |X_n|^2 < \infty$ to ensure the convergence of the time series almost everywhere toward $x_t \in L^2\left(\left[0, \frac{1}{\nu_0}\right]\right)$, which in accordance with the Paley–Zygmund theorem [9]. The initial phases φ_n are assumed iid with a uniform distribution over the interval $[-\pi, \pi]$ and are independent of ϵ_n.

The simulated results are based on the average of 100 Monte Carlo simulations for each model. At least 8 IMFs are obtained using the EMD method, and at maximum,

16 IMFs are extracted using the VMD, EWT, SSA, and STL. Our findings are illustrated in Figure 1 for an OP length $d = 3$ for both PE and PPE.

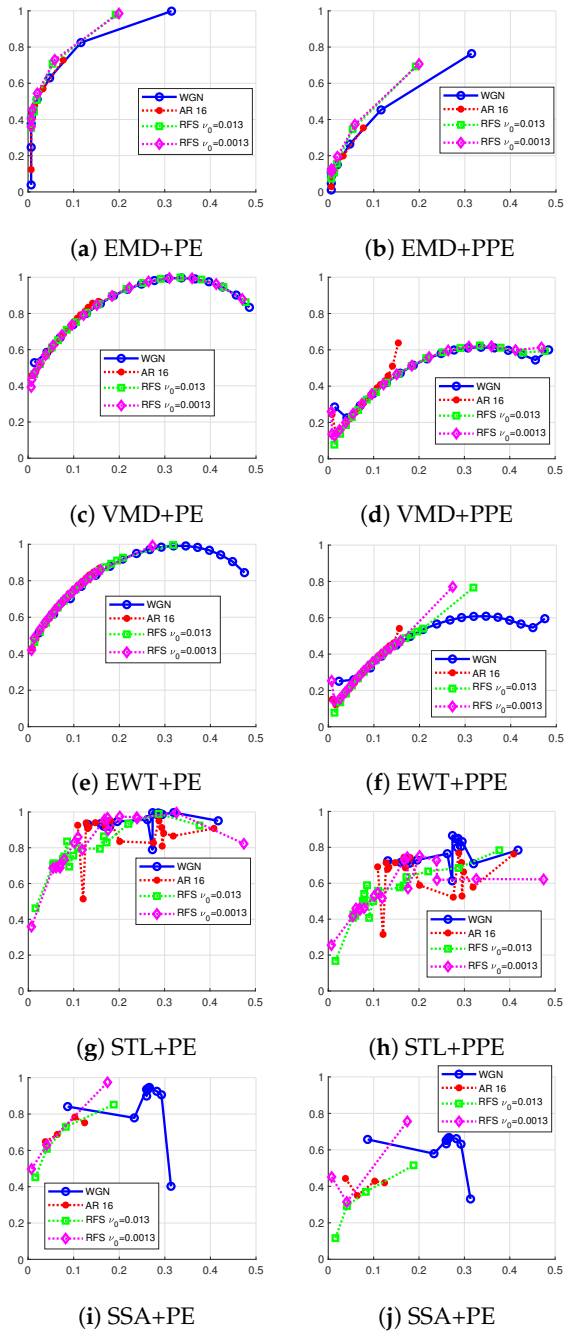


Figure 1. Data-driven signal decomposition combined with PE (left column) and PPE (right column) applied to 4 signal-generating models: WGN, an AR(16) model, and 2 random Fourier series (RFS) with reduced fundamental frequencies, namely, $\nu_0 = 0.013$ and 0.0013. The simulated results are based on the average of 100 Monte Carlo simulations for each model. IMF PE and PPE (y-axis) are displayed as a function of the reduced frequency (x-axis) of each obtained IMF.

The combination of data-driven signal decomposition-based techniques with PE measures yields a multiscale PE approach characterized by a single parameter to adjust, namely, the OP length. This combination effectively eliminates the inherent limitations of conventional linear filtering-based multiscale PEs [3]. Nonetheless, as emphasized in [3] and illustrated in Figure 1a,c,e,g,i, the typical variation in IMF PE observed for each generating process can primarily be solely linked to the spectrum shifts of these IMFs, which is par-

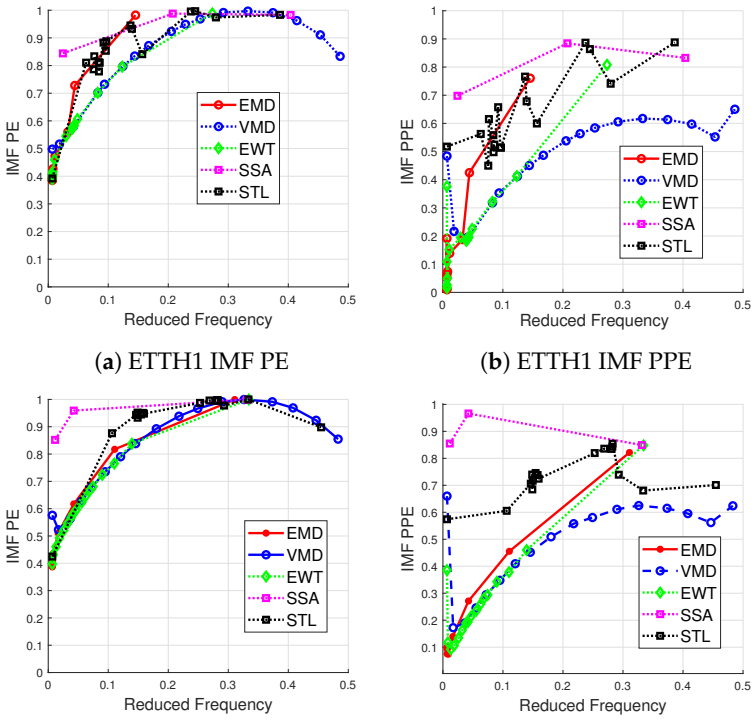
ticularly evident when using methods like VMD, EWT, and STL. Notably, the majority of IMF PE values closely align with those of WGN IMFs. Few exceptions are noted in IMFs obtained using strategies such as EMD and SSA.

In Figure 1b,d,f,h,j, we depict the variations in PPE values across the IMFs obtained using all the aforementioned data-driven methods and signal-generating models. Noticeable deviations in PPE values from those of WGN processes are observed for specific IMFs, except those obtained using the STL strategy. The combination of EMD or EWT with PPE effectively distinguishes between random Fourier series (RFS) and Gaussian processes (WGN and AR(16)). Moreover, combining VMD with PPE reveals higher PPE values in the IMFs with the highest frequency extracted from AR processes compared to other generating models. Similarly, SSA combined with PPE highlights discernible differences between WGN IMFs and those derived from other processes.

This study emphasizes the highest sensitivity of PPE to the specific characteristics of each decomposition method, given their distinct approaches to extracting AM-FM components. This observation underscores the importance of selecting an appropriate decomposition method when evaluating PPE, especially when dealing with real data. Additionally, the combination of data-driven signal decomposition methods with PPE appears to enhance sensitivity to signal-generating models. It appears that EMD and SSA are particularly beneficial for enhancing PPE's discriminate power among different signal-generating models.

3.2. Real Data

In Figure 2, we compare the impact of data-driven signal decomposition on PE and PPE of real signals: ETTH1 and M3Forcast time series from the M3 competition dataset. Our analysis reveals that employing EMD or SSA as preprocessing steps enhances the ability of both PE and PPE to discriminate between these two time series. Notably, this discrimination is particularly pronounced when using PPE. Conversely, IMFs obtained using VMD and EWT do not significantly aid in distinguishing between the ETTH1 and M3Forecast time series.



(**a**) ETTH1 IMF PE (**b**) ETTH1 IMF PPE

(**c**) M3Forecast IMF PE (**d**) M3Forecast IMF PPE

Figure 2. Data-driven signal decomposition combined with PE (left column) and PPE (right column) applied to real data: (**a**,**b**) EHHT1 data and (**c**,**d**) M3Forecast data from M3 competition dataset.

4. Conclusions

Our study represented a comprehensive exploration of the impact of various data-driven signal decomposition methods, including EMD, VMD, EWT, SSA, and STL, as preprocessing steps before calculating the complexity measure, Phase Permutation Entropy (PPE). Among these considered decomposition methods, SSA and EMD emerge as particularly interesting preprocessing approaches for enhancing the discriminatory power of PPE when analyzing econometric data.

Author Contributions: Conceptualization, methodology, software, validation, formal analysis, investigation: M.J. and E.P.; Writing—original draft preparation: E.P. and M.J.; Resources, review editing, visualization, supervision, project administration, funding acquisition: M.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available at https://forecasters.org/resources/time-series-data/m3-competition/ (accessed on 1 July 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bandt, C.; Pompe, B. Permutation Entropy: A Natural Complexity Measure for Time Series. *Phys. Rev. Lett.* **2002**, *88*, 174102. [CrossRef]
2. Kang, H.; Zhang, X.; Zhang, G. Phase permutation entropy: A complexity measure for nonlinear time series incorporating phase information. *Phys. Stat. Mech. Its Appl.* **2021**, *568*, 125686. [CrossRef]
3. Jabloun, M.; Ravier, P.; Buttelli, O. On the Genuine Relevance of the Data-Driven Signal Decomposition-Based Multiscale Permutation Entropy. *Entropy* **2022**, *24*, 1343. [CrossRef] [PubMed]
4. Zhang, J.; Feng, F.; Marti-Puig, P.; Caiafa, C.F.; Sun, Z.; Duan, F.; Solé-Casals, J. Serial-EMD: Fast empirical mode decomposition method for multi-dimensional signals based on serialization. *Inf. Sci.* **2021**, *581*, 215–232. [CrossRef]
5. Lu, Y.; Ma, H.; Zhang, Z.; Jiang, L.; Sun, Y.; Song, Q.; Liu, Z. Real-time chatter detection based on fast recursive variational mode decomposition. *Int. J. Adv. Manuf. Technol.* **2023**, *130*, 3275–3289. [CrossRef]
6. Fan, G.F.; Peng, L.L.; Hong, W.C. Short-term load forecasting based on empirical wavelet transform and random forest. *Electr. Eng.* **2022**, *104*, 4433–4449. [CrossRef]
7. Ouyang, Z.; Ravier, P.; Jabloun, M. STL Decomposition of Time Series Can Benefit Forecasting Done by Statistical Methods but Not by Machine Learning Ones. *Eng. Proc.* **2021**, *5*, 42. [CrossRef]
8. Harmouche, J.; Fourer, D.; Auger, F.; Borgnat, P.; Flandrin, P. The Sliding Singular Spectrum Analysis: a Data-Driven Non-Stationary Signal Decomposition Tool. *IEEE Trans. Signal Process.* **2017**, *66*, 251–263. [CrossRef]
9. Paley, R.E.A.C.; Zygmund, A. On some series of functions, (3). *Math. Proc. Camb. Philos. Soc.* **1932**, *28*, 190–205. [CrossRef]
10. Kahane, J. *Some Random Series of Functions*; Cambridge Studies in Advanced Mathematics; Cambridge University Press: Cambridge, UK, 1985.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.