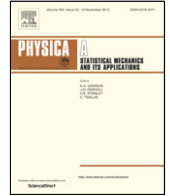




Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Comparing different approaches to compute Permutation Entropy with coarse time series

Francisco Traversaro ^{a,b}, Nicolás Ciarrocchi ^c, Florencia Pollo Cattaneo ^d,
Francisco Redelico ^{e,f,*}

^a CONICET - Universidad Nacional de Lanús, Grupo de Investigación en sistemas de información, 29 de Septiembre 3901, B1826GLC Lanús, Buenos Aires, Argentina

^b Pontifical Catholic University of Argentina (UCA), Av. Alicia Moreau de Justo 1300 - (C1107AAZ) Ciudad Autónoma de Buenos Aires, Argentina

^c Hospital Italiano de Buenos Aires, Servicio de Terapia Intensiva Adultos, Perón 4190 - (C1199ABB) Ciudad Autónoma de Buenos Aires, Argentina

^d Universidad Tecnológica Nacional, Facultad Regional Buenos Aires, Programa de Maestría en Ingeniería en Sistemas de Información y Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS) - Av. Medrano 951, (C1179AAQ) Ciudad Autónoma de Buenos Aires, Argentina

^e CONICET - Hospital Italiano de Buenos Aires, Departamento de Informática en Salud, Perón 4190 - (C1199ABB) Ciudad Autónoma de Buenos Aires, Argentina

^f Universidad Nacional de Quilmes - Departamento de Ciencia y Tecnología, Roque Sáenz Peña 352 (B1876BXD), Bernal Buenos Aires, Argentina

HIGHLIGHTS

- A novel method for imputation of tied values is presented.
- This method is evaluated against other presented in the literature.
- The proposed method outperform the compared one.

ARTICLE INFO

Article history:

Received 5 March 2018

Received in revised form 23 June 2018

Available online xxxx

Keywords:

Permutation entropy

Coarse time series

Non linear dynamics

ABSTRACT

Bandt and Pompe introduced Permutation Entropy as a complexity measure and has been widely used in time series analysis and in many fields of nonlinear dynamics. In theory these time series come from a process that generates continuous values, and if equal values exists in a neighborhood, $x_{t^*} = x_t, t^* \neq t$, they can be neglected with no consequences because their probability of occurrence is insignificant. Since then, this measure has been modified and extended, in particular in cases when the amount of equal values in the time series is large due to the observational method, and cannot be neglected. We test the new Data Driven Method of Imputation that cope with this type of time series without modifying the essence of the Bandt and Pompe Probability Distribution Function and compare it with the Modified Permutation Entropy, a complexity measure that assumes that equal values are not from artifacts of observations but they are typical of the data generator process. The Data Driven Method of Imputation proves to outperform the Modified Permutation Entropy.

© 2018 Elsevier B.V. All rights reserved.

* Corresponding author at: CONICET - Hospital Italiano de Buenos Aires, Departamento de Informática en Salud, Perón 4190 - (C1199ABB) Ciudad Autónoma de Buenos Aires, Argentina.

E-mail address: francisco.redelico@hospitalitaliano.org.ar (F. Redelico).

1. Introduction

In the seminal article of Bandt and Pompe [1], when introducing Permutation Entropy (PE) it is stated as a condition for the estimation that the data of the time series is continuous, hence the probability of two equal values is equal to zero. Despite equal values are common in periodic signals, due to the process of sampling and discretization these equal values do not occur. In the rare event that ties exists, the suggestion is whether to ignore the patterns with ties, or to add a small random perturbation. Unfortunately, that condition does not hold for many time series. Discrete time series with repeated values may occur by two reasons. The first reason is that the time series comes from a discrete data generator process, for example the Monthly Polio Counts in USA [2] or the Daily Asthma Presentations that are Poisson autoregressive processes where the marginal probability distribution follows a Poisson member family [3]. The second reason is when the data of the generator process is continuous but only a coarse version of the actual realization is observed as a consequence of lack of precision of the data measuring device, for example in Heart Rate Variability or Arterial Pressure Variability time series [4]. This contribution deals with the changes necessary to analyze both kind of data, but emphasizes in the case when the Bandt and Pompe hypothesis of continuous data holds despite the deficiencies in the observational process. Several papers suggest modification in the estimation of Permutation Entropy to cope with repeated values. They either do this extending the symbolic alphabet presented by Bandt and Pompe, [5] or giving new rules to sort data [6]. We test the new strategy, presented in [7,8], the Data Driven Method of Imputation, to cope with this issue. This methodology uses the information of the actual time series to deal with patterns with ties. It assumes that this patterns are actually coming from suitable versions of the original patterns without ties and so they will contribute to the count of these patterns using an *a priori* probability distribution derived for the time series at hand. It can be seen as a method of imputation of missing data.

All efforts to understand the statistical properties of PE are important since this complexity measure was widely applied in Applied Science. Permutation Entropy was used in many different applied science for example Mechanics Engineering [9,10], Epilepsy [11–13], Anesthesia [14], Cardiology [15,6,16], Finance [17], Climate Change [18]. We refer to [19,20] for a comprehensive review of the Permutation Entropy application and its evolution through years.

The objective of this contribution is to compare the Data Driven Method of Imputation versus the Modified Permutation Entropy. The paper reads as follows, Section 2 presents both strategies, Section 3 is devoted to the Numerical Simulation, Section 4 reviews the performance of both strategies, and finally Section 5 is devoted to Conclusions.

2. Permutation entropy

The Shannon Informational Entropy [21] is defined as:

$$\mathcal{H} = -\frac{1}{K} \sum_{i=1}^N p_i \log(p_i), \quad (1)$$

where N is the number of possible states of the system, K is a constant and p_i , $i = \{1 \dots N\}$ is the probability of the system being in the state i .

Now, let $\{X_t\}_{t \in T}$ be a realization of a data generator process in form of a real valued time series of length T , at first assuming $P(x_t = x_{t^*}) = 0 \forall t \neq t^*$ with $t, t^* \in T$ (i.e there is not equal values in the time series).

If the $\{X_t\}_{t \in T}$ attain infinitely many values, it is common to replace them by a symbol sequence $\{(\pi_i)_t\}_{t \in T}$, $i = \{1, \dots, N\}$ with finitely many symbols $-N-$, and calculate source entropy for the $\{(\pi_i)_t\}$, $i = \{1, \dots, N\}$ [1].

Let $X_m(t) = (x_t, x_{t+1}, \dots, x_{t+m-1})$ with $0 \leq t \leq T - m + 1$ and $m \geq 2$ be the embedded vectors of length m of the time series $\{X_t\}_{t \in T}$. Let $S_{m \geq 2}$ the symmetric group of order $m!$ form by all possible permutation of order m , $\pi_i = (i_1 i_2 \dots i_m) \in S_m$ ($i_j \neq i_k \forall j \neq k$ so every element i in π_i is unique). We will call an element π_i in S_m a symbol. Then $X_m(t)$ can be mapped to a symbol π_i in S_m . This mapping should be defined in a way that preserves the desired relation between the elements x_t in $X_m(t)$; and all $X_m(t)$, $t \in T$ that shares this pattern has to mapped to the same element of $\pi_i \in S_m$.

For a given but otherwise arbitrary t , the m number of real values $X_m(t) = (x_t, x_{t+1}, \dots, x_{t+m-1})$ can be rearranged in increasing order respect to their amplitude. In order to do the mapping to $\pi_i = (i_1 i_2 \dots i_m) \in S_m$, $(i_1 i_2 \dots i_m)$ must comply that

$$x_{t+i_1-1} < x_{t+i_2-1} < \dots < x_{t+i_m-1}$$

thus, the time indexes are ordered according to their amplitude. The complete alphabet is all the possible permutations of these chronological indexes. As an example: $X_3(1) = (4, 7, 9)$ and $X_3(2) = (7, 9, 10)$ represents the permutation $\pi = 123$ since $x_{t+1} < x_{t+2} < x_{t+3}$. $X_3(3) = (9, 10, 6)$ and $X_3(4) = (6, 11, 3)$ correspond to the permutation $\pi = 312$ since $x_{t+3} < x_{t+1} < x_{t+2}$. This is a chronological index permutation mapping and one simply maps each value x_{t+i-1} in $X_m(t)$ ordering its time index $i \in \{1, 2, \dots, m\}$ according to the increasing amplitude (rank) of each x_{t+i-1} in $X_m(t)$. In Fig. 1 an illustrative drawing of this mapping for all alternatives in $m = 3$ is presented. It can be seen that the indexes of the time axis are fixed in chronological order, and they are mapped onto the vertical (amplitude) axis. The resultant symbol can be obtained reading the labels in the vertical axis from bottom to top (in the direction of the increasing amplitude). This method is used by [1,6,5] among others. A mapping from $X_m(t)$ to $\pi_i \in S_m$ is made for $\forall t \in T + m - 1$ according to the strategy

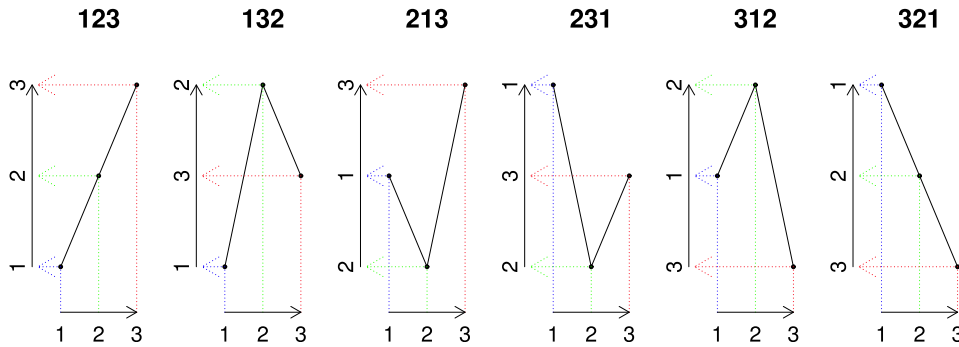


Fig. 1. All symbols for $m = 3$ are shown. This order based alphabet simply maps each value x_{t+i} in $X_m(t)$ ordering its index $i \in \{1, 2, \dots, m\}$ according to the increasing amplitude (rank) of each x_{t+i} in $X_m(t)$. It can be seen that the indexes of the time axis are fixed in chronological order, and they are mapped onto the vertical (amplitude) axis. For each pattern $X_3(t) = (x_t, x_{t+1}, x_{t+2})$, the resultant symbol $\pi_i \in S_3$ can be obtained reading the labels in the vertical axis from the bottom to the top (in the direction of the increasing amplitude).

suggested below. $\{X_t\}_{t \in T}$ will led to $T - m + 1$ symbols (π) so:

$$p(\pi_i) = \frac{\#\{t \leq T - m + 1 \mid (x_t, x_{t+1}, \dots, x_{t+m-1}) \text{ has type } \pi_i\}}{T - m + 1} \tag{2}$$

The Probability Distribution Function of the symbols π_i is called the Bandt and Pompe Probability Distribution Function (BP PDF).

The Shannon Informational Entropy (Eq. (1)) evaluated in the BP PDF is then the Permutation Entropy:

$$H(m) = - \frac{1}{\log(m!)} \sum_{i=1}^{m!} p(\pi_i) \log(p(\pi_i)). \tag{3}$$

where $\frac{1}{\log(m!)}$ is a normalization constant.

With the condition $P(x_t = x_{t^*}) = 0 \forall t \neq t^*$ with $t, t^* \in T$ all the embedding vectors $X_m(t)$ has m unique values (no ties), but that condition may not comply in several real world time series, so a substantial amount of embedding vectors $X_m(t)$ of these time series could have tied values and the mapping from these $X_m(t)$ to $\pi_i \in S_m$ cannot be made in the way presented in this Section. For this reason, different methodologies were developed to handle with time series with tied values.

In essence there are two strategies of dealing with the issue of ties. One strategy, used by the Data Driven Imputation Methodology (DDMI), assumes that the process is indeed continuous so the patterns with ties are in fact observations with missing data deriving from non tied values patterns wrongly observed. The other one makes no such assumption and extends the alphabet in order to ignore the restriction $i_k \neq i_j \forall i \neq j$ for $\pi_i = (i_1 i_2 \dots i_m) \in S_m$ and lets $i_k = i_j$ for $i \neq j$, leading to the Modified Permutation Entropy presented in [5].

2.1. Extending the alphabet: modified permutation entropy

If equal values may represent a feature state of the system under study (and are not due to observational artifacts), mapping equal values in $\{x_t\}$ to equal representation in a symbol π could be considered.

In [5] the Modified Permutation Entropy is presented. First like the original alphabet, the values of $X_m(t)$ can be sorted in increasing order: $x_{t+i_1-1} \leq x_{t+i_2-1} \leq \dots \leq x_{t+i_m-1}$. Normally, if all the values in $X_m(t)$ are different each x_{t+i_k-1} is represented by i_k when it is mapped to $\pi_i = (i_1 i_2 \dots i_m)$.

But now, in the case $x_{t+i_{j_1}-1} = x_{t+i_{j_2}-1}$ and $i_{j_1} < i_{j_2}$, both $x_{t+i_{j_1}-1}$ and $x_{t+i_{j_2}-1}$ are represented by i_{j_1} in the symbol π_i . The corresponding permutation symbol of the pattern $X_m(t)$ is defined as: $\pi_i = (i_1, i_2, \dots, i_{j_1}, i_{j_1}, \dots, i_m)$.

For example, let us take the series:

$$X_t = (2, 5, 1, 2, 7, 1, 1, 3, 1), T = 9 \tag{4}$$

and take the vector $X_5(1) = (2, 5, 1, 2, 7)$ this led to the symbol $\pi_i = (31125)$. $X_3(6) = (1, 1, 3)$ and $X_3(7) = (1, 3, 1, 1)$ map into the symbol $\pi_i = (113)$ and $\pi_i = (1112)$ respectively. In Fig. 2 all the symbols for $m = 3$ are shown for this extended alphabet.

This extended alphabet results in more possible symbols for each length motif m so it characterizes more system states than the original Bandt and Pompe method, see Table 1.

This alphabet has more symbols than $m!$,

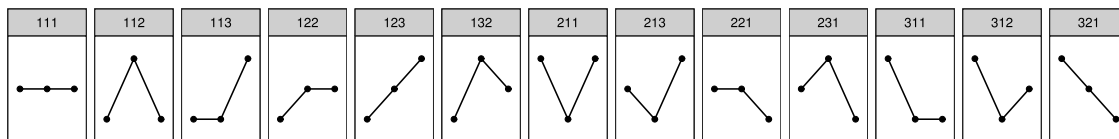


Fig. 2. The extended alphabet for $m = 3$ used to calculate the Modified Permutation Entropy in [5].

Table 1

Number of symbols π_i per length motiv m for each alphabet.

	$m=3$	$m=4$	$m=5$	$m=6$
Regular alphabet	6	24	120	720
Extended alphabet	13	73	501	4051

2.2. Imputing values to missing data: a data-driven method of imputation

In [1], they recommend to numerically break equalities by adding small perturbations at random. Other simple technique that deals with this problem is simply to ignore patterns with ties and eliminate them, along with the information they may provide. It is shown that in most situations simple techniques for handling missing data (such as the previously mentioned) in other research areas produce biased results [22], and there are more sophisticated techniques that give much better results. With these latter techniques, missing data for a subject are imputed by a value that is predicted using the subject's other, known characteristic. In [8] we propose a method that instead of adding a random perturbation that maps pattern with ties to each suitable symbol with the same probability, this probabilities are originated with a previous known Probability Distribution Function, and are not necessary uniformly distributed. The PDF proposed as a priori distribution is the one resulting of computing the $p^*(\pi_i) \forall i \in \{1, \dots, m!\}$ ignoring the patterns with ties.

In order to estimate the BP PDF with the DDMI methodology there are seven steps to follow:

1. Define the length motiv m . That leads to $S_m = \{\pi_1, \dots, \pi_{m!}\}$ all the $m!$ possible permutations of $(1\ 2 \dots m)$.
2. Map $X_m(t) \forall t \in \{1, \dots, T - m + 1\}$ to their correspondent $\pi_i \in S_m$. (Section 2).
3. If there is any ties in $X_m(t)$ for any $t \in \{1, \dots, T - m + 1\}$, eliminate the vector.
4. Calculate the $p^*(\pi_i)$ Eq. (2) for all the non eliminated patterns.
5. Repeat the procedure of mapping every $X_m(t) \forall t \in \{1, \dots, T - m + 1\}$ to their correspondent $\pi_i \in S_m$, but do not eliminate the vectors $X_m(t)$ with repeated values.
6. For each vector $X_m(t)$ with repeated values do the mapping to every compatible π_i but with probability $p^*(\pi_i)$ for each π_i .
7. Calculate the new $p(\pi_i)$

In order to show this methodology we propose the following illustrative example:

$$X_t = (2, 5, 1, 2, 7, 1, 1, 3, 1, 2, 4, 5, 1, 3, 2, 4, 4, 2, 2, 1, 0)$$

The DDMI methodology to estimate the BP PDF for this particular time series is illustrated in Fig. 3.

3. Numerical simulation

In this Section both methodologies, the Modified Permutation Entropy (MPE) and the Data Driven Method of Imputation (DDMI) are evaluated using data from simulated chaotic processes. In order to get a reproducible set of time series, all maps presented in [23] were simulated using the initial conditions presented therein. All those series presented none, or negligible amount, patterns $X_m(t)$ with ties. These time series will be referred as the original time series. After that, each original time series was truncated up to two decimal resolution, leading to a coarse version of those original time series. Due to that finite resolution, this coarse versions have an amount of patterns with tied values that cannot be considered negligible.

This yields to quantify how well each strategy estimates the actual PE $-H(m)$ for the different chaotic processes using the PE estimation $-\hat{H}(m)$ obtained for each methodology.

3.1. Experimental design

The simulation consists in 44 time series generated by different chaotic processes in every dimension, previously analyzed by Rosso and coworkers in [23]. They are 11 noninvertible maps (logistic map, sine map, tent map, linear congruential generator, cubic map, Ricker's population model, Gauss map, Cusp map, Pinchers map, Spence map, sinecircle map), 9 dissipative maps (Henon map, Lozi map, delayed logistic map, Tinkerbell map, Burgers' map; Holmes cubic map, dissipative standard map, Ikeda map, Sinai map, discrete predator-prey map) and 5 conservative maps (Chirikov standard map, Henon

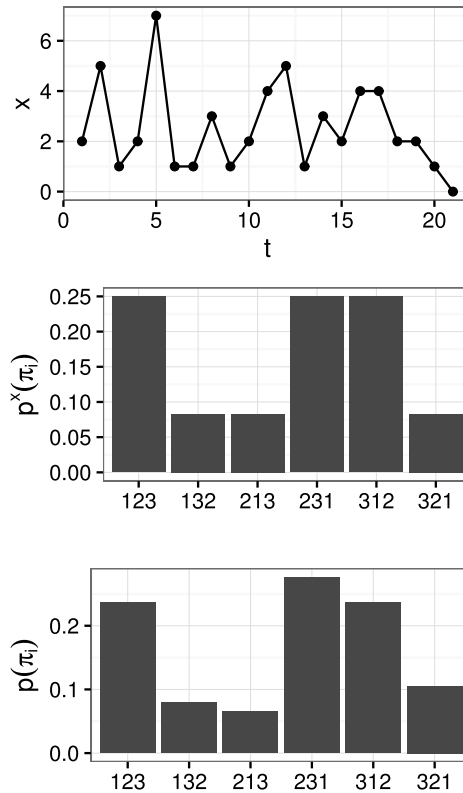


Fig. 3. The above plot shows the complete time series $X_t = (2, 5, 1, 2, 7, 1, 1, 3, 1, 2, 4, 5, 1, 3, 2, 4, 4, 2, 2, 1, 0)$, the middle plot computed the Bandt Pompe's Probability Distribution Function (PDF) eliminating patterns with repeated values for $m = 3$ ($p^*(\pi_i)$, $i \in \{1, \dots, 6\}$), using the remaining 12 complete cases. The third plot shows the resulting PDF with all 19 cases, imputing to the patterns with ties a suitable symbol. But the probability of choosing that symbol is according to the previously calculated PDF using Complete Cases methodology.

area-preserving quadratic map, Arnold's cat map, Gingerbreadman map, chaotic web map, Lorenz three dimensional chaotic map). We use the initial conditions used in [24] and refer to this book for a comprehensive analysis of those maps. Each map was simulated for a length of 10^6 and then starting in the position 10^5 take the following T values with $T = 10^5$ the time span. They were produced with IEEE 754 double precision floating point numbers. All those series presented none, or a negligible amount, patterns $X_m(t)$ with ties, and for each series the PE, $H(m)$, was calculated for length motif $m = \{3, 4, 5, 6\}$. For those initial condition, as the series is deterministic, the $H(m)$, $m = \{3, 4, 5, 6\}$ are the original PE of the processes. Next, these series are truncated up to a two decimal resolution, so the amount of patterns $X_m(t)$ with ties are not negligible any more. Both strategies enunciated in Sections 2.1 and 2.2, along with the computation of the PE eliminating patterns with ties are used separately for each truncated series to compute the $p(\pi_i)$, and their respective estimation of the PE $\hat{H}(m)$, $m = \{3, 4, 5, 6\}$.

4. Results

Fig. 4 shows, for each simulated map, the difference between the estimated PE for the coarse versions $\hat{H}(m)$ versus the PE of the original time series $H(m)$ for $m = 6$ and $T = 10^5$. This difference measures how much each methodology drifts apart from the real PE, so we define the Error as the Bias of the estimation:

$$\text{Error} = \hat{H}(m) - H(m) \tag{5}$$

This Figure shows that MPE consistently underestimates the PE, and it is outperformed by the DDMI, that is generally more accurate than simply eliminating the patterns with ties with the loss of information that this represents. This behavior is repeated for length motifs $m = 3, 4, 5, 6$. All these Errors were grouped by each methodology (Complete stands for eliminating patterns with ties), and the results are shown for each length motif m in Fig. 5. As the length motif increases the probability of finding repeated values in a pattern $X_m(t)$ also increases, and so the estimating error. For every length motif m the DDMI estimation is better than the MPE, and the variability of the Error of the DDMI is considerably lower than the MPE. The improvement in the estimation made by DDMI over the methodology of eliminating these patterns with ties can be seen as the gain of the information given by this methodology.

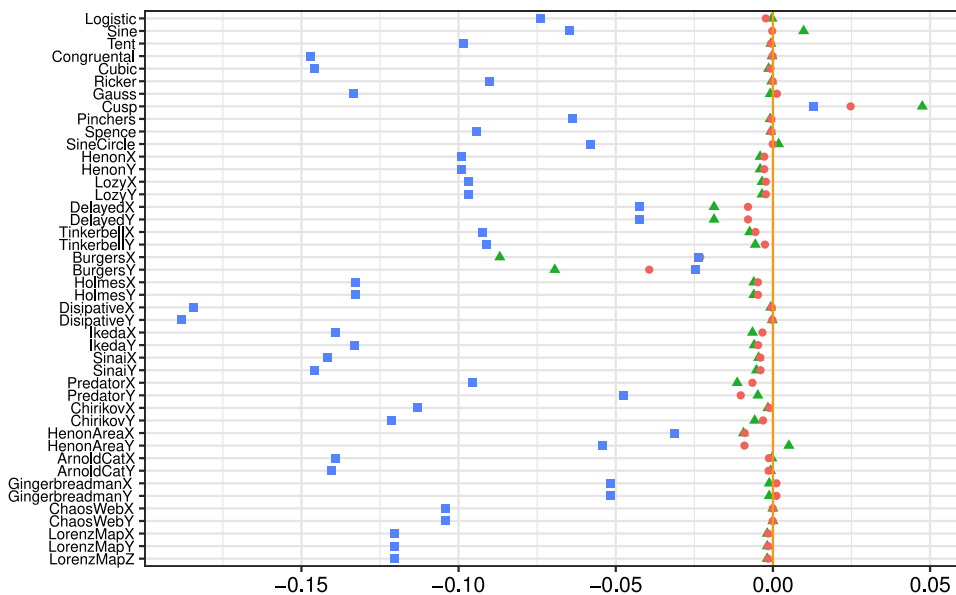


Fig. 4. For each simulated map, the difference between the estimated PE for the coarse versions $\hat{H}(m)$ versus the PE of the original time series $H(m)$ for $m = 6$ and $T = 10^5$. This difference, $\hat{H}(m) - H(m)$, measures how much each methodology drifts apart from the real PE, so it can be seen as the *Error* of each methodology. The Modified Permutation Entropy, in blue squares, consistently underestimates the PE (the vertical orange line), and it is outperformed by the Data Driven Method of Imputation, the red circles, that is generally more accurate than simply eliminating the patterns with ties, plotted as green triangles, with the loss of information that this represents. For length motif $m = 3, 4, 5$ similar results are achieved.

Another interesting property to notice is how well each methodology performs for different amounts of patterns $X_m(t)$ with ties. In Fig. 6 the performance measured by the Error (Eq. (5)) for each methodology and for different levels of repeated ratio (i.e. patterns $X_m(t)$ with ties over all the patterns $X_m(t)$) is shown for length motif $m = 6$. As might be expected, the error increases with the repeated ratio, but the DDMI is always better than the MPE. It should be noted that for the highest level of repeated ratio, at least 20% of the information is lost by eliminating this patterns and the DDMI retrieves some of this information lost improving the estimation.

5. Discussion and conclusions

When quantifying complexity for a given time series $\{X_t\}_{t \in T}$ entropy measures are an excellent choice, but common techniques neglect any effects stemming from the temporal order of the values x_t in $\{X_t\}_{t \in T}$. In order to take account this causal information, the time series can be encoded into a sequences of symbols as proposed by Bandt & Pompe in [1]. In this original approach, it was assumed that x_t in $\{X_t\}_{t \in T}$ has a continuous probability distribution function $\forall t \in T$, thus equal values appear with probability 0. In a variety of time series stemming of real life processes, e.g. Heart Rate Variability (HRV) series, equal values in a pattern $X_t(m)$ appear often and cannot be neglected without consequences. This increment in the frequency of patterns with equal values may occur by two major reasons: either the realization is not from a process with a continuous probability distribution function, so the repeated values represent the dynamic of the process, or indeed the assumptions of the process are complied but repeated values appear with high frequency due to a limited resolution in the data collection process. In order to deal with this issue, [5] proposed the Modified Permutation Entropy, that uses an extended alphabet that would take account of the repeated values on the symbolic representation $\pi_i = (i_1, i_2, \dots, i_m)$ by ignoring the restriction $i_j \neq i_k \forall j \neq k$ (see Section 2.1). Various points should be considered in relation to this methodology.

One major drawback of this methodology is that with the use of any measure of entropy, the equilibrium state of the system should represent the maximum value of entropy. Usually a completely random sequence represents total uncertainty and thus, it is the equilibrium state. If we take a random integer sequence, all the probabilities $p(\pi_i)$ in Eq. (3) must be equal, i.e. a discrete uniform distribution. The MPE does not achieve this premise since there are states more likely to happen because of the structure of the extended alphabet (See Tables 2 and 3).

In addition, if the repeated values of the time series $\{X_t\}$ under study are supposed to be due to low resolution, the augmented number of states of the extended alphabet (see Table 1) does not represent the states of the process. Even more, if the repeated ratio is not large enough, this methodology will greatly underestimate the real Permutation Entropy of the process because there will much more states with little representation, not because of the nature of the process but because of the fictitious states incorporated by the extension of the alphabet.

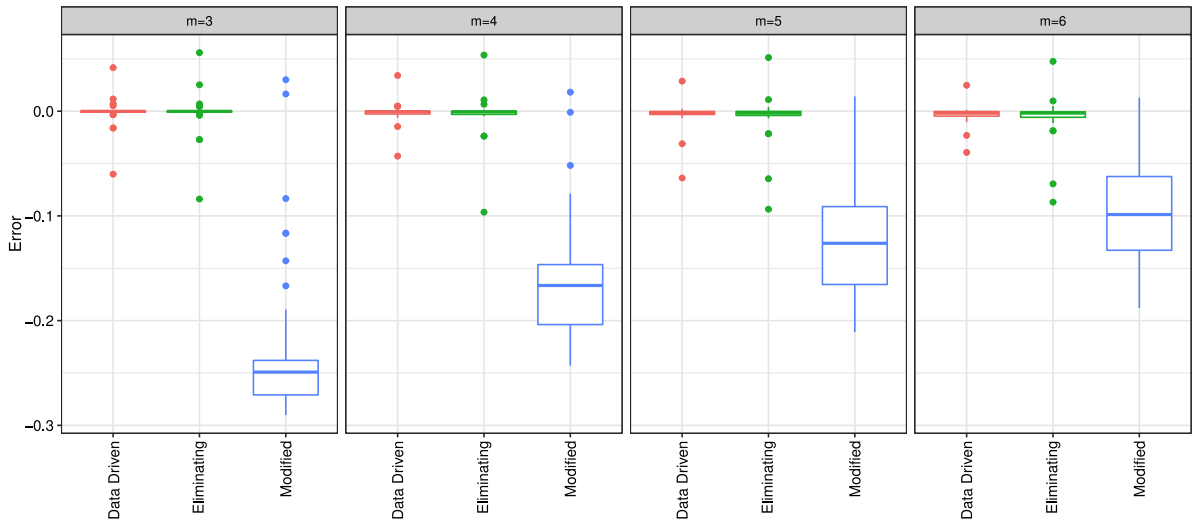


Fig. 5. The Errors $\hat{H}(m) - H(m)$ were grouped by each methodology (Eliminating stands for eliminating patterns with ties), and the results are shown for each length motif m and for $T = 10^5$. As the length motif increases the probability of finding repeated values in a pattern $X_m(t)$ increases, and so the estimating error. For every length motif m the DDMI is better than the MPE, and the variability of the Error of the DDMI is considerably lower than the MPE. The improvement in the estimation made by DDMI over the methodology of eliminating this patterns with ties can be seen as the gain of the information given by this methodology.

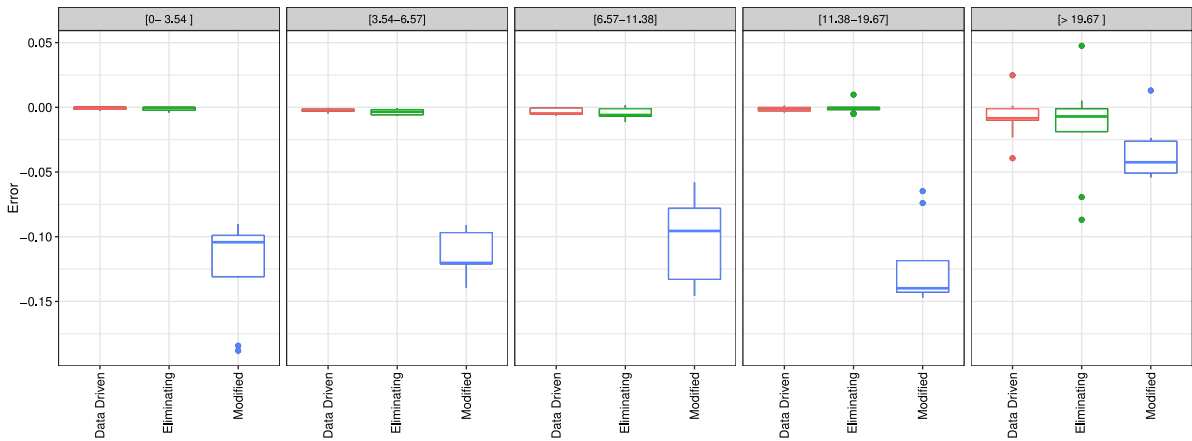


Fig. 6. The performance for each methodology, measured by the error $\hat{H}(m) - H(m)$, for different levels of repeated ratio (i.e patterns $X_m(t)$ with ties over all the patterns $X_m(t)$) is shown for length motif $m = 6$. As might be expected, the error increases with the repeated ratio, but the DDMI is always better than the MPE. It should be noted that for the highest level of repeated ratio, at least 20% of the information is lost eliminating this patterns and the DDMI retrieves some of this information lost improving the estimation.

As shown recently by Amigó et al. [25,26], in the case of deterministic one-dimensional maps not all the possible ordinal patterns can be effectively materialized into orbits, which in a sense makes these patterns “forbidden”. Indeed, the existence of these forbidden ordinal patterns becomes a persistent dynamical property. That is, for a fixed pattern length (length motif) m the number of forbidden patterns present in the time series (unobserved patterns) is independent of the series length T [27] and has a strong relation with the Permutation Entropy. In fact, the methodology used in [5] increases the number of forbidden patterns, but they are fictitious states and do not represent this persistent dynamics. The Modified Permutation Entropy proves not to be a fine tool in order to estimate the Permutation Entropy of the original series, at least if one assume that the repeated values are due to observational problems, so when dealing with Bandt & Pompe Permutation Entropy, values with ties should be treated as missing values. Eliminating all the patterns with ties is a good estimator of $H(m)$ (although much of the information is lost) as long as the length of the time series is much larger than the amount of deleted cases. So even though those cases are not negligible, its elimination does not affect the estimation. For all the stated above, the Data Driven Method of Imputation has all the good properties of the original Permutation Entropy, because it uses the information contained in the time series in analysis preserving its autocorrelation structure and the number of forbidden patterns, and also retrieves some of the information lost because of artifacts in the observation process.

Table 2
All the possible outcomes for a random sequence of an alphabet of length 3 and their probabilities $p(X_t(3))$ for $t \in T$ with T large enough (as each symbol has the same probability of occurrence every vector of length 3 has the same probability) for $m = 3$ and their correspondent symbolization in the extended alphabet.

$X_t(3)$			$p(X_t(3))$	Extended alphabet symbol
x_1	x_2	x_3		
1	1	1	$\frac{1}{27}$	111
1	1	2	$\frac{1}{27}$	113
1	1	3	$\frac{1}{27}$	113
1	2	1	$\frac{1}{27}$	131
1	2	2	$\frac{1}{27}$	122
1	2	3	$\frac{1}{27}$	123
1	3	1	$\frac{1}{27}$	131
1	3	2	$\frac{1}{27}$	132
1	3	3	$\frac{1}{27}$	122
2	1	1	$\frac{1}{27}$	311
2	1	2	$\frac{1}{27}$	212
2	1	3	$\frac{1}{27}$	213
2	2	1	$\frac{1}{27}$	221
2	2	2	$\frac{1}{27}$	111
2	2	3	$\frac{1}{27}$	113
2	3	1	$\frac{1}{27}$	231
2	3	2	$\frac{1}{27}$	131
2	3	3	$\frac{1}{27}$	122
3	1	1	$\frac{1}{27}$	311
3	1	2	$\frac{1}{27}$	312
3	1	3	$\frac{1}{27}$	212
3	2	1	$\frac{1}{27}$	321
3	2	2	$\frac{1}{27}$	311
3	2	3	$\frac{1}{27}$	212
3	3	1	$\frac{1}{27}$	221
3	3	2	$\frac{1}{27}$	221
3	3	3	$\frac{1}{27}$	111

Table 3
All the probabilities for the symbols for the extended alphabet for the random integer sequence showed in 2. An entropy measure should be maximum for a completely random sequence and all the probabilities $p(\pi_i)$ in Eq. (3) must be equal. Modified Permutation Entropy proposed by [5] does not comply with important premise.

Extended alphabet symbol	Frequency	$p(\pi_i)$
$\pi_1 = 111$	3	$\frac{1}{9}$
$\pi_2 = 113$	3	$\frac{1}{9}$
$\pi_3 = 122$	3	$\frac{1}{9}$
$\pi_4 = 123$	1	$\frac{1}{27}$
$\pi_5 = 131$	3	$\frac{1}{9}$
$\pi_6 = 132$	1	$\frac{1}{27}$
$\pi_7 = 212$	3	$\frac{1}{9}$
$\pi_8 = 213$	1	$\frac{1}{27}$
$\pi_9 = 221$	3	$\frac{1}{9}$
$\pi_{10} = 231$	1	$\frac{1}{27}$
$\pi_{11} = 311$	3	$\frac{1}{9}$
$\pi_{12} = 312$	1	$\frac{1}{27}$
$\pi_{13} = 321$	1	$\frac{1}{27}$

References

- [1] C. Bandt, B. Pompe, Permutation entropy: A natural complexity measure for time series, *Phys. Rev. Lett.* 88 (17) (2002) 174102.
- [2] S.L. Zeger, A regression model for time series of counts, *Biometrika* 75 (4) (1988).
- [3] E. McKenzie, Some ARMA models for dependent sequences of Poisson counts, *Adv. Appl. Probab.* (1988) 822–835.
- [4] M. Pagani, F. Lombardi, S. Guzzetti, O. Rimoldi, R. Furlan, P. Pizzinelli, G. Sandrone, G. Malfatto, S. Dell’Orto, E. Piccaluga, Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog, *Circ. Res.* 59 (2) (1986) 178–193.
- [5] C. Bian, C. Qin, Q.D. Ma, Q. Shen, Modified permutation-entropy analysis of heartbeat dynamics, *Phys. Rev. E* 85 (2) (2012) 021906.
- [6] U. Parlitz, S. Berg, S. Luther, A. Schirdewan, J. Kurths, N. Wessel, Classifying cardiac biosignals using ordinal pattern statistics and symbolic dynamics, *Comput. Biol. Med.* 42 (3) (2012) 319–327.
- [7] F. Traversaro, F.O. Redelico, M.R. Risk, A.C. Frery, O.A. Rosso, Bandt-Pompe symbolization dynamics for time series with tied values: A data-driven approach, *Chaos* 28 (7) (2018) 075502.
- [8] F. Traversaro, M. Risk, O. Rosso, F. Redelico, An empirical evaluation of alternative methods of estimation for permutation entropy in time series with tied value, *ArXiv e-Prints* (2017).
- [9] R. Yan, Y. Liu, R.X. Gao, Permutation entropy: A nonlinear statistical measure for status characterization of rotary machines, *Mech. Syst. Signal Process.* 29 (2012) 474–484.
- [10] F.O. Redelico, F. Traversaro, N. Oyarzabal, I. Vilaboa, O.A. Rosso, Evaluation of the status of rotary machines by time causal information theory quantifiers, *Physica A* 470 (2017) 321–329.
- [11] F.O. Redelico, F. Traversaro, M.d.C. García, W. Silva, O.A. Rosso, M. Risk, Classification of normal and Pre-Ictal EEG signals using permutation entropies and a generalized linear model as a classifier, *Entropy* 19 (2) (2017) 72.
- [12] E. Olofsen, J. Sleigh, A. Dahan, Permutation entropy of the electroencephalogram: A measure of anaesthetic drug effect, *Br. J. Anaesth.* 101 (6) (2008) 810–821.
- [13] F. Traversaro, F.O. Redelico, Confidence intervals and hypothesis testing for the permutation entropy with an application to epilepsy, *Commun. Nonlinear Sci. Numer. Simul.* 57 (2018) 388–401.
- [14] D. Jordan, G. Stockmanns, E.F. Kochs, S. Pilge, G. Schneider, Electroencephalographic order pattern analysis for the separation of consciousness and unconsciousness: an analysis of approximate entropy, permutation entropy, recurrence rate, and phase coupling of order recurrence plots, *J. Am. Soc. Anesthesiol.* 109 (6) (2008) 1014–1022.
- [15] B. Frank, B. Pompe, U. Schneider, D. Hoyer, Permutation entropy improves fetal behavioural state classification based on heart rate analysis from biomagnetic recordings in near term fetuses, *Med. Biol. Eng. Comput.* 44 (3) (2006) 179.
- [16] W. Legnani, F. Traversaro, F.O. Redelico, L.J. Cymberknop, R.L. Armentano, O.A. Rosso, Analysis of ischaemic crisis using the informational causal entropy-complexity plane, *Chaos* 28 (7) (2018) 075518.
- [17] M. Matilla-García, M.R. Marín, Detection of non-linear structure in time series, *Econom. Lett.* 105 (1) (2009) 1–6.
- [18] L.C. Carpi, P.M. Saco, A. Figliola, E. Serrano, O.A. Rosso, Analysis of an EL nino-southern oscillation proxy record using information theory quantifiers, *Concepts Recent Adv. Gen. Inf. Meas. Stat.* (2013) 3.
- [19] M. Riedl, A. Müller, N. Wessel, Practical considerations of permutation entropy, *Eur. Phys. J. Spec. Top.* 222 (2) (2013) 249–262.
- [20] M. Zanin, L. Zunino, O.A. Rosso, D. Papo, Permutation entropy and its main biomedical and econophysics applications: A review, *Entropy* 14 (8) (2012) 1553–1577.
- [21] C.E. Shannon, A mathematical theory of communication, *ACM SIGMOBILE Mob. Comput. Commun. Rev.* 5 (1) (2001) 3–55.
- [22] A.R.T. Donders, G.J. van der Heijden, T. Stijnen, K.G. Moons, Review: A gentle introduction to imputation of missing values, *J. Clin. Epidemiol.* 59 (10) (2006) 1087–1091.
- [23] O.A. Rosso, F. Olivares, L. Zunino, L. De Micco, A.L. Aquino, A. Plastino, H.A. Larrondo, Characterization of chaotic maps using the permutation Bandt-Pompe probability distribution, *Eur. Phys. J. B* 86 (4) (2013) 116.
- [24] J.C. Sprott, J.C. Sprott, *Chaos and Time-Series Analysis*, vol. 69, Citeseer, 2003.
- [25] J. Amigo, S. Zambrano, M. Sanjuan, True and false forbidden patterns in deterministic and random dynamics, *Europhys. Lett.* 79 (5) (2007) 50001.
- [26] J. Amigo, S. Zambrano, M. Sanjuan, Combinatorial detection of determinism in noisy time series, *Europhys. Lett.* 83 (6) (2008) 60005.
- [27] L.C. Carpi, P.M. Saco, O. Rosso, Missing ordinal patterns in correlated noises, *Phys. A* 389 (10) (2010) 2020–2029.