Using Permutations for Hierarchical Clustering of Time Series

Jose S. Cánovas * [ID], **Antonio Guillamón** [ID] and **María Carmen Ruiz-Abellón** [ID]

Departamento de Matemática Aplicada y Estadística, Universidad Politécnica de Cartagena, 30202 Cartagena, Spain; antonio.guillamon@upct.es (A.G.); maricarmen.ruiz@upct.es (M.C.R.-A.)

*   Correspondence: jose.canovas@upct.es

Abstract: Two distances based on permutations are considered to measure the similarity of two time series according to their strength of dependency. The distance measures are used together with different linkages to get hierarchical clustering methods of time series by dependency. We apply these distances to both simulated theoretical and real data series. For simulated time series the distances show good clustering results, both in the case of linear and non-linear dependencies. The effect of the embedding dimension and the linkage method are also analyzed. Finally, several real data series are properly clustered using the proposed method.

Keywords: time series clustering; permutation entropy; time series dependency; hierarchical clustering; mutual information

1. Introduction and Main Definitions

The goal of time series clustering is to split a set of time series into homogeneous groups, that is, similar time series should lie in the same cluster. However, there are many distances to measure the degree of similarity between two time series, depending on the clustering objectives. The most popular are the Euclidean and correlation-based distances. However, if we want to cluster time series in shape, it has been proved that Dynamic Time Warping distance (see [1]) is more appropriate. Some other distance measures are the short time series given in [2], the Kullback-Leibler studied in [3] or the recent copula-based distance introduced in [4].

Interesting surveys in the field can be found in [5,6], where three different types of clustering approaches are distinguished: shape-based, feature-based and model-based approach. In most cases, the procedures only take into account univariate features of the time series and do not consider the possible relationships among them. Therefore, these methods are useful in the case of independent time series and when the objective is to group them looking at the similarity of their univariate models. It is well known that the selection of a suitable distance measure mainly depends on the objective of clustering. Typically three objectives can be distinguished [5]: finding similar time series in time, finding similar time series in shape and finding similar time series in change (structural similarity). In the first case, the Euclidean distance on raw time series or the Wavelet Transform distance are proper for the first goal. The distances based on Pearson's correlation are usually considered of this type, even though the aim is to put the time series that are correlated in the same cluster. In the second case, where the time of occurrence of patterns in not important, the use of elastic methods like the Dynamic Time Warping distance is highly encouraged. In the third case, where the aim is to cluster time series with similar structure, autocorrelation-based measures are appropriate. All these similarity (dissimilarity) measures have been applied to some simulated data and compared to our proposed measures to illustrate that they are not proper for clustering time series by dependency. Many of these clustering methods are available in the R package TSclust, developed by Montero and Vilar [7].

Clustering time series by dependency has recently been analyzed in [8–10]. In these works, the goal is to cluster the time series according to their degree of dependency. The two former assume that the vector of time series is generated by a Dynamic Factor Model where some factors affect different groups of series. The latter proposes the generalized cross correlation as a general measure of linear association between two time series.

There are a wide range of papers dealing with the applications of ordinal patterns in time series analysis. For example, the utility of permutations to detect structural changes in time series has been widely studied and applied to real data such as speech signals [11], electroencephalogram signals [12–15] or economic and seismic data [16]. On the other hand, the discriminative power of ordinal pattern statistics and symbolic dynamics to classify cardiac biosignals has been evaluated in [17], whereas [18] uses ordinal patterns to detect and locate change points in the time series and classify the segments with similar dynamics. In this context, Ref. [19] proposes a new metric called Ordinal Synchronization to evaluate the level of synchronization between time series by means of a projection into ordinal patterns.

The aim of the present paper is to illustrate the utility of symbolic dynamic (the time series are codified by means of permutations) in clustering time series by dependency, where the main contribution respect to the works mentioned above are the absence of assumptions and the detection of linear and non-linear dependencies. For that, we have based on the work developed by Ruiz-Abellon et al. [20], where the next three aspects are combined: the time series codification by means of permutations, the distance measures among time series and different linkages for hierarchical clustering.

In this paper we will consider the labeling by permutations technique jointly with some measures which allow us to group different time series, or at least to state what time series are closer in the sense of the considered measure. Let us introduce the basic notation.

Several authors have pointed out that permutations can be a good tool to study time series. In [21] (see also [22–24]) the notion of *permutation entropy* has been used for analyzing time series. Let us recall a few notions on permutation labeling of a time series.

Let $(x_n)_{n=1}^T$, $T \in \mathbb{N} \cup \{\infty\}$, be a sequence which comes from real or simulated data. For $m \geq 2$, let $\mathcal{S}_m$ be the group of permutations of length m, whose cardinality is $m!$. Let $x_m(r) = (x_r, x_{r+1}, ..., x_{r+m-1})$, $1 \leq r < T - m + 1$, be a sliding window taken from the sequence $(x_n)_{n=1}^T$. We say that the window $x_m(r)$ is of π–type, $\pi \in \mathcal{S}_m$, if and only if $\pi = (i_1, i_2, ..., i_m)$ is the unique element of $\mathcal{S}_m$ satisfying the two following conditions:

(c1) $x_{r+i_1} \leq x_{r+i_2} \leq ... x_{r+i_m}$.
(c2) $i_{s-1} < i_s$ if $x_{r+i_{s-1}} = x_{r+i_s}$, with $2 \leq s \leq m$.

The positive integer m is usually known as *embedding dimension*. Fix $r < T - m + 1$. For each $\pi \in \mathcal{S}_m$, the probability of occurrence of π is given by:

$$p(\pi) = \frac{\#\{x_m(j), j = 1, 2, ..., r : x_m(j) \text{ is of } \pi\text{–type}\}}{r}.$$

A permutation $\pi \in \mathcal{S}_m$ with $p(\pi) > 0$ for some $r < T - m + 1$ is called an *admissible permutation* of $(x_n)_{n=1}^T$. It is clear that permutations are linked to data series complexity. For instance, a periodic or increasing time series has at most a finite number of admissible permutations which are bounded for any embedding dimension m. Conversely, it is also clear that a big enough i.i.d noise should admit any permutation of length m. On the other hand, for piecewise monotone maps, it is proved that topological entropy, a useful tool to decide whether a deterministic time series is complicated, can be computed by using permutations [25,26].

The above-described codification can be extended in a direct way when we have two dimensional time series. Let $(x_n)_{n=1}^T$ and $(y_n)_{n=1}^T$ be two real time series and let $(z_n)_{n=1}^T$ be the corresponding two-dimensional time series with $z_n = (x_n, y_n)$, for all $n = 1, ..., T$. Let $z_m(r) = (x_m(r), y_m(r))$, $1 \leq r < T - m + 1$, be a two-dimensional sliding window taken from the sequence $(z_n)_{n=1}^T$. The window $z_m(r)$ is said to be of type $\pi_i \times \pi_j$ if and only if $x_m(r)$ is a π_i–type and $y_m(r)$ is of π_j–type. After the codifying

process, all of the empirical information is collected in a contingency table, where $O_{i,j}$ denotes the observed frequency of the symbol $\pi_i \times \pi_j$ and as usual, the relative frequency is given by:

$$p(\pi_i \times \pi_j) = \frac{O_{i,j}}{T - m + 1},$$

for $1 \leq i, j \leq m!$. The contingency table is shown below.

$(x_n)/(y_n)$	π_1–type	π_2–type	$\ldots$	$\pi_{m!}$ –type	
π_1–type	$O_{1,1}$	$O_{1,2}$	$\ldots$	$O_{1,m!}$	$O_{1\bullet}$
π_2–type	$O_{2,1}$	$O_{2,2}$	$\ldots$	$O_{2,m!}$	$O_{2\bullet}$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$\pi_{m!}$–type	$O_{m!,1}$	$O_{m!,2}$	$\ldots$	$O_{m!,m!}$	$O_{m!\bullet}$
	$O_{\bullet1}$	$O_{\bullet2}$	$\ldots$	$O_{\bullet m!}$	

The above contingency table was used in [27] to decide whether two time series were independent or not. Among others, the Pearson's chi–square statistic was used for the above contingency table given by:

$$\chi^2 = \sum_{i=1}^{m!} \sum_{j=1}^{m!} \frac{(O_{i,j} - e_{i,j})^2}{e_{i,j}},$$

$e_{i,j}$ denotes the expected frequencies under the hypothesis of independence and is given by:

$$e_{i,j} = \frac{O_{i\bullet} \, O_{\bullet j}}{T - m + 1}.$$

Hence, we can define the Crammer's V measure as follows, see [28]:

$$V(x_n, y_n) = \sqrt{\frac{\chi^2}{(T - m - 1)(m! - 1)}}.$$

Values of Cramer's V close to zero means no association (independency) and close to one mean strong association (dependency). Cramer's V measure allows us to define the "distance":

$$D_V(x_n, y_n) = 1 - V(x_n, y_n).$$

It is unclear whether D_V is a metric, but as we will show later, it is a good help in clustering data series. Before that, we will introduce another measure based on mutual information measures.

Given two discrete random variables X and Y, the mutual information coefficient (see [29]) is defined by:

$$I(X, Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p_1(x_i) p_2(y_j)}$$

where $p(x_i, y_j)$ is the join probability function of (X, Y) and $p_1(x_i)$ and $p_2(y_j)$ are the marginal probability functions of X and Y, respectively. The mutual information coefficient can be computed using the concept of Shannon entropy (see [30]) by:

$$I(X, Y) = H(X) + H(Y) - H(X, Y)$$

where

$$H(X) = - \sum_{i=1}^{n} p_1(x_i) \log p_1(x_i)$$

and

$$H(Y) = - \sum_{j=1}^{m} p_2(y_j) \log p_2(y_j)$$

are the Shannon entropies of X and Y, respectively, and

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) \log p(x_i, y_j)$$

is the Shannon entropy of (X,Y). The mutual information coefficient is a dependency measure because $I(X,Y) = 0$ if and only if X and Y are independent. Moreover, it is symmetric and non-negative, but there is not a fixed upper bound. Hence, we can consider the metric given in [29]

$$D(X,Y) = 1 - \frac{I(X,Y)}{\max\{H(X), H(Y)\}},$$

which is a metric because it is non–negative, symmetric and holds the triangular property. Applied to the codified time series it reads as:

$$D(x_n, y_n) = 1 - \frac{\sum_{\pi_i \in \mathcal{S}_m} \sum_{\pi_j \in \mathcal{S}_m} \frac{O_{i,j}}{T-m+1} \log \frac{(T-m+1)O_{i,j}}{O_{i\bullet}O_{j\bullet}}}{\max\left\{-\sum_{\pi_i \in \mathcal{S}_m} \frac{O_{i\bullet}}{T-m+1} \log \frac{O_{i\bullet}}{T-m+1}, -\sum_{\pi_j \in \mathcal{S}_m} \frac{O_{j\bullet}}{T-m+1} \log \frac{O_{j\bullet}}{T-m+1}\right\}}$$

Remark that the efficiency of the Pearson's chi-square statistic (applied to symbolic dynamic) to detect linear and non-linear dependencies between two time series was illustrated in [27]. Therefore, a dissimilarity measure based on this statistic can be useful to develop a new clustering approach. Analogously, the efficiency of the mutual information coefficient for detecting dependencies was shown, among others, in [29], so a dissimilarity measure combining symbolic dynamic and the mutual information coefficient can be proposed as a good tool to cluster time series by dependency.

In the next section, different scenarios are simulated: linear dependency among time series using several models (the logistic map, uniformly distributed noise and autoregressive process); non-linear dependency for deterministic systems and non-linear dependency for non-deterministic systems. Additionally, simulations were carried out for different embedding dimensions and three hierarchical linkages (single, complete and average) to analyze their effect on the resulting dendrograms.

2. Synthetic Experiments

2.1. Linear Dependence

We consider data series by following the system of difference equations as follows:

$$x_{n+1}^i = \sum_{j=1}^{k} \lambda_{ij} f_j(x_n^j), \tag{1}$$

where $n \geq 0$, $k \geq 3$, f_j are real functions for $1 \leq j \leq k$, and for $1 \leq i, j \leq k$, we have that $\lambda_{ij} \geq 0$ and

$$\sum_{j=1}^{k} \lambda_{ij} \leq 1.$$

These maps have been introduced as a model for migration in population dynamics (see e.g., [31] and some references therein). In principle, if $\lambda_{ij} = \lambda_{ji} = 0$, then there is no relationship between sequences x_n^i and x_n^j. Of course, we are assuming that some of them are not completely independent. Note that if both sequences are independent and are generated by the same deterministic map f with the same initial condition, then they are the same sequence indeed.

The experiments we have done are as follows. We generate several data series and apply D_V and D to cluster them. The smaller values of D_V or D we get, the closer the two data series are in the sense of these measures.

Below, we summarize the experiments we have done.

We take $k = 5$ and consider the matrix:

$$\Lambda_1 = (\lambda_{ij}) = \begin{pmatrix} 0.1 & 0.9 & 0 & 0 & 0 \\ 0.9 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 & 0 \\ 0 & 0 & 0.8 & 0.2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

According to this matrix, sequences x_n^1 and x_n^2 are linked, as well as x_n^3 and x_n^4, while sequence x_n^5 is isolated. We consider several possibilities for f_j, $1 \leq j \leq 5$. Firstly, we take $f_j(x) = 4x(1 - x)$ (the logistic map) and random initial conditions. The data length is 10,000 and we consider different embeddings $m = 2, 3, 4$. We consider distances D_V and D and three different linkages for the hierarchical method (single, complete and average). This is the general procedure along the paper. As we can see, there are not significative differences in applying the three different hierarchical linkages. In this example, and as it is expected, x_n^3 and x_n^4 are linked together as well as x_n^1 and x_n^2, but the result shows a deep link between variables x_n^3 and x_n^4. Figures 1 and 2 show the results we obtain in detail.
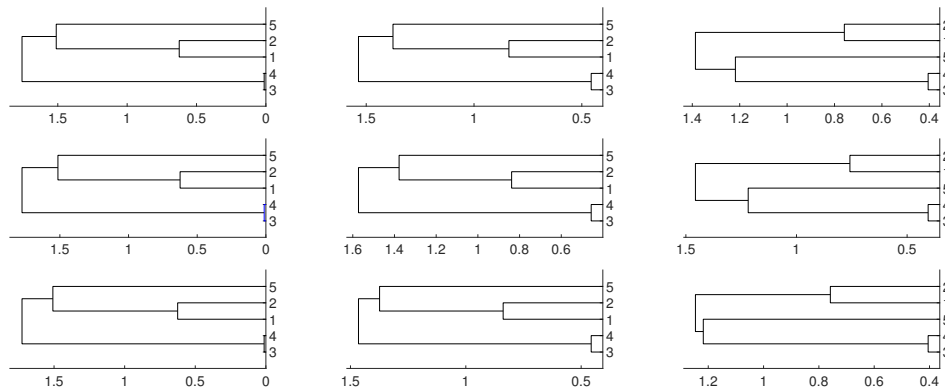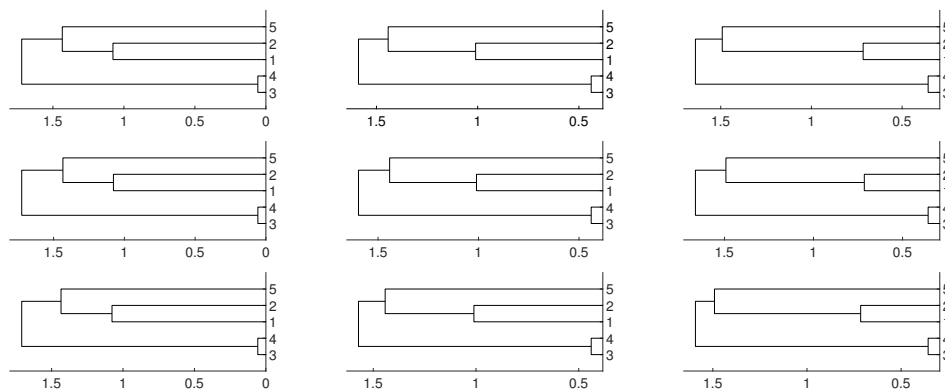

Figure 1. Dendrograms for distance D_V of the logistic time series for embedding dimension $m = 2$ (**left** column), $m = 3$ (**central** column) and $m = 4$ (**right** column), and three linkages, average (**first** row), complete (**central** row) and single (**bottom** row). To improve the visibility we write 1 for x_n^1 and so on. We see a closed connection between x_n^3 and x_n^4 as expected, then between x_n^1 and x_n^2. The difference appears with x_n^5 which is more close to x_n^1 and x_n^2 for $m = 2, 3$ and to x_n^3 and x_n^4 for $m = 4$.

Figure 2. Dendrograms for distance D of the logistic time series for embedding dimension $m = 2$ (**left** column), $m = 3$ (**central** column) and $m = 4$ (**right** column), and three linkages, average (**first** row), complete (**central** row) and single (**bottom** row). To improve the visibility we write 1 for x_n^1 and so on. We see a closed connection between x_n^3 and x_n^4 as expected, then between x_n^1 and x_n^2. In all cases x_n^5 is closer to x_n^1 and x_n^2.

In order to compare the results of the proposed approach with some traditional ones, Figure 3 shows the dendrograms obtained for system (1) and the logistic time series using the Wavelet Transform, Pearson's correlation and Dynamic Time Warping distances. For the correlation-based distance the right clustering is achieved as expected, because the dependency among the time series given by system (1) is linear. However, the Wavelet Transform distance is not able to detect the correct relationships (recall that the objective of this distance is to find similar time series in time). Even though the Dynamic Time Warping distance provides the expected clustering results in this case, we will see later an example where it does not.
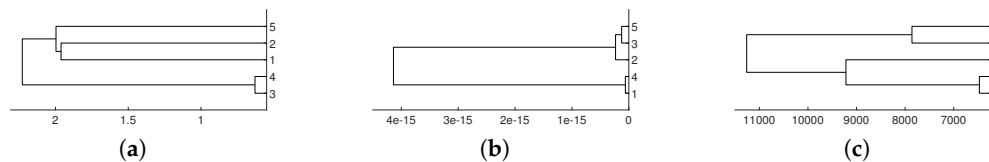
Figure 3. Dendrograms for the logistic time series given by the system defined in (1), average linkage and distances: (**a**) Pearson's correlation, (**b**) Wavelet Transform and (**c**) Dynamic Time Warping. To improve the visibility we write 1 for x_n^1 and so on.

Then, we repeat the experiments when f_j is an i.i.d. uniformly distributed noise for $1 \leq j \leq 5$. To avoid repetitive graphics we only use the average linkage in Figure 4 for obtaining the dendrograms. No significative differences are obtained when the other two linkages (single and complete) are used instead of the average one. Figure 4 shows the clustering results obtained for both distances D_V and D. We obtain the same results as in the previous purely deterministic case for $m = 3$ and $m = 4$. However, for $m = 2$ the actual relationships are not properly detected.
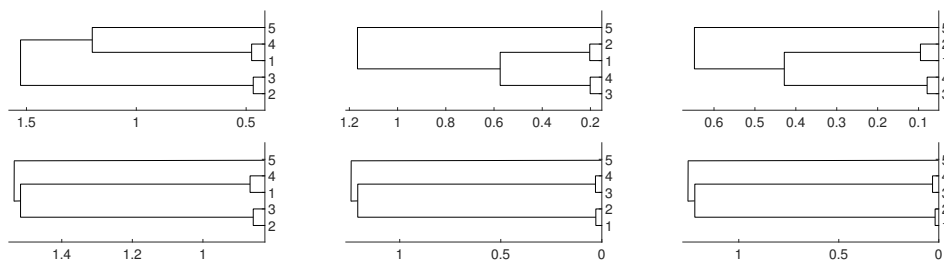

Figure 4. Dendrograms for distance D (first row) of a uniform i.i.d. time series for embedding dimension $m = 2$ (**left** column), $m = 3$ (**central** column) and $m = 4$ (**right** column), and average as linkage method. In the second row we show the same for distance D_V. To improve the visibility we write 1 for x_n^1 and so on.

Again, we repeat the experiment when $f_j(x) = 0.9x + \varepsilon$, where ε is an i.i.d. uniformly distributed noise, that is, f_j is an autoregressive process. The dendrograms shown in Figure 5 give us the same result as in the two previous cases.
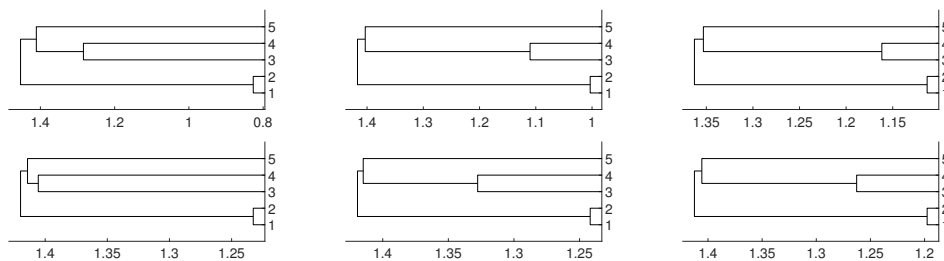

Figure 5. Dendrograms for distance D (first row) of an autoregressive time series for embedding dimension $m = 2$ (**left** column), $m = 3$ (**central** column) and $m = 4$ (**right** column), and average linkage method. In the second row we show the same for distance D_V. To improve the visibility we write 1 for x_n^1 and so on.

2.2. Non Linear Dependence: Deterministic Systems

Now, we consider non linear time series constructed by deterministic systems. We consider the system:

$$
\begin{cases}
x_{n+1}^1 = 2.25x_n^1(1 - x_n^2) \\
x_{n+1}^2 = x_n^1 \\
x_{n+1}^3 = x_n^4 \\
x_{n+1}^4 = x_n^3(1 - x_n^4) \\
x_{n+1}^5 = 4x_n^5(1 - x_n^5)
\end{cases}
\tag{2}
$$

with initial conditions 0.45 for all the variables. We consider a sample of 10,000 points and show the results with the average linkage method. We obtain that x_n^1 and x_n^2 are linked as expected, then x_n^3 and x_n^4, as one can expect given the system shape. Dendrograms are shown in Figure 6. Note that for the D distance, the expected results are obtained using different embeddings, whereas for the D_V distance, only with $m = 4$.
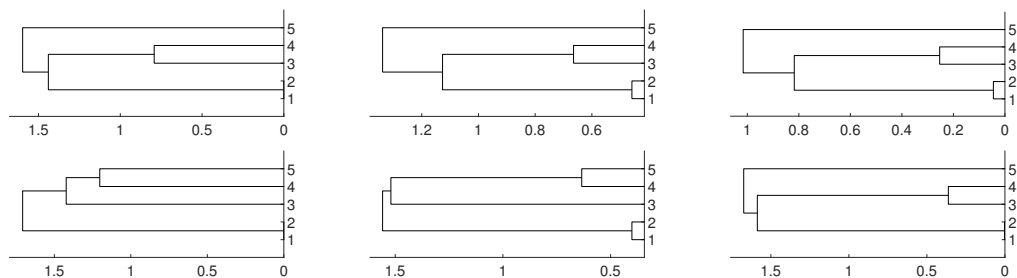
Figure 6. Dendrograms of system (2) for distance D (first row) for embedding dimension $m = 2$ (**left** column), $m = 3$ (**central** column) and $m = 4$ (**right** column), and average linkage method. In the second row we show the same for distance D_V. To improve the visibility we write 1 for x_n^1 and so on.

Next, we consider another nonlinear deterministic time series given by the system:

$$
\begin{cases}
x_{n+1}^1 = 2.25x_n^1(1 - x_n^2) \\
x_{n+1}^2 = x_n^1 \\
x_{n+1}^3 = x_n^4 \\
x_{n+1}^4 = x_n^5 \\
x_{n+1}^5 = 4x_n^5(1 - x_n^3)
\end{cases}
\tag{3}
$$

with initial conditions 0.45 for all the variables. Dendrograms are shown in Figure 7, where as usual we have considered a sample of 10,000 points and the average method. We see that x_n^1 and x_n^2 are connected first and then x_n^3, x_n^4 and x_n^5 are clustered together as one may expect for the system shape. Note that the appropriate clustering is obtained when $m = 3$ and $m = 4$, but not for the case $m = 2$.
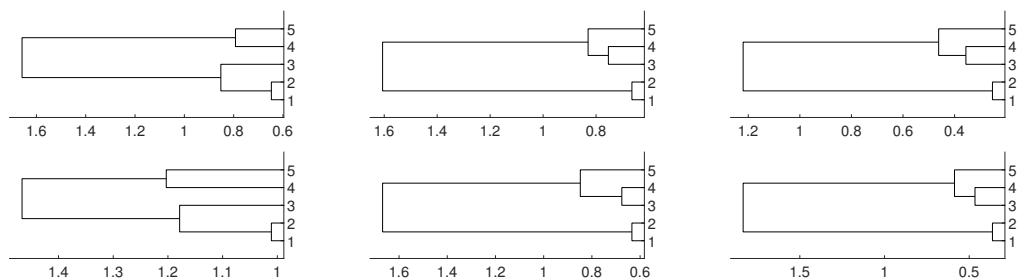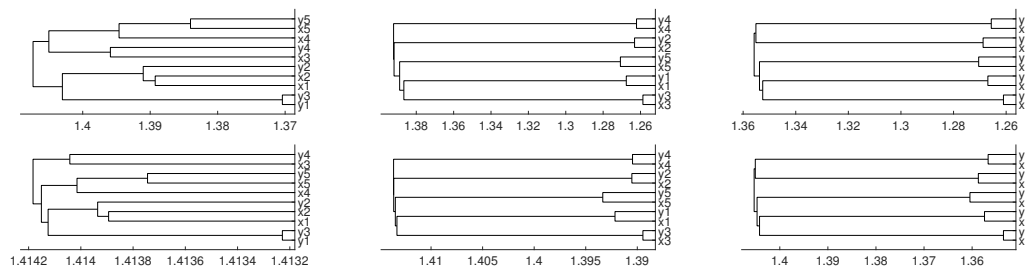

Figure 7. Dendrograms of system (3) for distance D (first row) for embedding dimension $m = 2$ (**left** column), $m = 3$ (**central** column) and $m = 4$ (**right** column), and average linkage method. In the second row we show the same for distance D_V. To improve the visibility we write 1 for x_n^1 and so on.

2.3. Non Linear Dependence: Non Deterministic Systems

First, we consider the time series generated by the difference equations:

$$\begin{cases} x_{n+1} = u_n \\ y_{n+1} = 0.8(x_n)^2 + v_n \end{cases} \tag{4}$$

where u_n and v_n are i.i.d. $N(0,1)$ variables. We consider 0.45 as initial conditions and generate samples of 60,000 points for each variable. Then we divide them into five time series of 12,000 points labeled by $x_1,\ldots, x_5, y_1,\ldots, y_5$ in Figure 8, where the results are presented with the average linkage method. We see that embedding dimensions $m = 3$ and 4 give us the expected results.

Figure 8. Dendrograms of system (4) for distance D (first row) for embedding dimension $m = 2$ (**left** column), $m = 3$ (**central** column) and $m = 4$ (**right** column), and average linkage method. In the second row we show the same for distance D_V.

Next, we consider a similar example generating 60,000 data points and dividing them into data series of 12,000 points labeled by $x_1,\ldots, x_5, y_1,\ldots, y_5$. The points are generated by the system of difference equations:

$$\begin{cases} x_{n+1} = 0.6x_n + u_n \\ y_{n+1} = 0.6(u_n)^2 + v_n \end{cases} \tag{5}$$

where u_n and v_n are i.i.d. $N(0,1)$ variables. We consider 0.45 as initial conditions and the results are given in Figure 9 for average method. We see how embedding dimensions $m = 3$ and 4 give us the results that one can expect.
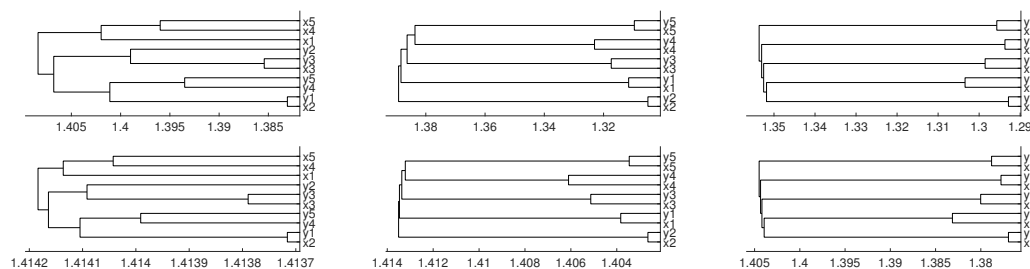

Figure 9. Dendrograms of system (5) for distance D (first row) for embedding dimension $m = 2$ (**left** column), $m = 3$ (**central** column) and $m = 4$ (**right** column), and average linkage method. In the second row we show the same for distance D_V.

Simulated time series given by (5) have also been clustered using traditional distance measures. Figure 10 depicts the results obtained for the Euclidean, Pearson's correlation, autocorrelation-based, Wavelet Transform and Dynamic Time Warping distances. In the first case, the Euclidean distance does not provide the expected clustering results, because the goal of this distance is to find similar time series in time. In the second case, the correlation-based distance is not able to detect the non-linear relationships among the time series. Regarding the autocorrelation-based distance, which computes

the dissimilarity between two time series as the distance between their estimated simple or partial autocorrelation coefficients, two groups are distinguished (one cluster formed by the time series split from x_n and the other formed by the time series split from y_n). Once again, the Wavelet Transform distance does not achieve the correct clustering results. Finally, the Dynamic Time Warping distance is not appropriate for clustering time series by dependency (recall that this distance leads to a shape-based approach in time series clustering, where the time series with similar shape are put together in the same cluster; however, two time series with different shapes can be strongly dependent).
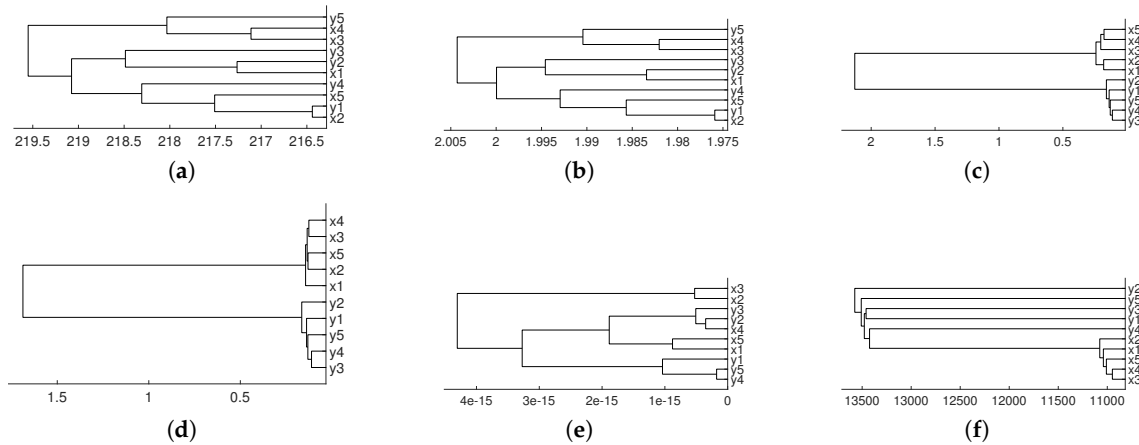
Figure 10. Dendrograms for time series given by the system defined in (5), average linkage and distances: (**a**) Euclidean, (**b**) Pearson's correlation, (**c**) autocorrelation, (**d**) partial autocorrelation, (**e**) Wavelet Transform and (**f**) Dynamic Time Warping.

3. Real Data Experiments

The above section introduced several simulated time series proving that the distances D and D_V are useful tools to establish what time series are closer for both linear and nonlinear dependencies. In this section we apply the theoretical framework for real time series.

3.1. Latin American Exchange Rate Dependencies

We consider the log–returns of daily exchange rates of six Latin American currencies vs. US dollar. Namely, we consider the currency of Argentina (ARS), Brasil (BRL), Chile (CLP), Colombia (COP), Mexico (MXN) and Peru (PEN), from 22 June 2005 to 25 April 2012. The data length is 1754, and it has been already considered in paper [32]. In that paper, authors evaluate the level of contagion among the exchange rates of the previous six Latin American countries, and using copulas they conclude that two blocs were distinguished: the first bloc consists of Brazil, Colombia, Chile and Mexico, whose exchange rates exhibit the largest dependence coefficients, and the second bloc consists of Argentina and Peru, whose exchange rate dependence coefficients with other Latin American countries are low. Figure 11 shows that the same conclusion can be obtained using the method proposed in this paper with embedding dimension $m = 3$ and complete linkage.
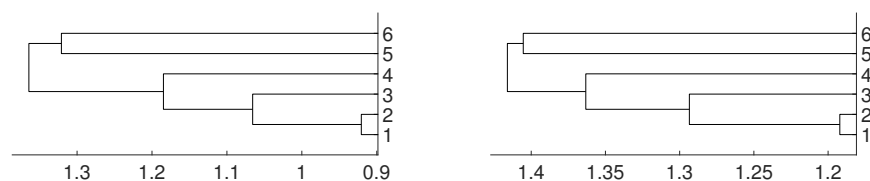

Figure 11. Dendrograms of six Latin American currencies for distance D (left) and D_V for embedding dimension $m = 3$. The currencies are labeled by 1 for BRL, 2 for MXN, 3 for CLP, 4 for COP, 5 for PEN and 6 for ARS.

3.2. Tumor Clustering According to RNA Sequences

The data analyzed below consists of 50 time series of length 20,531 (RNA-sequence). Each series belongs to one of the five types of tumors which can be found and downloaded at the web page https://archive.ics.uci.edu/ml/datasets/gene+expression+cancer+RNA-Seq (see also https://www.synapse.org/#!Synapse:syn2812961). In fact, the original dataset consists of 800 time series, but we have selected the first 10 series of each type of tumor (a total of 50 time series) in order to get a simpler dendrogram. This collection of data is part of the RNA-Seq (HiSeq) PANCAN data set, it is a random extraction of gene expressions of patients having different types of tumors: BRCA (breast carcinoma), KIRC (kidney renal clear-cell carcinoma), COAD (colon adenocarcinoma), LUAD (lung adenocarcinoma) and PRAD (prostate adenocarcinoma).

Our results show that distance D_V clusters properly each data series with its tumor group for embedding dimensions $m = 3$ and average linkage, although the most similar pair of tumors (BRCA and LUAD) cannot be properly grouped. With the same conditions for distance D, a completely correct clustering is obtained (see Figures 12 and 13).
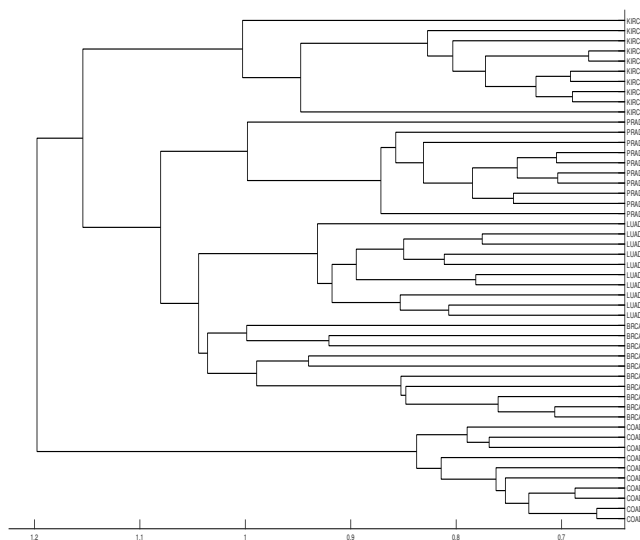

Figure 12. Dendrograms of the 50 time series of RNA-sequence using distance D, embedding dimension $m = 3$ and average linkage method.
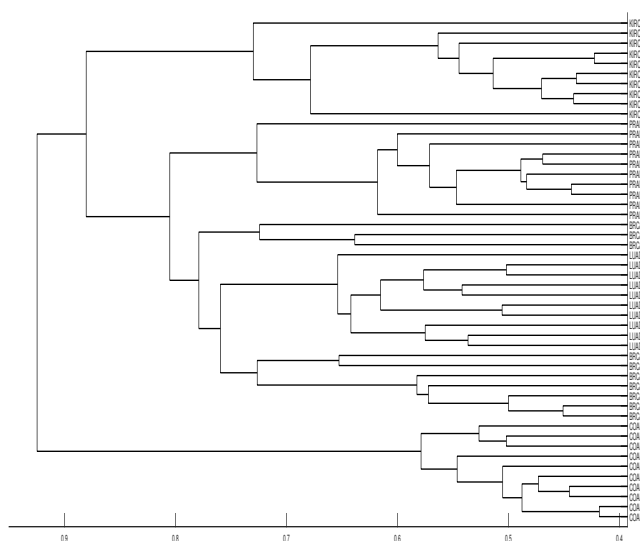

Figure 13. Dendrograms of the 50 time series of RNA-sequence using distance D_V, embedding dimension $m = 3$ and average linkage method.

3.3. Evolution of Spanish IBEX35 Banks

An interesting application consists in analyzing the evolution of log-returns of Spanish banks at IBEX35 Spanish index. In Figure 14 we show the index evolution and the days we use to divide the data. Namely, before the 2008 financial crisis, from 2005 to February of 2009, Banco Popular (POP), Banco Santander (SAN), Banco de Sabadell (SAB) and Bankinter (BKT) were grouped while Banco Bilbao Vizcaya (BBV) was not grouped with them as Figures 15 and 16 show. After the 2008 financial crisis, from February of 2009 to February of 2013, Banco Santander (SAN) was close to Banco Bilbao Vizcaya (BBV), while Popular (POP), Banco de Sabadell (SAB) and Bankinter (BKT) were in another group as Figures 15 and 16 show. For the last period from February 2013 to December 2016, no changes were found.
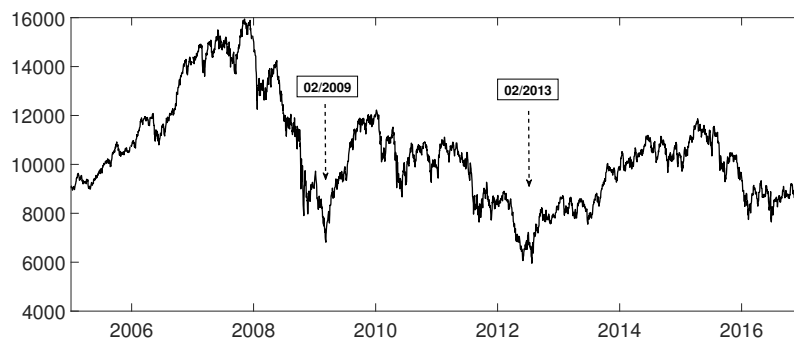

Figure 14. Spanish Ibex35 index evolution from January 2005 to December 2016. The two minimums shown are used to divide the time series.
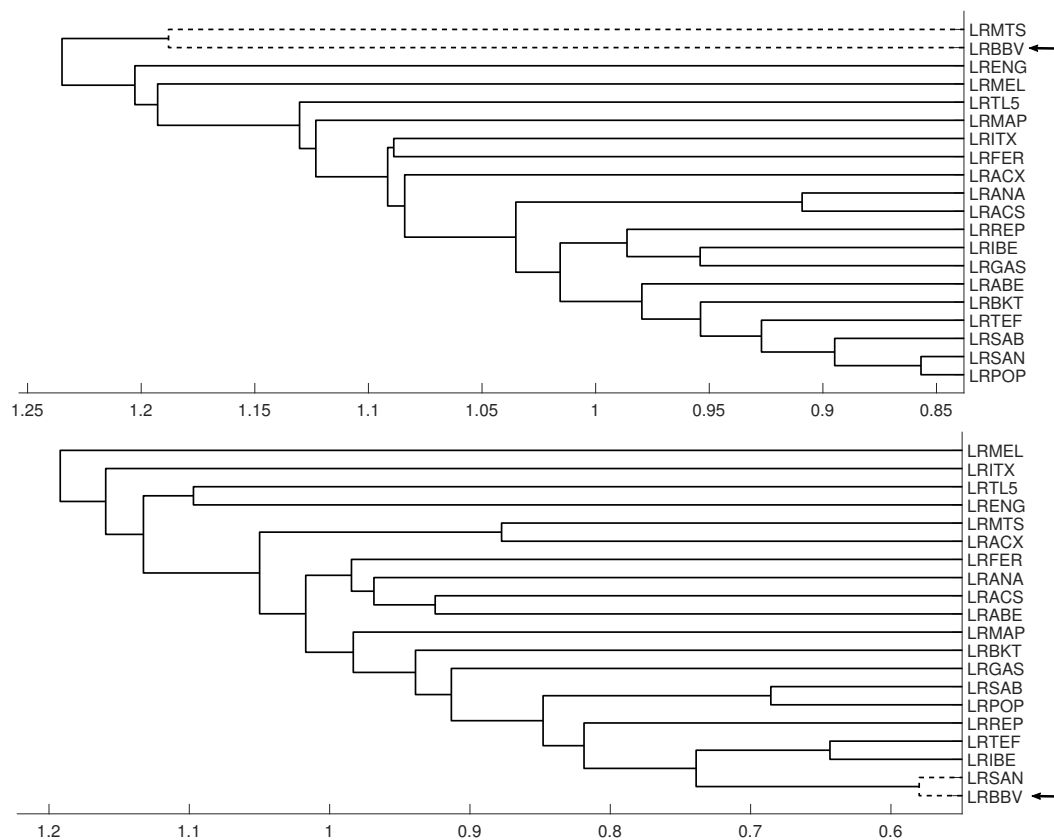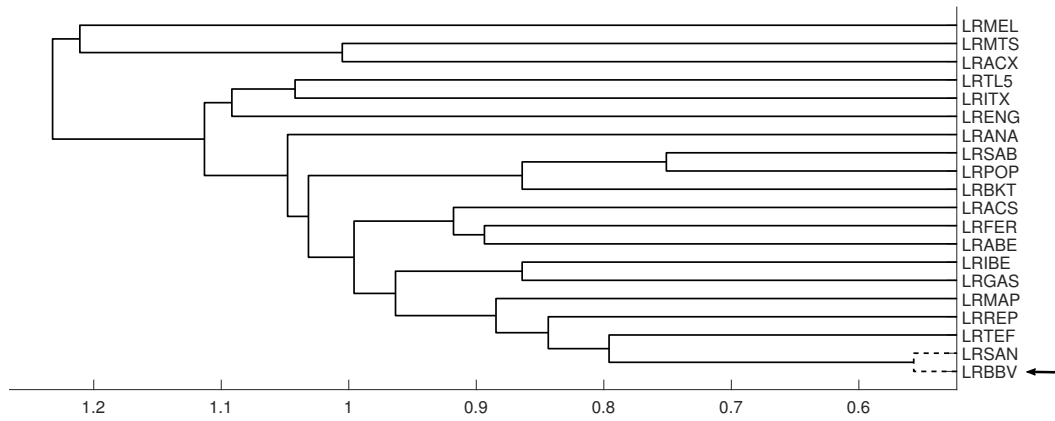

Figure 15. *Cont.*

Figure 15. For measure D, embedding $m = 3$ we show the dendrograms in three different periods, from January 2005 to February 2009 (**top**), from: February 2009 to February 2013 (**middle**) and for February 2013 to December 2106 (**down**). Note the evolution of log-returns (LRBBV) with respect to the log-returns (LRSAN).
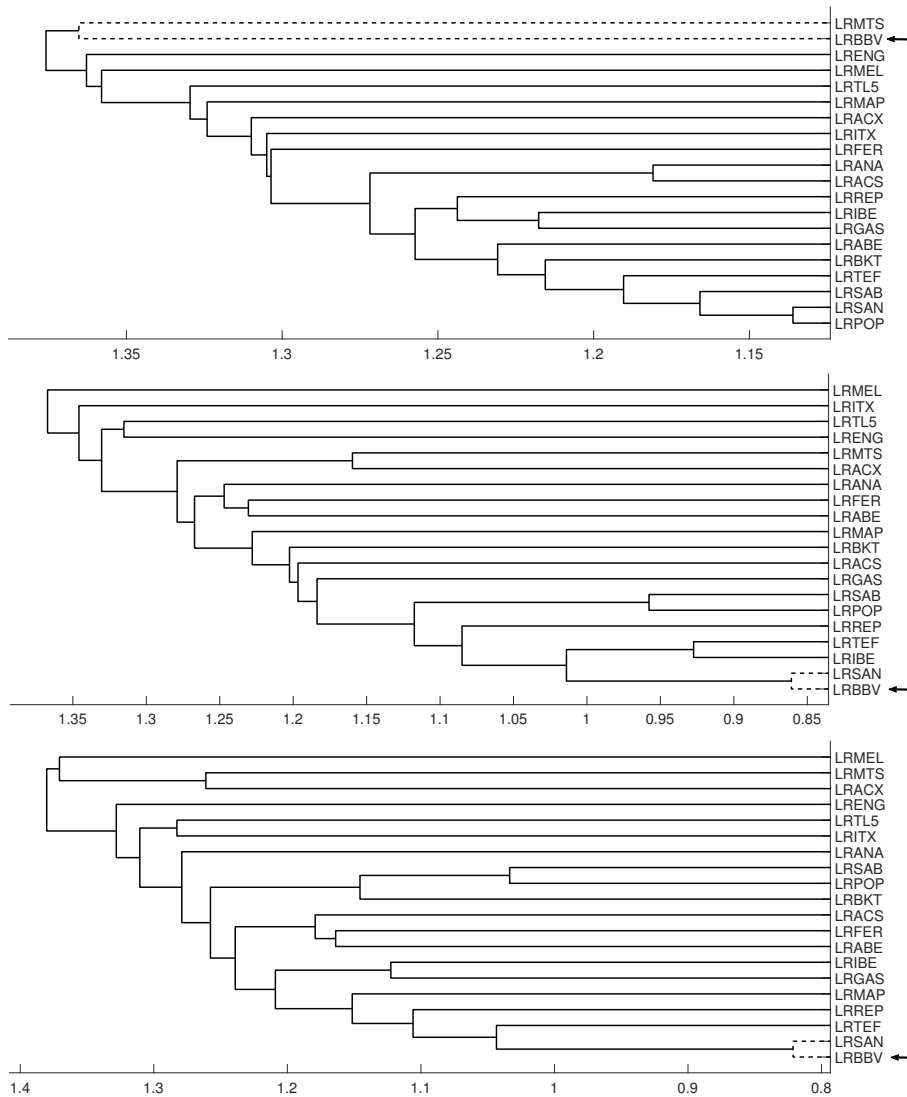

Figure 16. For measure D_V, embedding $m = 3$ we show the dendrograms in three different periods, from January 2005 to February 2009 (**top**), from: February 2009 to February 2013 (**middle**) and for February 2013 to December 2106 (**down**).

4. Conclusions

This paper proposes the using of permutations as an efficient tool in time series clustering. Although traditional approaches (shape-based, feature-based and model-based) are useful to cluster time series that are not related among them, we show that permutations play an important role to cluster time series according to their degree of dependency.

Two distances based on permutations have been considered for the simulations, as well as three different embedding dimensions and three linkages methods for the hierarchical procedure. Simulation results demonstrate that:

- The proposed clustering approach is able to detect linear and non-linear dependencies among time series.
- In some cases, a very small embedding dimension like $m = 2$ is not enough to detect dependencies among time series, thus a greater embedding dimension is required.
- The distance measure based on the mutual information has revealed a better performance than the distance measure based on the Crammer's V statistic.
- There are not significant differences with respect to the selected linkage method.

The necessity of the proposed approach to detect dependencies among time series has been shown using simulated data, by comparing the results obtained with some traditional distances. Furthermore, the performance of this clustering approach has also been validated using real data in the fields of Finance and Health Science.

References

1. Izakian, H.; Pedrycz, W.; Jamal, I. Fuzzy clustering of time series data using dynamic time warping distance. *Eng. Appl. Artif. Intell.* **2015**, *39*, 235–244. [CrossRef]
2. Möller-Levet, C.S.; Klawonn, F.; Cho, K.H.; Wolkenhauer, O. Fuzzy clustering of short time-series and unevenly distributed sampling points. *Adv. Intell. Data Anal.* **2003**, 330–340.
3. Foster, E.D. State Space Time Series Clustering Using Discrepancies Based on the Kullback-Leibler Information and the Mahalanobis Distance. Ph.D. Thesis, University of Iowa, Iowa City, IA, USA, 2012.
4. Zhang, B.; An, B. Clustering time series based on dependence structure. *PLoS ONE* **2018**, *13*, e0206753. [CrossRef]
5. Aghabozorgi, S.; Shirkhorshidi, A.S.; Wah, T.Y. Time-series clustering—A decade review. *Inf. Syst.* **2105**, *53*, 16–38. [CrossRef]
6. Liao, T.W. Clustering of time series data—A survey. *Pattern Recognit.* **2005**, *38*, 1857–1874. [CrossRef]
7. Montero, P.; Vilar, J. TSclust: An R package for time series clustering. *J. Stat. Softw.* **2014**, *62*, 1–43. [CrossRef]
8. Ando, T.; Bai, J. Panel data models with grouped factor structure under unknown group membership. *J. Appl. Econ.* **2016**, *31*, 163–191. [CrossRef]
9. Ando, T.; Bai, J. Clustering huge number of financial time series: A panel data approach with high-dimensional predictor and factor structures. *J. Am. Stat. Assoc.* **2017**, *112*, 1182–1198. [CrossRef]
10. Alonso, A.M.; Peña, D. Clustering time series by linear dependency. *Stat. Comput.* **2018**. [CrossRef]
11. McClellan, S.; Gibson, J. Spectral entropy: An alternative indicator for rate allocation? In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'94), Adelaide, SA, Australia, 19–22 April 1994; pp. 201–204.

12. Bruhn, J.; Ropcke, H.; Hoeft, A. Approximate entropy as an electroencephalographic measure of anesthetic drug effect during desflurane anesthesia. *Anesthesiology* **2000**, *92*, 715–726. [CrossRef]

13. Bruzzo, A.A.; Gesierich, B.; Santi, M.; Tassinari, C.A.; Birbaumer, N.; Rubboli, G. Permutation entropy to detect vigilance changes and preictal states from scalp EEG in epileptic patients. A preliminary study. *Neurol. Sci.* **2008**, *29*, 3–9. [CrossRef] [PubMed]

14. Olofsen, E.; Sleigh, J.W.; Dahan, A. Permutation entropy of the electroencephalogram: A measure of anaesthetic drug effect. *Br. J. Anaesth.* **2008**, *6*, 810–821. [CrossRef] [PubMed]

15. Quian-Quiroga, R.; Blanco, S.; Rosso, O.A.; Garcıa, H.; Rabinowicz, A. Searching for hidden information with Gabor transform in generalized tonic–clonic seizures, Electroencephalography and Clinical. *Neurophysiology* **1997**, *103*, 434–439.

16. Cánovas, J.S.; Garcia-Clemente, G.; Mu noz-Guillermo, M. Comparing permutation entropy functions to detect structural changes in time series. *Phys. A* **2018**, *507*, 153–174. [CrossRef]

17. Parlitz, U.; Berg, S.; Luther, S.; Schirdewan, A.; Kurths, J.; Wessel, N. Classifying cardiac biosignals using ordinal pattern statistics and symbolic dynamics. *Comput. Biol. Med.* **2012**, *42*, 319–327. [CrossRef]

18. Sinn, M.; Keller, K.; Chen, B. Segmentation and classification of time series using ordinal pattern distributions. *Eur. Phys. J. Spec. Top.* **2013**, *222*, 587–598. [CrossRef]

19. Echegoyen, I.; Vera-Avila, V.; Sevilla-Escoboza, R.; Martinez, J.H.; Buldua, J.M. Ordinal synchronization: Using ordinal patterns to capture interdependencies between time series. *Chaos Solitons Fractals* **2019**, *119*, 8–18. [CrossRef]

20. Ruiz-Abellón, M.C.; Guillamón, A.; Gabaldón, A. Dependency-aware clustering of time series and its application on Energy Markets. *Energies* **2016**, *9*, 809. [CrossRef]

21. Bandt, C.; Pompe, B. Permutation entropy—A natural complexity measure for time series. *Phys. Rev. Lett.* **2002**, *88*, 174102. [CrossRef]

22. Amigó, J.M.; Zambrano, S.; Sanjuán, M.A.F. True and false forbidden patterns in deterministic and random dynamics. *Europhys. Lett. EPL* **2007**, *79*, 50001. [CrossRef]

23. Amigó, J.M.; Kennel, M.B. Topological permutation entropy. *Phys. D Nonlinear Phenom.* **2007**, *231*, 137–142. [CrossRef]

24. Matilla, M.; Ruíz, M. A non–parametric independence test using permutation entropy. *J. Econom.* **2008**, *144*, 139–155. [CrossRef]

25. Bandt, C.; Keller, G.; Pompe, B. Entropy of interval maps via permutations. *Nonlinearity* **2002**, *15*, 1595–1602. [CrossRef]

26. Cánovas, J.S. Estimating topological entropy from individual orbits. *Int. J. Comput. Math.* **2009**, *86*, 1901–1906. [CrossRef]

27. Cánovas, J.S.; Guillamón, A.; del Ruíz, M. Using permutations to detect dependence between time series. *Phys. D Nonlinear Phenom.* **2011**, *240*, 1199–1204. [CrossRef]

28. Wallis, S. *Measures of Association for Contingency Tables*; University College London: London, UK, 2012.

29. Kraskov, A.; Stögbauer, H.; Andrzejak, R.G.; Grassberger, P. Hierarchical Clustering Based on Mutual Information. *arXiv* **2005**, arXiv:q-bio/0311039.

30. Walters, P. *An Introduction to Ergodic Theory*; Springer: New York, NY, USA, 1982.

31. Herrera, A.R. Analysis of dispersal effects in metapopulation models. *J. Math. Biol.* **2016**, *72*, 683–698. [CrossRef] [PubMed]

32. Gómez-González, J.E.; Melo-Velandia, L.F.; Maya, R.L. Latin American Exchange Rate Dependencies: A Regular Vine Copula Approach. *Contemp. Econ. Policy* **2015**, *33*, 535–549.