



An MDS-based unifying approach to time series K-means clustering: application in the dynamic time warping framework

J. Fernando Vera¹ · José M. Angulo¹

Accepted: 11 May 2023 / Published online: 30 May 2023
© The Author(s) 2023

Abstract

Partitioning algorithms, and in particular K-means clustering, are widely used in time series analysis. K-means clustering is intrinsically related to the use of the Euclidean distance as a measure of dissimilarity. When other dissimilarity measures, such as dynamic time warping, are involved, K-means clustering is usually replaced by the optimisation of a sums-of-the-stars clustering criterion, giving rise to an algorithm other than that of K-means, such as K-medoids. Another common restriction in the implementation of K-means concerns the need to estimate the average as the cluster prototype, which may represent a drawback for this method in time series when elastic measures such as dynamic time warping are used. In this paper, we propose a multidimensional scaling based K-means clustering algorithm that enables the use of K-means clustering together with any dissimilarity measure, and in particular with dynamic time warping, without requiring us to estimate cluster prototypes for the time series. This procedure is a true K-means clustering algorithm that searches for the partition in an equivalent auxiliary configuration, usually in a dimension lower than the time series length. The approach proposed is of particular interest when dynamic time warping is used in the analysis of series of unequal length and/or when some data are missing, and hence Euclidean distances cannot be used. The performance of our procedure is tested by conducting an extensive Monte Carlo experiment, comparing the results with those obtained by K-medoids. The procedure is also illustrated with the analysis of carbon dioxide emissions from 133 countries.

Keywords K-means clustering · K-medoids · Time series · Dynamic time warping · Cluster prototype · Multidimensional scaling

1 Introduction

Cluster analysis is a popular data analysis tool that has been widely used in diverse fields, for different types of data sets; one such is the analysis of time series (Liao 2005). Many studies have sought to identify homogeneous group structures in time-varying data, and recent increases in storage and manipulation capacities for large real-world

data sets have enabled significant technical developments in this context. Applications of time series clustering can be found, for instance, in the determination of singular patterns (Sfetsos and Siriopoulos 2004), in pattern discovery (Pavlidis et al. 2006) and in the detection of similar dynamic changes (He et al. 2012). A comprehensive review of time series clustering and its applications is given in Aghabozorgi et al. (2015).

In addition to computational difficulties inherent to the size of the data sets considered in time series analysis, comparisons in terms of dissimilarity measures are subject to certain difficulties, related to the characteristics of the measurement scale, the uniformity of the sample and the length of the series. Various dissimilarity measures have been proposed to overcome these issues, which may also affect the clustering procedure adopted (see Montero and Vilar 2014). Several methods have been employed for whole time series clustering. Among many partition clustering approaches, the traditional K-means clustering

J. Fernando Vera and José M. Angulo contributed equally to this work.

✉ J. Fernando Vera
jfvera@ugr.es

José M. Angulo
jmangulo@ugr.es

¹ Department of Statistics and Operations Research, University of Granada, Campus Fuente Nueva s/n, 18071 Granada, Spain

technique stands out, together with optimisation algorithms related to a generalised K-means procedure, in which Euclidean distances are replaced by dissimilarities and/or the centroid is replaced by an exemplar or cluster prototype (see Liao 2005, for a comprehensive review).

The K-means algorithm (MacQueen 1967; Hartigan and Wong 1979), produces a partition of the rows of an $N \times p$ matrix \mathbf{X} into a specified number K of non-overlapping groups, on the basis of their (squared) Euclidean distances. For application in this framework, the rows of \mathbf{X} are considered as N time series of equal length p , and each observation is classified into the cluster with the nearest mean (centroid) value. The useful data dispersion properties of the K-means algorithm are well known (Anderson 1958). These properties are intrinsically related to the use of the group means or centroids as the cluster exemplars, together with Euclidean distances (see Vera and Macías 2021). In time series analysis, K-means clustering can be performed either directly on the raw data, or with the prior use of feature-based approaches to reduce the high dimensionality of the data, in which the same time units are used to consider the data being sampled.

However, as noted above, the presence of series of unequal length, feature-value pair vectors or transition matrices is common in this framework, which means that the Euclidean distance cannot always be used. On many occasions, this means that a sums-of-the-stars clustering criterion must be optimised (see Everitt et al. 2011, Chapter 5) based on a (non-Euclidean) measure of dissimilarity between the series. Such algorithms usually allocate a time series into the group whose exemplar or prototype (not necessarily the centroid) is nearest in terms of the dissimilarity measure. The group exemplars or medoids (Kaufman and Rousseeuw 1990) are not necessarily the centroids, but a time series within the data set, giving rise to a non-K-means algorithm, such as the well-known technique of partitioning around medoids or the PAM algorithm (Kaufman and Rousseeuw 1990). In another approach, avoiding the need to estimate cluster prototypes, Witten and Tibshirani (2010) proposed a generalised criterion for clustering in a dissimilarity matrix, pointing out that their criterion is equivalent to K-means only when the dissimilarities are considered to be squared Euclidean distances (see also Hastie et al. 2009, Section 14.3.6).

Various approaches to perform time series partitioning have been suggested. None of these, however, although related to K-means, correspond to a true K-means algorithm. For instance, Kakizawa et al. (1998) minimised the lack-of-homogeneity criterion (Everitt et al. 2011) replacing the squared Euclidean distances by measures of divergence between series. Previously, in a feature-based framework, Wilpon and Rabiner (1985) described a procedure that

combined the substitution of Euclidean distances by one based on log-likelihood, with the determination of the centroids by a minimax or pseudo-average procedure to determine the centres of the clusters. In another proposal, Vlachos et al. (2003) suggested using wavelets together with a procedure to progressively select the initial centroids in K-means.

In many of the above cases, the primary consideration is that, in some experimental situations, the estimation of the cluster centroid for time series is a challenge in itself, particularly in the presence of time series of unequal length (Gupta et al. 1996), or when the similarities between series are based on their shape. This problem has been addressed in relation to the use of elastic approaches such as dynamic time warping (DTW) and longest common sub-sequence (LCSS). In this respect, for example, Niennattrakul and Ratanamahatana (2007) and Hautamaki et al. (2008) have highlighted the existence of convergence problems in K-means because of the difficulty of accurately reflecting the average shape of time series. Aghabozorgi et al. (2015) have also observed the fact that elastic approaches have sometimes been experimentally eluded when related to the need for the use of cluster prototypes. Specifically, Niennattrakul and Ratanamahatana (2007) pointed out that dynamic time warping should not be used in conjunction with K-means clustering because it may fail to average the shape of the time series, but that K-medoids with dynamic time warping gives satisfactory results.

Without involving the estimation of the cluster exemplars, Vera and Macías (2017) proposed a method to perform (generalised) K-means clustering and to select the number of clusters when the information is given in terms of a dissimilarity matrix. More recently, Vera and Macías (2021) demonstrated the usefulness of multidimensional scaling (MDS) in relation to K-means clustering in a dissimilarity matrix. One of the most interesting outcomes of this study is the conclusion that minimising a lack-of-homogeneity criterion in terms of the squared dissimilarities is equivalent to applying a true K-means clustering procedure, whilst avoiding the need to estimate cluster prototypes in the raw data set. In addition, these authors identify an auxiliary embedding space with respect to which the clustering partition is equivalent, and an efficient procedure to determine the number of clusters using any existing criteria for rectangular data.

In this paper we propose a K-means clustering procedure for the analysis of time series of equal or unequal length. As it makes use of any appropriate dissimilarity measure, this procedure does not require the estimation of the cluster representatives in the raw data set. Neither is it exposed to the above-mentioned problems in the application of K-means clustering in this framework (Aghabozorgi et al. 2015, Section 4), in particular, those related to the

inaccuracies of K-means in conjunction with a distance time warping procedure (Niennattrakul and Ratanamahatana 2007). Our method is based on the use of MDS in full dimension together with a suitable translation of the off-diagonal elements of the matrix of squared dissimilarities. A further benefit is that this methodology allows the use of any traditional criterion for determining the number of clusters in K-means for time series.

The rest of the paper is organised as follows. In the next section, we describe the dynamic time warping method for comparing time series in shape, and discuss the use of traditional K-means clustering, generalised K-means and K-medoids procedures for time series. In Sect. 3, we introduce a procedure to perform K-means clustering without estimating cluster prototypes, and consider criteria for determining the number of clusters derived from a dissimilarity matrix. Section 4 then details the proposed procedure for K-means clustering in time series. In Sect. 5, a Monte Carlo experiment is conducted to determine and analyse the performance of this procedure, for well-known real and simulated data sets. In addition, its application to the case of carbon dioxide emissions is described. In the final section, we discuss the results obtained and summarise the main conclusions drawn.

2 Dynamic time warping and generalised K-means clustering

Dynamic time warping is a well known elastic method to obtain the dissimilarity in shape between two time series of equal or unequal length. Broadly speaking, the idea is to align the two time series by finding an optimal mapping on which to calculate the minimum distance, usually Euclidean, between the series.

For two time series $\mathbf{a} = (a_1, \dots, a_n)$ and $\mathbf{b} = (b_1, \dots, b_m)$, of lengths n and m , respectively, first the $n \times m$ matrix of distances $d(a_i, b_j)$, $i = 1, \dots, n$, $j = 1, \dots, m$, is calculated. Then, a warping path $\mathbf{w} = (w_1, \dots, w_R)$ of pairs $w_r = (i_r, j_r)$, $i_r \in \{1, \dots, n\}$, $j_r \in \{1, \dots, m\}$, $r = 1, \dots, R$, such that minimises the normalized cumulative distance, i.e.

$$d_{DTW}(\mathbf{a}, \mathbf{b}) = \min_{\mathcal{P}} \frac{1}{R} \sum_{r=1}^R d_{w_r}, \tag{1}$$

where $d_{w_r} = d(a_{i_r}, b_{j_r})$, $r = 1, \dots, R$, is found within the set \mathcal{P} of warping paths, defined under the following conditions: (1) the path must start and finish in the diagonally opposite cells (1, 1) and (n, m), respectively; (2) the allowable steps of the path must be to adjacent cells in a non-reversal time direction, that is, with a maximum increment of one positive unit in each index. These

constraints ensure that all the elements of both time series are included (at least once) in the path, with the length R satisfying $\max(n, m) \leq R \leq m + n - 1$.

Operationally, the warping path is usually calculated by dynamic programming, based on the recursive evaluation of elements of the $n \times m$ matrix of optimal partial cumulative distances, as follows:

$$d_{cum}(i, j) = d(a_i, b_j) + \min\{d_{cum}(i - 1, j - 1), d_{cum}(i - 1, j), d_{cum}(i, j - 1)\}. \tag{2}$$

As mentioned in Sect. 1, dynamic time warping is widely used in conjunction with clustering procedures, due to its suitability for comparing the shapes of two time series, particularly when they are of unequal length. However, and as noted above, estimating a representative of the cluster as the average shape of the time series in that cluster, according to the optimal mapping between them, may produce inaccuracies. Difficulties with the estimation of the average shape have led to K-means clustering in time series being rejected in favour of alternative partitioning algorithms such as K-medoids, in which the prototype is one of the group’s time series, the medoid (see Aghabozorgi et al. 2015, for an extensive overview).

2.1 K-means, generalised K-means and K-medoids clustering procedures

Let us denote by \mathbf{X} an $N \times m$ matrix representing a set of N time series of equal length m , $\mathbf{x}_i = \{x_{i1}, \dots, x_{im}\}$, arranged by rows, and all observed at the same time periods. Assume a partition of the rows of \mathbf{X} into K clusters, and denote by \mathbf{E} an $N \times K$ indicator matrix, whose elements e_{ik} are equal to one if x_i belongs to cluster k and zero otherwise. The dispersion of \mathbf{X} measured by the trace of its covariance matrix \mathbf{S} can be written in terms of the Euclidean distances as

$$tr \mathbf{T} = tr \mathbf{N}\mathbf{S} = tr \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' = \sum_{i=1}^N d(\mathbf{x}_i, \bar{\mathbf{x}})^2. \tag{3}$$

where $d(\mathbf{x}_i, \bar{\mathbf{x}})^2$ is the squared Euclidean distance of the i th time series to the overall time series average. Hence, performing K-means clustering on \mathbf{X} consists in finding \mathbf{E} , by minimising the within-cluster dispersion given by (see Everit et al. 2011)

$$\min_E W(K) = \min_E \sum_{k=1}^K \sum_{i=1}^N e_{ik} d_{ik}^2, \tag{4}$$

where $d_{ik}^2 = (\mathbf{x}_i - \bar{\mathbf{x}}_k)'(\mathbf{x}_i - \bar{\mathbf{x}}_k)$ is the Euclidean distance of each time series to the average of the time series in each cluster k , denoted by $\bar{\mathbf{x}}_k$.

When the time series are not comparable in times, or the series are of different length, the use of Euclidean distances is usually preceded by a transformation stage (Keogh and Kasetty 2003). Otherwise, a generalised K-means procedure is normally used.

Various partitioning algorithms have been proposed to generalise K-means, although none can be properly considered a K-means clustering procedure, and so the latter’s useful properties in terms of dispersion (Anderson 1958) are not available. A currently popular approach in time series is to apply the K-means subroutine by simply replacing the squared Euclidean distances with another measure of dissimilarity, and the cluster prototypes accordingly (see, e.g., Liao 2005, or Aghabozorgi et al. 2015). This method has been used with dynamic time warping on raw time series data by minimising (Niennattrakul and Ratanamahatana 2007)

$$\min_E \sum_{k=1}^K \sum_{i=1}^N e_{ik} d_{DTW}(\mathbf{x}_i, \mathbf{c}_k), \tag{5}$$

where \mathbf{c}_k is the average shape of the time series in cluster k according to the dynamic time warping optimal mapping between these series (see, e.g., Aghabozorgi et al. 2015).

Alternatively, the K-medoids procedure (Kaufman and Rousseeuw 1990), in addition to considering in (4) any measure of dissimilarity instead of the squared Euclidean distance, takes \mathbf{x}_k as the nearest time series \mathbf{x}_k in the cluster to all the others in terms of such a dissimilarity, which for dynamic time warping is

$$\min_E \sum_{k=1}^K \sum_{i=1}^N e_{ik} d_{DTW}(\mathbf{x}_i, \mathbf{x}_k). \tag{6}$$

This method does not suffer from the inaccuracies that can arise when using the averaged shape as a cluster prototype, but even for Euclidean distances, it does not represent a true K-means clustering procedure. However, the problem with the representatives of the groups in time series can be avoided since a true K-means clustering algorithm can be properly applied with any dissimilarity measure without the need to estimate cluster prototypes.

3 K-means clustering without estimating prototypes of time series groups

Contrary to many views that have been expressed in the time series literature (see, e.g., Aghabozorgi et al. 2015, Section 5.2), the application of K-means does not require us to estimate the raw representatives of the clusters. Instead, the optimal partition E can be obtained in a K-means framework by minimising (see, e.g., Vera and Macías 2017, Section 2)

$$W(K) = \sum_{k=1}^K \sum_{i=1}^N e_{ik} d_{ik}^2 = \sum_{k=1}^K \frac{1}{2N_k} \sum_{i=1}^N \sum_{j=1}^N e_{ik} e_{jk} d_{ij}^2, \tag{7}$$

where $d_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j)$ represents the squared Euclidean distance between the rows \mathbf{x}_i and \mathbf{x}_j of X , and N_k is the cardinality of cluster k . This problem is also equivalent to maximising

$$B(K) = \frac{1}{2N} \sum_{i=1}^N \sum_{j=1}^N d_{ij}^2 - \sum_{k=1}^K \frac{1}{2N_k} \sum_{i=1}^N \sum_{j=1}^N e_{ik} e_{jk} d_{ij}^2, \tag{8}$$

which represents the between-group dispersion in terms of pairwise squared Euclidean distances, since the first component on the right-hand side of (8) represents the dispersion ($tr T$), and there is no need to consider the time series average. Whether (7) or (8) is used, Euclidean distances are considered together with K-means to group similar time series in time. However, as noted above, there are many practical situations in which the time series comparison is performed in terms of other dissimilarity measures, and for this purpose a generalised K-means clustering algorithm is often used.

In general, for any matrix of dissimilarities between time series, $\Delta = (\delta_{ij})$, the above-described expressions allow us to perform generalised K-means clustering without the need to estimate cluster prototypes, for instance, by minimising the lack of homogeneity criterion in which (7) is formulated in terms of dissimilarities instead of squared Euclidean distances. It is important to note that this criterion is equivalent to K-means clustering only when dissimilarities are considered to be squared Euclidean distances (Hastie et al. 2009, Section 14.3.6).

3.1 Selecting the number of clusters

As shown by Vera and Macías (2017), since any partition E induces a block-shaped partition of Δ , the overall dispersion of a dissimilarity matrix can be expressed as

$$\sum_{k \leq l} \sum_{i=1}^N \sum_{j=1}^N e_{ik} e_{jl} w_{ij} (\delta_{ij} - \bar{\delta})^2 = \sum_{k \leq l} \sum_{i=1}^N \sum_{j=1}^N e_{ik} e_{jl} w_{ij} (\delta_{ij} - \bar{\delta}_{kl})^2 + \sum_{k \leq l} \hat{w}_{kl} (\bar{\delta}_{kl} - \bar{\delta})^2, \tag{9}$$

where the coefficients w_{ij} represent weights (usually $w_{ij} = 1$, and $w_{ij} = 0$ for missing dissimilarities, but their values can be set by the researcher), $\bar{\delta}$ is the overall mean of dissimilarities, and $\bar{\delta}_{kl}$ and \hat{w}_{kl} are the averaged dissimilarity and the number of dissimilarities in the block Δ_{kl} , respectively. The first component on the right-hand side of (9) represents the within-block dispersion with $(N(N - 1) - K(K + 1))/2$ degrees of freedom, whereas

the second component represents the between-block dispersion with $K(K + 1)/2$ degrees of freedom. The within-block dispersion and the between-block dispersion are respectively denoted as

$$\begin{aligned}
 W^*(K) &= \sum_{k \leq l} \sum_{i=1}^N \sum_{j=1}^N e_{ik} e_{jl} w_{ij} (\delta_{ij} - \bar{\delta}_{kl})^2, \\
 B^*(K) &= \sum_{k \leq l} \hat{w}_{kl} (\bar{\delta}_{kl} - \bar{\delta})^2,
 \end{aligned}
 \tag{10}$$

which enables us to formulate variance-based criteria to select the number of clusters. In this respect, Vera and Macías (2017) proposed two criteria, CH^* and H^* , as an alternative to the Calinski–Harabasz index (Calinski and Harabasz 1974) and to Hartigan’s rule (Hartigan 1975), respectively, in terms of the block-shaped partition of the dissimilarity matrix. In the first case, the optimum number of clusters, K , is associated with a large value of

$$CH^*(K) = \frac{B^*(K)/(K(K + 1)/2)}{W^*(K)/([N(N - 1) - K(K + 1)]/2)}, \tag{11}$$

where, for $K = 1$, $B^*(1) = 0$ and thus $CH^*(1) = 0$. Therefore, the maximum will never be reached for $K = 1$. For the second criterion, these authors proposed selecting the number of clusters when $H^*(K) \leq 5N$, where

$$\begin{aligned}
 H^*(K) &= \left[\frac{W^*(K)}{W^*(K + 1)} - 1 \right] \\
 &\quad \times (([N(N - 1) - K(K + 1)]/2) - 1).
 \end{aligned}
 \tag{12}$$

4 A unified approach to K-means clustering in time series

We now consider squared dissimilarities instead of squared Euclidean distances, and denote by Δ^2 the matrix of squared dissimilarities, and by $\tilde{\Delta}^2 = \Delta^2 + c(\mathbf{I}\mathbf{I}' - \mathbf{I})$ a linear transformation of the off diagonal elements of Δ^2 , with $c = -2\lambda_N$, where $\lambda_1 \geq \dots \geq \lambda_N$ are the eigenvalues of $\mathbf{B} = -\frac{1}{2}\mathbf{H}\Delta^2\mathbf{H}$, and $\mathbf{H} = \mathbf{I} - \frac{1}{N}\mathbf{I}\mathbf{I}'$ is the centring matrix. Note that $c \geq 0$, and $c = 0$ when Δ is a matrix of Euclidean distances. Then $\tilde{\Delta}^2$ is a matrix of squared Euclidean distances (Lingoes 1971; Vera and Macías 2021). Vera and Macías (2021) have also shown that performing K-means clustering on the classical multidimensional scaling configuration \mathbf{X}_c in dimension $p = \text{rank}(\mathbf{B}) \leq N - 2$, related to the Euclidean distance matrix $\tilde{\Delta}$, is equivalent to doing so by minimising (7) directly using Δ^2 .

These results have important implications for time series clustering, as this approach allows the application of K-

means clustering relative to any dissimilarity matrix between time series. In addition, it does not require the estimation of representative time series for each group. This outcome is achieved by means of an auxiliary matrix \mathbf{X}_c , whose dimension $p \leq (N - 2)$ only depends on the number of time series, N , which is usually considerably smaller than the maximum length m of the time series. Hence, this procedure avoids the two main drawbacks in the application of K-means to time series. The proposed dissimilarity-based K-means clustering algorithm for time series can be summarised as follows (see Fig. 1 for a schematic version):

1. A dissimilarity matrix Δ of dimension $N \times N$ is obtained between the time series for each particular experimental situation. In particular, dynamic time warping can be used to obtain dissimilarities between series of unequal length. For asymmetric patterns in dynamic time warping matrices, the dissimilarity matrix is symmetrized (see, for example, Vera and Rivera 2014, and Vera and Mair 2019), for a more detailed description of asymmetric dissimilarity analysis in MDS), or the lower triangular matrix is considered.
2. The minimum eigenvalue λ_N of the matrix $\mathbf{B} = -\frac{1}{2}\mathbf{H}\Delta^2\mathbf{H}$, with $\mathbf{H} = \mathbf{I} - \frac{1}{N}\mathbf{I}\mathbf{I}'$, is obtained, and the matrix of Euclidean distances $\tilde{\Delta}^2 = \Delta^2 - 2\lambda_N(\mathbf{I}\mathbf{I}' - \mathbf{I})$ is calculated.
3. The classical scaling solution in full dimension $\mathbf{X}_c = \Gamma\tilde{\Delta}^{-1/2}$ for the spectral decomposition of $\tilde{\mathbf{B}} = -(1/2)\mathbf{H}\tilde{\Delta}^2\mathbf{H} = \Gamma\tilde{\Lambda}\Gamma'$ is obtained. Therefore, the time series are represented exactly with respect to their K-means partition in terms of the dissimilarities in a Euclidean space of dimension $p \leq (N - 2)$, while the dispersion is preserved (except for a constant $-2\lambda_N(N - 1) \geq 0$, which does not depend on K).
4. K-means clustering is performed on \mathbf{X}_c for a selected value of K , and the partition matrix \mathbf{E} is estimated accordingly. To address the problem of determining the number of clusters, different partitions $\mathbf{E}(K)$ can be estimated for different values of K using this procedure. Hence, any criterion formulated directly in terms of the corresponding partition of Δ can be used to determine the number of clusters. Another advantage of the MDS is that it also allows us to use any classical criterion on the full dimensional scaling configuration (Vera and Macías 2021). Since there is no unique criterion to determine the number of clusters, this also confers an important added value on the proposed procedure.

Dissimilarities: Calculate dynamic time warping dissimilarities $\Delta_{N \times N}$
Symmetrise: $\Delta_{N \times N} = 0.5(\Delta_{N \times N} + \Delta'_{N \times N})$, or use the triangular part
Additive constant: Calculate: $\mathbf{H} = \mathbf{I} - (1/N)\mathbf{1}\mathbf{1}^t$, $\Delta^2 = (\delta_{ij}^2)$, $\mathbf{B} = -0.5\mathbf{H}\Delta^2\mathbf{H}$
Eigenvalues: Obtain singular values of \mathbf{B} ; $\lambda_1 \geq \dots \geq \lambda_N$
Additive constant: $c = -2\lambda_N$
Squared Euclidean distances: $\tilde{\Delta}^2 = \Delta^2 - 2\lambda_N(\mathbf{1}\mathbf{1}^t - \mathbf{I})$
Classical MDS: Calculate: $\tilde{\mathbf{B}} = -(1/2)\mathbf{H}\tilde{\Delta}^2\mathbf{H}$
Eigendecomposition: $\tilde{\mathbf{B}} = \mathbf{\Gamma}\tilde{\mathbf{\Lambda}}\mathbf{\Gamma}'$
Auxiliary configuration: $\mathbf{X}_c = \mathbf{\Gamma}\tilde{\mathbf{\Lambda}}^{1/2}$
K-means: Estimate the partition \mathbf{E} using any classical algorithm on \mathbf{X}_c ;
equivalently, using $\mathbf{D}^2(\mathbf{X}_c)$, squared Euclidean distances on \mathbf{X}_c
Cluster prototypes: Calculate cluster representatives, if required, using any approach

Fig. 1 K-means clustering algorithm for dissimilarities

5 Experimental results

The performance of the proposed procedure was analysed for both simulated and real data sets using R Statistical Software (v4.2.2; R Core Team 2022), and the results obtained were compared with those achieved by the K-medoids procedure using the `pam` function implemented in the `cluster` R package (v2.1.4; Maechler et al. 2022). The dissimilarity between each pair of time series was calculated by dynamic time warping using the `dtw` package (v1.23.1; Giorgino 2009), and each data set was analysed considering 50 random starts and a maximum of 100 iterations for each procedure. It is important to note that different packages in R may implement different dynamic time warping algorithms. For example, when using the lower bound technique, or when large time series matrices are involved, the `dtwclust` package (v5.5.11; Sardá-Espinoza 2019, 2022) is a more advisable alternative. Where possible, the accuracy of the resulting classification was measured using the cluster similarity index defined by Gavrilov et al. (2000), as implemented in the `TSclust` package (v1.3.1; Montero and Vilar 2014; see also Díaz and Vilar 2010), which is given by

$$\text{Sim}(E, E') = \frac{1}{K} \sum_{k=1}^K \max_{1 \leq t \leq K} \frac{|J_k \cap J'_t|}{|J_k| + |J'_t|}, \quad (13)$$

where $E = \{J_1, \dots, J_k\}$ and $E' = \{J'_1, \dots, J'_k\}$ are two partitions for the same data set, and $|\cdot|$ denotes the cardinality. This index varies between zero meaning total dissimilarity, and one, which indicates that the two classifications are identical. As a criterion to determine if two classifications are comparable, values below 0.7 are considered here as not advisable.

5.1 UCR data sets

Fifteen well-known data sets from the UCR Time Series Classification Archive (Chen et al. 2015) were considered,

including both real and simulated data. The `GesturePebbleZ1` and `GesturePebbleZ2` sets of series of unequal length, together with the `CBF`, `Mallat`, `SmoothSubspace`, `SyntheticControl`, and `TwoPattern` sets of series of equal length, were analysed. Each set was divided into a training sample and a test sample, and together with the complete set, all were analysed separately, except for the `Mallat` and `TwoPattern` sets, due to their large size. In addition, the eight real data sets `BasicMotions`, `DiatomSizeReduction`, `DistalPOAgeGroup`, `FaceFour`, `InsectEPGRegular`, `OliveOil`, `Plane` and `Trace`, were considered. In this experiment, only the data sets producing an accuracy greater than 0.7 in at least one of the procedures used were taken into account.

Table 1 shows the characteristics of these data sets, together with the results obtained with the procedures using dynamic time warping. For the proposed K-means procedure, the dimensionality of the MDS auxiliary configuration is also shown. As can be appreciated, in all simulated data sets the proposed K-means procedure outperformed the K-medoids procedure in terms of the quality of the classification obtained, while the two procedures presented similar levels of performance in terms of CPU time. Comparable results were obtained for the real data sets, although the differences were less apparent. Note that the dimensionality of the MDS auxiliary configuration, in most of the data sets analysed, was significantly lower than the time series length. Only when the size of the data set, that is, the number of time series in the set, was larger than the maximum length of the series, was the dimensionality larger.

5.2 Simulated time series of unequal length

We also performed Monte Carlo simulations to compare the efficiency of the proposed K-means procedure with that of the K-medoids procedure for dynamic time warping in grouping time series of unequal length. Specifically, we

Table 1 Results for the UCR data sets using dynamic time warping

Data set	Size	Length	K	K-means			K-medoids	
				Sim	Dim	Time (s)	Sim	Time (s)
<i>Simulated series of unequal length</i>								
GesturePebbleZ1(TEST)	172	455	6	0.70	88	60	0.62	60
GesturePebbleZ2(TRAIN)	146	455	6	0.72	75	72	0.52	72
<i>Simulated series of equal length</i>								
CBF	930	128	3	0.88	549	1187	0.87	1196
Mallat(TRAIN)	55	1024	8	0.95	32	98	0.80	98
SmoothSubspace	300	15	3	0.90	149	47	0.73	47
SyntheticControl	600	60	6	0.98	335	246	0.94	252
TwoPatterns(TRAIN)	1000	128	4	0.99	507	1047	0.97	1059
<i>Real data sets with series of equal length</i>								
BasicMotions	80	100	4	0.78	51	9	0.77	9
DiatomSizeReduction	322	345	4	0.81	167	645	0.77	71
DistalPOAgeGroup	439	80	3	0.70	230	283	0.62	285
FaceFour	102	350	4	0.73	71	80	0.73	71
InsectEPGRegular	311	601	3	1	148	1221	1	1221
OliveOil	60	570	4	0.80	43	55	0.80	55
Plane	210	144	7	1	110	66	0.98	69
Trace	200	275	4	0.76	101	141	0.76	142

The size (number of instances), length of the series, number of clusters and similarity index (13) values, together with the CPU time for the proposed K-means and the K-medoids procedures, are shown. For the proposed procedure, the dimensionality of the MDS auxiliary configuration is also shown. For the first two sets of series of unequal length, the largest length value is shown. The complete data set was analysed, except for the very large sets, in which case only the training or test subsets were considered. The results for the data sets with an accuracy greater than 0.7 are shown

simulated ten time series for each of the time series lengths of 100, 200, 300, 400, 500, and each of the nine different processes described below.

For each pair of length combinations, ten time series were generated using the corresponding process, for each of the five process pairs described below. Thus, we obtain for each situation a set of twenty series in which there are two groups, one for each process. Therefore, 125 sets of time series were analyzed, each one with 20 series belonging to two groups of 10 time series related to the process and length of the corresponding pairs. The following comparisons were then made:

1. AR(1), $\phi = 0.9$, versus AR(1), $\phi = 0.5$;
2. AR(1), $\phi = 0.95$, versus ARIMA(0,1,0);
3. AR(2), $\phi_1 = 0.6$, $\phi_2 = -0.3$, versus MA(2), $\theta_1 = -0.6$, $\theta_2 = 0.3$;
4. ARFIMA(0,0.45,0), versus white noise;
5. ARMA(1,1), $\phi = 0.95$, $\theta = 0.74$, versus white noise.

In all cases, the models are considered to be driven by Gaussian white noise. The first situation compares low-order models of similar type and autocorrelation function structure. In the second case, a nonstationary process and a near-nonstationary AR process are compared. In the third,

we compare the performance of two selected second-order ARMA processes in order to deal with peak spectra. In the fourth and fifth situations, near-nonstationary long and short-memory processes, respectively, are compared with a white noise process.

For each of the five pairs of processes, 5×5 sets of time series corresponding to all combinations of the sizes were analysed. This procedure was repeated ten times, and therefore 1250 sets of time series were analysed using the proposed K-means and the K-medoids procedures. The resulting classifications were then compared with the original.

Table 2 shows the average values of the similarity index (13), when the classifications obtained with the MDS-based K-means procedure and with the *pam* algorithm are compared with the true classification. In general, the proposed K-means procedure outperforms the K-medoids procedure in all data sets where the partitioning may be considered as acceptable (similarity index above 0.70). This was the case in practically all the comparisons except for AR(1), $\phi = 0.95$, versus ARIMA(0,1,0), although the performance of both procedures, in general, improves with larger series.

Table 2 Average value of similarity indices for the classifications obtained with K-means and K-medoids corresponding to the simulated time series

Groups	K-means					K-medoids				
	(100)	(200)	(300)	(400)	(500)	(100)	(200)	(300)	(400)	(500)
Lengths	(100)									
AR1,0.9 versus AR1,0.5	0.73	0.98	1.00	1.00	1.00	0.73	0.92	0.99	1.00	1.00
AR1,0.95 versus ARIMA010	0.66	0.66	0.67	0.69	0.67	0.67	0.68	0.68	0.69	0.67
AR2 versus MA2	0.63	1.00	1.00	1.00	1.00	0.62	0.87	1.00	1.00	1.00
ARFIMA versus ARIMA000	0.60	1.00	1.00	1.00	1.00	0.61	0.94	1.00	1.00	1.00
ARMA11 versus ARIMA000	0.62	1.00	1.00	1.00	1.00	0.59	0.91	1.00	1.00	1.00
Lengths	(200)									
AR1,0.9 versus AR1,0.5	0.73	0.83	0.99	1.00	1.00	0.73	0.79	0.92	0.97	1.00
AR1,0.95 versus ARIMA010	0.67	0.67	0.67	0.66	0.67	0.70	0.67	0.67	0.69	0.68
AR2 versus MA2	1.00	0.66	0.88	1.00	1.00	0.99	0.65	0.68	0.83	0.98
ARFIMA versus ARIMA000	1.00	0.65	1.00	1.00	1.00	0.95	0.59	0.78	0.99	1.00
ARMA11 versus ARIMA000	1.00	0.59	0.97	1.00	1.00	0.96	0.61	0.82	1.00	1.00
Lengths	(300)									
AR1,0.9 versus AR1,0.5	0.78	0.81	0.92	0.99	1.00	0.81	0.76	0.81	0.90	0.99
AR1,0.95 versus ARIMA010	0.67	0.67	0.67	0.67	0.67	0.69	0.70	0.69	0.69	0.69
AR2 versus MA2	1.00	1.00	0.69	0.73	0.99	1.00	0.94	0.63	0.65	0.62
ARFIMA versus ARIMA000	1.00	0.99	0.60	0.95	1.00	1.00	0.87	0.63	0.75	0.99
ARMA11 versus ARIMA000	1.00	0.98	0.60	0.95	1.00	0.99	0.82	0.60	0.77	0.96
Lengths	(400)									
AR1,0.9 versus AR1,0.5	0.83	0.89	0.91	0.97	1.00	0.86	0.81	0.76	0.80	0.91
AR1,0.95 versus ARIMA010	0.72	0.71	0.72	0.72	0.72	0.76	0.75	0.74	0.75	0.72
AR2 versus MA2	1.00	1.00	1.00	0.67	0.67	1.00	1.00	0.91	0.64	0.62
ARFIMA versus ARIMA000	1.00	1.00	0.94	0.62	0.84	1.00	1.00	0.73	0.62	0.65
ARMA11 versus ARIMA000	1.00	1.00	0.90	0.63	0.87	1.00	0.99	0.82	0.63	0.68
Lengths	(500)									
AR1,0.9 versus AR1,0.5	0.89	0.94	0.92	0.97	0.99	0.86	0.86	0.86	0.87	0.90
AR1,0.95 versus ARIMA010	0.68	0.65	0.65	0.65	0.65	0.72	0.72	0.68	0.65	0.66
AR2 versus MA2	1.00	1.00	1.00	1.00	0.68	1.00	1.00	1.00	0.82	0.60
ARFIMA versus ARIMA000	1.00	1.00	1.00	0.85	0.62	1.00	1.00	0.95	0.66	0.60
ARMA11 versus ARIMA000	1.00	1.00	1.00	0.85	0.61	1.00	1.00	0.96	0.73	0.62

For each pair of processes (Groups), twenty series were generated for each pair of sizes, thus ten time series per process and size. For each pair of processes, and each combination of two sizes, the twenty corresponding series were analysed. The experiment was repeated ten times

5.3 Analysis of carbon dioxide emissions from 133 countries

To illustrate the performance of the proposed method, a real data set consisting of 133 time series of unequal length was analysed. These series correspond to the carbon dioxide emissions ($MtCO_2e$) from the use of coal, oil and gas (combustion and industrial processes), the process of gas flaring and the manufacture of cement from 1960 to 2019 for 133 countries (GCP 2019). The data set is freely available at the online platform Climate Watch 2020. All the series were matched in length to 60, and the emission

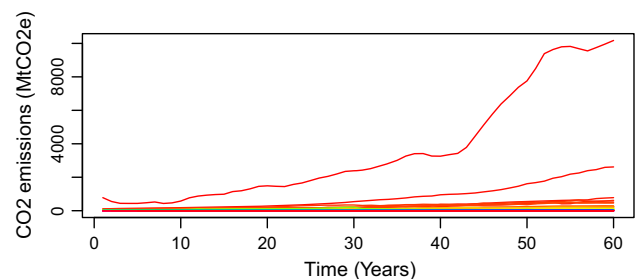


Fig. 2 Carbon dioxide emissions ($MtCO_2e$) for 133 countries from 1960 to 2019

data which in some cases were not provided were considered missing values. Figure 2 shows the emissions for the 60-year period. Strikingly, the curves corresponding to China and India differ sharply from the rest.

With the proposed procedure, K-means clustering is first performed using the values of $K = 2, \dots, 10$ to determine the number of clusters. Working directly on the dynamic time warping dissimilarity matrix, the CH^* criterion suggests three clusters, while the H^* criterion indicates four clusters, and together with a closer inspection of the augmented Fig. 2, $K = 4$ clusters were selected. First, the proposed K-means clustering procedure was performed on the basis of the full MDS-based configuration in 65 dimensions, obtaining four clusters of sizes $n_1 = 1$, $n_2 = 1$, $n_3 = 15$, and $n_4 = 116$. The first cluster corresponds to *China*, the second to *India*, the third to *Iran, Indonesia, Saudi Arabia, South Africa, Brazil, Thailand, Malaysia, Pakistan, Vietnam, Egypt, Iraq, United Arab Emirates, Argentina, Algeria* and *Venezuela*, and the remaining countries form the last cluster.

Figure 3 shows the time series clusters obtained with the MDS-based K-means procedure. The series that belong to group 3 are relatively homogeneous in shape, with some small differences over the sixty years represented. The largest cluster is the fourth, in which, according to its shape, *North Korea* (dashed line) appears to be misclassified. Although some of the shapes within this group vary somewhat, the emission levels are similar in every case.

This data set was also analysed with K-medoids for $K = 4$ clusters using the *pam* function in the *cluster R* package. Four clusters, of sizes $n_1 = 1$, $n_2 = 6$, $n_3 = 20$, $n_4 = 106$, were found with this method. The first cluster, as in K-means, corresponds to China. The second contains *India, Iran, Indonesia, Saudi Arabia, South Africa* and *Brazil*. The countries belonging to the third cluster are *Thailand, Malaysia, Pakistan, Vietnam, Egypt, Iraq, United Arab Emirates, Argentina, Algeria, Philippines, Nigeria, Venezuela, Qatar, Kuwait, Colombia, Bangladesh, Turkmenistan, Chile, Morocco* and *North Korea*, and the remaining countries are classified in the fourth cluster. Hence, different classifications are found according to whether the K-means or the K-medoids procedure is used. Figure 4 shows the times series in each cluster. As can be seen, the second group obtained with K-medoids is fairly inaccurate, since the time series corresponding to *India* seems to be misclassified, while for the third cluster the time series for *North Korea* (dashed line) also shows a different shape, and so this too seems to be misclassified. The remaining series are grouped within the fourth cluster. Here, in general, the time series all have similar shapes, with the notable exception of *North Korea*, as was the case with K-means. A comparison of the classifications provided by both methods reveals a cluster similarity index

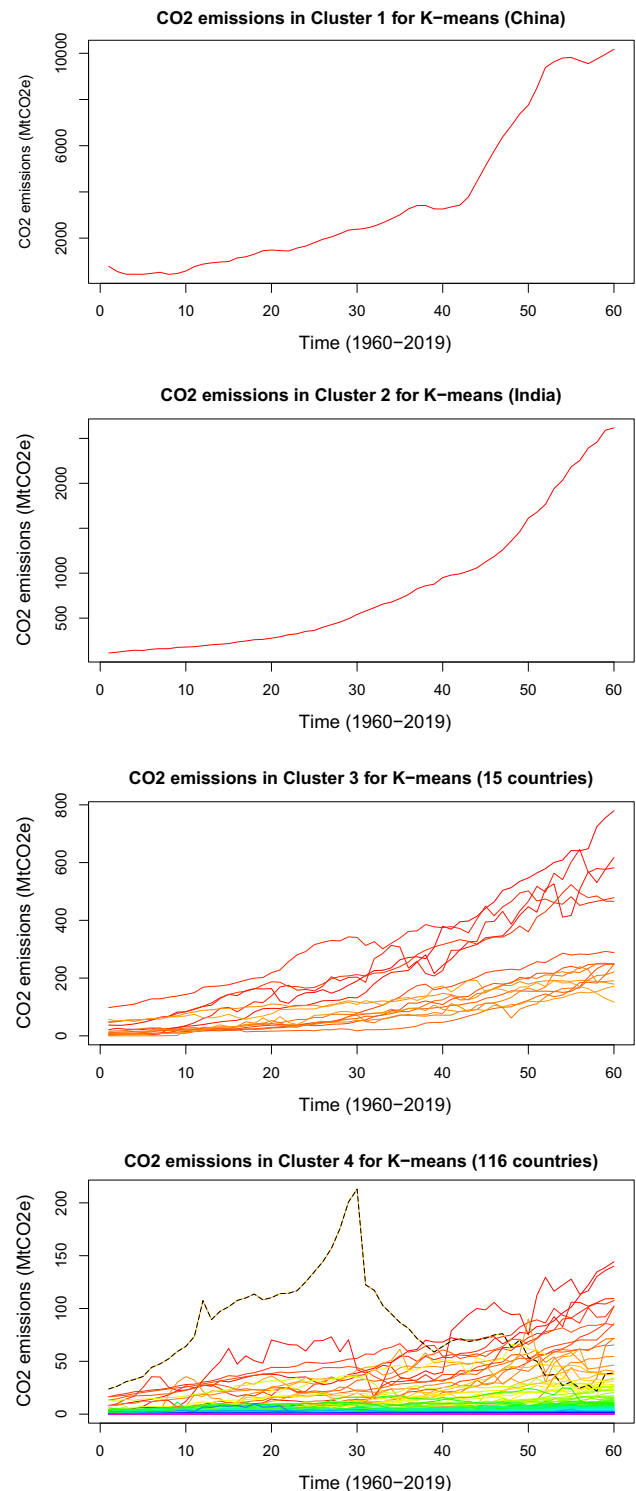


Fig. 3 Carbon dioxide emissions ($MtCO_2e$) for the four clusters found using K-means

value of 0.7, which confirms a comparable behaviour for both procedures to a certain extent, as expected. In general, better results are obtained by the K-means procedure than

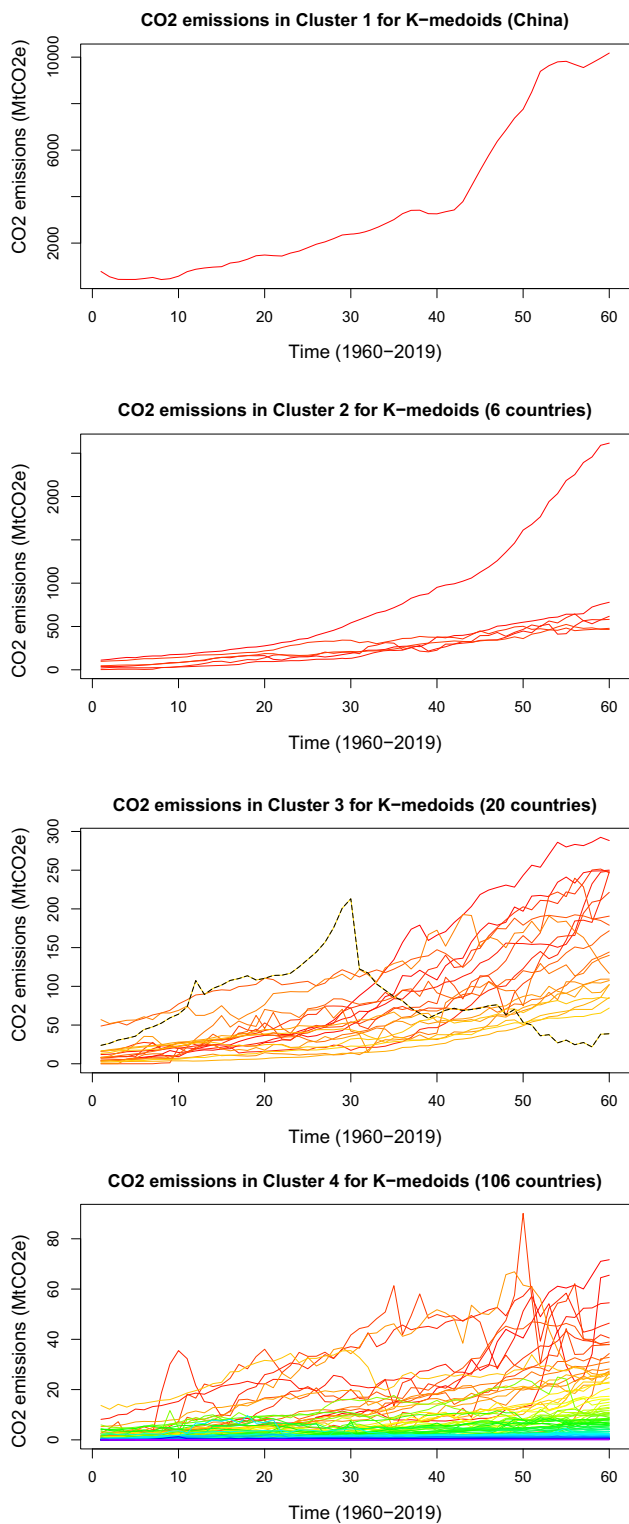


Fig. 4 Carbon dioxide emissions ($MtCO_2e$) for the four clusters found using K-medoids

by K-medoids, particularly with regard to the time series for China and India.

In addition to illustrating the behavior of both procedures for the complete data set using the objective criterion of $K = 4$ clusters, we have also considered two situations for which much closer results could be expected between both procedures. First of all, and despite the fact that K-medoids are assumed to be a more robust procedure than K-means, we have analyzed the data by previously eliminating the time series for China and India, considering $K = 3$ clusters. On the other hand, we have analyzed the complete data set considering $K = 5$ clusters subjectively, according to what can be seen at first sight in cluster 3 of Fig. 3 and cluster 2 of Fig. 4. For $K = 5$, both clustering procedures show identical results in the first three clusters, with cluster 1 for China, cluster 2 for India, and cluster 3 for Iran, Indonesia, Saudi Arabia, South Africa, and Brazil. The differences between both procedures appear in the two remaining clusters, with sizes of 18 and 108 for K-means, and 20 and 106 for K-medoids, where the two time series that represent these differences correspond to Bangladesh and Morocco. For $K = 3$, the results are identical to those of the last three groups for $K = 5$, both for K-means and for K-medoids. Figure 5 shows the classification of the two

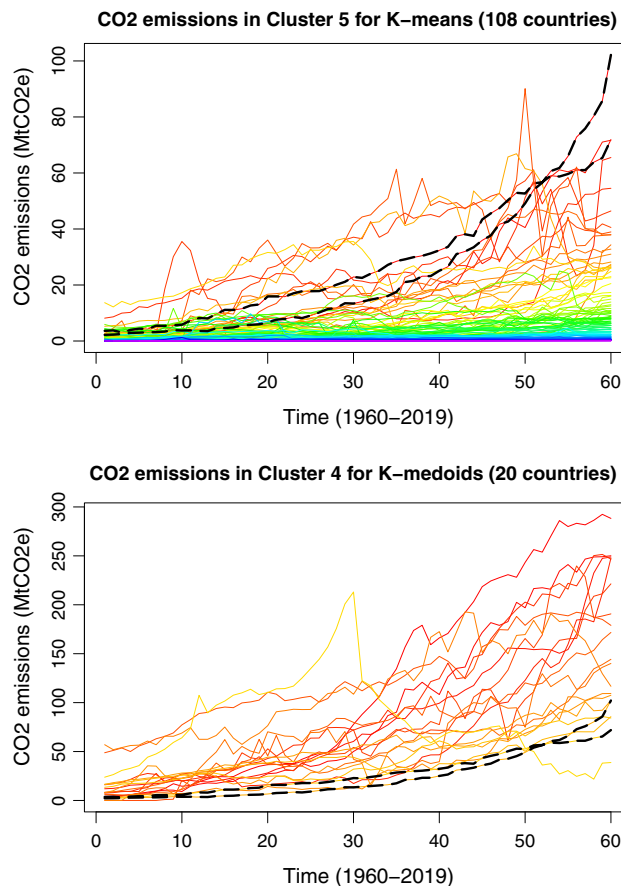


Fig. 5 Classification of Bangladesh and Morocco time series (in dotted lines) by the two procedures for $K = 5$

time series in the clusters obtained by the K-means (cluster 5) and K-medoids (cluster 4) procedures.

In terms of the dynamic time warping dissimilarity matrix, let us now consider the average of the 18 dissimilarity vectors related to the elements of cluster 4, as well as that corresponding to the 106 elements of cluster 5, after excluding the vectors for Bangladesh and Morocco. Then, the Euclidean distance between the dissimilarity vectors of both countries and the average vectors of clusters 4 (312.4 and 309.4) and cluster 5 (220.6 and 199.9) are considerably smaller for cluster 5 in both countries. Therefore, as in the previous situation, the proposed K-means procedure provides the best results in this data set.

6 Conclusions

In this paper, we present a true K-means clustering procedure for time series that, based on multidimensional scaling in full dimension, makes use of any appropriate measure of dissimilarity, particularly when dealing with dynamic time warping. This method is especially useful when the time series are of unequal length and/or when some values are missing, as frequently occurs when elastic measures of dissimilarity are used.

In the proposed procedure, K-means clustering is performed using a particular auxiliary configuration, generally of a dimension less than the greatest length of the time series, the main characteristic of which is that the optimal partition is equivalent to what can be found by directly applying K-means clustering to the dissimilarity matrix (Vera and Macías 2021). This novel methodology does not suffer from the problems mentioned above regarding the application of K-means in this framework (see Aghabozorgi et al. 2015, Section 4), in particular, as concerns the inaccuracy of K-means clustering in reflecting the average shape of time series in conjunction with a distance time warping procedure (Niennattrakul and Ratanamahatana 2007). In addition, our method enables the use of any K-means algorithm for the analysis of time series together with any dissimilarity matrix, as well as the application of any classical criterion to determine the number of clusters in the time series. Furthermore, cluster prototypes can be determined afterwards using any criteria on the K-means partition obtained.

The performance of the proposed procedure is analysed together with that of the K-medoids method, for both simulated and real data sets. In this analysis, we first consider simulated and well-known real data sets from the UCR Time Series Classification Archive (Chen et al. 2015). In addition, we conduct a Monte Carlo experiment to investigate the performance of each procedure, analysing the two groups of simulated time series and comparing

pairs of processes with different characteristics. In general, the solutions obtained with the proposed procedure outperform those obtained with K-medoids in terms of classification quality and also, albeit slightly, in terms of CPU time.

The proposed methodology was applied to analyse carbon dioxide emissions ($MtCO_2e$) from the use of coal, oil and gas (combustion and industrial processes), the process of gas flaring and the manufacture of cement from 1960 to 2019, using data from 133 countries (GCP 2019). The results of this analysis were compared with those obtained using K-medoids. These results vary according to the criteria employed to determine the number of clusters. While most of the classical criteria only indicate two clusters, the extended criteria of Vera and Macías (2017) indicate $K = 3$ for CH^* and $K = 4$ for H^* . For $K = 4$, the proposed K-means procedure generates a partition that seems to be more in agreement with the actual emissions recorded, especially with respect to the second group, for which the K-medoids procedure clearly obtains a deficient classification.

The problem of estimating the cluster prototype as the average when dynamic time warping is used together with K-means clustering, can be extended to the framework of spatiotemporal processes. In this respect, we are currently investigating the behaviour of the proposed procedure, together with various elastic measures of dissimilarity in this context. Another aspect of interest for practical considerations concerns the CPU time growth under different alternatives when the proposed procedure is used with large data sets.

Acknowledgements The authors are grateful to the associate editor and reviewers for their constructive comments and suggestions, which led to improving and clarifying various aspects in the revision of the original manuscript.

Funding Funding for open access publishing: Universidad de Granada/CBUA.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted

use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aghabozorgi S, Shirkhorshidi A, Wah T (2015) Time-series clustering—a decade review. *Inf Syst* 53:16–38
- Anderson TW (1958) An introduction to multivariate statistical analysis. Wiley, New York
- Caiado J, Crato N, Peña D (2009) Comparison of times series with unequal length in the frequency domain. *Commun Stat Simul Comput* 38:527–540
- Calinski RB, Harabasz J (1974) A dendrite method for cluster analysis. *Commun Stat Theory Methods* 3:1–27
- Chen Y, Keogh BH, Begum N, Bagnall A, Mueen A, Batista G (2015) The UCR time series classification archive. <http://www.time-seriesclassification.com/index.php>
- Climate Watch 2020 Washington, DC: World Resources Institute. <https://www.climatewatchdata.org>
- Díaz SP, Vilar JA (2010) Comparing several parametric and nonparametric approaches to time series clustering: a simulation study. *J Classif* 27:333–362
- Everitt BS, Landau S, Leese M, Stahl D (2011) Cluster analysis. Wiley series in probability and statistics, 5th edn. Wiley, Chichester
- Gavrilov M, Anguelov D, Indyk P, Motwani R (2000) Mining the stock market: which measure is best? In: Proceedings of the 6th international conference on knowledge discovery and data mining (KDD'00), pp 487–496
- GCP data: Global Carbon Project (2019) Carbon budget and trends 2019. www.globalcarbonproject.org/carbonbudget, published on 4 December 2019. <https://doi.org/10.18160/gcp-2020>
- Giorgino T (2009) Computing and visualizing dynamic time warping alignments in R: the DTW package. *J Stat Softw* 31:1–24
- Gupta L, Molfese DL, Tammana R, Simos PG (1996) Nonlinear alignment and averaging for estimating the evoked potential. *IEEE Trans Biomed Eng* 43:348–56
- Hartigan JA (1975) Clustering algorithms. Wiley, New York
- Hartigan JA, Wong MA (1979) Algorithm AS 136: a K-means clustering algorithm. *Appl Stat* 28:100–108
- Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction. Springer, New York
- Hautamaki V, Nykanen P, Franti P (2008) Time-series clustering by approximate prototypes. In: Proceedings of 19th International Conference on Pattern Recognition (ICPR2008), no. D 1–4
- He W, Feng G, Wu Q, He T, Wan S, Chou J (2012) A new method for abrupt dynamic change detection of correlated time series. *Int J Climatol* 32:1604–1614
- Kakizawa Y, Shumway RH, Taniguchi M (1998) Discrimination and clustering for multivariate time series. *J Am Stat Assoc* 93:328–340
- Kaufman L, Rousseeuw PJ (1990) Finding groups in data: an introduction to cluster analysis. Wiley Series in Probability and Statistics. Wiley, Hoboken
- Keogh E, Kasetty S (2003) On the need for time series data mining benchmarks: a survey and empirical demonstration. *Data Min Knowl Discov* 7:349–371
- Liao W (2005) Clustering of time series data—a survey. *Pattern Recognit* 38:1857–1874
- Lingoes JC (1971) Some boundary conditions for a monotone analysis of symmetric matrices. *Psychometrika* 36:195–203
- Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K (2022) Cluster: cluster analysis basics and extensions. R Package Version 2(1):4
- MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In: Fifth Berkeley symposium on mathematical statistics and probability, vol II, pp 281–297
- Montero P, Vilar J (2014) TSclust: an R package for time series clustering. *J Stat Softw* 62:1–43
- Niennattrakul V, Ratanamahatana CA (2007) Inaccuracies of shape averaging method using dynamic time warping for time series data. In: Shi Y, van Albada GD, Dongarra J, Sloat PMA (eds) Computational science—ICCS 2007. Lecture notes in computer science, vol 4487. Springer, Berlin, Heidelberg, pp 513–520
- Pavlidis NG, Plagianakos VP, Tasoulis DK, Vrahatis MN (2006) Financial forecasting through unsupervised clustering and neural networks. *Oper Res* 6:103–127
- R Core Team (2022) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Sardá-Espinosa A (2019) Time-series clustering in R using the dtwclust package. *R J* 11:22–43
- Sardá-Espinosa A (2022) dtwclust: time series clustering along with optimizations for the dynamic time warping distance. R package version 5.5.11. <https://CRAN.R-project.org/package=dtwclust>
- Sfetsos A, Siriopoulos C (2004) Time series forecasting with a hybrid clustering scheme and pattern recognition. *IEEE Trans Syst Man Cybern Part A Syst Hum* 34:399–405
- Steinley D (2006) K-means clustering: a half-century synthesis. *Br J Math Stat Psychol* 59:1–34
- Vera JF, Macías R (2017) Variance-based cluster selection criteria in a K-means framework for one-mode dissimilarity data. *Psychometrika* 82:275–294
- Vera JF, Macías R (2021) On the behaviour of K-means clustering of a dissimilarity matrix by means of full multidimensional scaling. *Psychometrika* 86:489–513
- Vera JF, Mair P (2019) An R package for structural equation multidimensional scaling. *Struct Equ Model Multidiscip J* 26:803–818
- Vera JF, Rivera CD (2014) A structural equation multidimensional scaling model for one-mode asymmetric dissimilarity data. *Struct Equ Model Multidiscip J* 21:54–62
- Vlachos M, Lin J, Keogh E, Gunopulos D (2003) A wavelet-based anytime algorithm for K-means clustering of time series. In: Proceeding of workshop on clustering high dimensionality data and its applications. 3rd SIAM international conference on data mining, pp 23–30
- Wilpon JG, Rabiner RL (1985) Modified K-means clustering algorithm for use in isolated word recognition. *IEEE Trans Acoust Speech Signal Process* 33:587–594
- Witten DM, Tibshirani R (2010) A framework for feature selection in clustering. *J Am Stat Assoc* 105:713–726

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.