



---

*Research article*

## **A robust alternating least squares K-means clustering approach for times series using dynamic time warping dissimilarities**

**J. Fernando Vera-Vera\*** and **J. Antonio Roldán-Nofuentes**

Department of Statistics and O.R., University of Granada, Faculty of Sciences, Fuentenueva s/n, 18071, Granada, Spain

\* **Correspondence:** Email: [jfvera@ugr.es](mailto:jfvera@ugr.es); Tel: +34 958243157; Fax: +34-958243267.

**Abstract:** Time series clustering is a usual task in many different areas. Algorithms such as K-means and model-based clustering procedures are used relating to multivariate assumptions on the datasets, as the consideration of Euclidean distances, or a probabilistic distribution of the observed variables. However, in many cases the observed time series are of unequal length and/or there is missing data or, simply, the time periods observed for the series are not comparable between them, which does not allow the direct application of these methods. In this framework, dynamic time warping is an advisable and well-known elastic dissimilarity procedure, in particular when the analysis is accomplished in terms of the shape of the time series. In relation to a dissimilarity matrix, K-means clustering can be performed using a particular procedure based on classical multidimensional scaling in full dimension, which can result in a clustering problem in high dimensionality for large sample sizes. In this paper, we propose a procedure robust to dimensionality reduction, based on an auxiliary configuration estimated from the squared dynamic time warping dissimilarities, using an alternating least squares procedure. The performance of the model is compared to that obtained using classical multidimensional scaling, as well as to that of model-based clustering using this related auxiliary linear projection. An extensive Monte Carlo procedure is employed to analyze the performance of the proposed method in which real and simulated datasets are considered. The results obtained indicate that the proposed K-means procedure, in general, slightly improves the one based on the classical configuration, both being robust in reduced dimensionality, making it advisable for large datasets. In contrast, model-based clustering in the classical projection is greatly affected by high dimensionality, offering worse results than K-means, even in reduced dimension.

**Keywords:** K-means; dynamic time warping; model-based clustering; time series; multidimensional scaling; ALSICAL

---

## 1. Introduction

Time series clustering is a recurring problem in real-world applications that has been addressed in many different areas. Many clustering procedures have been proposed, with partition clustering being one of the most used methods, its objective being to determine separate homogeneous groups of time series (see e.g., [1, 2] for a comprehensive review). Applications can be found, for example, related to the analysis of smart environments based on IoT [3], biomedical measurements like blood pressure, electrocardiogram [4], image processing as in brain activity [5], or diverse applications evolving microarray time series [6], among others. Usually, these datasets are characterized by a high dimensionality compared to the sample size, and the data usually needs to be preprocessed before clustering. A notable situation is when the times in which the event are observed are not directly comparable between different time series, since this characteristic does not allow the consideration of an intrinsic coordinate system in the observed dataset. Here we are interested in partition clustering procedures when traditional methods cannot be directly applied to the raw observed dataset. This situation may arise, for instance, when the time series are of different lengths, when there is missing data, and, in general, when the information for clustering is focused on some dissimilarity measure between the time series. A particularly useful and well-known dissimilarity measure in this framework is dynamic time warping (see e.g., [2]), which enables comparisons between series in terms of their shape, with or without the same length, and, in general, with non-synchronized observations.

Traditional partitioning methods as K-means [7] or model-based clustering procedures [8] are related to inherent assumptions on the observed datasets, as the consideration of Euclidean distance as the dissimilarity measure between time series in the first (see e.g., [9]) or a probabilistic distribution assumption on the raw data [10]. When the information of interest is given in terms of dissimilarities between the time series that are not Euclidean distances, procedures as K-means clustering cannot be directly applied. Furthermore, class-specific distribution patterns in the observed time series are difficult to assume and scalability problems may also arise [1], which makes probabilistic clustering procedures also not directly applicable.

Several general dissimilarity measures can be considered between time series further apart of Euclidean distances (see e.g., [11–13]), for an extensive description. Among these, here we focus on comparing the time series with respect to their shape using elastic dissimilarity measures [2], in particular, dynamic time warping. Although generalized K-means clustering procedures have been proposed for a dissimilarity matrix (see e.g., [14]), none of them can be properly considered a K-means clustering procedure if the dissimilarities are not Euclidean distances. Recently, Vera and Macías [15] have shown a procedure to perform real clustering of K-means in a dissimilarity matrix using an auxiliary configuration that is a linear projection estimated from a matrix of Euclidean distances using (classical) multidimensional scaling (MDS) in full dimension. One advantage of this procedure is that the estimated partition using traditional K-means on this configuration is equivalent to that estimated using the observed dissimilarity matrix in a true K-means clustering procedure. Based on these results, Vera and Angulo [16] have shown a unified approach for performing K-means clustering in time series analysis, in particular, when dynamic time warping dissimilarities are used. The procedure does not need to estimate time series cluster prototypes and has shown superior results compared to K-medoids [17] in the data analyzed in this framework.

In the above described procedures, K-means clustering can be performed on the auxiliary configu-

ration using any traditional algorithm. However, it is well known that K-means may perform poorly when a large number of variables and/or samples is present. The fact that multidimensional scaling is carried out in full dimension may result in a practical limitation since the dimensionality of the estimated configuration can turn out to be large and close to the size of the sample (see [18]). Several methods have been proposed to reduce dimensionality combined with time series clustering [1]. In particular, for K-means clustering, principal components analysis has been used to reduce the number of variables, but it is well known that tandem analysis, which first reduces the number of variables and then performs K-means grouping, generally it doesn't work well (see, for example, [19]). Although an extensive simulation experiment has not been carried out in this regard, the main drawback of this procedure seems to be that the use of a very reduced space obtained from principal component analysis does not guarantee the definition of a subspace that is informative with respect to the true underlying structure of the data, possibly because Euclidean distances are altered [20]. However, since the auxiliary configuration in the procedure of Vera and Angulo [16] is naturally oriented in the directions of the principal axes, the partition should not be greatly affected by a reduction of dimensionality while preserving the representation (as for MDS).

The auxiliary configuration in the Vera and Angulo [16] procedure exactly represents the time series in terms of the related squared Euclidean distances using classical MDS [15], for which it uses full dimension. However, a least squares efficient MDS method that directly approximates squared dissimilarities by squared Euclidean distances is that given by the alternating least squares scaling (ALSCAL) procedure of Takane, Young and de Leeuw [21], which generally outperforms the classical configuration, and is usually used as the initial solution in the alternating iterative procedure, in low dimension. Bailey and Gower [22] showed the maximum dimensionality in ALSCAL is the same as the classical MDS procedure on the (non-squared) dissimilarities [15]. Since ALSCAL starts from Euclidean distances using an additive constant, it is interesting to study if the auxiliary configuration estimated using ALSCAL also preserves the K-means partition, being the nonlinear solution that best approximates the squared dissimilarities by squared Euclidean distances. In addition, the classical MDS configuration is nested, that is, the solution in any dimension simply is an extension of the solution in the previous dimensions (a solution in lower dimension is a linear projection of that in full dimension). However, in the alternating least squares approach, the solution in each dimension is estimated optimally and independently of the previous solutions. Hence, the ALSCAL configuration may result an interesting alternative approach to perform K-means clustering.

In addition to analyzing the performance of K-means clustering under alternative conditions for the auxiliary configuration, probabilistic formulations in MDS, and, in particular, the well-known Hefner [23] model, we also motivate the consideration of a model-based clustering procedure in the auxiliary configuration. The Hefner model induces Gaussian distributions of the dimensions in the MDS configuration [24] and, therefore, the performance of model-based clustering under the consideration of a Gaussian mixture approach [25] is an interesting alternative, taking into account its well-known relationship between a Gaussian mixture and K-means clustering [19]. In this paper we also investigate whether the results persist for time series clustering in this particular framework of dynamic time warping, under different scenarios and dimensionality reduction in this probabilistic model. The main motivation regarding the previous findings is that the auxiliary configuration is in the orientation of the principal axes, while exactly preserving the original partition structure of the dynamic time warping dissimilarities related to a translation of the squared dissimilarities.

In general, it is well known that the additive constant problem considerably increases the dimensionality of the solution, but of very smooth shape. This means that a greater number of dimensions is needed to explain the same percentage of variability compared to the untranslated solution, and more dimensions are needed to go from one percentage to a higher one. Thus, proper dimension reduction should not greatly alter the structure of the data in this situation, while gaining computational efficiency. In this paper, we show that time series K-means clustering is robust to dimension reduction in a dynamic time warping framework. In addition, a flexible and dimensionally robust procedure is proposed to perform K-means clustering using ALSCAL together with the squared Euclidean distances resulting from the additive constant procedure on the observed squared dissimilarities. Therefore, the performance of the K-means procedure is studied when considering variations in the configuration using ALSCAL, in the estimation method considering mixtures of Gaussian distributions, as well as in dimensionality reduction, all within the framework of dynamic time warping dissimilarities, in which large datasets are usually involved. It is important to keep in mind that the dimensionality reduction in classical MDS is related to the optimal linear projection, while in ALSCAL, this configuration is obtained as the best approximation in reduced dimension in a least squares framework and, therefore, is less restrictive.

The remainder of the paper is structured as follows. In the following section, K-means clustering in a dissimilarity matrix is described, along with two alternative clustering procedures, one related to the ALSCAL configuration and the other one in a related probabilistic framework. In Section 3 an extensive Monte Carlo experiment is described to analyze the behavior of the analyzed clustering procedures, both for real and simulated datasets. In the final section, we discuss the results obtained and summarize the main conclusions drawn.

## 2. Materials and methods

Let us consider two time series  $\mathbf{a} = (a_1, \dots, a_n)$  and  $\mathbf{b} = (b_1, \dots, b_m)$  of lengths  $n$  and  $m$ , respectively. Dynamic time warping is a method that allows us to determine the dissimilarities between both series in terms of shape, even when there is missing data or the series are of unequal lengths (see [26] for a detailed description of the method in those different scenarios). A warping path  $\mathbf{w} = (w_1, \dots, w_R)$  of pairs  $w_r = (i_r, j_r)$ ,  $i_r \in \{1, \dots, n\}$ ,  $j_r \in \{1, \dots, m\}$ ,  $r = 1, \dots, R$ , is estimated minimizing the normalized cumulative distance, i.e.,

$$d_{DTW}(\mathbf{a}, \mathbf{b}) = \min_{\mathcal{P}} \frac{1}{R} \sum_{r=1}^R d_{w_r}, \quad (2.1)$$

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, usually by means of dynamic programming. The warping path must satisfy three conditions. First,  $d_{w_1} = d(a_1, b_1)$ , and  $d_{w_R} = d(a_n, b_m)$ , that is, the path must start and finish in diagonally opposite corner cells of the matrix of distances. Second, the allowable steps of the path must be adjacent cells of the distance matrix, and third, the points in the warping path must be monotonically spaced in time, with the length  $R$  satisfying  $\max(n, m) \leq R \leq m + n - 1$ . Let us assume we observe  $N$  time series and denote by  $\Delta$  the obtained dynamic time warping symmetric dissimilarity matrix between the  $N(N - 1)/2$  pairs of comparisons. Note that in general, dynamic time warping dissimilarities may not be symmetric (see [26]), but without loss of generality, in MDS the dissimilarity matrix will be assumed to be symmetric considering only the triangular part, or it is transformed to be symmetric (see [27, 28]).

MDS gives us a representation of the time series in a Euclidean space of dimension at most  $N - 1$ , so that the Euclidean distances are as close as possible to the observed dissimilarities. For K-means clustering based on a dissimilarity measure, Vera and Macías [15] have shown a procedure based on the additive constant problem that allows obtaining a configuration that exactly preserves the partition of K-means in the original dissimilarities. In particular, the procedure establishes how to perform clustering of K-means in time series in a dynamic time warping framework [16], without the need to estimate clustering prototypes, avoiding the problems previously noted in this regard [1].

From  $\Delta^2$ , the matrix with entries and the squared dissimilarities, K-means clustering can be performed considering  $\tilde{\Delta}^2 = \Delta^2 + c(\mathbf{1}\mathbf{1}' - \mathbf{I})$ , a linear transformation of the off diagonal elements of  $\Delta^2$ , with  $c = -2\lambda_N \geq 0$ , where  $\lambda_1 \geq \dots \geq \lambda_N$  are the eigenvalues of  $\mathbf{B} = -\frac{1}{2}\mathbf{H}\Delta^2\mathbf{H}$ , and  $\mathbf{H} = \mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}'$  is the centering matrix. The intercept is  $c = 0$  only when  $\Delta$  is a matrix of Euclidean distances. Since  $\tilde{\Delta}^2$  is a matrix of squared Euclidean distances [15], performing K-means clustering on the classical multidimensional scaling configuration  $\mathbf{X}$  in dimension  $p = \text{rank}(\mathbf{B}) \leq N - 2$ , related to the Euclidean distance matrix  $\mathbf{D}(\mathbf{X}) = \tilde{\Delta}$ , is equivalent to doing so by minimizing

$$W(K) = \sum_{k=1}^K \sum_{i=1}^N e_{ik} d_{ik}^2 = \sum_{k=1}^K \frac{1}{2N_k} \sum_{i=1}^N \sum_{j=1}^N e_{ik} e_{jk} d_{ij}^2, \quad (2.2)$$

where  $d_{ik}^2$  is the squared Euclidean distance between the  $i$ th row of  $\mathbf{X}$  and the center of the rows of  $\mathbf{X}$  related to the  $k$ th cluster of time series,  $d_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j)$  represents the squared Euclidean distance between the rows  $\mathbf{x}_i$  and  $\mathbf{x}_j$  of  $\mathbf{X}$ , and  $N_k$  is the cardinality of cluster  $k$ . Vera and Macías [15] have shown that the estimated partition matrix minimising (2.2) is equivalent to that given when  $\Delta$  is used instead of  $\mathbf{D}$  in 2.2. Since the configuration  $\mathbf{X}$  is given in the direction of the principal axes, the partition can be expected to be preserved despite a certain reduction in dimensionality. This formulation has the additional implication of performing K-means clustering without the need to compute cluster prototypes throughout the procedure, which represents an additional advantage in time series clustering (see [16]). However, once the clusters are estimated, the cluster representative can be calculated using any procedure according to each specific practical application.

## 2.1. Two related clustering approaches

The K-means procedure described above has been shown to be adequate in relation to an exact representation of the squared dissimilarities in terms of squared Euclidean distances, and, to this end, the classic MDS solution in full dimension  $p = N - 2$  is used. However, this model is naturally related to two alternative procedures. On the one hand, there is the estimation of the optimal partition by model-based clustering on the basis of the MDS solution. On the other hand, we consider the direct estimation of the configuration in a least squares model that approximates the squared dissimilarities through squared Euclidean distances, without the need for a prior additive constant. In both situations, it is also interesting to investigate how dimensionality reduction affects the quality of the solution obtained.

### 2.1.1. Model-based clustering on the linear projection

Each dimension in the above estimated configuration  $\mathbf{X}$  is represented by the corresponding eigenvector to one of the nonzero eigenvalues of  $\mathbf{B}$ , thus being the columns of  $\mathbf{X}$  orthogonal between them.

In this situation, the specific probabilistic model proposed by Hefner [23] can be assumed, based on which model-based clustering can be performed on  $\mathbf{X}$ . Here, we follow the extension of this model proposed by Zinnes and Mackay [24], and therefore, it is assumed each row of  $\mathbf{X}$  is characterized, in each dimension by a localization and a variability parameter. The row is considered a  $p$  dimensional random vector, the component of which is normally and independently distributed with variance varying on each dimension. This assumption is also in line with that assumed by Oh and Raftery [25], although associated with a Bayesian estimation approach for model-based clustering on dissimilarities. In a mixture model framework, the probability  $f_t(\cdot)$  of a row  $\mathbf{x}_i$ ,  $i = 1, \dots, N$  that belongs to cluster  $t$ , for  $t = 1, \dots, K$ , is given by the multivariate normal distribution of mean vector  $\boldsymbol{\mu}_t$  and covariance matrix  $\boldsymbol{\Sigma}_t$ ,

$$f_t(\mathbf{x}_i | \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t) = (2\pi)^{p/2} |\boldsymbol{\Sigma}_t|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu}_t)^T \boldsymbol{\Sigma}_t^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_t)\right). \quad (2.3)$$

Because we do not know in advance which latent class a row belongs to, the p.d.f. ( $g(\cdot)$ ) of the random variable  $\mathbf{x}_i$  becomes a finite mixture of Gaussian densities, also in consonance with the assumption in Oh and Raftery [25] given by (2.3), adopting the expression:

$$g(\mathbf{x}_i | \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{t=1}^K \gamma_t f_t(\mathbf{x}_i | \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t), \quad (2.4)$$

where  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_K)'$ ,  $0 \leq \gamma_t \leq 1$ , and  $\sum_t \gamma_t = 1$ . So, the log-likelihood function to be a maximized subject to the above constraints can be written as

$$\log L = \sum_{i=1}^N \log \sum_{t=1}^K \gamma_t f_t(\mathbf{x}_i | \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t). \quad (2.5)$$

Given the maximum likelihood estimators  $\hat{\boldsymbol{\mu}}$  and  $\hat{\boldsymbol{\Sigma}}_t$ , the posterior probability that an element  $\mathbf{x}_i$  belongs to latent class  $t$  is calculated by means of Bayes' theorem as follows:

$$\pi_{it}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}_t) = \frac{\hat{\gamma}_t f_t(\mathbf{x}_i | \hat{\boldsymbol{\mu}}_t, \hat{\boldsymbol{\Sigma}}_t)}{g(\mathbf{x}_i | \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})}. \quad (2.6)$$

Hence, from the maximum likelihood estimators, a row  $\mathbf{x}_i$  will be assigned to the class that it is most likely to belong to given these posterior probabilities, and for the parameter estimation, the expectation-maximization (EM) algorithm [37] can be employed. When  $\boldsymbol{\Sigma}_k = \sigma^2 \mathbf{I}$ ,  $k = 1, \dots, K$ , the above estimation procedure reduces to the K-means clustering algorithm [19]. The procedure is summarized in Figure 1.

### 2.1.2. A K-means clustering approach using an alternating least squared configuration

In a least squares framework, classical MDS can be seen as the problem of minimizing (minus) the square of doubly residuals between squared distances and squared dissimilarities [29], i.e., the distortion,

---

**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:  $\mathbf{H} = \mathbf{I} - (1/N)\mathbf{1}\mathbf{1}'$ ,  $\Delta^2 = (\delta_{ij}^2)$ ,  $\mathbf{B} = -0.5\mathbf{H}\Delta^2\mathbf{H}$   
**Eigenvalues:** Obtain singular values of  $\mathbf{B}$ ;  $\lambda_1 \geq \dots \geq \lambda_N$   
**Additive constant:**  $c = -2\lambda_N$   
**Squared Euclidean distances:**  $\tilde{\Delta}^2 = \Delta^2 - 2\lambda_N(\mathbf{1}\mathbf{1}' - \mathbf{I})$   
**Classical MDS:** Calculate:  $\tilde{\mathbf{B}} = -(1/2)\mathbf{H}\tilde{\Delta}^2\mathbf{H}$   
**Eigendecomposition:**  $\tilde{\mathbf{B}} = \Gamma\tilde{\Lambda}\Gamma'$   
**Auxiliary configuration:**  $X_c = \Gamma\tilde{\Lambda}^{1/2}$   
**Model-based clustering: Hefner:** Each row  $x_i$  of  $X_c$  follows  $g(x_i | \mu, \Sigma) = \sum_{t=1}^K \gamma_t f_t(x_i | \mu_t, \Sigma_t)$   
**Maximise using EM:**  $\max_{\gamma, \mu, \Sigma} \log L = \sum_{i=1}^N \log \sum_{t=1}^K \gamma_t f_t(x_i | \mu_t, \Sigma_t)$   
**Calculate**  $\pi_{it} = \frac{\tilde{\gamma}_t f_t(x_i | \mu_t, \Sigma_t)}{g(x_i | \hat{\mu}, \hat{\Sigma})}$   
**Estimate**  $E$  by  $\max_t \pi_{it}$ ;  
**Cluster prototypes:** Calculate cluster representatives, if required, using any approach

---

**Figure 1.** Model-based clustering algorithm for dissimilarities.

$$\min_{\hat{X}} \|\mathbf{B} - \hat{X}\hat{X}'\|^2 = \min_{\hat{X}} \left\| -\frac{1}{2}\mathbf{H}(\Delta^2 - \mathbf{D}^2(\hat{X}))\mathbf{H} \right\|^2. \quad (2.7)$$

On the other hand, the ALSCAL model of Takane, Young and de Leeuw [21] is a general multidimensional scaling procedure stating that for the given  $N \times N$  dissimilarity matrix  $\Delta$ , it enables the direct estimation of a configuration  $\mathbf{Y}$  such that the matrix of squares dissimilarities  $\Delta^2$  is directly approached by the matrix of squared Euclidean distances  $\mathbf{D}^2(\mathbf{Y})$  by minimizing the S-Stress criterion given by,

$$SSTRESS(\mathbf{Y}) = \min_{\mathbf{Y}} \|(\Delta^2 - \mathbf{D}^2(\mathbf{Y}))\|^2. \quad (2.8)$$

Bailey and Gower [22] showed that the least squares residuals between squared dissimilarities and squared distances (without double centering), result in an equivalent symmetric matrix approximation problem through a positive semi-definite matrix with diagonal entries weighted by  $w = (N + 2)/4$ , with respect to the off-diagonal elements, that is,

$$\|\Delta^2 - \mathbf{D}^2(\hat{\mathbf{Y}})\|^2 = \frac{N + 2}{4} \sum_{i=1}^N (b_{ii} - \hat{b}_{ii})^2 + \sum_{i \neq j} (b_{ij} - \hat{b}_{ij})^2, \quad (2.9)$$

where the distances are assumed to be related to a centred configuration  $\hat{\mathbf{Y}}$  in dimension  $p \leq \text{rank}(\mathbf{B})$ ,  $\hat{\mathbf{B}} = \hat{\mathbf{Y}}\hat{\mathbf{Y}}'$ , and normalized such that  $\sum_i \sum_j d_{ij}^2 = \sum_i \sum_j \delta_{ij}^2$ . Therefore, here we propose a K-means clustering procedure for dynamic time warping dissimilarities based on this configuration. As shown for ALSCAL, its maximum dimensionality will be, at most, that of the classical method. The K means clustering procedure based on this configuration  $\mathbf{Y}$  aims at minimizing (2.2) in terms of the Euclidean distances  $\mathbf{D}^2$  when this well approximates the squared dissimilarities  $\tilde{\Delta}^2$ . We name this procedure *KmALSCAL*. Figure 2 summarizes this 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  
**ALSCAL:** Minimise:  $SS\text{tress}(Y) = \|(\tilde{\Delta}^2 - D^2(Y))\|^2$ , which  $\tilde{\Delta}^2$  defined as in Figure 1  
**Auxiliary configuration:**  $\hat{Y} = \min_Y SS\text{tress}(Y)$ , centered, and normalised:  $\sum_{i < j} d_{ij}^2 = \sum_{i < j} \tilde{d}_{ij}^2$   
**K-means:** Estimate the partition  $E$  using any classical algorithm on  $Y$ ;  
 equivalently, using  $D^2(Y)$ , squared Euclidean distances on  $Y$   
**Cluster prototypes:** Calculate cluster representatives, if required, using any approach

---

**Figure 2.** K-means clustering algorithm using ALSCAL.

### 3. Results

The performance of the KmALSCAL and model-based clustering procedures in this framework were tested and the given results were compared with those obtained using the K-means procedure on the classical solution, following a method similar to that described in Vera and Angulo [16]. Fifteen well-known time series datasets from the UCR Time Series Classification Archive [30] database were considered, as well as time series sets simulated using different scenarios in an extensive Monte Carlo experiment. In addition, the influence of dimension reduction on the behavior of the three procedures was also investigated.

The analysis was performed using R Statistical Software (v4.3.1; R Core Team 2023), using the functions *cmdscale* and *kmean* for classical MDS and K-means clustering, the function *alscal* implemented in the *smacofx* R package (v0.6.6; Rusch et al. [31]) for ALSCAL MDS, and the function *Mclust* implemented in the *mclust* R package [32] for model-based clustering. The dissimilarity between each pair of time series was calculated by dynamic time warping using the *dtw* package (v1.23.1; Giorgino [26]), and each dataset was analyzed considering 50 random starts and a maximum of 100 iterations for each procedure. The accuracy of the resulting classification was measured using the cluster similarity index defined by Gavrilov et al. [33], as implemented in the *TSclust* package (v1.3.1; Montero and Vilar [11]), which is given by

$$\text{Sim}(E, E') = \frac{1}{K} \sum_{k=1}^K \max_{1 \leq t \leq K} \frac{|J_k \cap J'_t|}{|J_k| + |J'_t|}, \quad (3.1)$$

where  $E = \{J_1, \dots, J_k\}$  and  $E' = \{J'_1, \dots, J'_k\}$  are two partitions for the same dataset, and  $|\cdot|$  denotes the cardinality. This index takes into account the well-known labeling problem and varies between zero, meaning total dissimilarity, and one, which indicates that the two classifications are identical (see [33]). As in Vera and Angulo [16], in this experiment, only the datasets producing an accuracy greater than 0.7 in at least one of the procedures used were taken into account.

#### 3.1. UCR time series archive datasets

The real and simulated datasets from the UCR Time Series Classification Archive [30], previously studied by Vera and Angulo [16] in comparison to  $K$ -medoids [17], have been analyzed in this context. In particular, 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,



**Table 1.** Results for the UCR datasets using dynamic time warping. The size (number of instances), length of the series, number of clusters and similarity index (3.1) values for the K-means, KmALSCAL and the mclust procedures, are shown. The dimensionality of the classical MDS auxiliary configuration is also shown. For the first two sets of series of unequal length, the largest length value is shown. The complete dataset was analysed, except for the very large sets, in which case only the training or test subsets were considered. The results for the datasets with an accuracy greater than 0.7 for at least one method are shown.

Data set	Size	Length	K-means			KmALSCAL	mclust
			K	Sim	Dim	Sim	Sim
Simulated series of unequal length							
GesturePebbleZ1(TEST)	172	455	6	0.70	88	0.68	0.28
GesturePebbleZ2(TRAIN)	146	455	6	0.72	75	0.69	0.28
Simulated series of equal length							
CBF	930	128	3	0.88	549	0.88	0.50
Mallat(TRAIN)	55	1024	8	0.95	32	0.95	0.26
SmoothSubspace	300	15	3	0.90	149	0.94	0.50
SyntheticControl	600	60	6	0.98	335	0.97	0.28
TwoPatterns(TRAIN)	1000	128	4	0.99	507	0.99	0.40
Real data sets with series of equal length							
BasicMotions	80	100	4	0.78	51	0.80	0.40
DiatomSizeReduction	322	345	4	0.81	167	0.76	0.39
DistalPOAgeGroup	439	80	3	0.70	230	0.68	0.46
FaceFour	102	350	4	0.73	71	0.73	0.40
InsectEPGRegular	311	601	3	1	148	1	0.75
OliveOil	60	570	4	0.80	43	0.74	0.36
Plane	210	144	7	1	110	1	0.25
Trace	200	275	4	0.76	101	0.54	0.40

were analyzed. 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 datasets BasicMotions, DiatomSizeReduction, DistalPOAgeGroup, FaceFour, InsectEPGRegular, OliveOil, Plane and Trace, were considered. In this experiment, only the datasets producing an accuracy greater than 0.7 in at least one of the procedures used were taken into account, and in all situations, the performance of K-means clustering was superior to that given by the K-medoids procedure.

Table 1 shows the resulting classification accuracy in terms of the *Sim* values for the K-means, KmALSCAL, and mclust procedures in full dimension when compared to the true classification. As can be seen, the model-based clustering procedure showed very poor results on all datasets. On the contrary, the K-means clustering procedure on the classical and ALSCAL solutions showed similar results in full dimension.

To investigate the influence that a dimensionality reduction in the auxiliary configuration could have on the precision of the three procedures, the data have been analyzed when the auxiliary configuration preserves 60%, 70%, 80%, and 90% of the variability of the exact solution. For a fixed percentage of variability, the selected dimension has been the smallest value  $p$ , so that the sum of the  $p$  eigenvalues ordered from largest to smallest of the matrix  $B = -0.5HD^2H$ , divided by the total sum, is at least the indicated percentage. Table 2 shows the results obtained for the UCR datasets, showing

**Table 2.** Similarity index values (3.1) for the UCR datasets using the K-means, KmALSCAL and, mclust procedures when the dimension for the auxiliary configuration explains the 60%, 70%, 80%, and 90% of the variability according to classic MDS.

	K-means				KmALSCAL				mclust			
	60%	70%	80%	90%	60%	70%	80%	90%	60%	70%	80%	90%
Simulated series of unequal length												
GesturePebbleZ1(TEST)	0.70	0.67	0.66	0.67	0.67	0.68	0.65	0.65	0.28	0.28	0.28	0.28
GesturePebbleZ2(TRAIN)	0.72	0.69	0.65	0.70	0.78	0.76	0.76	0.76	0.28	0.28	0.28	0.28
Simulated series of equal length												
CBF	0.72	0.69	0.65	0.70	0.78	0.76	0.76	0.76	0.28	0.28	0.28	0.28
Mallat(TRAIN)	0.87	0.85	0.87	0.85	0.95	0.95	0.95	0.95	0.23	0.51	0.25	0.22
SmoothSubspace	0.94	0.94	0.91	0.91	0.94	0.94	0.94	0.94	0.50	0.50	0.50	0.50
SyntheticControl	0.83	0.92	0.77	0.91	0.98	0.98	0.97	0.97	0.29	0.29	0.29	0.29
TwoPatterns(TRAIN)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.40	0.40	0.40	0.40
Real data sets with series of equal length												
BasicMotions	0.78	0.75	0.75	0.75	0.78	0.78	0.78	0.78	0.72	0.64	0.62	0.40
DiatomSizeReduction	0.78	0.77	0.77	0.77	0.76	0.76	0.76	0.76	0.39	0.39	0.39	0.39
DistalPOAgeGroup	0.69	0.69	0.69	0.70	0.68	0.68	0.68	0.68	0.46	0.46	0.46	0.46
FaceFour	0.73	0.73	0.73	0.73	0.73	0.73	0.73	0.73	0.40	0.40	0.40	0.40
InsectEPGRegular	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.48	0.48	0.49	0.49
OliveOil	0.74	0.80	0.80	0.80	0.74	0.74	0.74	0.74	0.77	0.80	0.65	0.39
Plane	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.25	0.25	0.25	0.25
Trace	0.54	0.55	0.54	0.54	0.53	0.53	0.53	0.54	0.40	0.40	0.40	0.40

few differences between the dimensions tested when the classic or ALSCAL configurations are used, these being somewhat more in favor of the KmALSCAL procedure for the set of simulated time series. As can be appreciated, the mclust procedure does not seem to perform well in any of the dimensions tested. This is an interesting finding, since the assumptions of a normal mixture distribution based on the classical MDS configuration has been indirectly assumed by several models. On the contrary, except for the Trace dataset, for which the ALSCAL procedure did not show good results even in full dimension, the K-means procedure showed good behavior in reduced dimensionality, both for the classical configuration and for that of ALSCAL, with the KmALSCAL procedure showing slightly better performance.

### 3.2. Simulated time series of unequal length

We also analyzed with the extended procedures the 125 datasets in a Monte Carlo experiment similar to that performed in Vera and Angulo [16], each one with 20 series belonging to two groups of 10 time series related to a stochastic process and length of the corresponding pairs. The datasets consisted on the simulation of 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 situation, a set of twenty series are thus analyzed in which there are two groups, one for each process. 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 a similar type and autocorrelation function structure. In the second case, a nonstationary process and a near-nonstationary autoregressive (AR) process are compared. In the third, we compare the performance of two selected second-order autoregressive moving-average (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.

**Table 3.** Average value of similarity indices for the classifications obtained with K-means, KmALSCAL, and mclust for the simulated time series. For each pair of processes (Pairs), twenty series were generated for each pair of lengths, thus ten time series per process and length. For each pair of processes and each combination of two lengths, the twenty corresponding series were analyzed. The experiment was repeated ten times.

Length Pairs	K-means					KmALSCAL					mclust				
	(100)	(200)	(300)	(400)	(500)	(100)	(200)	(300)	(400)	(500)	(100)	(200)	(300)	(400)	(500)
AR1-AR1	0.73	0.98	1.00	1.00	1.00	0.75	0.97	1.00	1.00	1.00	0.66	0.66	0.66	0.83	0.93
AR1-ARI	0.66	0.66	0.67	0.69	0.67	0.65	0.68	0.67	0.68	0.67	0.66	0.66	0.66	0.66	0.66
AR2-MA2	0.63	1.00	1.00	1.00	1.00	0.62	1.00	1.00	1.00	1.00	0.66	0.66	0.66	0.69	0.66
ARF-ARI	0.60	1.00	1.00	1.00	1.00	0.61	1.00	1.00	1.00	1.00	0.66	0.66	0.66	0.66	0.66
ARM-ARI	0.62	1.00	1.00	1.00	1.00	0.60	1.00	1.00	1.00	1.00	0.66	0.66	0.66	0.66	0.66
Length Pairs	(200)														
AR1-AR1	0.73	0.83	0.99	1.00	1.00	0.78	0.87	0.99	1.00	1.00	0.66	0.66	0.66	0.66	0.90
AR1-ARI	0.67	0.67	0.67	0.66	0.67	0.69	0.66	0.67	0.67	0.67	0.66	0.66	0.66	0.66	0.66
AR2-MA2	1.00	0.66	0.88	1.00	1.00	1.00	0.69	0.85	1.00	1.00	0.66	0.66	0.66	0.66	0.66
ARF-ARI	1.00	0.65	1.00	1.00	1.00	1.00	0.64	1.00	1.00	1.00	0.66	0.66	0.66	0.66	0.66
ARM-ARI	1.00	0.59	0.97	1.00	1.00	1.00	0.59	0.96	1.00	1.00	0.66	0.66	0.66	0.66	0.66
Length Pairs	(300)														
AR1-AR1	0.78	0.81	0.92	0.99	1.00	0.81	0.85	0.94	0.98	1.00	0.66	0.66	0.66	0.66	0.69
AR1-ARI	0.67	0.67	0.67	0.67	0.67	0.69	0.68	0.67	0.68	0.68	0.66	0.66	0.66	0.66	0.66
AR2-MA2	1.00	1.00	0.69	0.73	0.99	1.00	1.00	0.70	0.72	0.99	0.69	0.66	0.66	0.65	0.65
ARF-ARI	1.00	0.99	0.61	0.95	1.00	1.00	0.99	0.61	0.94	1.00	0.66	0.66	0.66	0.65	0.66
ARM-ARI	1.00	0.98	0.60	0.95	1.00	1.00	0.98	0.59	0.96	1.00	0.66	0.66	0.66	0.66	0.66
Length Pairs	(400)														
AR1-AR1	0.83	0.89	0.91	0.97	1.00	0.86	0.90	0.91	0.97	1.00	0.66	0.66	0.66	0.66	0.66
AR1-ARI	0.72	0.71	0.72	0.72	0.72	0.74	0.73	0.73	0.73	0.73	0.65	0.66	0.66	0.66	0.66
AR2-MA2	1.00	1.00	1.00	0.67	0.67	1.00	1.00	1.00	0.72	0.66	0.72	0.66	0.66	0.66	0.66
ARF-ARI	1.00	1.00	0.94	0.62	0.84	1.00	1.00	0.95	0.61	0.89	0.66	0.66	0.66	0.66	0.66
ARM-ARI	1.00	1.00	0.90	0.63	0.87	1.00	1.00	0.94	0.61	0.89	0.66	0.66	0.66	0.66	0.66
Length Pairs	(500)														
AR1-AR1	0.89	0.94	0.92	0.97	0.99	0.87	0.94	0.92	0.97	1.00	0.66	0.66	0.66	0.66	0.66
AR1-ARI	0.68	0.65	0.65	0.65	0.65	0.73	0.67	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66
AR2-MA2	1.00	1.00	1.00	1.00	0.68	1.00	1.00	1.00	1.00	0.75	1.00	0.66	0.66	0.65	0.65
ARF-ARI	1.00	1.00	1.00	0.85	0.62	1.00	1.00	1.00	0.85	0.63	0.66	0.66	0.66	0.66	0.66
ARM-ARI	1.00	1.00	1.00	0.85	0.61	1.00	1.00	1.00	0.88	0.62	0.66	0.66	0.66	0.66	0.66

For each of the five pairs of processes,  $5 \times 5$  sets of time series corresponding to all combinations of the sizes were analyzed. This procedure was repeated ten times, and therefore, 1250 sets of time series were analyzed using the proposed K-means, the KmALSCAL, and mclust procedures. The resulting classifications were then compared with the original. Table 3 shows the given results in full dimension for the three tested procedures. Again, mclust showed the worst results in all combinations of lengths and pairs of processes, except in the comparisons between a nonstationary process and a near-nonstationary AR process, for which all procedures, in general, offered low results. This is consistent with the fact that for those processes, the MDS configuration obtained from the dynamic time warping dissimilarities showed a large degree of overlap between the two groups and, as is well known, K-means performance decreases in the presence of overlap between groups [9]. Regarding the performance of K-means when using ALSCAL or the classic MDS solution in full dimension, both procedures offered very similar results in all comparisons, with the performance being somewhat higher the more the lengths of the series of both groups differ and slightly better for KmALSCAL.

**Table 4.** Average value of the resulting similarity indices for K-means, KmALSCAL and mclust, for the simulated pairs of time series in the dimensions that explain 60%, 70%, 80%, 90% and 100% of the variability according to classical MDS. The length of the first process is set to 100.

Length Pairs	K-means					KmALSCAL (100)					mclust				
	(100)	(200)	(300)	(400)	(500)	(100)	(200)	(300)	(400)	(500)	(100)	(200)	(300)	(400)	(500)
<b>AR1-AR1</b>															
60%	0.73	0.97	1.00	1.00	1.00	0.76	0.97	1.00	1.00	1.00	0.86	0.96	0.97	0.98	1.00
70%	0.73	0.97	1.00	1.00	1.00	0.75	0.97	1.00	1.00	1.00	0.72	0.92	0.97	1.00	1.00
80%	0.73	0.97	1.00	1.00	1.00	0.75	0.97	1.00	1.00	1.00	0.65	0.72	0.97	0.97	1.00
90%	0.72	0.98	1.00	1.00	1.00	0.75	0.97	1.00	1.00	1.00	0.66	0.68	0.90	0.97	1.00
100%	0.73	0.98	1.00	1.00	1.00	0.75	0.97	1.00	1.00	1.00	0.67	0.72	0.69	0.83	0.93
<b>AR1-ARI</b>															
60%	0.64	0.67	0.67	0.69	0.68	0.64	0.67	0.67	0.67	0.67	0.65	0.65	0.68	0.78	0.72
70%	0.64	0.67	0.67	0.68	0.66	0.65	0.67	0.67	0.67	0.67	0.65	0.66	0.65	0.66	0.65
80%	0.64	0.67	0.66	0.67	0.66	0.64	0.68	0.68	0.68	0.67	0.66	0.66	0.65	0.66	0.66
90%	0.65	0.68	0.67	0.68	0.68	0.64	0.69	0.69	0.69	0.68	0.66	0.66	0.66	0.66	0.66
100%	0.66	0.66	0.67	0.69	0.67	0.65	0.68	0.67	0.68	0.67	0.66	0.66	0.66	0.66	0.66
<b>AR2-MA2</b>															
60%	0.62	1.00	1.00	1.00	1.00	0.61	1.00	1.00	1.00	1.00	0.64	0.77	0.99	1.00	1.00
70%	0.63	1.00	1.00	1.00	1.00	0.64	1.00	1.00	1.00	1.00	0.64	0.67	0.92	0.98	0.99
80%	0.62	1.00	1.00	1.00	1.00	0.63	1.00	1.00	1.00	1.00	0.65	0.66	0.72	0.79	0.90
90%	0.63	1.00	1.00	1.00	1.00	0.63	1.00	1.00	1.00	1.00	0.65	0.66	0.66	0.66	0.72
100%	0.62	1.00	1.00	1.00	1.00	0.61	1.00	1.00	1.00	1.00	0.65	0.69	0.66	0.69	0.69
<b>ARF-ARI</b>															
60%	0.61	1.00	1.00	1.00	1.00	0.61	1.00	1.00	1.00	1.00	0.62	0.68	1.00	1.00	1.00
70%	0.61	1.00	1.00	1.00	1.00	0.61	1.00	1.00	1.00	1.00	0.64	0.65	0.83	0.86	0.99
80%	0.59	1.00	1.00	1.00	1.00	0.60	1.00	1.00	1.00	1.00	0.65	0.65	0.69	0.72	0.79
90%	0.59	1.00	1.00	1.00	1.00	0.60	1.00	1.00	1.00	1.00	0.66	0.66	0.66	0.66	0.66
100%	0.60	1.00	1.00	1.00	1.00	0.61	1.00	1.00	1.00	1.00	0.66	0.72	0.66	0.66	0.69
<b>ARM-ARI</b>															
60%	0.64	1.00	1.00	1.00	1.00	0.61	1.00	1.00	1.00	1.00	0.63	0.75	1.00	1.00	0.98
70%	0.61	1.00	1.00	1.00	1.00	0.60	1.00	1.00	1.00	1.00	0.65	0.65	0.86	0.93	0.98
80%	0.63	1.00	1.00	1.00	1.00	0.60	1.00	1.00	1.00	1.00	0.66	0.65	0.72	0.76	0.83
90%	0.62	1.00	1.00	1.00	1.00	0.60	1.00	1.00	1.00	1.00	0.66	0.65	0.66	0.66	0.72
100%	0.62	1.00	1.00	1.00	1.00	0.60	1.00	1.00	1.00	1.00	0.66	0.66	0.69	0.66	0.69

We again investigate the influence of dimensionality on the clustering procedure when the auxiliary configuration is considered in low dimension. For each dataset, the analysis was performed considering the corresponding auxiliary configuration in the dimension that explains the 60%, 70%, 80%, 90%, and 100% of the variability according to the exact representation, as indicated for the UCR datasets. For the sake of preserving space, we show here only the tables with the length size combinations with 100, 300, and 500 (all the given results are available as supplementary material).

**Table 5.** Average value of the resulting similarity indices for K-means, KmALSCAL and mclust, for the simulated pairs of time series in the dimensions that explain 60%, 70%, 80%, 90% and 100% of the variability according to classical MDS. The length of the first process is set to 300.

Length Pairs	K-means					KmALSCAL (300)					mclust				
	(100)	(200)	(300)	(400)	(500)	(100)	(200)	(300)	(400)	(500)	(100)	(200)	(300)	(400)	(500)
<b>AR1-AR1</b>															
60%	0.79	0.79	0.93	0.99	1.00	0.81	0.85	0.94	0.98	1.00	0.95	0.98	1.00	0.98	0.99
70%	0.79	0.81	0.93	0.99	1.00	0.81	0.85	0.94	0.98	1.00	0.87	0.97	0.99	0.97	0.99
80%	0.78	0.81	0.92	0.99	1.00	0.81	0.85	0.94	0.98	1.00	0.72	0.85	0.75	0.91	0.90
90%	0.79	0.81	0.92	0.99	1.00	0.81	0.85	0.94	0.98	1.00	0.66	0.66	0.66	0.65	0.69
100%	0.78	0.80	0.92	0.99	1.00	0.81	0.83	0.94	0.98	1.00	0.66	0.65	0.66	0.66	0.72
<b>AR1-ARI</b>															
60%	0.66	0.65	0.65	0.65	0.65	0.67	0.65	0.65	0.67	0.67	0.67	0.67	0.72	0.68	0.75
70%	0.66	0.65	0.65	0.65	0.65	0.68	0.66	0.65	0.67	0.67	0.67	0.66	0.66	0.65	0.66
80%	0.66	0.67	0.65	0.65	0.65	0.67	0.67	0.65	0.67	0.67	0.66	0.66	0.66	0.66	0.66
90%	0.67	0.66	0.65	0.65	0.65	0.69	0.66	0.66	0.68	0.69	0.66	0.66	0.66	0.66	0.66
100%	0.66	0.67	0.67	0.67	0.66	0.67	0.68	0.65	0.68	0.68	0.70	0.66	0.71	0.66	0.69
<b>AR2-MA2</b>															
60%	1.00	1.00	0.72	0.76	0.99	1.00	1.00	0.72	0.75	0.99	1.00	0.72	0.67	0.67	0.89
70%	1.00	1.00	0.70	0.75	0.99	1.00	1.00	0.70	0.76	0.99	0.97	0.65	0.66	0.65	0.68
80%	1.00	1.00	0.70	0.76	0.99	1.00	1.00	0.70	0.74	0.98	0.83	0.65	0.65	0.65	0.65
90%	1.00	1.00	0.71	0.72	0.99	1.00	1.00	0.70	0.70	0.98	0.72	0.66	0.66	0.66	0.66
100%	1.00	1.00	0.68	0.73	0.99	1.00	1.00	0.68	0.72	0.99	0.72	0.66	0.65	0.65	0.65
<b>ARF-ARI</b>															
60%	1.00	0.99	0.61	0.96	1.00	1.00	0.99	0.60	0.96	1.00	1.00	0.67	0.66	0.63	0.64
70%	1.00	0.99	0.61	0.94	1.00	1.00	0.99	0.59	0.95	1.00	0.91	0.65	0.65	0.65	0.65
80%	1.00	0.99	0.59	0.95	1.00	1.00	0.99	0.60	0.95	1.00	0.69	0.65	0.65	0.65	0.66
90%	1.00	0.99	0.61	0.94	1.00	1.00	0.99	0.59	0.95	1.00	0.66	0.66	0.65	0.66	0.66
100%	1.00	0.99	0.60	0.95	1.00	1.00	0.99	0.61	0.94	1.00	0.79	0.66	0.66	0.65	0.66
<b>ARM-ARI</b>															
60%	1.00	0.98	0.59	0.94	1.00	1.00	0.98	0.59	0.94	1.00	1.00	0.66	0.62	0.64	0.66
70%	1.00	0.98	0.58	0.96	1.00	1.00	0.98	0.59	0.96	1.00	0.86	0.65	0.65	0.65	0.65
80%	1.00	0.98	0.59	0.95	1.00	1.00	0.98	0.58	0.96	1.00	0.72	0.66	0.65	0.66	0.65
90%	1.00	0.98	0.58	0.94	1.00	1.00	0.98	0.59	0.96	1.00	0.66	0.65	0.66	0.66	0.65
100%	1.00	0.98	0.60	0.95	1.00	1.00	0.98	0.59	0.96	1.00	0.72	0.66	0.66	0.66	0.69

Table 4 shows the results of the comparison index  $Sim$  when the length of the simulated time series for the first process of each pair was set to 100, which represents the first situation of greater discrepancy in the length of the pairs. The performance of the K-means procedure when the auxiliary matrix is estimated in low dimension offered the same efficiency for all the pairs of tested processes, even when only the 60% of the variability was explained by the auxiliary configuration. This means the proposed K-means procedure of Vera and Angulo [16] is robust against the dimensionality of the solution. This

is an interesting finding in relation to the proposed procedure, since dimensionality reduction is one of the most used procedures when performing time series clustering [1]. On the other hand, the mclust procedure offered good results when applied to the MDS configuration in low dimension, and much better than those obtained in full dimension. In general, its performance decreases as the number of dimensions increases, which is in line with the well-known decrease in performance of model-based clustering procedures in high-dimensional applications, as a result of over-parameterization [34].

**Table 6.** Average value of the resulting similarity indices for K-means, KmALSCAL and mclust, for the simulated pairs of time series in the dimensions that explain 60%, 70%, 80%, 90% and 100% of the variability according to classical MDS. The length of the first process is set to 500.

Length Pairs	K-means					KmALSCAL (500)					mclust				
	(100)	(200)	(300)	(400)	(500)	(100)	(200)	(300)	(400)	(500)	(100)	(200)	(300)	(400)	(500)
<b>AR1-AR1</b>															
60%	0.88	0.94	0.92	0.97	0.99	0.87	0.94	0.92	0.97	0.99	0.99	1.00	1.00	1.00	0.99
70%	0.89	0.94	0.92	0.97	0.99	0.87	0.94	0.91	0.97	1.00	0.96	1.00	1.00	0.99	1.00
80%	0.89	0.94	0.92	0.97	0.99	0.87	0.94	0.92	0.97	1.00	0.86	0.93	0.83	0.87	0.83
90%	0.89	0.94	0.92	0.97	0.99	0.87	0.94	0.92	0.97	1.00	0.66	0.69	0.69	0.69	0.66
100%	0.89	0.94	0.92	0.97	0.99	0.87	0.94	0.91	0.96	1.00	0.66	0.68	0.68	0.72	0.66
<b>AR1-ARI</b>															
60%	0.68	0.65	0.64	0.64	0.64	0.72	0.67	0.66	0.66	0.66	0.70	0.71	0.75	0.77	0.71
70%	0.68	0.66	0.64	0.64	0.64	0.72	0.67	0.66	0.66	0.66	0.65	0.71	0.68	0.67	0.71
80%	0.67	0.65	0.65	0.65	0.64	0.72	0.68	0.66	0.66	0.65	0.66	0.66	0.65	0.65	0.68
90%	0.67	0.66	0.65	0.64	0.64	0.73	0.67	0.66	0.67	0.66	0.66	0.66	0.66	0.66	0.66
100%	0.68	0.65	0.65	0.65	0.65	0.73	0.67	0.66	0.67	0.66	0.66	0.66	0.68	0.65	0.67
<b>AR2-MA2</b>															
60%	1.00	1.00	1.00	1.00	0.69	1.00	1.00	1.00	1.00	0.73	1.00	0.99	0.80	0.74	0.74
70%	1.00	1.00	1.00	1.00	0.70	1.00	1.00	1.00	1.00	0.74	1.00	0.96	0.71	0.69	0.68
80%	1.00	1.00	1.00	1.00	0.70	1.00	1.00	1.00	1.00	0.73	1.00	0.76	0.65	0.65	0.65
90%	1.00	1.00	1.00	1.00	0.68	1.00	1.00	1.00	1.00	0.70	0.97	0.72	0.66	0.65	0.65
100%	1.00	1.00	1.00	1.00	0.68	1.00	1.00	1.00	1.00	0.75	1.00	0.72	0.66	0.65	0.65
<b>ARF-ARI</b>															
60%	1.00	1.00	1.00	0.79	0.60	1.00	1.00	1.00	0.82	0.61	1.00	0.90	0.68	0.64	0.64
70%	1.00	1.00	1.00	0.85	0.62	1.00	1.00	1.00	0.85	0.61	0.97	0.79	0.64	0.65	0.65
80%	1.00	1.00	1.00	0.85	0.61	1.00	1.00	1.00	0.82	0.64	0.83	0.66	0.66	0.66	0.66
90%	1.00	1.00	1.00	0.85	0.61	1.00	1.00	1.00	0.85	0.62	0.69	0.66	0.66	0.66	0.66
100%	1.00	1.00	1.00	0.85	0.62	1.00	1.00	1.00	0.85	0.63	0.69	0.69	0.66	0.66	0.66
<b>ARM-ARI</b>															
60%	1.00	1.00	1.00	0.83	0.60	1.00	1.00	1.00	0.88	0.62	1.00	0.95	0.65	0.65	0.62
70%	1.00	1.00	1.00	0.84	0.59	1.00	1.00	1.00	0.88	0.64	1.00	0.79	0.66	0.66	0.65
80%	1.00	1.00	1.00	0.85	0.62	1.00	1.00	1.00	0.89	0.61	0.79	0.69	0.66	0.66	0.66
90%	1.00	1.00	1.00	0.83	0.61	1.00	1.00	1.00	0.88	0.62	0.69	0.66	0.66	0.66	0.66
100%	1.00	1.00	1.00	0.85	0.62	1.00	1.00	1.00	0.88	0.63	0.76	0.66	0.66	0.66	0.65

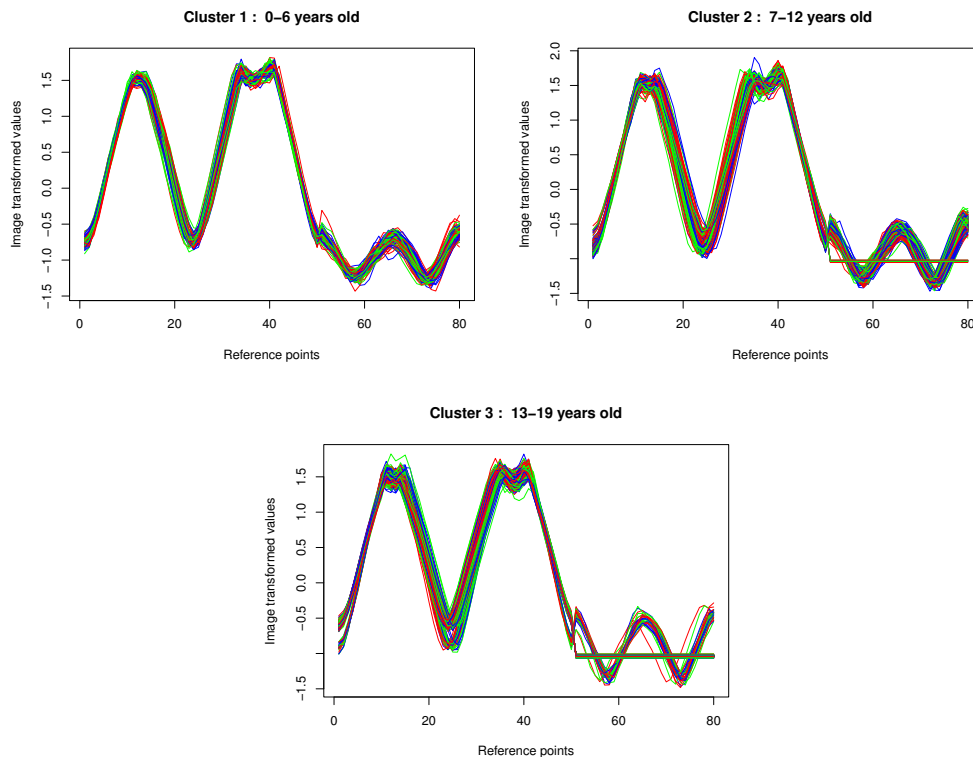
Table 5 shows the results given when the first group is formed by series of length 300. In general, the differences in lengths between the time series in both groups are the most similar in the experiment, with the groups by time series of generally large lengths. Again, the results when using classic MDS or ALSCAL in K-means are practically the same when the configurations are considered in full dimension or in reduced dimension, explaining only 60% of the variability. As in previous situations, the differences are more evident for the mclust procedure, which generally decreases its performance

as the dimension increases. This discrepancy is more evident when comparing the group of series of length 300 with that of time series of length 100, while when compared to the group of length 500, the results are not as good. Therefore, it is observed that mclust offers worse results both for large dimensions and when the length of the series in both groups is medium or large.

Table 6 shows the results when one of the groups is conformed by the larger times series of length 500. For dispersions up to 70%, mclust showed good results, sometimes slightly better than K-means in the first three pairs of processes analyzed, its performance decreasing as the dimension increases. The two K-means procedures, as before, did not show significant differences regarding their performance in low dimension in all analyzed datasets, being good in general in all situations.

### 3.3. Hand images

To illustrate the performance of the three clustering procedures a well known set of time series generated from hand images was analyzed [35]. The original problem was to study the effectiveness of hand and bone contour detection in predicting bone age [36]. In particular, we analyze the proximal phalanx contour age group dataset, which is focused on the problem of using the contour of one of the phalanges to predict whether the subject belongs to one of three age groups: 0–6 years, 7–12 years, and 13–19 years old. This dataset is also available in the UCR time series archive. The dataset has a total size of 605 instances, with all time series being of equal length 80. Figure show the three times series groups in this dataset.



**Figure 3.** Representation of the proximal phalanx time series for the age groups.

First, the  $605 \times 605$  dissimilarity matrix was calculated using dynamic time warping, and classical MDS was performed in full dimension 603, then, K-means clustering was performed for  $K = 3$  clusters on the given MDS configuration, following the procedure of Vera and Angulo [16]. A *Sim* value of 0.77 was obtained when comparing the given partition with the real one, then K-means was conducted for  $K = 3$  on the configuration obtained using ALSCAL, applying it directly to the dynamic time squared dissimilarity dissimilarities  $\Delta^2$ . The classification obtained was similar to that obtained using the classic MDS configuration, with a value of *Sim* = 0.771. Finally, the Gaussian mixture model on the classical MDS configuration in full dimension was estimated, obtaining the best results for the model with covariance matrix diagonal and equal volume and shape in the three clusters, after all the possible models estimated by the *Mclust* function were analyzed. However, a poor solution compared to the true classification was obtained, for a *Sim* value of 0.49.

The analysis was performed also in the four dimensions 303, 373, 446, and 520, related to the percentages of estimated variability of 60%, 70%, 80%, and 90% of the auxiliary matrix  $X$ . For the K-means procedure using the classical MDS solution and the ALSCAL solution, similar classification results were found at all dispersion values, again indicating that the procedure is robust to dimensionality reduction. In contrast, the *mclust* solution was practically the same in all dimensions, showing bad behavior. To investigate the low performance of the model based clustering procedure, the Gaussian mixture was also estimated for  $K = 1, \dots, 9$  components (the default option in the *Mclust* function). The best solution found was that of a single group in all dimensions tested, which indicates the poor performance of the model-based clustering model when dealing with a high-dimensional dataset.

Finally, we analyzed the differences in the K-means clustering results both using the classical and the ALSCAL multidimensional scaling configurations. Both procedures obtained identical results for the three clusters except for the time series number 132, classified in different clusters by both procedures, being classified in the right cluster when using ALSCAL.

#### 4. Discussion

In this paper, we revisit the problem of K-means clustering in time series of equal or unequal length using dynamic time warping dissimilarities [16]. The classical method is carried out in full dimension, based on the equivalence in K-means between the partition of the dissimilarity matrix and the partition of an auxiliary configuration estimated by means of classical MDS. The auxiliary configuration is thus oriented in the direction of the principal axes while exactly preserving the K-means partition in the dynamic time warping dissimilarities. Here, the robustness of the procedure is discussed in terms of an approximate configuration, estimated by alternating least squares from the squared dynamic time warping dissimilarities (with additive constant) using ALSCAL, as well as using model-based clustering in the classical solution.

In addition, since K-means may perform poorly for a large number of dimensions, we investigate the robustness of the procedure under dimensionality reduction, as this is one of the most used and controversial strategies when performing time series clustering [1]. The main motivation comes from the fact that this model should not be greatly affected by a dimension reduction, as long as the representation of the time series is adequately preserved, which facilitates its application on large datasets. Given that the classical configuration is nested and therefore the possible low-dimensional approximations are linear projections of the full dimensional solution, the use of an alternative auxiliary approximation in low



dimension in a nonlinear framework is proposed using ALSCAL. Alternatively, the relationship between K-means clustering, the Gaussian distribution in MDS [23] and clustering based on mixtures of Gaussian distributions [25], make it necessary to analyze the behavior when performing model-based clustering based on the classical auxiliary configuration.

The three aspects are discussed therefore, a nonlinearly estimated auxiliary configuration alternative to the classic MDS solution using ALSCAL, the dimensionality reduction, and a related probabilistic estimation procedure that would allow performing inference, for instance, to select the number of clusters. A Monte Carlo experiment similar to that described by Vera and Angulo [16] has been performed for time series of equal and unequal lengths, now considering, in addition to the classical MDS solution, the performance of model-based clustering in this configuration, and the efficiency of the method for the estimated configuration using ALSCAL. Furthermore, the influence of the dimensionality of the auxiliary configuration on the recovery of the clusters is studied, analyzing four scenarios that explain 60, 70, 80, 90 and 100 percent of the total variability of the configuration that accurately represents the time series. In general, in all the scenarios analyzed, the performance of the K-means clustering, both for the classical and for the ALSCAL configurations, was good and similar in both procedures in full dimension, although some differences were found in low dimension in favor of the KmALSCAL procedure.

Specifically, for the UCR datasets, small differences were found for the solutions obtained along the different dimensions tested. In these datasets, in general, the model-based clustering procedure showed very poor results, which showed the performance of this procedure being greatly affected by large dimensionality problems. On the contrary, the K-means procedure showed a generally good performance when using the classical and the ALSCAL auxiliary configurations, also for reduced dimensionality (except in one dataset), these being somewhat in favor of the KmALSCAL procedure. For the simulation experiment, the K-means clustering again showed similar results for both the classical and ALSCAL configurations, with performance increasing as the differences in time series length between each group increased, and somewhat a favor of the KmALSCAL procedure. In terms of the dimensionality, similar efficiency was found for both K-means procedure, showing almost the same results in all dimensions tested. For model-based clustering, the best results were obtained in the lower dimension and much better than those obtained in the full dimension, and, in general, they offered worse results when the series length in both groups was medium or large, showing a bad behavior in high dimensionality.

Finally, the performance in recovering the right partition of a well-known image-related time series dataset was analyzed using all the three procedures and scenarios for  $K = 3$  clusters. While K-means clustering showed the best results for both configurations, the model based clustering performance was poor in any dimension. The consequence of a higher dimensionality, even in the minor dispersion scenery (303 dimensions), was that only one cluster was the estimated response when the procedure was run from one to nine clusters, in all dimension tested. Differences in the classification for K-means using the classical or the ALSCAL solutions was only for the time series in row 132, which was well classified using KmALSCAL.

In summary, the performance of KmALSCAL was good in all scenarios compared to K-means clustering using the classical configuration, with the results being somewhat better than the latter in reduced dimension for some datasets. Given that both procedures used the Euclidean distances obtained from the translated squared dissimilarities, the good performance of KmALSCAL was more evident

in low dimension, since the classical solution is restricted to a linear projection. Therefore, both procedures were robust to dimension reduction, since the estimated configuration in reduced dimension is still informative about the true underlying structure of the data, unlike the typical performance of a two-step clustering procedure [20]. In addition to preserving the partition in the dissimilarities, the proposed KmALSCAL procedure also enables the possibility of estimating the auxiliary configuration considering weights for the time series and usual transformations in MDS, in this case, of the squared Euclidean distances, which allows including constraints in the configuration.

Although an elastic measure of dissimilarity such as dynamic time warping allows comparisons between series of different lengths, the presence of a large amount of missing data influences the observed values. It is interesting to investigate the behavior of these procedures when other dissimilarity measures are used and/or the series are of equal length, which is currently being investigated by the authors. Furthermore, the use of ALSCAL to estimate the auxiliary configuration also allows considering a non-metric MDS procedure based on isotonic regression, which is less restrictive since it only preserves the order of the original dissimilarities and therefore it does not alter the distances for K-means. The performance of such a model is also currently being investigated by the authors.

### Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

### Acknowledgments

This work has been partially supported by grant B-CTS-184-UGR20 funded by ERDF, EU / Ministry of Economic Transformation, Industry, Knowledge and Universities of Andalusia (J. F. Vera); grant PID2021-126095NB-100 funded by Spanish Ministry of Science and Innovation, MCIN/AEI/10.13039/501100011033 and by “ERDF A way of making Europe” (J. A. Roldán).

### Conflict of interest

The authors declare no conflict of interest.

### References

1. S. Aghabozorgi, A. Shirkorshidi, T. Wah, Time-series clustering—A decade review, *Inf. Syst.*, **53** (2015), 16–38, <https://doi.org/10.1016/j.is.2015.04.007>
2. W. Liao, Clustering of time series data—A survey, *Pattern Recognit.*, **38** (2005), 1857–1874. <https://doi.org/10.1016/j.patcog.2005.01.025>
3. H. Li, J. Tong, A novel clustering algorithm for time-series data based on precise correlation coefficient matching in the IoT, *Math. Biosci. Eng.*, **16** (2019), 6654–6671. <https://doi.org/10.3934/mbe.2019331>
4. S. Policker, A. B. Geva, Nonstationary time series analysis by temporal clustering, *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, **30** (2000), 339–343. <https://doi.org/10.1109/3477.836381>

5. C. Goutte, P. Toft, E. Rostrup, F. A. Nielsen, L. K. Hansen, On clustering fMRI time series, *Neuroimage*, **9** (1999), 298–310. <https://doi.org/10.1006/nimg.1998.0391>
6. N. Subhani, L. Rueda, A. Ngom, C. J. Burden, Multiple gene expression profile alignment for microarraytime-series data clustering, *Bioinformatics*, **26** (2010), 2281–2288. <https://doi.org/10.1093/bioinformatics/btq422>
7. J. McQueen, Some methods for classification and analysis of multivariate observations, in *Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. II*, (eds. M. Lucien, C. Le, N. Jerzy), Statistical Laboratory of the University of California, Berkeley, (1967), 281–297.
8. J. D. Banfield, A. E. Raftery, Model-based Gaussian and non-Gaussian clustering, *Biometrics*, **49** (1993), 803–821. <https://doi.org/10.2307/2532201>
9. B. S. Everitt, S. Landau, M. Leese, D. Stahl, *Cluster analysis*, 5th edition, Wiley series in probability and statistics, Wiley, Chichester, 2011. <https://doi.org/10.1002/9780470977811>
10. H. H. Bock, Model-based clustering methods for time series, in *German-Japanese Interchange of Data Analysis Results. Studies in Classification, Data Analysis, and Knowledge Organization*, (eds. W. Gaul, A. Geyer-Schulz, Y. Baba, A. Okada), Springer, Cham, (2013), 3–12. [https://doi.org/10.1007/978-3-319-01264-3\\_1](https://doi.org/10.1007/978-3-319-01264-3_1)
11. P. Montero, J. Vilar, TSclust: An R package for time series clustering, *J. Stat. Softw.*, **62** (2014), 1–43. <https://doi.org/10.18637/jss.v062.i01>
12. P. Ortega-Jiménez, M. A. Sordo, A. Suárez-Llorens, Stochastic comparisons of some distances between random variables, *Mathematics*, **9** (2021), 981. <https://doi.org/10.3390/math9090981>
13. J. F. Vera, Clustering and representation of time series. Application to dissimilarities based on divergences, in *Trends in Mathematical, Information and Data Sciences. Studies in Systems, Decision and Control*, (eds. N. Balakrishnan, M. A. Gil, N. Martín, D. Morales, M. C. Pardo), Springer, Cham, **445** (2023), 243–251. [https://doi.org/10.1007/978-3-031-04137-2\\_22](https://doi.org/10.1007/978-3-031-04137-2_22)
14. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Science and Business Media, New York, (2009). <https://doi.org/10.1007/978-0-387-84858-7>
15. J. F. Vera, R. Macías, On the behaviour of K-means clustering of a dissimilarity matrix by means of full multidimensional scaling, *Psychometrika*, **89** (2021), 489–513. <https://doi.org/10.1007/s11336-021-09757-2>
16. J. F. Vera, J. M. Angulo, An MDS-based unifying approach to time series K-means clustering: application in the dynamic time warping framework, *Stoch. Environ. Res. Risk Assess.*, **37** (2023), 4555–4566. <https://doi.org/10.1007/s00477-023-02470-9>
17. L. Kaufman, P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley Series in Probability and Statistics, Wiley, Hoboken, NJ, USA, 1990. <https://doi.org/10.1002/9780470316801>
18. J. C. Lingo, Some boundary conditions for a monotone analysis of symmetric matrices, *Psychometrika*, **36** (1971), 195–203. <https://doi.org/10.1007/BF02291398>
19. D. Steinley, K-means clustering: A half-century synthesis, *Br. J. Math. Stat. Psychol.*, **59** (2006), 1–34, <https://doi.org/10.1348/000711005X48266>

20. M. Vichi, H. A. L. Kiers, Factorial K-means analysis for two-way data, *Comput. Stat. Data Anal.*, **37** (2001), 49–64. [https://doi.org/10.1016/S0167-9473\(00\)00064-5](https://doi.org/10.1016/S0167-9473(00)00064-5)
21. Y. Takane, F. W. Young, J. de Leeuw, Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features, *Psychometrika*, **42** (1977), 7–67. <https://doi.org/10.1007/BF02293745>
22. R. Bailey, J. Gower, Approximating a symmetric matrix, *Psychometrika*, **55** (1990), 665–675. <https://doi.org/10.1007/BF02294615>
23. R. A. Hefner, *Extension of the Law of Comparative Judgment to Discriminable and Multidimensional Stimuli*, PhD. thesis, University of Michigan, 1958.
24. J. L. Zinnes, D. B. Mackay, Probabilistic multidimensional scaling: Complete and incomplete data, *Psychometrika*, **48** (1983), 27–48. <https://doi.org/10.1007/BF02314675>
25. M. S. Oh, A. E. Raftery, Model-based clustering with dissimilarities: A Bayesian approach, *J. Comput. Graph. Stat.*, **16** (2007), 559–585. <https://doi.org/10.1198/106186007X236127>
26. T. Giorgino, Computing and visualizing dynamic time warping alignments in R: The dtw package, *J. Stat. Softw.*, **31** (2009), 1–24. <https://doi.org/10.18637/jss.v031.i07>
27. J. F. Vera, C. D. Rivera, A structural equation multidimensional scaling model for one-mode asymmetric dissimilarity data, *Struct. Equation Modell. Multidiscip. J.*, **21** (2014), 54–62. <https://doi.org/10.1080/10705511.2014.85669>
28. J. F. Vera, P. Mair, SEMDS: An R package for structural equation multidimensional scaling, *Struct. Equation Modell. Multidiscip. J.*, **26** (2019), 803–818. <https://doi.org/10.1080/10705511.2018.1561292>
29. K. V. Mardia, Some properties of classical multi-dimensional scaling, *Commun. Stat.- Theory Methods*, **7** (1978), 1233–1241. <https://doi.org/10.1080/03610927808827707>
30. Y. Chen, B. H. Keogh, N. Begum, A. Bagnall, A. Mueen, G. Batista, *The UCR Time Series Classification Archive*, 2015. Available from: <http://www.timeseriesclassification.com/index.php>.
31. T. Rusch, J. de Leeuw, L. Chen, P. Mair, *smacofx: Flexible Multidimensional Scaling and 'smacof' Extensions. R Package Version 0.6-6*, 2003. Available from: <https://CRAN.R-project.org/package=smacofx>.
32. L. Scrucca, M. Fop, T. B. Murphy, A. E. Raftery, mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models, *R J.*, **8** (2016), 289–317. <https://doi.org/10.32614/RJ-2016-021>
33. M. Gavrilov, D. Anguelov, P. Indyk, R. Motwani, Mining the stock market: Which measure is best, in *Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining (KDD'00)*, (2000), 487–496. <https://doi.org/10.1145/347090.347189>
34. C. Bouveyron, C. Brunet-Saumard, Model-based clustering of high-dimensional data: A review, *Comput. Stat. Data Anal.*, **71** (2013), 52–78. <https://doi.org/10.1016/j.csda.2012.12.008>
35. L. Davis, *Predictive Modelling of Bone Ageing*, PhD. thesis, University of East Anglia, UK, 2013.
36. A. Bagnall, L. Davis, Predictive modelling of bone age through classification and regression of bone shapes, preprint, arXiv:1406.4781v1, 2014. <https://doi.org/10.48550/arXiv.1406.4781>

- 
37. A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood estimation from incomplete data via the EM algorithm, *J. R. Stat. Soc. B*, **39** (1977), 1–38. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>



AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)