Clustering and Geodesic Scaling of Dissimilarities on the Spherical Surface

J. Fernando Vera, Ricardo Subiabre, and Rodrigo Macías

Spherical embedding is an important tool in several fields of data analysis, including environmental data, spatial statistics, text mining, gene expression analysis, medical research and, in general, areas in which the geodesic distance is a relevant factor. Many data acquisition technologies are related to massive data acquisition, and these high-dimensional vectors are often normalised and transformed into spherical data. In this representation of data on spherical surfaces, multidimensional scaling plays an important role. Traditionally, the methods of clustering and representation have been combined, since the precision of the representation tends to decrease when a large number of objects are involved, which makes interpretation difficult. In this paper, we present a model that partitions objects into classes while simultaneously representing the cluster centres on a spherical surface based on geodesic distances. The model combines a partition algorithm based on the approximation of dissimilarities to geodesic distances with a representation procedure for geodesic distances. In this process, the dissimilarities are transformed in order to optimise the radius of the sphere. The efficiency of the procedure described is analysed by means of an extensive Monte Carlo experiment, and its usefulness is illustrated for real data sets.

Supplementary material to this paper is provided online.

Key Words: Multidimensional scaling; Dissimilarity; Geodesic distance; Cluster analysis; Spherical surface.

1. INTRODUCTION

Given a dissimilarity matrix between N objects, multidimensional scaling (MDS) aims to estimate a configuration of points in a low-dimensional space p, such that the distances between them, usually Euclidean, are as close as possible to the observed dissimilarities between the objects. Since the seminal work of Sampson and Guttorp (1992), MDS plays an important role in spatial deformation modelling and estimation techniques, particularly in

J. Fernando Vera (✉) Department of Statistics and O.R. Faculty of Sciences, University of Granada, 18071 Granada, Spain
(E-mail: *jfvera@ugr.es*). R. Subiabre, Centro de Docencia Superior en Ciencias Básicas, Universidad Austral de Chile, Puerto Montt, Chile. R.Macías, Centro de Investigación en Matemáticas, Unidad de Monterrey, Monterrey, Mexico.

nonparametric approaches to the analysis of the covariance structure of the spatiotemporal processes underlying environmental studies. However, in applications involving the problem of estimating the covariance structure, the localisation of environmental phenomena on the surface of the sphere rather than the planar representation is of increasing interest, allowing us to capture both the spatiotemporal dependencies between observations, as well as the development of accurate predictions about planet Earth (see, for example, Alegría et al. 2018).

Although one of the main goals of MDS is to facilitate the visual interpretation of relationships between objects, when the number of objects is large, this representation of the results tends to be rather confusing, making it difficult to identify structures in the data. In addition, the approximation error between the dissimilarities and the estimated distances increases with the number of points, in particular in low dimensions. This problem is usually addressed by the joint application of MDS and cluster analysis. However, while the latter approach is optimal in the original space, this may not be the case in a reduced space, where performing cluster analysis and then representing the objects in a low-dimensional space can lead to interpretation errors (Heiser and Groenen 1997). To enhance the interpretation of the MDS solution and/or to obtain an adequate fit of the model when the number of objects is too large, cluster-MDS models have proven useful, both in the classical context and in the least squares framework (Bock 1986, 1987; Heiser 1993; Heiser and Groenen 1997; Vera et al. 2008). In a probabilistic framework, Vera et al. (2009a,b) have proposed latent class multidimensional scaling models for dissimilarity data in Euclidean spaces.

In some experimental situations, it is desirable to impose constraints on the MDS configuration to best represent the singularities between the relations of objects, usually obtaining a representation in a specific parametric space. In particular, constraints related to quadratic surfaces such as the sphere arise, for example, in the analysis of data for dissimilarities between geographical regions, or distances between cities (or countries), or in the measurement of large-scale environmental phenomena and/or those affecting different locations on the planet, and in general when the representation of the MDS solution in three dimensions tends to have a spherical shape. Other applications related to quadratic surface embedding have been proposed as parabolas or ellipses, but in the present case we are interested in spherical surfaces. On the other hand, a different situation from the one we are dealing with here is when the true configuration of the objects is known in high dimensionality and the objective is to embed it in the surface of the sphere. In these cases, in addition to local MDS, there are other very efficient procedures that allow the representation of points while preserving local structures, such as SRCA (Luo et al. 2023), t-SNE (Van Der Maaten and Hinton 2008) or UMAP (McInnes et al. 2018). However, it is important to note that these methods do not perform clustering, and in general, are not designed to preserve global structures (Wang et al. 2021).

Let us consider as a running example some study data concerning the monthly mean near-surface air temperature dataset (Harris et al. 2020a,b) of the Climatic Research Unit of the University of East Anglia (CRU TS 4.02). A random sample of 500 locations was analysed from the total of 67,420 locations, during the $T = 12$ months of 2017. The complete database can be found at Harris et al. (2020a) Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. Here we focus on the study of the spherical spatial

representation of locations based on their dissimilarities related to temperature information using MDS, in particular, on the combined use of clustering and spherical MDS to facilitate the interpretation of the representation and reduce the number of parameters to be estimated. In addition, the estimation accuracy of a spatial deformation procedure using spherical spline interpolation is analysed in relation to the representation error in MDS.

Our analysis considers the situation in which the MDS configuration must fall on a spherical surface. This case requires a more appropriate measure of the distance between points in this framework than the Euclidean one, in order to approximate the dissimilarities. One such measure is the geodesic distance, that is, the length of the shortest geodesic along the surface. There exists a monotonic relation between Euclidean and geodesic distances, and various approaches to this problem have been proposed (see, for example, De Leeuw and Mair 2009, for further details). One of the most used methods for this purpose consists of applying MDS with quadratic restrictions in the configuration, with both Euclidean (MDS-Q) and geodesic (geodesic MDS-Q) distances. This model can be viewed as a weakly constrained MDS (Borg and Groenen 2005).

Gnanadesikan (1977) proposed a two-step method to impose constraints in MDS as a form of nonlinear component analysis, and Bookstein (1979) and Fitzgibbon et al. (1999), later proposed improvements to this method. In general, in the two-step procedure, first an unrestricted MDS solution is found and then the best quadratic surface is fitted. In a direct approach for Euclidean distances with spherical constraints on the configuration, Bentler and Weeks (1978) applied Gauss–Newton methods with linear constraints, while De Leeuw and Heiser (1980) developed a general theory of MDS with restrictions on the configuration. For the geodesic distance, Cox and Cox (1991) proposed circular and spherical MDS non-metric models, and Elad et al. (2005) solved the metric spherical MDS problem for geodesic distances, minimising the stress by means of a gradient method with line search.

De Leeuw and Mair (2009) proposed a majorisation-based methodology for quadratic surfaces in a least squared framework, based on two approaches, the primal or quadratic multidimensional scaling (Q-MDS) method and the dual method. In the first, the quadratic constraints are incorporated in parametric form directly into the loss function based on Euclidean distances, while in the second, the constraints are imposed at convergence by means of penalty or Lagrangian terms. Various dual methods have been proposed by Borg and Lingoes (1980) in which the constraints are directly imposed on the distances.

However, regardless of the estimation procedure used, the problem of achieving visualisation with a large number of objects is even more acute in the case of a spherical representation. This problem was addressed by Dzwinel et al. (2005), who used multi-resolutional clustering and nonlinear MDS separately to cluster the time events in the feature space and inspected the resulting clustering structures using three-dimensional MDS. For example, in Lopes et al. (2014), dissimilarities were obtained between fifty subjectively chosen zones or clusters, with the intention of investigating the behaviour of more than three million seismic occurrences around the world, and the clusters were represented by MDS.

The usefulness of a model that enables clusters to be estimated while the centre of each cluster is represented on the sphere is thus evident in a wide variety of practical applications. This approach is interesting, for instance, for the problem of estimating non-stationary spatial covariance when a broad domain of the spherical surface is involved (see for example Vera

et al. 2008, 2009a). In this type of applications, reducing the representation to clusters while preserving the structure between them also contributes to reducing the number of underlying parameters to be estimated, in addition to facilitating interpretation. This also helps to avoid known problems such as those of estimation in oversampled domains (Kovitz and Christakos 2004) or reduce the appearance of non-injective mappings as a result of a "nugget effect" or folding problems (Sampson and Guttorp 1992; Vera et al. 2008, 2009a).

Additionally, since the dimensionality in this particular MDS model is fixed a priori, the use of criteria to select the number of clusters directly from the dissimilarity matrix is also advisable (see Vera and Macías 2017).

In the present paper, we propose a model that, given a dissimilarity matrix between a set of objects, obtains its classification in homogeneous clusters while simultaneously, not the objects themselves but the centres of the clusters are represented on a spherical surface using the geodesic distance. The parameter estimation is performed using an alternating estimation procedure for which, given a classification, the cluster centres are represented using the well-known monotonic relation between the geodesic and the Euclidean distances. For the representation step, we propose a Q-MDS approach for geodesic distances that estimates the configuration in the sphere surface, in combination with a ratio transformation that allows us to estimate the optimal radius of the sphere assuming that it is centred at the origin.

In the next section, we formulate the combined clustering and spherical representation model. Section 3 then describes the alternating estimation procedure together with the criterion applied to determine the number of clusters. In Sect. 4, we analyse the behaviour of the model, based on a Monte Carlo experiment, and in Sect. 5 its performance for empirical data is illustrated and compared with that of a two-step estimation procedure based on K-means clustering and a quadratic constraints MDS algorithm. In addition, the adjustment of a spatial deformation between the geographic configuration and that of the MDS by means of spherical splines is illustrated, and the interpolation and location errors of the estimated points on the sphere are compared, with or without clustering. Finally, we discuss the results obtained and present the main conclusions drawn.

2. THE CLUSTERING SPHERICAL SCALING (CSS) MODEL

Let us denote by $\mathbf{\Delta}$ a dissimilarity matrix between N objects o_i, with entries δ_{ij}, $i, j = 1, \ldots, N$. It is assumed that the objects are grouped into K disjoint clusters, and thus, we denote by $\mathbf{E}$ a matrix of $N \times K$ binary e_{ik} entries, where $e_{ik} = 1$ if the ith object belongs to the kth cluster, and zero otherwise.

We denote by $\mathbf{X}$ a configuration of K points representing the cluster centres on the surface of a sphere of radius μ, and denote by d_{kl}, the Euclidean distance between the points $\mathbf{x}_k$, $\mathbf{x}_l \in \mathcal{S}(\mu)$, and by $\breve{d}_{kl}$, the related geodesic great-circle distance given by

$$\breve{d}_{kl} = \breve{d}(\mathbf{x}_k, \mathbf{x}_l) = \mu \arccos\left(\frac{\mathbf{x}_k' \mathbf{x}_l}{\mu^2}\right) = \mu \arccos\left(\frac{2\mu^2 - d_{kl}^2}{2\mu^2}\right). \tag{1}$$

Like the Euclidean distance, the geodesic distance suffers from scale indeterminacy, since for $\mathbf{x}_k, \mathbf{x}_l \in \mathcal{S}(\mu)$, and $b > 0$, $b\breve{d}(\mathbf{x}_k, \mathbf{x}_l) = \breve{d}(b\mathbf{x}_k, b\mathbf{x}_l)$, where $b\mathbf{x}_k, b\mathbf{x}_l \in \mathcal{S}(b\mu)$, both distances being monotonically related. Without loss of generality, we can consider weights w_{ij} for the dissimilarities, which also enables us to deal with missing data. The loss function (stress) in this model is given by

$$\sigma_\leq (\mathbf{X}, \mathbf{E}) = \sum_{k \leq l} \sum_{i=1}^{N} \sum_{j=1}^{N} e_{ik} e_{jl} \omega_{ij} \left(\delta_{ij} - \breve{d}_{kl} (\mathbf{X}) \right)^2, \tag{2}$$

and our aim, hence, is to minimise Eq. (2) in terms of a classification $\mathbf{E}$ and a configuration $\mathbf{X}$ on the surface of a sphere of radius μ.

According to least squares orthogonality (Heiser and Groenen 1997), the stress function Eq. (2) can be decomposed considering geodesics distances as follows:

$$\sigma_\leq (\mathbf{X}, \mathbf{E}) = \sum_{k \leq l} \sum_{i=1}^{N} \sum_{j=1}^{N} e_{ik} e_{jl} \omega_{ij} \left(\delta_{ij} - \widetilde{\delta}_{kl} \right)^2 + \sum_{k \leq l} \widetilde{\omega}_{kl} \left(\widetilde{\delta}_{kl} - \breve{d}_{kl} (\mathbf{X}) \right)^2 \tag{3}$$

where $\widetilde{\delta}_{kl}$ is the Sokal–Michener dissimilarity (Sokal and Michener 1958) given by

$$\widetilde{\delta}_{kl} = \sum_{i=1}^{N} \sum_{j=1}^{N} e_{ik} e_{jl} \frac{\omega_{ij} \delta_{ij}}{\widetilde{\omega}_{kl}} \quad , \text{ whith } \quad \widetilde{\omega}_{kl} = \sum_{i=1}^{N} \sum_{j=1}^{N} e_{ik} e_{jl} \omega_{ij}. \tag{4}$$

The first term in Eq. (3) depends only on the classification $\mathbf{E}$, while the last term depends on both the classification and the representation. Therefore, the parameter estimation in this model can be performed in an alternating least squares procedure.

3. PARAMETER ESTIMATION

The algorithm consists of two stages that alternate iteratively between an allocation step and a representation step, until the convergence criterion holds. The iterative cycle continues until the difference between two consecutive values of the stress function Eq. (2) is below a small value previously set by the investigator. The algorithm starts with an initial classification (allocation step). This can be set by the investigator, or taken randomly. Then, the Sokal–Michener dissimilarities are calculated, and the related configuration in a sphere for the cluster centres is estimated (representation step). The alternating procedure continues iteratively until the loss function is minimised (see Fig. 1). The two estimation phases are described in detail below.

3.1. ALLOCATION STEP

Given an estimated value for $\mathbf{X} \in \mathcal{S}(\mu)$, we classify each object into the cluster to which the corresponding dissimilarities are nearest to the related geodesic distances between the

corresponding cluster centres. To this end, the following loss function is minimised (Heiser and Groenen 1997),

$$\min_E \kappa^2 (E|\mathbf{X}, \mathbf{\Delta}) = \sum_i \sum_k e_{ik} \|\mathbf{a}_i - \mathbf{b}_k^{(i)}\|^2, \tag{5}$$

where $\|a_i - b_k^{(i)}\|^2$ denotes the squared Euclidean distance between the ith row of the matrix $\mathbf{A} = \{a_{ij}\}$ of order $N \times (N-1)$ and the kth row of the matrix $\mathbf{B}^{(i)} = \{b_{kr}^{(i)}\}$ of order $K \times (N-1)$, and with elements of $\mathbf{A}$ and $\mathbf{B}^{(i)}$, $i = 1, \ldots, n$, which is specified as $a_{ir} = \delta_{ir}^*$ and $b_{kr}^{(i)} = \breve{d}_{kr}^*$, where $\delta_{ir}^* = \delta_{is}$ and $\breve{d}_{kr}^* = \sum_l e_{sl} \breve{d}_{kl} (\mathbf{X})$, for $r = 1, \ldots, N-1$, with $s = r$ if $r < i$, and $s = r + 1$ if $r \geq i$ and where e_{lk} is a binary variable which is equal to one if object i is an element of cluster k in the matrix $\mathbf{E}$.

The optimal classification stage consists of a nested iterative cycle of NK iterations, after which a new classification matrix $\mathbf{E}$ is obtained, from which the weights $\widetilde{\omega}_{kl}$, and the Sokal–Michener dissimilarities $\widetilde{\delta}_{kl}$ are updated. Heiser and Groenen (1997) have shown that the assignment step can be viewed as a K-means clustering procedure. A simple convergent algorithm is possible if we proceed row by row finding

$$\min_{\mathbf{e}_i} \sum_k e_{ik} \|\mathbf{a}_i - \mathbf{b}_k^{(i)}\|^2, \tag{6}$$

being $\mathbf{e}_i = (e_{i1}, \ldots, e_{iK})'$, the ith row of matrix $\mathbf{E}$ and maintaining the allocation of the other objects $j \neq i$, fixed. The minimum value is achieved in some row κ of $B^{(i)}$ such that if κ does not change the assignment, the algorithm moves to the next object, otherwise object i is reallocated and the ith row of $\mathbf{E}$ is adjusted first; then the next object is considered until the position of all objects have been examined, concluding the allocation phase.

3.2. Geodesic MDS-Q Step

For a given classification $\mathbf{E}$, the corresponding Sokal–Michener dissimilarities $\widetilde{\delta}_{kl}$ are calculated, and the configuration for the cluster centres is estimated in a *geodesic MDS-Q step* by minimising,

$$\min_{\mathbf{X}, \mu} \sum_{k<l} \widetilde{\omega}_{kl} \left(\widetilde{\delta}_{kl} - \mu \breve{d}_{kl} (\mathbf{X}) \right)^2 \quad, \quad \text{with } \mathbf{X} \in \mathcal{S}(1). \tag{7}$$

The configuration $\mathbf{X}$ is estimated using Euclidean distances in the smacof framework using the Guttman transform, after which the points are projected onto the surface of a sphere for an optimal radius μ (De Leeuw and Mair 2009). Here, we present a refinement of the MDS-Q algorithm, considering the estimation of the optimal radius of the sphere, not in terms of the projection from Euclidean space to the surface of the sphere, but to minimise the stress function Eq. (2) at each step. This task is performed by an alternating estimation procedure in the MDS-Q step that estimates the radius as the slope of a linear transformation of the MDS-Q geodesic distances, given its linear invariance property. The algorithm at this

representation step can be summarised as follows (see also De Leeuw and Mair 2009, for further details).

1. Given $\mathbf{E}^{(s)}$ at the sth iteration in the geodesic step, take $\mathbf{Z} = \mathbf{Y}^{(s)}$ the classical MDS solution in three dimensions for the Sokal–Michener dissimilarities.

2. Calculate the Guttman transform which improves the configuration in terms of Euclidean distances minimising the second terms in (2). $\mathbf{Y}^{(s)} = \mathbf{V}^{+}\mathbf{B}(\mathbf{Z})\mathbf{Z}$, where $\mathbf{V}^{+}$ is the Moore–Penrose inverse of $\mathbf{V}$, $\mathbf{V}^{+} = \left(\mathbf{V} + N^{-1}\mathbf{1}\mathbf{1}'\right)^{-1} + N^{-1}\mathbf{1}\mathbf{1}'$, where $\mathbf{V}$ is given by $\mathbf{V} = \sum_{k<l} \widetilde{w}_{kl}\mathbf{M}_{kl}$, with $\mathbf{M}_{kl} = (\mathbf{m}_k - \mathbf{m}_l)(\mathbf{m}_k - \mathbf{m}_l)'$, with elements $m_{kk} = m_{ll} = 1$, $m_{kl} = m_{lk} = -1$, and 0 elsewhere, and $\mathbf{B}$ is given by $\mathbf{B}(\mathbf{Z}) = \sum_{k<l} \widetilde{w}_{kl}\, s_{kl}(\mathbf{Z})\,\mathbf{M}_{kl}$ with $s_{kl}(\mathbf{Z}) = \widetilde{\delta}_{kl}/d_{kl}(\mathbf{Z})$ if $d_{kl}(\mathbf{Z}) > 0$ and $s_{kl}(\mathbf{Z}) = 0$ if $d_{kl}(\mathbf{Z}) = 0$.

3. Obtain the projection $\mathbf{X}^{(s)} \in \mathcal{S}(\lambda^{(s)})$ of $\mathbf{Y}^{(s)}$,

$$\min_{\lambda^{(s)}, \mathbf{X}^{(s)} \in \mathcal{S}(\lambda^{(s)})} tr(\mathbf{X}^{(s)} - \mathbf{Y}^{(s)})'\mathbf{V}(\mathbf{X}^{(s)} - \mathbf{Y}^{(s)}).$$

Then consider $\mathbf{X}^{(s)} = \mathbf{X}^{(s)}/\lambda^{(s)} \in \mathcal{S}(1)$ (see De Leeuw and Mair 2009 for details on convergence).

4. Calculate $\breve{d}(\mathbf{X}^{(s)})$ the geodesic distances Eq. (1), and estimate $\mu^{(s)}$ (see Appendix) by minimising (7), at the fixed value of $\mathbf{X}^{(s)}$.

5. Set $\mathbf{X}^{(s)} = \mu\mathbf{X}^{(s)}$, and evaluate the stress $\sigma^{(s)}(\mathbf{X}^{(s)}|\mathbf{E})$.

6. For $s > 0$, and $\varepsilon > 0$ small enough, if $(\sigma^{(s-1)} - \sigma^{(s)}) \leq \varepsilon$, then $\mathbf{X} = \mathbf{X}^{(s)} \in \mathcal{S}(\mu^{(s)})$ is found and the algorithm stop. Otherwise, update the iteration index and return to 2.

Finally, the geodesic distances $\breve{d}(\mathbf{X})$ are calculated and the algorithm continues with the allocation step, minimising Eq. (2). The overall alternating estimation procedure converges to a local minimum. The allocation step is a nested iterative loop of N assignments (a total of NK iterations), while the MDS step is essentially a K-order ratio MDS procedure, for which the Guttman transformation can be seen as a steepest descent step with a fixed stepsize parameter (see de Leeuw 1988, for details on the convergence of smacof).

3.3. Selecting the Number of Clusters

Several criteria can be employed to select the number of clusters for a dissimilarity matrix (see Vera and Macías 2017, 2021). Here, we consider the adapted version of Hartigan's criterion H^* proposed by Vera and Macías (2017), which has been experimentally shown to obtain good results even when some degree of overlap is present. Since the clustering is optimal in the entire space, the criterion is used directly on the original dissimilarity matrix without imposing geometric restrictions. This is defined by

$$H^*(K) = \left[\frac{W^*(K)}{W^*(K+1)} - 1\right] (([N(N-1) - K(K+1)]/2) - 1), \tag{8}$$

Input values: Dissimilarity matrix $\mathbf{\Delta}_{N \times N}$ and K
Initialise parameters: $t_1 = 0$, $\mathbf{X}_{K \times 3}^{(t_1)} \in \mathcal{S}(1)$, $\mathbf{E}^{(t_1)}$, ϵ_1
Calculate: $\sigma_{\leq}^2(\mathbf{X}^{(t_1)})$ by Eq. (2)
Repeat:

$(t_1 = t_1 + 1)$
Allocation Step:
update $\mathbf{E}^{(t_1)}$ by Eq. (6)
MDS-Q Step:
Initialise parameters:
$t_2 = 0$, cluster dissimilarity matrix $\widetilde{\mathbf{\Delta}}_{K \times K}^{(t_2)} \leftarrow \mathbf{E}^{(t_1)}$, $\mathbf{X}^{(t_2)} \in \mathcal{S}(\mu^{(t_2)})$, ϵ_2
Calculate: $\sigma^2(\mathbf{X}^{(t_2)})$ by Eq. (7)
Repeat:

$(t_2 = t_2 + 1)$
Calculate: $\bar{\mathbf{X}}^{(t_2)} = \mathbf{V}^+ B\left(\mathbf{X}^{(t_2-1)}\right)\mathbf{X}^{(t_2-1)}$
and obtain: $\mathbf{X}^{(t_2)} \in \mathcal{S}(\lambda^{(t_2)})$ of $\bar{\mathbf{X}}^{(t_2)}$ by **Projection step** (see Appendix).
Normalise: $\mathbf{X}^{(t_2)^*} = \mathbf{X}^{(t_2)}/\lambda^{(t_2)} \in \mathcal{S}(1)$
Radius adjustment:
Adjust: $\mathbf{X}^{(t_2)^*}$ and $\mu^{(t_2)}$ by Eq. (7)
and update: $\mathbf{X}^{(t_2)} \leftarrow \mu^{(t_2)}\mathbf{X}^{(t_2)^*}$
Evaluate: $\sigma^2(\mathbf{X}^{(t_2)})$ by Eq. (7)
Until: $|\sigma^2(\mathbf{X}^{(t_2)}) - \sigma^2(\mathbf{X}^{(t_2-1)})| < \epsilon_2$
Update: $\mathbf{X}^{(\mathbf{t_1})} = \mathbf{X}^{(t_2)} \in \mathcal{S}(\mu^{(t_2)})$
Evaluate: $\sigma_{\leq}^2(\mathbf{X}^{(t_1)})$ by Eq. (2)
Until: $|\sigma_{\leq}^2(\mathbf{X}^{(t_1-1)}) - \sigma_{\leq}^2(\mathbf{X}^{(t_1)})| < \epsilon_1$ or $(t_1 = t_{1\max})$

Figure 1.   Pseudocode of the CSS Model algorithm.

where,

$$W^*(K) = \sum_{k \leq l} \sum_{i=1}^{N} \sum_{j=1}^{N} e_{ik} e_{jl} w_{ij} \left(\delta_{kl} - \widetilde{\delta}_{kl}\right)^2, \tag{9}$$

and where $\widetilde{\delta}_{kl}$ is given by Eq. (4). According to the selection rule proposed by Vera and Macías (2017), the values of $H^*(t)$ are calculated for $t = 1, \ldots, T$, where T is usually predetermined by the investigator. Then, the estimated number of clusters is the smallest value $K \leq T$ such that $H^*(K) \leq 5N$.

4. MONTE CARLO EXPERIMENT

To test the performance of the proposed model, artificially clustered data sets were generated in the unit sphere, with mixtures of the well-known von Mises–Fisher distribution (Banerjee et al. 2005). To this end, we used the rmixvfm function of the Directional R package (v.6.0; Tsagris et al. 2022). All statistical analyses were performed in R v.4.1.3 (R Development Core Team 2023), working on an Intel(R) Core(TM) i5-6200U CPU 2.40 GHz computer with 6 GB of RAM.

The data sets were generated considering a structure of $K = 4, 6, 8, 10$ clusters with equal prior probabilities in the unit sphere. The distribution of points within each cluster was selected considering different concentration indices $k_c = 6, 3, 1.5, 0.75$ for the mixtures of von Mises–Fisher distributions, taking into account that the higher the concentration value, the more separated the clusters will be from each other, and the lower this value, the more

Table 1. Average values of ARI, congruence coefficient (CC) and CPU time (T) from simulation experiment with equal probability for each cluster, these being well separated ($k_c = 6$) and moderately separated ($k_c = 3$)

K	N	$k_c = 6$				$k_c = 3$			
		50	100	250	500	50	100	250	500
4	ARI	0.983	0.982	0.983	0.981	0.868	0.858	0.895	0.879
	CC	0.945	0.952	0.932	0.955	0.950	0.946	0.946	0.953
	T	0.196	0.797	0.634	2.665	0.402	0.839	0.665	3.135
6	ARI	0.991	0.995	0.984	0.989	0.891	0.895	0.916	0.926
	CC	0.950	0.953	0.942	0.958	0.953	0.954	0.946	0.953
	T	0.290	0.259	1.030	2.387	0.610	0.265	1.441	3.581
8	ARI	0.910	0.952	0.960	0.962	0.811	0.880	0.902	0.886
	CC	0.928	0.929	0.931	0.933	0.925	0.930	0.936	0.929
	T	0.330	0.246	1.240	3.169	0.594	0.495	0.981	6.048
10	ARI	0.815	0.846	0.875	0.898	0.715	0.734	0.763	0.768
	CC	0.886	0.888	0.897	0.897	0.897	0.900	0.887	0.894
	T	0.568	0.662	1.201	6.190	0.549	0.729	3.349	9.079

the clusters will overlap (see, for example, Chevallier et al. 2022 for alternative distributions on the sphere surface. For each combination of the above factors, ten clustered sets of points were generated in the unit sphere, for each of the sizes $N = 50, 100, 250, 500$. Hence, a total of 640 data sets were analysed with the proposed model for the true number of clusters, and the results obtained were compared to the original ones in terms of the classification using the Adjusted Rand Index (ARI) of Hubert and Arabie (1985). The Tucker congruence coefficient (Tucker 1951) was calculated between the simulated data sets and the outcome of the Procrustes transformation on the estimated configuration. For both indices, a value close to one indicates good performance.

Table 1 shows the averaged values for each set of ten simulated matrices, for the ARI, the congruence index and the CPU time in seconds, according to the sample size, number of clusters and the larger concentration index values of $k_c = 6, 3$ producing non-overlapping clusters. Almost all of these averaged values indicate excellent performance by the proposed procedure, both in terms of the quality of the classification obtained (ARI) and of recovering the cluster centres (CC). In general, as the size increased and the number of clusters decreased, the model performed better, as expected. In terms of CPU time, the procedure was very efficient for all data sets analysed, with the time increasing as the size of the data set and the number of clusters grew.

Since the performance of K-means clustering is known to decrease for overlapping clusters, we investigated the performance of the proposed procedure when a degree of overlapping was present. Table 2 shows the corresponding averaged results for the lower concentration index values of $k_c = 1.5, 0.75$, which are related to clusters that are somewhat overlapping. In general, the average ARI values decreased as the number of clusters K increased, and the concentration index k_c decreased, as expected. However, the congruence coefficient (CC index) values remained high, since the average values of dissimilarities between clusters did not seem to be greatly altered by the misclassified points.

Table 2. Average values of ARI, congruence coefficient (CC) and CPU time (T) from simulation experiment with equal probability for each cluster, these being somewhat overlapping ($k_c = 1.5$) and very overlapping ($k_c = 0.75$)

K	N	$k_c = 1.5$				$k_c = 0.75$			
		50	100	250	500	50	100	250	500
4	ARI	0.645	0.644	0.630	0.653	0.345	0.330	0.392	0.361
	CC	0.958	0.959	0.959	0.947	0.968	0.953	0.965	0.967
	T	0.257	0.970	3.171	6.254	2.097	0.359	4.388	6.898
6	ARI	0.643	0.682	0.738	0.716	0.421	0.381	0.399	0.383
	CC	0.957	0.956	0.954	0.959	0.947	0.957	0.963	0.945
	T	0.352	0.375	2.954	3.797	3.064	0.365	0.945	15.028
8	ARI	0.601	0.670	0.714	0.702	0.379	0.366	0.364	0.411
	CC	0.929	0.933	0.932	0.928	0.941	0.939	0.940	0.937
	T	0.287	0.399	1.301	4.804	0.403	0.424	4.372	6.654
10	ARI	0.511	0.555	0.556	0.556	0.334	0.337	0.366	0.349
	CC	0.911	0.901	0.907	0.900	0.906	0.912	0.908	0.907
	T	0.561	0.912	2.467	8.405	0.284	3.690	8.733	7.852

Table 3. Average values of ARI, congruence coefficient (CC) and CPU time (T) from simulation experiment with unequal probabilities for each cluster, these being well separated ($k_c = 6$) and moderately separated ($k_c = 3$)

K	N	$k_c = 6$				$k_c = 3$			
		50	100	250	500	50	100	250	500
4	ARI	0.953	0.984	0.978	0.971	0.838	0.863	0.834	0.840
	CC	0.950	0.947	0.955	0.959	0.953	0.944	0.948	0.953
	T	0.155	0.264	0.675	2.305	0.152	0.400	4.138	3.365
6	ARI	0.908	0.965	0.988	0.989	0.831	0.887	0.903	0.930
	CC	0.950	0.950	0.956	0.949	0.945	0.965	0.954	0.955
	T	0.145	0.335	0.733	2.947	0.402	2.008	1.092	4.327
8	ARI	0.828	0.901	0.935	0.927	0.734	0.769	0.849	0.848
	CC	0.928	0.935	0.939	0.930	0.920	0.925	0.925	0.929
	T	0.339	0.664	1.939	4.896	1.206	0.832	2.891	9.676
10	ARI	0.779	0.860	0.871	0.835	0.653	0.653	0.714	0.708
	CC	0.888	0.884	0.885	0.893	0.892	0.899	0.882	0.885
	T	0.625	2.300	4.073	8.100	0.884	0.938	2.365	5.341

We have also investigated the performance of the model when the groups are unbalanced, for which the two previous experiments were carried out but now considering unequal probabilities for the simulation of the data. For each value of k, the kth group ($k = 1, \ldots, K$) was weighted using the values $p_k = k / \sum_k k$. Tables 3 and 4 show the given results. As can be appreciated, in general, the performance of the procedure was good when the clusters are well separated, without much influence of the different sizes of the clusters, becoming poorer for the ARI coefficient as the degree of overlap increases, as expected.

Finally, the convergence rates of the model were tested for three large data sets to also analyse the scalability of the model, considering the values of $K = 15$, $k_c = 2$ and $N = 500, 1000, 2000, 4000, 6000$. Figure 2 shows the convergence plots in terms of the

Table 4. Average values of ARI, congruence coefficient (CC) and CPU time (T) from simulation experiment with unequal probabilities for each cluster, these being somewhat overlapping ($k_c = 1.5$) and very overlapping ($k_c = 0.75$)

K	N	$k_c = 1.5$				$k_c = 0.75$			
		50	100	250	500	50	100	250	500
4	ARI	0.504	0.606	0.518	0.540	0.251	0.293	0.260	0.268
	CC	0.971	0.956	0.956	0.951	0.960	0.964	0.951	0.950
	T	5.873	0.470	3.098	4.766	2.666	6.302	15.899	4.689
6	ARI	0.638	0.631	0.691	0.657	0.404	0.403	0.370	0.404
	CC	0.945	0.947	0.950	0.952	0.956	0.957	0.953	0.945
	T	0.167	0.379	2.639	12.844	0.357	0.619	1.321	4.500
8	ARI	0.634	0.606	0.616	0.612	0.326	0.415	0.425	0.396
	CC	0.927	0.923	0.929	0.924	0.934	0.933	0.938	0.930
	T	0.562	1.130	12.171	6.608	2.002	1.411	2.125	6.486
10	ARI	0.527	0.504	0.522	0.539	0.375	0.376	0.410	0.371
	CC	0.895	0.897	0.903	0.901	0.902	0.918	0.905	0.902
	T	1.072	1.157	1.532	7.577	0.445	3.741	3.864	24.357

normalised stress value given for each main iteration, starting from the second iteration to better appreciate the differences. In terms of iterations, the stress value decreased rapidly for the overall iterative procedure, even for the largest data sets, indicating that most of the work is done in each of the stages of the alternating estimation process. The above experiment was repeated ten times, and the average execution times (rounded) were of 14, 42, 189, 1094 and 2932 s, respectively. In terms of CPU time, the cost increases as data size increases, as expected, although this is still competitive for large data sets. It seems evident that part of the efficiency of the procedure is due to the well-known speed of convergence towards a local minimum of the smacof procedure (see, for example, De Leeuw and Mair 2009), which also here is performed only for $K << N$ points.

5. ILLUSTRATIVE EXAMPLE

We now analyse the time series data of temperature measured at 500 locations introduced in Sect. 1. We first show the results of our new approach, the CSS model. We then show the results obtained using a two-step procedure that first performs clustering and then plots the cluster centres on the sphere. Finally, we illustrate the fitting behaviour of the spatial deformation estimation procedure using smoothing spline regression between geographic locations and the MDS spherical representation.

To perform MDS, the dissimilarities between time series within the sample were obtained using the cosine correlation-based dissimilarity measure. Considering the Pearson correlation coefficient between two time series $\mathbf{x}_i = \{x_{i1}, \ldots, x_{iT}\}$ and $\mathbf{x}_j = \{x_{j1}, \ldots, x_{jT}\}$,

$$\rho\left(\mathbf{x}_i, \mathbf{x}_j\right) = \frac{\sum_{t=1}^{T} \left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right)\left(\mathbf{x}_{jt} - \bar{\mathbf{x}}_j\right)}{\sqrt{\sum_{t=1}^{T} \left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right)^2}\sqrt{\sum_{t=1}^{T} \left(\mathbf{x}_{jt} - \bar{\mathbf{x}}_j\right)^2}}, \tag{10}$$
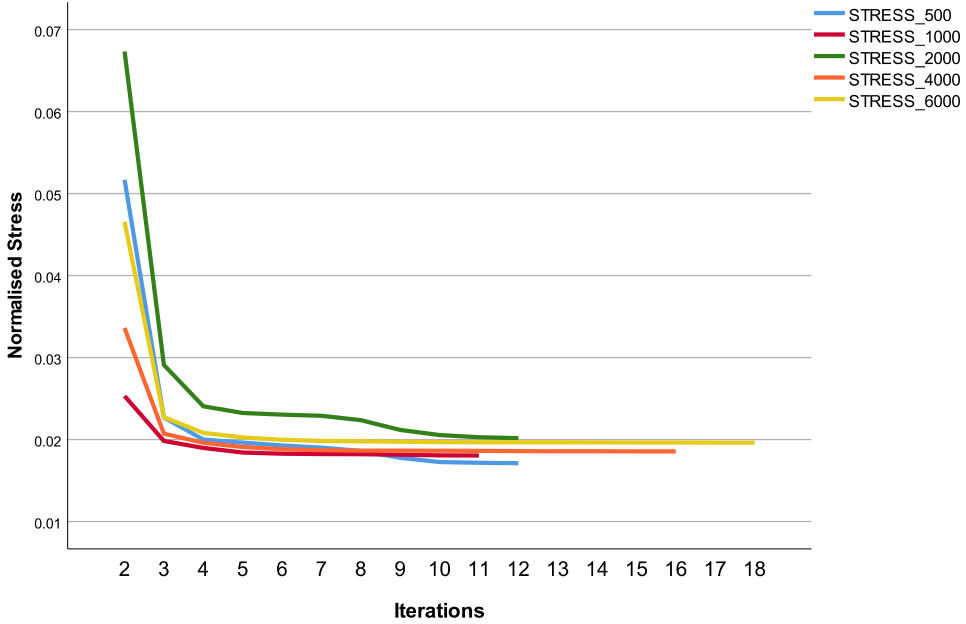
Figure 2. Convergence plots for each iteration for the normalised stress of the simulated data sets, considering $K = 6$, k_c=2 and $N = 500, 1000, 2000, 4000, 6000$.

with $\bar{\mathbf{x}}_i$ and $\bar{\mathbf{x}}_j$ the averaged temperature values of the time series realisations respectively, the dissimilarities are calculated as $\delta_{ij} = \sqrt{2\left(1 - \rho\left(\mathbf{x}_i, \mathbf{x}_j\right)\right)}$ (see, for example, Golay et al. 1998; Montero and Vilar 2014).

The number of clusters was selected from the dissimilarity matrix using the H^* index (Vera and Macías 2017), for values of $K = 1, 2, 3, 4, ..., 20$. For each value of K, the partition was estimated without imposing geometric constraints, using the allocation step and the Sokal–Michener dissimilarities between the clusters instead of the Euclidean distances, and the value of Eq. (8) was calculated. The lowest value of the $H^*(K)$ index was found for $K = 15$ clusters (see Vera and Macías 2017, for further details).

In view of these considerations, the CSS model was run for $K = 15$ clusters, obtaining a normalised stress value of $\sigma^2 = 0.02235079$. The top panel of Fig. 3 shows the configuration for the cluster centres given by the CSS model, except an invariant deformation for geodetic distances. Although related to the geographical location, the CSS configuration is not expected to exactly match the location of the geographic centres of the clusters. The bottom panel of Fig. 3 shows the geographic locations of the cluster centres together with the CSS configuration after applying the Procrustes transformation in three dimensions ($CC = 0.955686$), and projecting the optimally transformed configuration on the surface of the sphere. Despite the differences in some locations, in general the spatial relationship of the clusters of temperature time series can be appreciated.

The relationship is best appreciated for clusters in which the related stations are in the same hemisphere and therefore the centre is not deformed by curvature. For example, Fig. 4 shows the temperature time series for stations in groups 1 and 11. Although the average
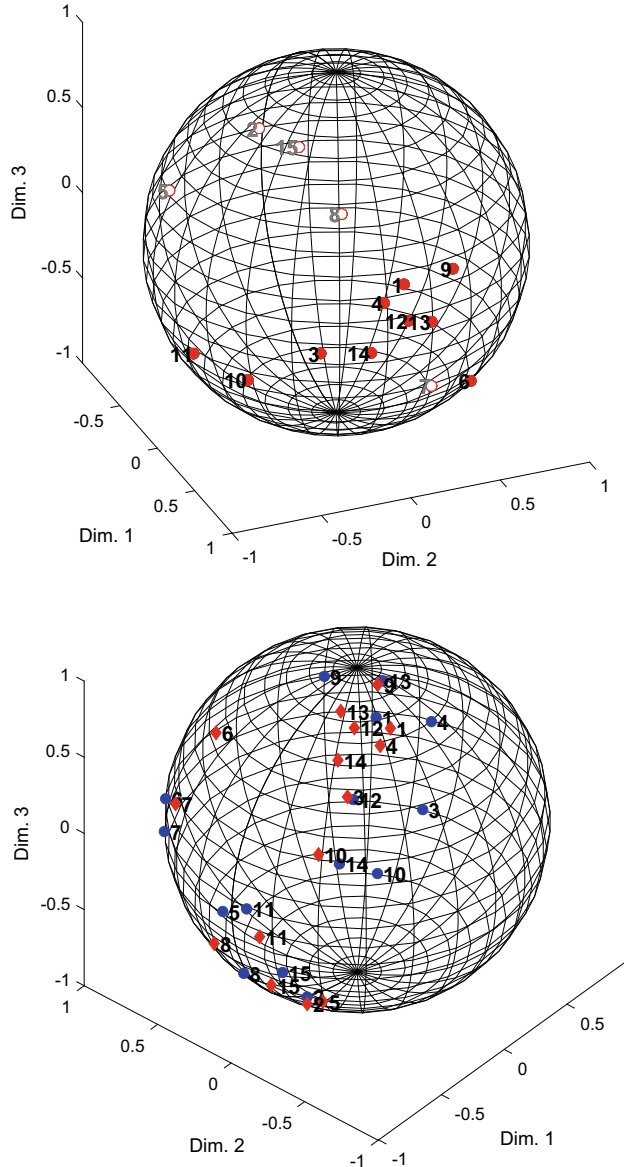
Figure 3. Cluster configuration on the sphere for the real data (top panel). The geographical coordinates of the cluster centres (blue) together with the CSS configuration after Procrustes (red) are shown in the bottom panel.

geographic location of the station locations in each cluster may be affected by the longitude effect, the averaged time series of both clusters are well related to their geographic centre.

Figure 5 shows the MDS-Q solution on the unit sphere surface, obtained using geodesic distances for all the sampling series (without clustering). Different colours and numbers represent the classification provided by the CSS model for $K = 15$. In general, the different cluster structures can be appreciated, some of which are quite close together, while others are sparse, but in general a clear interpretation is difficult due to the large number of points
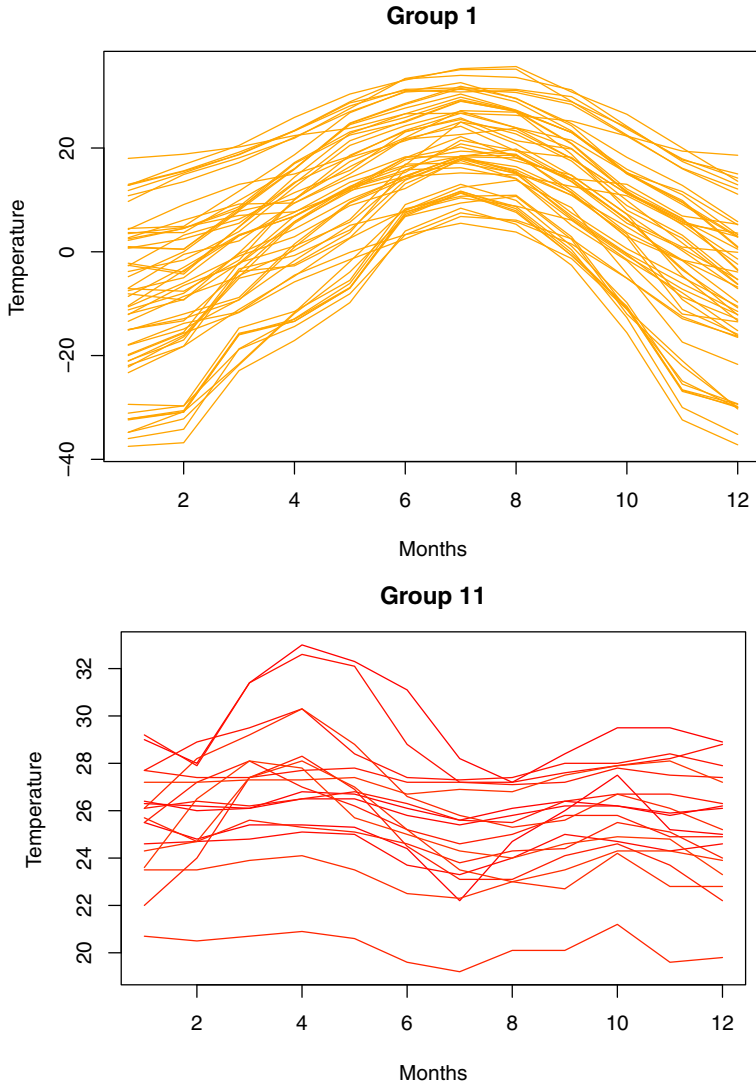
Figure 4.   Time series for groups 1 and 11.

represented. As an example, let us now illustrate how the representation error with or without clustering can influence the estimation of a spatial interpolator.

Mapping between geographic locations and the MDS representation is a widely used procedure, which, for example, allows interpolation in the context in which the covariance structure, expressed in terms of spatial dispersion, is stationary and isotropic, since the MDS representation only depends on distances. This is a very useful procedure in this framework, for example, when the aims is estimate the spatial covariance in a non-stationary process, although this aspect is beyond the scope of this paper (see Sampson and Guttorp 1992 for further details). Denote by $\mathbf{g} \in \mathcal{S}(1)$, the geographic location of a station, and consider $0 \leq \theta \leq 2\pi$, and $-\pi/2 \leq \phi \leq \pi/2$, the longitude and the latitude of $\mathbf{g}$, respectively. Let
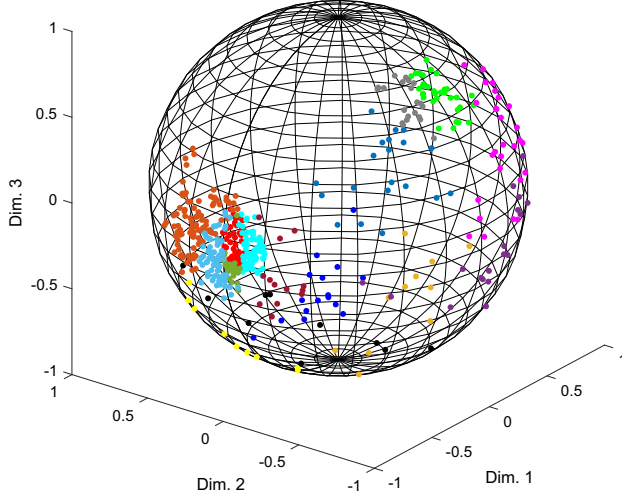
Figure 5.   MDS in the sphere with a geodesics distance solution for the 500 real data points, with colours assigned according to the classification provided by the CSS model.

denote by W_2^2 the model space for the spherical spline f of order 2 given by

$$W_2^2 = \left\{ f : \left| \int_{\mathcal{S}(1)} f \, \mathrm{d}\mathbf{g} \right| < \infty, J(f) < \infty \right\}, \tag{11}$$

where $J(f) = \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} (\Delta f)^2 \cos(\phi) \mathrm{d}\phi \mathrm{d}\theta$, and Δf is the surface Laplacian on the unit sphere (see Wang 2011 for further details). Then, denoting by ϑ_i and ψ_i the longitude and latitude of a point $\mathbf{x}_i$, $i = 1 \ldots, N$ in the MDS configuration on $\mathcal{S}(1)$, two smoothing spherical spline regression functions are estimated $f = (f_\vartheta, f_\psi)$, for the longitude and for the latitude values, respectively. For each component $v = \vartheta, \psi$, f_v is estimated by minimising

$$\frac{1}{N} \sum_{i=1}^N (v_i - f_v(\mathbf{g}_i))^2 + \lambda J(f), \tag{12}$$

where $\mathbf{g}_i = (\theta_i, \phi_i)$ are the geographical coordinates, and $v_i = \vartheta_i$ or $v_i = \psi_i, i = 1, \ldots, N$ for the MDS configuration. We analyse here the location error (LE) in terms of the average of the squared geodetic distances between the interpolated geographic locations and the MDS representation given by

$$LE = \frac{1}{N} \sum_{i=1}^N \breve{d}(\mathbf{x}_i, \hat{\mathbf{x}}_i)^2, \tag{13}$$

where $\hat{\mathbf{x}}_i = f(\mathbf{x}_i) = (f_\vartheta(\theta_i), f_\psi(\phi_i))$, is the image of f for the ith geographical location. The analysis has been performed using the ssr function at the assist R package (v2.1.8; Wang et al. 2022). A value of LE of 2.482347 was obtained using the CSS model, while this was of 6.11973 when the interpolation process was performed using the MDS spherical representation for the $N = 500$ stations. In addition, we consider the mean squares error

(MSE) of the fit for each component (longitude, latitude) after Procrustes, given by $MSE_v = \sum_i (v_i - f_v(\mathbf{g}_i))^2 / N$. For the longitude component, the MSE values of 0.4730863 and 0.8416269 where obtained for the CSS and spherical MDS (without clustering) respectively, while for the latitude, the values were of 0.8057067 and 1.262198, respectively. The results obtained show that the performance of the interpolation procedure is better with the CSS model, as expected.

Finally, we investigated whether a two-step procedure, consisting of an independent classification and representation stage, offered an adequate solution to the problem. Using the correlation-based dissimilarity matrix for the time series based on 500 locations, we first performed classical MDS in full dimension (dimension 499). Then, K-means clustering was performed for $K = 15$ using this configuration (see Vera and Macías 2021). The Sokal–Michener dissimilarity matrix was calculated, and the configuration of the cluster centres was estimated using the representation step described in Sect. 3.2. The normalised value of the total stress (2) given for the two-step procedure was 0.02449717, which is higher than that obtained with the proposed CSS model (0.02235079), as was hoped. In addition, the normalised value of the first term in (3) was of 0.009104791 for the two-step model and 0.006082755 for the CSS model, while the normalised stress values in (7) were 0.01077115, and 0.009900273, respectively, reflecting the good performance of the proposed CSS model.

6. DISCUSSION

This paper proposes a spherical-constrained cluster-MDS model for two-way one-mode dissimilarity data using geodesic distances. One of the main advantages of this model is that it enables us to address clustering and representation problems presenting a large number of points, particularly on a spherical surface, all based on any dissimilarity measure. Instead of a two-step procedure, and using only the information given by a dissimilarity matrix between a set of objects, the proposed model obtains a classification into homogeneous clusters while simultaneously, not the objects themselves but the centres of the clusters are represented on a spherical surface of optimal radius, using geodesic distances. Hence, for any partition, the dissimilarities are assumed to vary randomly within a cluster, while the corresponding distance is constant within the same cluster, whereas between clusters, differences in distance will reflect the tendency of the corresponding dissimilarities to vary systematically. Furthermore, when the representation of the points in high dimensionality is also known, any measure of dissimilarity can be used, and the proposed procedure, in addition to grouping, facilitates interpretation by preserving the global structure between the groups, unlike other spherical embedding-only procedures.

The parameter estimation is performed in an alternating step procedure which consists of a dissimilarity-based assignment step using geodesic distances, and a double representation step in which a ratio transformation is considered in a MDS model for a configuration that is constrained to the sphere surface using geodesic rather than Euclidean distances, in a metric approach. The search for the optimal object classification is formulated using a minimum distance procedure (Heiser and Groenen 1997), together with geodesic distances, which can be seen as a generalised K-means clustering procedure on the dissimilarity matrix in

terms of geodesic (instead of Euclidean) distances. For the representation step, the search for the overall optimal radius is performed introducing a ratio transformation in the primal quadratic constraints MDS algorithm proposed by De Leeuw and Mair (2009), also using an alternating estimation procedure.

The performance of the CSS model was analysed considering 640 artificial clustered data sets, in an extensive Monte Carlo experiment with different sizes, numbers of clusters, sizes of clusters and degrees of overlap. Among other aspects, we determined the quality of the classification obtained using the ARI (Hubert and Arabie 1985), and the degree to which the true configuration of the cluster centres was recovered using the congruence coefficient (Tucker 1951) after the Procrustes transformation. Regardless of whether the groups were balanced or not, the model showed good performance in all data sets considered with non-overlapping clusters. For somewhat overlapping groups, the ARI values worsened as the number of clusters increased, while for fairly overlapping groups, the ARI values were poor for both clusters of equal or different sizes. In all situations, the congruence coefficient values remained high, as the average values of the differences between the groups did not appear to be greatly altered by misclassified points.

To illustrate the performance of the model for real data sets, we analysed the monthly mean near-surface air temperature for 500 locations, worldwide. The results obtained were compared with those given by a two-step procedure in which first, K-means clustering was performed using the dissimilarity matrix (Vera and Macías 2021), after which the cluster centres were represented using the primal quadratic constraint MDS algorithm proposed by De Leeuw and Mair (2009), and implemented in the smacof package in R (Mair et al. 2021). As expected, our procedure outperformed the two-step algorithm. In addition to facilitating interpretation and reducing the number of parameters, the proposed procedure improves the fit when the interpolation procedure based on spherical spline is used on the centres of the clusters instead of on the complete data set.

The proposed model performs clustering based on any dissimilarity measure, but does not take into account the spatial proximity between objects, which may be a limitation for some practical applications. Clustering with constraints is necessary, for example, when we wish stations and clusters to retain their spatial relationships, and in this situation, additional spatial contiguity constraints are required (see Vera et al. 2008). Another limitation of the model is related to the assumption that the scale of the dissimilarities is metric. In many cases, dissimilarities are measured on an ordinal scale, that is, only the order between dissimilarities is preserved by distances, so in these cases a transformation based on monotonic regression must be considered when approximating them by geodesic distances. This results in a more flexible but imprecise clustering and representation model, more appropriate when it is not required to strictly preserve the global structure, but only the order of dissimilarities across the distances between group positions.

Spherical embedding is an important tool for data analysis in diverse areas of interest. Clustering and/or a planar representation using MDS have been considered for the problem of non-stationary spatial covariance structure estimation (Sampson and Guttorp 1992; Vera et al. 2008, 2009a; Vera et al. 2017). The model we present allows the extension of this procedure when the dissimilarities are determined by the location of the points on the

surface of a sphere. This model, together with the spatiotemporal processes defined on the sphere surface is currently being investigated by the authors.

ACKNOWLEDGEMENTS

Declarations
Conflict of interest The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

APPENDIX: PROJECTION STEP

At each iteration of the geodesic representation step, the Guttman transform in terms of Euclidean distances $\bar{\mathbf{X}}^{(s)}$ is projected onto the sphere surface by minimising $\mathbf{tr}\left(\mathbf{X}^{(s)} - \bar{\mathbf{X}}^{(s)}\right)' \mathbf{V}\left(\mathbf{X}^{(s)} - \bar{\mathbf{X}}^{(s)}\right)$, with $\mathbf{V}$ defined in Section (3.2) and $\mathbf{X}^{(s)} = \lambda^{(s)}\mathbf{Z}^{(s)}$, diag $\mathbf{Z}^{(s)}\mathbf{Z}^{(s)'} = \mathbf{I}$. Then, for fixed $\mathbf{Z}^{(s)}$, $\min_{\lambda^{(s)}} \mathbf{tr}\left(\lambda^{(s)}\mathbf{Z}^{(s)} - \bar{\mathbf{X}}^{(s)}\right)' \mathbf{V}\left(\lambda^{(s)}\mathbf{Z}^{(s)} - \bar{\mathbf{X}}^{(s)}\right)$ is given by $\lambda^{(s)} = \mathbf{tr}\bar{\mathbf{X}}^{(s)'}\mathbf{VZ}^{(s)}/\mathbf{trZ}^{(s)'}\mathbf{VZ}^{(s)}$, and $\mathbf{Z}$ is estimated by maximising $\rho(\mathbf{Z}^{(s)}) = (\mathbf{tr}\bar{\mathbf{X}}^{(s)'}\mathbf{VZ}^{(s)})^2/\mathbf{trZ}^{(s)'}\mathbf{VZ}^{(s)}$ (Further details can be found in De Leeuw and Mair 2009). Finally, the configuration $\mathbf{X}^{(s)} \in \mathcal{S}(\lambda^{(s)})$ obtained from the primal MDS-Q procedure is normalised on the sphere or radius one, simply by dividing $\mathbf{X}^{(s)}$ by the radius $\lambda^{(s)}$, which produces $\mathbf{X}^{(s)*} \in \mathcal{S}(1)$. Then, for the parameter estimates, the radius is again adjusted for the geodesic distances to minimise (2).

Radius Adjustment

Given the monotonic relation between Euclidean and geodesic distances (1), at each iteration, the radius of the sphere is globally optimised in terms of Eq. (2) in an alternating

estimation procedure. Expanding

$$
\sigma^2 \left(b^{(s)}(\mathbf{X}^s)^* \right) = \sum_{k<l} \widetilde{\omega}_{kl} \left(\frac{\widetilde{\delta}_{kl}}{\lambda^{(s)}} - b^{(s)} \breve{d}_{kl} \left(\mathbf{X}^{(s)*} \right) \right)^2
$$

$$
= \sum_{k<l} \widetilde{\omega}_{kl} \frac{\widetilde{\delta}_{kl}^2}{\lambda^{(s)2}} + b^{(s)2} \sum_{k<l} \widetilde{\omega}_{kl} \breve{d}_{kl}^2 \left(\mathbf{X}^{(s)*} \right) - 2b^{(s)} \sum_{k<l} \widetilde{\omega}_{kl} \frac{\widetilde{\delta}_{kl}}{\lambda^{(s)}} \breve{d}_{kl} \left(\mathbf{X}^{(s)*} \right),
$$

$$(A.1)$$

and setting the derivative of Eq. (A.1) with respect to $b^{(s)}$ equal to zero, we obtain

$$
b^{(s)*} = \frac{\sum_{k<l} \widetilde{\omega}_{kl} \frac{\widetilde{\delta}_{kl}}{\lambda^{(s)}} \breve{d}_{kl} \left(\mathbf{X}^{(s)*} \right)}{\sum_{k<l} \widetilde{\omega}_{kl} \breve{d}_{kl}^2 \left(\mathbf{X}^{(s)*} \right)}.
$$

Thus, $\mu^{(s)}$ in Eq. (7) is given by $\mu^{(s)} = \lambda^{(s)} b^{(s)*}$, and the coordinates of $\mathbf{X}^{(s)*}$ are multiplied by $\mu^{(s)}$ such that the update solution $\mathbf{X}^{(s)}$ is optimal for Eq. (2) in terms of the geodesic distances.

REFERENCES

Alegría A, Porcu E, Furrer R, Mateu Jorge (2018) Covariance functions for multivariate Gaussian fields evolving temporally over planet earth. Stoch Environ Res Risk Assess 33:1593–1608. https://doi.org/10.1007/s00477-019-01707-w. (**2019**)

Banerjee A, Dhillon IS, Ghosh J, Sra S (2005) Clustering on the unit hypersphere using von Mises–Fisher distributions. J Mach Learn Res 6:1345–1382

Bentler PM, Weeks DG (1978) Restricted multidimensional scaling models. J Math Psychol 17(2):138–151. https://doi.org/10.1016/0022-2496(78)90027-5

Bock HH (1986) Multidimensional scaling in the framework of cluster analysis. In: Hermes HJ, Optiz O, Degens PO (eds) Studien zur Klassifikation: [Classification and its environment], vol 17. INDEKS-Verlag, Frankfurt, pp 247–258

Bock HH (1987) On the interface between cluster analysis, principal component analysis, and multidimensional scaling. In: Bozdogan H, Gupta AK (eds) Multivariate statistical modeling and data analysis. Reidel, New York, pp 17–34

Bookstein F (1979) Fitting conic sections to scattered data. Comput Vis Graph 9:56–71. https://doi.org/10.1016/0146-664X(79)90082-0

Borg I, Groenen PJF (2005) Modern multidimensional scaling: theory and applications, 2nd edn. Springer, New York

Borg I, Lingoes JC (1980) A model and algorithm for multidimensional scaling with external constraints on the distances. Psychometrika 45:25–38. https://doi.org/10.1007/BF02293597

Chevallier E, Li D, Lu Y, Dunson D (2022) Exponential-wrapped distributions on symmetric spaces. SIAM J Math Data Sci 4:1347–1368. https://doi.org/10.1137/21M1461551

Cox TF, Cox MAA (1991) Multidimensional scaling on a sphere. Commun Stat Theory Methods 20(9):2943–2953. https://doi.org/10.1080/03610929108830679

de Leeuw J (1988) Convergence of the majorization method for multidimensional scaling. J Classif 5:163–180. https://doi.org/10.1007/BF01897162

De Leeuw J, Heiser WJ (1980) Multidimensional scaling with restrictions on the configuration. In: Krishnaiah PR (ed) Multivariate analysis. North-Holland, Amsterdam, pp 501–522

De Leeuw J, Mair P (2009) Multidimensional scaling using majorization: SMACOF in R. J Stat Softw 31(3):1–30. https://doi.org/10.18637/jss.v031.i03

Dzwinel W, Yuen DA, Boryczko K, Ben-Zion Y, Yoshioka S, Ito T (2005) Nonlinear multidimensional scaling and visualization of earthquake clusters over space, time and feature space. Nonlinear Proc Geophys 12:117–128. https://doi.org/10.5194/npg-12-117-2005

Elad A, Keller Y, Kimmel R (2005) Texture mapping via spherical multi-dimensional scaling. In: Kimmel R, Sochen NA, Weickert J (eds) Scale space and pde methods in computer vision. Scale-Space 2005. Lecture notes in computer science, vol 3459. Springer, Berlin, pp 443–455. https://doi.org/10.1007/11408031_38

Fitzgibbon A, Pilu M, Fisher RB (1999) Direct least square fitting of ellipses. IEEE Trans Pattern Anal Mach Intell 21(5):476–480. https://doi.org/10.1109/34.765658

Gnanadesikan R (1977) Methods for statistical data analysis of multivariate observations, 2nd edn. Wiley, New York

Golay X, Kollias S, Stoll G, Meier D, Valavanis A, Boesiger P (1998) A new correlation based fuzzy logic clustering algorithm for FMRI. Magn Reson Med 40(2):249–60. https://doi.org/10.1002/mrm.1910400211

Harris I, Osborn TJ, Jones P, Lister D (2020a) Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. Sci Data 7(109):1–18. https://doi.org/10.1038/s41597-020-0453-3

Harris I, Osborn TJ, Jones P, Lister D (2020b) Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.02. Accessed 24 Nov 2022

Heiser WJ (1993) Clustering in low-dimensional space. In: Opitz O, Lausen B, Klar R (eds) Information and classification. Studies in classification, data analysis and knowledge organization. Springer, Berlin, pp 162–173. https://doi.org/10.1007/978-3-642-50974-2_17

Heiser WJ, Groenen PJF (1997) Cluster differences scaling with a within-clusters loss component and a fuzzy successive approximation strategy to avoid local minima. Psychometrika 62:63–83. https://doi.org/10.1007/BF02294781

Hubert L, Arabie P (1985) Comparing partitions. J Classif 2:193–218. https://doi.org/10.1007/BF01908075

Kovitz J, Christakos G (2004) Spatial statistics of clustered data. Stoch Environ Res Risk Assess 18:147–166. https://doi.org/10.1007/s00477-003-0133-9

Lopes AM, Machado JAT, Pinto CMA (2014) Multidimensional scaling visualization of earthquake phenomena. J Seismol 18(1):163–179. https://doi.org/10.1007/s10950-013-9409-9

Luo H, Purvis JE, Li D (2023) Spherical rotation dimension reduction with geometric loss functions. arXiv:2204.10975v2. Stat.ML. https://doi.org/10.48550/arXiv.2204.10975

Mair P, De Leeuw J, Groenen PJF, Borg I (2021) smacof: multidimensional scaling. R package (Version 2.1-3), 2021. https://cran.r-project.org/web/packages/smacof/index.html. Accessed 24 Nov 2022

McInnes L, Healy J, Melville J (2018) UMAP: uniform manifold approximation and projection for dimension reduction. http://arxiv.org/abs/1802.03426

Montero P, Vilar JA (2014) TSclust: an R package for time series clustering. J Stat Softw 62(1):1–43. https://doi.org/10.18637/jss.v062.i01

Sampson PD, Guttorp P (1992) Nonparametric estimation of nonstationary spatial covariance structure. J Am Stat Assoc 87(417):108–119. https://doi.org/10.2307/2290458

Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. Univ Kansas Sci Bull 38(22):1409–1438

Tsagris M, Athineou G, Adam C, Sajib A, Amson E, Waldstein MJ (2022) Directional: a collection of functions for directional data analysis. R package (Version 5.5.). https://CRAN.R-project.org/package=Directional. Accessed 24 Nov 2022

Tucker LR (1951) A method for synthesis of factor analysis studies. Personnel research section report, No.984. Department of the Army, Washington, D.C

Van Der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579–2605

Vera J, Angulo J, Roldán JA (2017) Stability analysis in nonstationary spatial covariance estimation. Stoch Environ Res Risk Assess 31:815–828. https://doi.org/10.1007/s00477-016-1228-4

Vera J, Macías R (2017) Variance-based cluster selection criteria in a k-means framework for one-mode dissimilarity data. Psychometrika 82(2):275–294. https://doi.org/10.1007/s11336-017-9561-1

Vera J, Macías R (2021) On the behaviour of K-means clustering of a dissimilarity matrix by means of full multidimensional scaling. Psychometrika 86:489–513. https://doi.org/10.1007/s11336-021-09757-2

Vera J, Macías R, Angulo J (2008) Non-stationary spatial covariance structure estimation in oversampled domains by cluster differences scaling with spatial constraints. Stoch Environ Res Risk Assess 22(1):95–106. https://doi.org/10.1007/s00477-006-0100-3

Vera J, Macías R, Angulo J (2009a) A latent class MDS model with spatial constraints for non-stationary spatial covariance estimation. Stoch Environ Res Risk Assess 23:769–779. https://doi.org/10.1007/s00477-008-0257-z

Vera J, Macías R, Heiser WJ (2009b) A latent class multidimensional scaling model for two-way one-mode continuous rating dissimilarity data. Psychometrika 74:297–315. https://doi.org/10.1007/s11336-008-9104-x

Wang Y (2011) Smoothing splines: methods and applications, 1st edn. Chapman and Hall/CRC, London. https://doi.org/10.1201/b10954

Wang Y, Huang H, Rudin C, Shaposhnik Y (2021) Understanding how dimension reduction tools work: an empirical approach to deciphering t-SNE, UMAP, TriMAP, and PaCMAP for data visualization. J Mach Learn Res 22(1):9129–9201

Wang Y, Ke C, Moler C (2022) assist: a suite of R functions implementing spline smoothing techniques. R package version 3.1.8. https://CRAN.R-project.org/package=assist