

Model-based clustering for spherical and hyper-spherical data using elliptically symmetric distributions

Theodoros Perdikis

University of Piraeus

Nader Alharbi

King Saud bin Abdulaziz University for Health Sciences, King Abdullah International Medical Research Center

Michail Tsagris

mtsagris@uoc.gr

University of Crete

Research Article

Keywords: spherical data, model-based clustering, elliptical symmetry

Posted Date: August 5th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10084451/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Model-based clustering for spherical and hyper-spherical data using elliptically symmetric distributions

Theodoros Perdikis¹, Nader Alharbi², and Michail Tsagris^{*3}

¹Department of Tourism Studies, University of Piraeus, Greece, theoperdik@unipi.gr

²King Saud bin Abdulaziz University for Health Sciences, and , King Abdullah International Medical Research Center, Riyadh, Saudi Arabia, alharbina@ksau-hs.edu.sa

³Department of Economics, University of Crete, Greece, mtsagris@uoc.gr

June 18, 2026

Abstract

Model-based clustering for directional data has attracted a lot of interest, but most methods utilize rotationally symmetric distributions. This paper suggests the use of elliptically symmetric distributions, namely the elliptically symmetric angular Gaussian and the spherical elliptically symmetric projected Cauchy distributions that were recently proposed in the literature for modelling spherical data. We employ the expectation-maximization algorithm and the inclusion of covariates is also examined. We also study the case of hyper-spherical data after projecting them on the sphere. Simulation studies compare the two distributions in terms of choosing the optimal number of clusters and computational cost. We use the mixtures of these two distributions to cluster two datasets on the sphere (earthquake locations) and two hyper-spherical datasets.

Keywords: spherical data, model-based clustering, elliptical symmetry

1 Introduction

Directional data refers to multivariate data constrained to a unit norm, with the sample space represented as:

$$\mathbb{S}^d = \left\{ \mathbf{x} \in \mathbb{R}^{d+1} \mid \|\mathbf{x}\| = 1 \right\},$$

^{*}Corresponding author.

where $\|\cdot\|$ denotes the Euclidean norm. For $d = 1$, circular data reside on a circle, while for $d = 2$, spherical data are situated on a sphere. Directional data arise in a wide range of fields, including geology (Chang, 1986), environmental sciences (Heaton et al., 2014), image analysis (Straub et al., 2015), robotics (Bullock et al., 2014), and space (Kent et al., 2016), where direction plays a central role in understanding the underlying phenomena where directional information is essential to properly model these type of data.

Over the past decades several spherical (and hyper-spherical) distributions have been proposed. In the early 50’s and 70’s, Fisher (1953) and Kendall (1974) introduced two novel classes of distributions named as the von Mises–Fisher (vMF) and the projected normal distributions respectively. Recently, Kato and McCullagh (2020) proposed an extension of the Cauchy distribution to model spherical data, named as spherical Cauchy (SC) while Golzy and Markatou (2020) considered the design of the Poisson kernel-based (PKB) distribution. However, besides the novelty and contribution of these distributions, these distributions assume rotational symmetry, which may restrict their applicability in certain scenarios. To mitigate this constraint, Kent (1982) introduced an elliptically symmetric distribution that relaxes the rotational symmetry assumption. This distribution constitutes a special case of the Fisher–Bingham distribution (Mardia, 1975), and has proven valuable for modelling more sophisticated data structures. More recently, Paine et al. (2018) proposed the elliptically symmetric angular Gaussian (ESAG) distribution as an alternative to the Kent distribution, possessing some nice features (closed formula for the normalizing constant, straightforward random vector generation) and with contour lines that match closely those of Kent. ESAG emerges by projecting the Gaussian distribution onto the sphere and imposing two conditions on the covariance matrix in order to avoid indeterminacy. Along similar lines, Tsagris and Alzeley (2025) proposed the spherical elliptically symmetric projected Cauchy (SESPC) distribution, and Scealy and Wood (2019) proposed the scaled vMF, an elliptically symmetric distribution. These distributions, among others, have been applied to paleomagnetism (Scealy and Wood, 2019), geology (Dong and Melnykov, 2024), and in seismology to model earthquake data (Kagan, 2013, Kagan and Jackson, 2015, Walsh et al., 2009).

The application of model-based clustering to directional data has attracted increasing attention, driven by the need to uncover latent groups in possibly heterogeneous spherical datasets. Under the model-based clustering framework, the observed data are modeled as a finite mixture of distributions where the Expectation–Maximization (EM) algorithm (Dempster et al., 1977) is a standard approach in estimating finite mixture models (Fruhwirth-Schnatter et al., 2019, McLachlan, 2000, McNicholas, 2016, Yao and Xiang, 2024). Banerjee et al. (2005), Hornik and Grün (2014) proposed mixtures of vMF distributions for clustering of directional data. Ever since, many works have dealt with this problem using mixtures of VMF distributions (Barbaro and Rossi, 2021, Gopal and Yang, 2014, McGraw et al., 2006, Røge et al., 2017, Taghia et al., 2014). Golzy and Markatou (2020) proposed mixtures of PKB distributions, and Tsagris et al. (2025) proposed mixtures of SC distributions. To the best of our knowledge, Peel et al. (2001) is the only paper that proposed model-based clustering using an elliptically symmetric distribution (the Kent distribution), and more recently, Dong and Melnykov

(2024) robustified the Kent mixtures model to account for outliers. Let us visualize the effect of modelling spherical data via elliptically symmetric distributions in Figure 1. Suppose there are two known clusters of spherical data and we model each one of them using the vMF (rotationally symmetric) and the ESAG (elliptically symmetric) distributions. Evidently, in contrast to the two vMF distributions, the two ESAG distributions have fitted the two clusters adequately.

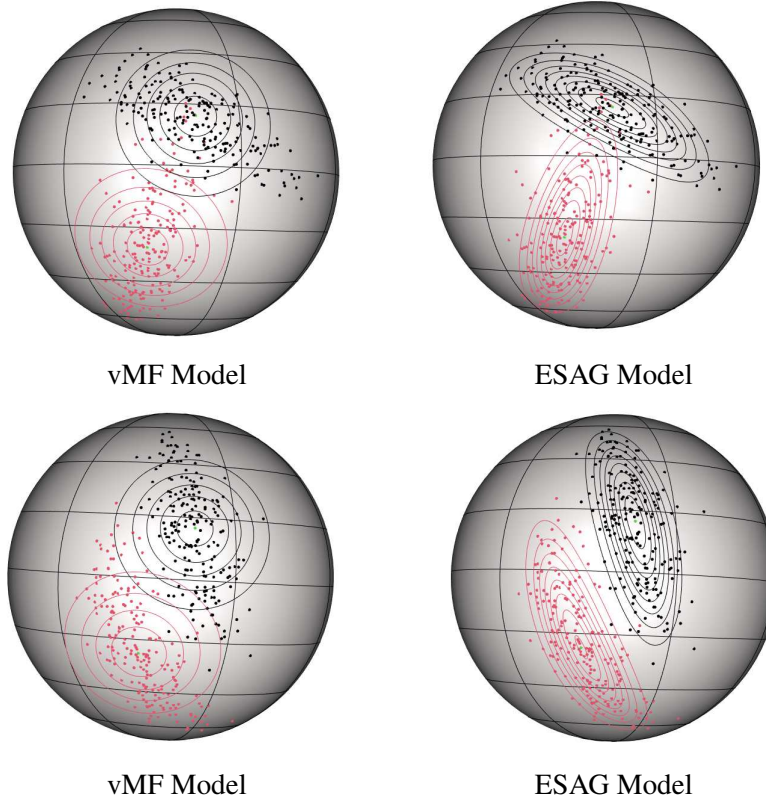


Figure 1: Contour plots for ESAG data: Modelled by the vMF (left column) and ESAG (right column).

Although model-based clustering for directional data has received attention by the research community in recent years, the vast majority of existing approaches rely on rotationally symmetric distributions, such as the vMF, SC, and PKB. As illustrated in Figure 1, rotationally symmetric distributions impose isotropic contours on the sphere, which may fail to capture the true shape of clusters that exhibit elliptical structure. This limitation motivates the need for mixture models based on elliptically symmetric distributions, which allow for a more flexible covariance structure and, consequently, a more faithful representation of the underlying cluster geometry. Furthermore, while most existing methods are restricted to spherical data ($d = 3$), many real-world applications involve data residing on the hyper-sphere ($d > 3$), a setting that has received comparatively less attention in the model-based clustering literature.

The present paper aims to fill both gaps by proposing and evaluating mixture models for both spherical and hyper-spherical data, relying on elliptically symmetry while retaining tractable para-

metric forms. We propose model-based clustering using mixtures of ESAG and SESPC distributions for spherical and hyper-spherical data. We employ the EM algorithm to fit the mixtures with a fixed number of components K , and utilize the integrated completed likelihood (ICL) criterion (Biernacki et al., 2000) to choose the appropriate number of K . We compare, via simulation studies, the two distributions in terms of number of clusters, selected according to ICL, and quality of the clusters, measured by the adjusted Rand index (ARI), but also in terms of computational cost. When it comes to data that reside on the hyper-sphere we project them onto the sphere and then employ the same mixture models. We use these mixture models to cluster the locations of earthquake locations from two regions (spherical), and wine quality and wholesales data (hyper-spherical).

The remainder of this paper is organised as follows. In Section 2, some preliminaries regarding the properties and parametrization of the two distributions under study are presented. In Section 3, a presentation of the EM algorithm being considered to estimate the mixture models is illustrated with and without the inclusion of covariates. Furthermore, Section 4, contains simulation studies offering an inter comparison of the two proposed mixture models and Section 5 illustrates their performance modelling the aforementioned (hyper-)spherical datasets. Finally, in Section 6, some concluding remarks and suggestions for future research are discussed.

2 Theoretical background of the distributions

2.1 The elliptically symmetric angular Gaussian distribution

Paine et al. (2018) proposed a new family of symmetrical distributions for spherical data. In fact, the ESAG distribution arises from projecting a multivariate normal on the sphere. Let $\mathbf{X}$ be a d -dimensional random variable from a d -variate Normal distribution with covariance matrix $\mathbf{V}$ and mean vector $\boldsymbol{\mu}$ which is projected on the circle/sphere/hyper-sphere, $\mathbf{Y} = \mathbf{X}/r$, where $r = \|\mathbf{X}\|$. The marginal distribution of $\mathbf{Y}$, denoted by angular Gaussian distribution (AG), arises by integrating out r

$$f(\mathbf{y}) = \int_0^\infty r^{d-1} f(r\mathbf{y}) dr, \quad (1)$$

resulting in

$$f_{\text{AG}}(\mathbf{y}) = \frac{C_d}{|\mathbf{V}|^{1/2}} \frac{1}{B^{d/2}} \exp \left[\frac{1}{2} \left(\frac{A^2}{B} - \Gamma^2 \right) \right] \mathcal{M}_{d-1} \left(\frac{A}{B^{1/2}} \right),$$

where $C_d = 1/(2\pi)^{(d-1)/2}$,

$$A = \mathbf{y}^\top \mathbf{V}^{-1} \boldsymbol{\mu}, \quad B = \mathbf{y}^\top \mathbf{V}^{-1} \mathbf{y}, \quad \text{and} \quad \Gamma^2 = \boldsymbol{\mu}^\top \mathbf{V}^{-1} \boldsymbol{\mu},$$

and, for real α ,

$$\mathcal{M}_{d-1}(\alpha) = \int_{u=0}^{\infty} u^{d-1} \frac{1}{(2\pi)^{1/2}} \exp \left\{ -(u - \alpha)^2 / 2 \right\} du.$$

Paine et al. (2018) showed that:

$$\mathcal{M}_0(\alpha) = \Phi(\alpha), \quad \mathcal{M}_1(\alpha) = \alpha\Phi(\alpha) + \phi(\alpha), \quad \mathcal{M}_2(\alpha) = (1 + \alpha^2)\Phi(\alpha) + \alpha\phi(\alpha),$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ correspond to the p.d.f and c.d.f of the standard normal distribution, respectively. Paine et al. (2018) extended (1) into a general family, termed elliptically symmetric angular Gaussian distribution (ESAG), by imposing the following two conditions

$$\mathbf{V}\boldsymbol{\mu} = \boldsymbol{\mu}, \tag{2}$$

$$|\mathbf{V}| = 1. \tag{3}$$

As a result, it follows that the p.d.f. of the AG distribution presented in (1) can be expressed as:

$$f_{\text{ESAG}}(\mathbf{y}) = \frac{C_d}{B^{d/2}} \exp \left[\frac{1}{2} \left\{ \frac{(\mathbf{y}^\top \boldsymbol{\mu})^2}{B} - \boldsymbol{\mu}^\top \boldsymbol{\mu} \right\} \right] \mathcal{M}_{d-1} \left\{ \frac{\mathbf{y}^\top \boldsymbol{\mu}}{B^{1/2}} \right\}. \tag{4}$$

From (2), it can be seen that the positive definite matrix $\mathbf{V}$ has a unit eigenvalue. By defining the other eigenvalues as

$$0 < \rho_1 \leq \dots \leq \rho_{d-1},$$

the inverse of covariance matrix $\mathbf{V}$ can be written as:

$$\mathbf{V}^{-1} = \xi_d \xi_d^\top + \sum_{j=1}^{d-1} \xi_j \xi_j^\top / \rho_j,$$

where $\xi_1, \dots, \xi_{d-1}$ and $\xi_d = \boldsymbol{\mu} / \|\boldsymbol{\mu}\|$ is a set of mutually orthogonal unit vectors and

$$\prod_{j=1}^{d-1} \rho_j = 1.$$

For $d = 3$, an efficient parameterization for the matrix $\mathbf{V}$ which satisfies the constraints (2) and (3) and has two free parameters. More specifically, let $\tilde{\xi}_1$ and $\tilde{\xi}_2$ be two unit vectors orthogonal to each other and to the mean direction $\xi_3 = \boldsymbol{\mu} / \|\boldsymbol{\mu}\|$:

$$\tilde{\xi}_1 = (-\mu_0^2, \mu_1\mu_2, \mu_1\mu_3)^\top / (\mu_0\|\boldsymbol{\mu}\|) \quad \text{and} \quad \tilde{\xi}_2 = (0, -\mu_3, \mu_2)^\top / \mu_0,$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)^\top$ and $\mu_0 = (\mu_2^2 + \mu_3^2)^{1/2}$. Additionally, the axes of symmetry ξ_1 and ξ_2 are

expressed as:

$$\xi_1 = \cos \psi \tilde{\xi}_1 + \sin \psi \tilde{\xi}_2 \quad \text{and} \quad \xi_2 = -\sin \psi \tilde{\xi}_1 + \cos \psi \tilde{\xi}_2,$$

where $\psi \in (0, \pi]$ is the angle of rotation. Finally, by defining the pair $\gamma = (\gamma_1, \gamma_2)^\top$ as

$$\gamma_1 = 2^{-1}(\rho^{-1} - \rho) \cos 2\psi \quad \text{and} \quad \gamma_2 = 2^{-1}(\rho^{-1} - \rho) \sin 2\psi,$$

where $\rho_1 = \rho$ and $\rho_2 = 1/\rho$ for $\rho \in (0, 1]$, $\mathbf{V}^{-1}$ can be expressed as:

$$\mathbf{V}^{-1} = \mathbf{I}_3 + \gamma_1 (\tilde{\xi}_1 \tilde{\xi}_1^\top - \tilde{\xi}_2 \tilde{\xi}_2^\top) + \gamma_2 (\tilde{\xi}_1 \tilde{\xi}_2^\top + \tilde{\xi}_2 \tilde{\xi}_1^\top) + \left\{ (\gamma_1^2 + \gamma_2^2 + 1)^{1/2} - 1 \right\} (\tilde{\xi}_1 \tilde{\xi}_1^\top + \tilde{\xi}_2 \tilde{\xi}_2^\top),$$

where $\mathbf{I}_3$ denotes the 3-dimensional identity matrix.

2.2 The spherical elliptically symmetric projected Cauchy distribution

Tsagris and Alzeley (2025) projected the 3-dimensional Cauchy distribution on the sphere and, by imposing the same conditions as in ESAG (2)–(3), introduced the spherical elliptically symmetric projected Cauchy distribution (SESPC) as:

$$f_{\text{SESPC}}(\mathbf{y}) = \frac{B(\|\boldsymbol{\mu}\|^2 + 1) \sqrt{E} \left[\arctan2(\sqrt{E}, -\mathbf{y}^\top \boldsymbol{\mu}) - \arctan2(\sqrt{E}, \mathbf{y}^\top \boldsymbol{\mu}) + \pi \right] + 2\mathbf{y}^\top \boldsymbol{\mu} E}{4\pi^2 B E^2},$$

where $E = B\|\boldsymbol{\mu}\|^2 + B - (\mathbf{y}^\top \boldsymbol{\mu})^2$ and $\arctan2(\cdot)$ is the 2-argument arc tangent. Similarly to the design of the ESAG distribution the corresponding inverse of the covariance matrix of $\mathbf{y}$, $\mathbf{V}$ is computed in the same exact manner.

3 Model-based clustering

Let K be the number of clusters of a population where $p_1, \dots, p_K$ are the clusters' unknown weights where $p_j > 0$ and $\sum_{j=1}^K p_j = 1$. Furthermore, we shall denote $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iK})^\top$, $i = 1, \dots, n$ as the latent allocation variable following the multinomial distribution, $\mathbf{Z}_i \sim \mathcal{M}(1; p_1, \dots, p_K)$, with K categories and success probabilities equal to $p_1, \dots, p_K$ which are the corresponding prior probabilities of selecting an observation from cluster j . The distribution of $\mathbf{Y}_i$ conditionally on $\mathbf{Z}_i = \mathbf{z}_i$, can be expressed as:

$$\mathbf{Y}_i | (\mathbf{Z}_i = \mathbf{z}_i) \sim f_{\mathbf{Y}_i | \mathbf{Z}_i}(\mathbf{y}_i | \mathbf{z}_i) = \prod_{j=1}^K f^{z_{ij}}(\mathbf{y}_i; \boldsymbol{\mu}_j, \gamma_j), \quad \text{independent for } i = 1, \dots, n \quad (5)$$

where $f(\cdot; \boldsymbol{\mu}_j, \gamma_j)$ denotes the probability density function of the 3-dimensional SESPC or ESAG distribution with parameter $\boldsymbol{\mu}_j, \gamma_j$ for $j = 1, \dots, K$. When $Z_{ij} = 1$, i.e. the j -th element of $\mathbf{Z}_i$ is equal to 1, the above equation can be written as $\mathbf{Y}_i | (Z_{ij} = 1) \sim f(\cdot; \boldsymbol{\mu}_j)$. Additionally, the joint

distribution of $\mathbf{Y}_i, \mathbf{Z}_i$ is:

$$(\mathbf{Y}_i, \mathbf{Z}_i) \sim f_{\mathbf{Y}_i, \mathbf{Z}_i}(\mathbf{y}_i, \mathbf{z}_i) = \prod_{j=1}^K \{p_j f(\mathbf{y}_i; \boldsymbol{\mu}_j, \gamma_j)\}^{z_{ij}}. \quad (6)$$

while the marginal distribution of $\mathbf{Y}_i$ is expressed as a mixture of K distributions:

$$\mathbf{Y}_i \sim f_{\mathbf{Y}_i}(\mathbf{y}_i) = \sum_{j=1}^K p_j f(\mathbf{y}_i; \boldsymbol{\mu}_j, \gamma_j), \quad \text{independent for } i = 1, \dots, n. \quad (7)$$

Furthermore, the conditional probability that an observation i is assigned to component j , given that $\mathbf{Y}_i = \mathbf{y}_i$, equals to

$$P(Z_{ij} = 1 | \mathbf{Y}_i = \mathbf{y}_i) = \frac{p_j f(\mathbf{y}_i; \boldsymbol{\mu}_j, \gamma_j)}{\sum_{\ell=1}^K p_\ell f(\mathbf{y}_i; \boldsymbol{\mu}_\ell, \gamma_\ell)} =: w_{ij}, \quad j = 1, \dots, K \quad (8)$$

or equivalently, $\mathbf{Z}_i | (\mathbf{Y}_i = \mathbf{y}_i) \sim \mathcal{M}(1, w_{i1}, \dots, w_{iK})$, independent for $i = 1, \dots, n$.

Finally, the observed log-likelihood is defined as

$$\ell(\boldsymbol{\mu}, \boldsymbol{\gamma}, \mathbf{p} | \mathbf{y}) = \sum_{i=1}^n \log \left\{ \sum_{j=1}^K p_j f(\mathbf{y}_i; \boldsymbol{\mu}_j, \gamma_j) \right\}, \quad (9)$$

where $\boldsymbol{\mu}, \boldsymbol{\gamma} \in \mathbb{R}^{4K}$, $\mathbf{p} \in \mathcal{P}_{K-1}$, where $\mathcal{P}_{K-1} = \{p_1, \dots, p_K : p_j > 0, j = 1, \dots, K, \sum_{j=1}^K p_j = 1\}$. It is clear that the goal is to maximize (9). In order to achieve that, similarly to [Tsagris et al. \(2025\)](#), the EM algorithm is considered where for a set of starting values, the expectation of the complete log-likelihood (E-step) is computed and then is maximized with respect to $\boldsymbol{\mu}, \mathbf{p}$ and $\boldsymbol{\gamma}$ (M-step). In particular, the complete log-likelihood is

$$\ell_c(\boldsymbol{\mu}, \boldsymbol{\gamma}, \mathbf{p} | \mathbf{y}, \mathbf{z}) = \sum_{i=1}^n \sum_{j=1}^K z_{ij} [\log p_j + \log f(\mathbf{y}_i; \boldsymbol{\mu}_j, \gamma_j)]. \quad (10)$$

E-step. At iteration t , given the current parameter estimates $\Theta^{(t)} = \{\boldsymbol{\mu}^{(t)}, \boldsymbol{\gamma}^{(t)}, \mathbf{p}^{(t)}\}$, the expected value of Z_{ij} is computed with respect to the conditional distribution $Z_i | (\mathbf{Y}_i = \mathbf{y}_i, \Theta^{(t)})$. The corresponding membership probabilities are derived as:

$$w_{ij}^{(t)} = \frac{p_j^{(t)} f(\mathbf{y}_i; \boldsymbol{\mu}_j^{(t)}, \gamma_j^{(t)})}{\sum_{\ell=1}^K p_\ell^{(t)} f(\mathbf{y}_i; \boldsymbol{\mu}_\ell^{(t)}, \gamma_\ell^{(t)})}, \quad i = 1, \dots, n, \quad j = 1, \dots, K. \quad (11)$$

M-step. At this step, the expected complete log-likelihood $Q(\Theta | \Theta^{(t)}) = \mathbb{E}[\ell_c | \mathbf{y}, \Theta^{(t)}]$ is max-

imised with respect to Θ and the mixing proportions are updated as the weighted averages of the posterior probabilities:

$$p_j^{(t+1)} = \frac{1}{n} \sum_{i=1}^n w_{ij}^{(t)}, \quad j = 1, \dots, K. \quad (12)$$

Due to the fact that neither the ESAG nor the SESPC density yields a closed-form maximizer for (μ_j, γ_j) , the weighted log-likelihood contribution of component j ,

$$Q_j(\mu_j, \gamma_j) = \sum_{i=1}^n w_{ij}^{(t)} \log f(y_i; \mu_j, \gamma_j), \quad (13)$$

is maximized numerically. Specifically, for each component $j = 1, \dots, K$, a general-purpose optimizer (e.g. Nelder-Mead) is applied to Q_j , and the procedure is restarted until the improvement in the objective falls below ϵ ($= 10^{-5}$), yielding the updated estimates $\mu_j^{(t+1)}$ and $\gamma_j^{(t+1)}$.

Regarding the optimal number of clusters the integrated complete likelihood (ICL) has been considered (Biernacki et al. (2003)). Initialization of the EM algorithm demands extra care in order to avoid convergence to local maxima. Some indicative works on this topic include Baudry and Celeux (2015), Biernacki et al. (2003), Fraley et al. (2005), Karlis and Xekalaki (2003), Michael and Melnykov (2016), Papastamoulis et al. (2016). In our implementation of the mixtures of ESAG and SESPC distributions, we tried two initialization schemes. One based on k -means starting values and one based on Gaussian mixture models (GMMs). The first strategy is adequate for rotationally symmetric distributions (Tsagris et al., 2025), but in our case GMMs that allow for a structured covariance was shown to yield better performance.

3.1 Inclusion of covariates

We may link the mixing probabilities p_j to some covariates $\mathbf{X}$ yielding the concomitant model

$$p_j(\mathbf{x}_i) = \frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta}_j)}{\sum_{\ell=1}^K \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_\ell)}. \quad (14)$$

In this case we substitute the p_j in Eq (12) with $p_j(\mathbf{x}_i)$ from Eq. (14), and hence the M-step requires an extra optimization. The $\boldsymbol{\beta}_{js}$ are obtained from maximization of the following quantity

$$D_j = \sum_{i=1}^{n_j} w_{ij}^{(t)} \log [p_j(\mathbf{x}_i)], \quad j = 1, \dots, K.$$

The maximization of D_j is performed independently of the maximization of the weighted log-likelihood in Eq. (13), and the observed log-likelihood (9) now becomes

$$\ell(\boldsymbol{\mu}, \boldsymbol{\gamma}, \mathbf{p}|\mathbf{y}) = \sum_{i=1}^n \log \left\{ \sum_{j=1}^K p_j(\mathbf{x}_i) f(\mathbf{y}_i; \boldsymbol{\mu}_j, \boldsymbol{\gamma}_j) \right\}.$$

We may also link the mean vector of the ESAG or SESPC to some covariates yielding the mixtures of regressions model, where $\boldsymbol{\mu}_j(\mathbf{x}_i) = \mathbf{x}_i^\top \boldsymbol{\beta}_j$. In this case, the observed log-likelihood becomes

$$\ell(\boldsymbol{\mu}, \boldsymbol{\gamma}, \mathbf{p}|\mathbf{y}) = \sum_{i=1}^n \log \left\{ \sum_{j=1}^K p_j f(\mathbf{y}_i; \boldsymbol{\mu}_j(\mathbf{x}_i^\top \boldsymbol{\beta}_j), \boldsymbol{\gamma}_j) \right\}.$$

Both the ESAG and the SESPC offer the possibility to relate their $\boldsymbol{\gamma}$ parameters to some covariates in the same linear manner, at the cost of increased computational cost. The last option is to add the covariates in both the mixing probabilities and the parameters of either distribution, but this is computationally expensive.

3.2 Visual inspection of the mixtures of ESAG and SESPC distributions

Our scope is to investigate the performance of each model assuming two scenarios where the data comes from the ESAG distribution and the data comes from the SESPC distribution. For each case, under the assumption that the true number of clusters is two, we will compare the performance of each model using simulation based on their corresponding ARI value. Before we proceed to any numerical investigations, it is important to examine the effect on the vector parameter $\boldsymbol{\gamma}$ on forming a cluster. Regarding the structure of the clusters, the corresponding mean vectors were defined as follows. For a given value of $\boldsymbol{\mu}_1$, by expressing it through its corresponding unit vector $\boldsymbol{\mu}'_1 = \frac{\boldsymbol{\mu}_1}{\|\boldsymbol{\mu}_1\|}$, we find a vector $\boldsymbol{\mu}'_2$ such that the angle between $\boldsymbol{\mu}'_1$ and $\boldsymbol{\mu}'_2$ is equal to ω . Lastly, we simulate a dataset forming two clusters by generating a random sample of size $n = 200$ from a mixture of ESAG (SESPC) distributions with approximately equal mixing probabilities, $p_1 \approx p_2$. For each cluster the corresponding mean vectors are considered to be equal to $\tau_1 \times \boldsymbol{\mu}'_1$ and $\tau_2 \times \boldsymbol{\mu}'_2$ respectively. As it will be shown hereafter, the values of (τ_1, τ_2) play the role of the concentration parameter.

In Figure 2 the corresponding shape of two clusters is presented under different values of $\boldsymbol{\gamma}$. For all cases, we set $\omega = 45$, $\boldsymbol{\mu}'_1 = (-0.927, -0.282, 0.249)$ and $\boldsymbol{\mu}'_2 = (-0.886, -0.008, -0.464)$, by generating $\boldsymbol{\mu}_1$ based on three random numbers from $N(0, 1)$. For illustration purposes we fix the values of (τ_1, τ_2) at $(5, 5)$. From Figure 2, it can be seen that the values of $\boldsymbol{\gamma}$ affect how the two clusters are formed. Note that, we intentionally considered these specific values of $\boldsymbol{\gamma}$ in order to investigate each model's performance under the scenarios where the two clusters are formed vertically or in an angle. Additionally, it can be observed that, considering exactly the same parameters for both distribution, the SESPC distribution is more spread than the ESAG but the corresponding relation

between the two clusters remains the same.

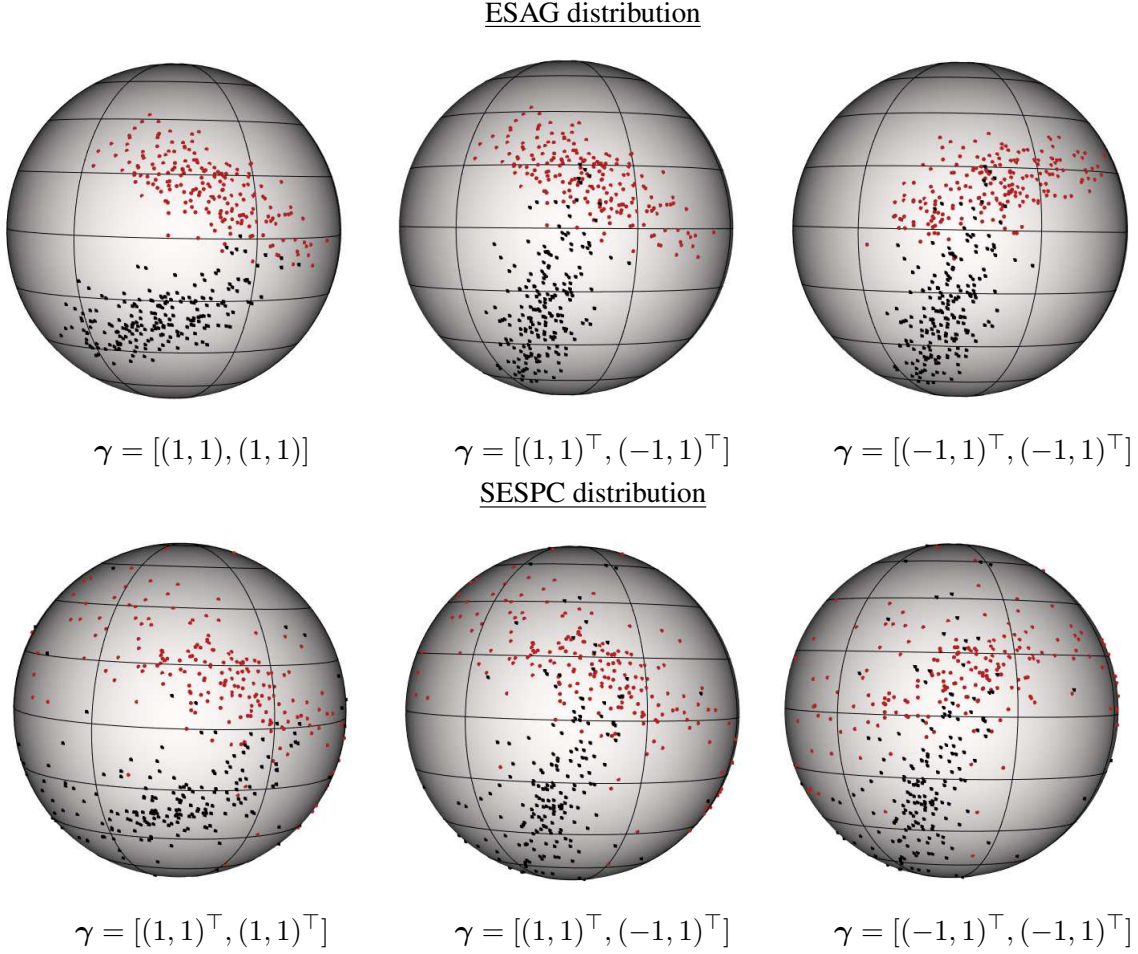


Figure 2: Effect of γ for ESAG distribution (top row, with $(\tau_1, \tau_2) = (5, 5)$) and SESP distribution (bottom row, with $(\tau_1, \tau_2) = (5, 5)$). For all cases, $\mu'_1 = (-0.927, -0.282, 0.249)^\top$, $\mu'_2 = (-0.886, -0.008, -0.464)^\top$.

3.3 Model-based clustering with hyper-spherical data

From Section 2, it is clear that the parameterization of ESAG and SESP distributions, as they were originally introduced in the works of [Paine et al. \(2018\)](#) and [Tsagris and Alzeley \(2025\)](#), can be considered only for spherical data. As a consequence, an interesting research question arises which is how practitioners could consider the use of Model-based clustering for hyper-spherical elliptical data. Recently, [Zehao and Xianzheng \(2024\)](#) proposed a novel parameterization of the ESAG distribution for hyper-spherical data. Motivated by their work, we consider the EM algorithm presented in Section 3. However, some preliminary results showed that the computational cost of the EM is quite high.

In order to tackle this limitation, we suggest to project the data onto the sphere using a geodesic principal component analysis (PCA) type of projection ([Fletcher et al., 2004](#), [Jung et al., 2012](#)) and

then apply the SESPC or ESAG mixture model. The algorithm first computes the intrinsic Fréchet mean of the data points on the sphere, then uses the logarithmic map to project all points into the tangent space at that mean. In this flat tangent space, standard PCA is applied to find the top principal directions. The function then constructs a low-dimensional geodesic subspace (a 3-dimensional sphere) by orthonormalizing the mean vector with the principal tangent directions, and finally projects the original spherical data onto this subspace, yielding points on the sphere.

In Figure 3 several sphere plots are illustrated of the projected 10-dimensional data into the sphere. The corresponding parameter γ_1 , γ_2 and μ_1 , μ_2 are presented in Figures 18 and 19, respectively, in the Appendix.

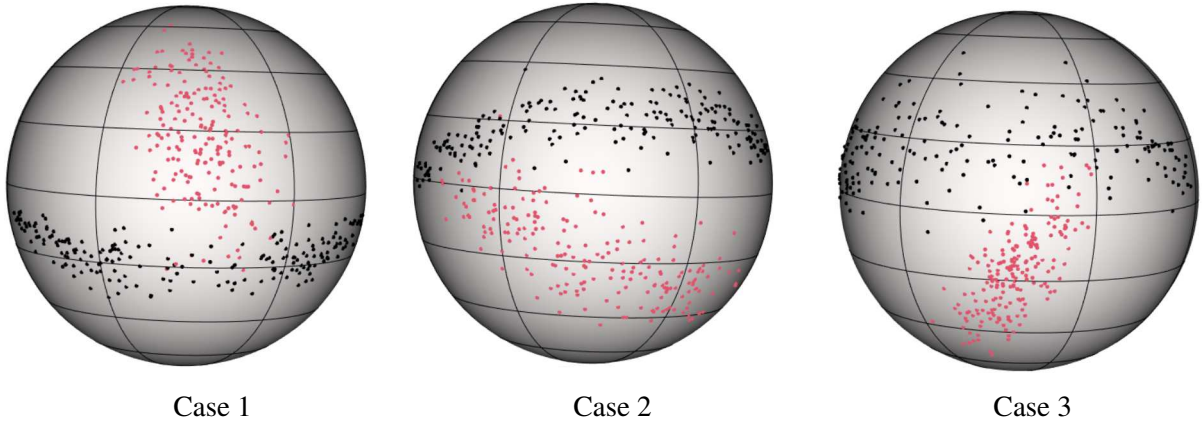


Figure 3: Hyper-spherical data for the projected 10-dimensional ESAG distribution. The values of γ_1 , γ_2 and μ_1 , μ_2 are presented in Figures 18 and 19, respectively, in the Appendix.

4 Simulation studies

4.1 Numerical analysis for spherical data

Firstly, the performance of each method, assuming almost equal sizes for the two clusters, is investigated (total size $n = 200$). In Figure 4, the performance of the two models are presented. In particular, for $R = 50$ iterations (i.e. simulating 50 different samples consisting of two clusters) the boxplots of the Adjusted Rand index (ARI) metric are presented. For simplicity, in the results presented below we shall denote as ARI^{ESAG} (ARI^{SESPC}) the ARI values obtained considering the ESAG (SESPC) model. We may see that:

- ESAG distribution: When the model being considered is the ESAG, it can create the correct number of clusters in all the cases where large values of ARI indicate the correct number of predicted clusters. Additionally, when the model is the SESPC, the performance does not significantly change. In particular, for the case where $\gamma = [(1, 1)^\top, (1, 1)^\top]$ ($\gamma = [(1, 1)^\top, (-1, 1)^\top]$)

), the boxplots of ARI^{ESAG} 's are narrowed comparing to the boxplots of ARI^{SESPC} with corresponding medians close to 0.95 (1). Lastly, when $\gamma = [(-1, 1)^\top, (-1, 1)^\top]$ the boxplots of ARI^{ESAG} 's are wider comparing to the boxplots of ARI^{SESPC} with corresponding medians close to 0.9. In summary, for all the investigated scenarios, both models can detect the correct number of clusters and give similar ARI values larger than 0.8 while the worst performance is achieved under the third case.

- **SESPC distribution:** Using the SESPC model can effectively define the correct number of clusters in most of the cases. Furthermore, for all the investigated cases, the boxplots of ARI^{SESPC} do not significantly differ. On the other hand, using the ESAG model the efficiency on detecting the correct number of clusters is worse and the boxplots of ARI^{ESAG} indicate poor performance with large variability in the last case.
- **Performance comparisons:** The performance of each model is better when the data comes from the ESAG distribution, which is an expected result since in both cases the same values of $\tau = (5, 5)$ are considered. Let us recall the fact that the SESPC distribution is more spread in the tails compared to the ESAG (see Figure 2).

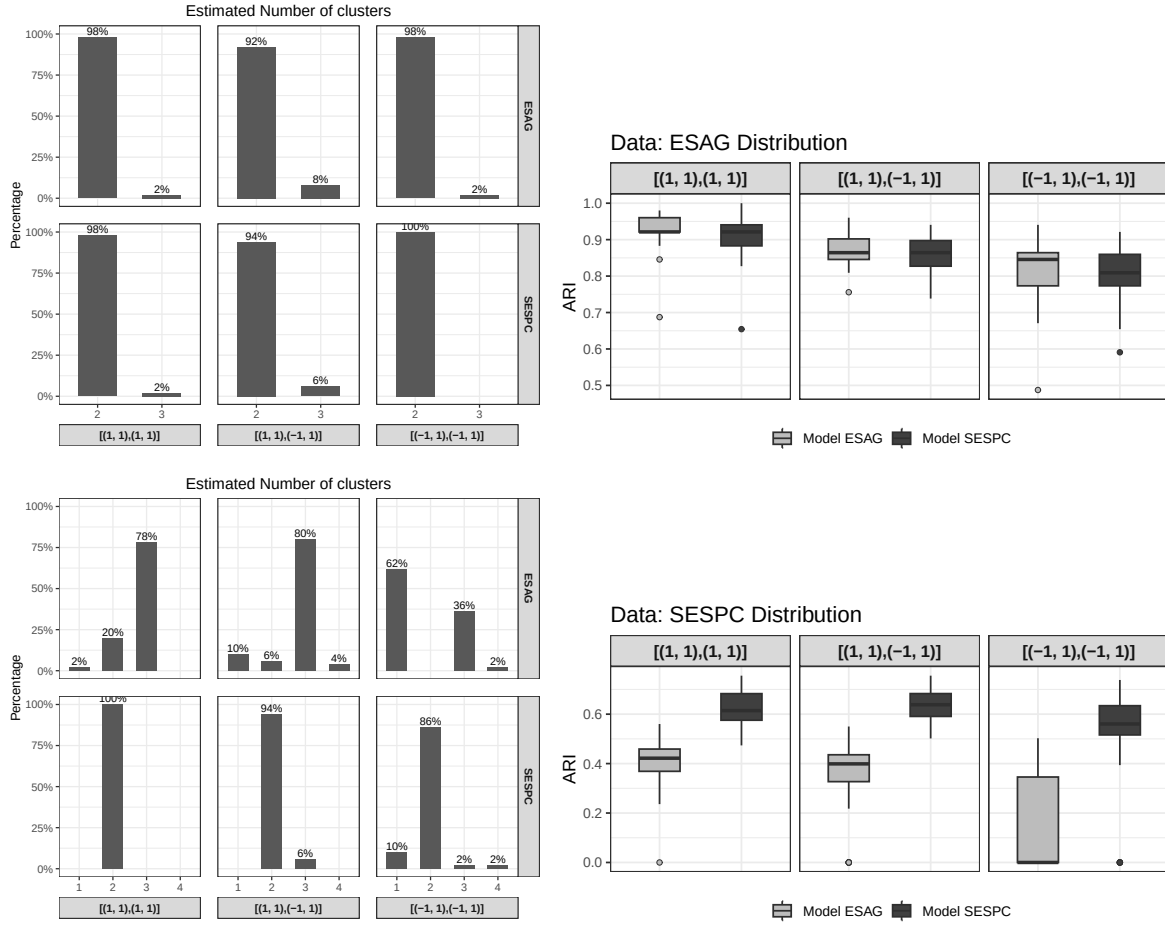


Figure 4: ESAG and SESPC generated data with equal clusters: $n = 200$, $\omega = 45^\circ$, $\mu'_1 = (-0.927, -0.282, 0.249)^\top$, $\mu'_2 = (-0.886, -0.008, -0.464)^\top$, $(\tau_1, \tau_2) = (5, 5)$, and various γ combinations.

In Figure 5, some additional performance comparisons between the two models are illustrated but now the probabilities of an observation to belong in a cluster is randomly generated from the 2-dimensional Dirichlet distribution. Note that, we intentionally increase the number of iterations on $R = 100$ in order to capture many scenarios with respect to the size of each cluster. From Figure 5 we observe that:

- **ESAG distribution:** When the model being considered is the ESAG, it can create the correct number of clusters in almost all the cases and the corresponding boxplots of ARI^{ESAG} do not significantly differ. On the other hand, when the model is the SESPC, the performance does not significantly change for the first two cases while for the third one, the variability of ARI^{SESPC} is very large.
- **SESPC distribution:** The SESPC model can effectively define the correct number of clusters in most of the cases. The smallest variability in the boxplots ARI^{SESPC} is achieved for the

case where $\gamma = [(1, 1)^\top, (1, 1)^\top]$ while for the other two the variability is quite large. It seems that large difference between the size of the clusters might affect the performance of the SESPC model in some cases. Nevertheless, it should be noted that, using the ESAG model, the efficiency on detecting the correct number of clusters is significantly low and the boxplots of ARI^{ESAG} indicate worse performance comparing to the SESPC model.

- Performance comparisons: The performance of each model is better when the data comes from the ESAG distribution, which is an expected result since in both cases the same values of $\tau = (5, 5)$ are considered and the SESPC is more spread in the tails compared to the ESAG distribution (see, Figure 2). In summary, under the assumption that the sizes of the clusters are balanced, the performance between the two models is more or less the same with the difference that when the data come from the SESPC distribution the ESAG model performs worse.

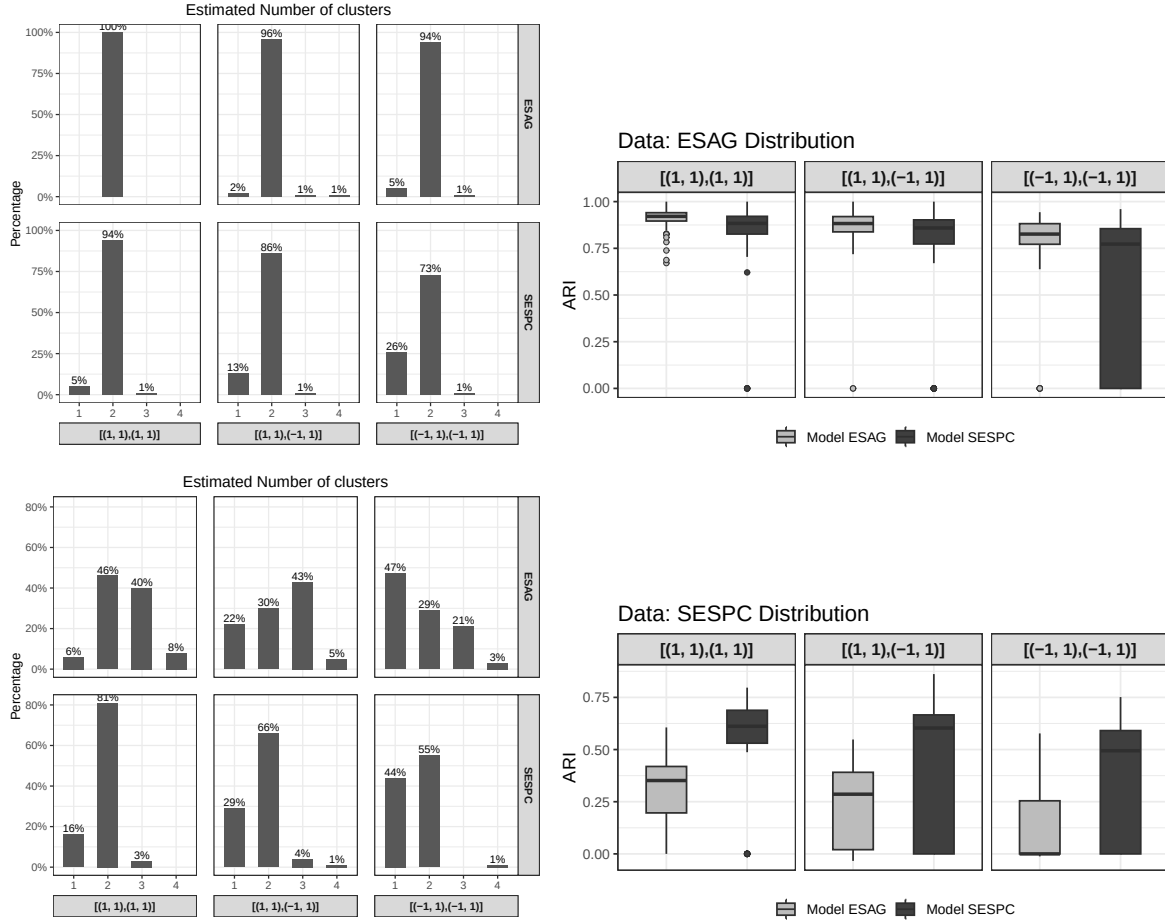


Figure 5: ESAG and SESPC generated data with unequal clusters: $n = 200$, $\omega = 45^\circ$, $\mu'_1 = (-0.927, -0.282, 0.249)^\top$, $\mu'_2 = (-0.886, -0.008, -0.464)^\top$, $(\tau_1, \tau_2) = (5, 5)$, and various γ combinations.

Figure 6 illustrates the effect of sample size on the performance of each model for $n \in \{100, 200, 500\}$. Without loss of generality we assume that, across all cases, the probabilities for an observation to belong at each cluster are the same. When data are generated from the ESAG distribution, the performance of both models remains relatively stable in terms of their ARI values across all the investigated sample sizes, suggesting robustness with respect to n . In contrast, when the data arise from the SESPC distribution, the sample size appears to have an effect on both the ability to recover the correct number of clusters and the models' performance as measured by the ARI. Specifically, an increase in sample size leads to higher with less variability ARI values, indicating improved clustering performance as n grows.

Additional results are presented in Figure 7 regarding the effect of the concentration parameter τ . In general, small (large) values of τ_1, τ_2 result in small (large) spread in the corresponding clusters. This, as expected, affects the models' performance. In particular for small combination of $(\tau_1, \tau_2) = (2, 2)$ both models do not perform well since are not able to detect the correct number of clusters in most of the cases. As the magnitude of the concentration increases the performance of each models is improved.

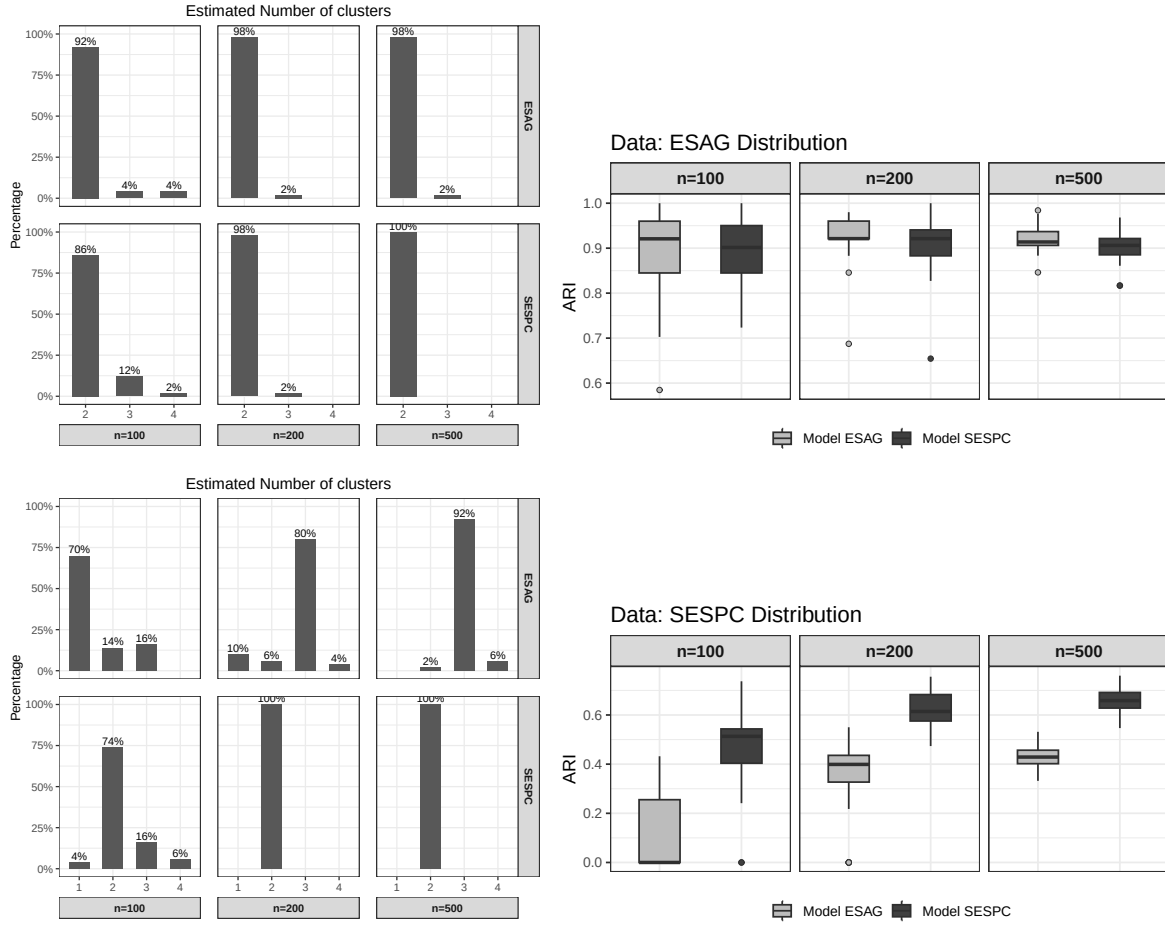


Figure 6: ESAG and SESPC generated data with equal clusters: effect of n . $\gamma = [(1, 1)^\top, (1, 1)^\top]$, $\omega = 45^\circ$, $\mu'_1 = (-0.927, -0.282, 0.243)^\top$, $\mu'_2 = (-0.886, -0.008, -0.464)^\top$ and various γ combinations.

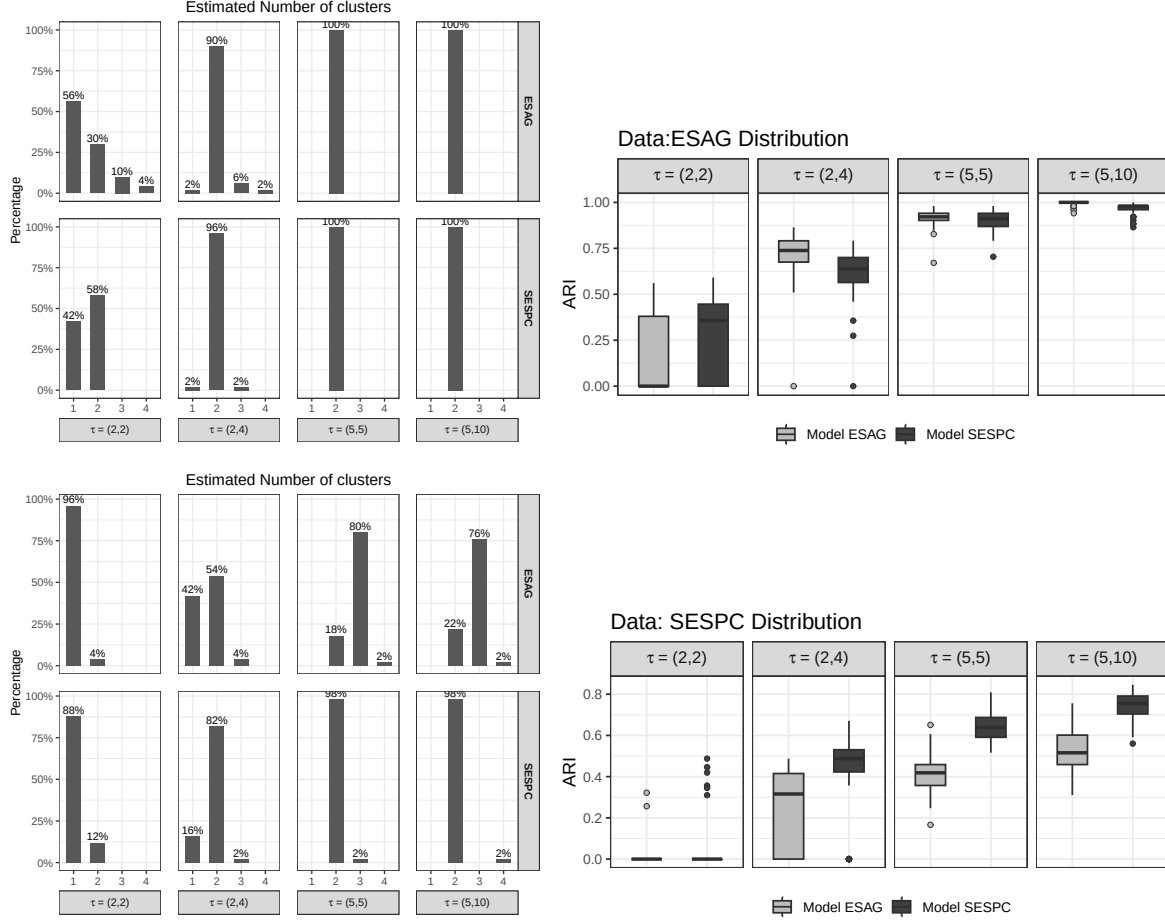


Figure 7: ESAG and SESPC generated data with equal clusters: effect of τ . $n = 200$, $\gamma = [(1, 1)^\top, (1, 1)^\top]$, $\omega = 45^\circ$, $\mu'_1 = (-0.927, -0.282, 0.243)^\top$, $\mu'_2 = (-0.886, -0.008, -0.464)^\top$ and $(\tau_1, \tau_2) = [(2, 2), (4, 4), (5, 5), (5, 10)]$.

4.2 Numerical analysis for hyper-spherical data

The remainder of this section investigates the performance of the ESAG and SESPC models on hyper-spherical data of dimension $d = 10$, considering the scenarios presented in Figure 3. Specifically, we examine the effect of sample size, cluster size, and cluster configuration on the performance of each model. In Figure 3, analogously to the spherical case ($d = 3$) examined in the previous section, equal cluster probabilities are assumed. From Figure 10 it can be seen that, for $n = 500$, both models have similar performance. Additionally, for $n = \{100, 200\}$ the boxplots of the ARI values corresponding to the SESPC model have smaller variability and lie inside the boxplots which corresponds to the ESAG model. Lastly, the boxplots of ARIs regarding the performance of each model is presented under the assumption of having equal (Figure 8) and unequal (Figure 9) cluster sizes. In both figures, we may observe that the relation between the performance of each model is roughly the same, with the difference that when having unequal clusters, as expected, the width of the boxplots is larger.

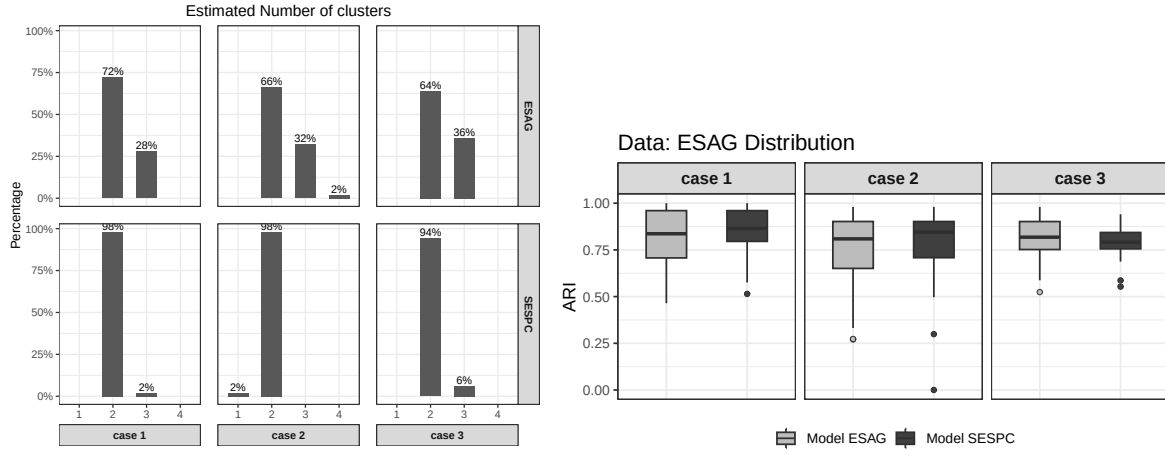


Figure 8: Hyper-spherical ESAG generated data with equal clusters: $n = 200$, $\omega = 45^\circ$, The values of γ_1, γ_2 and μ_1, μ_2 are presented in the Appendix.

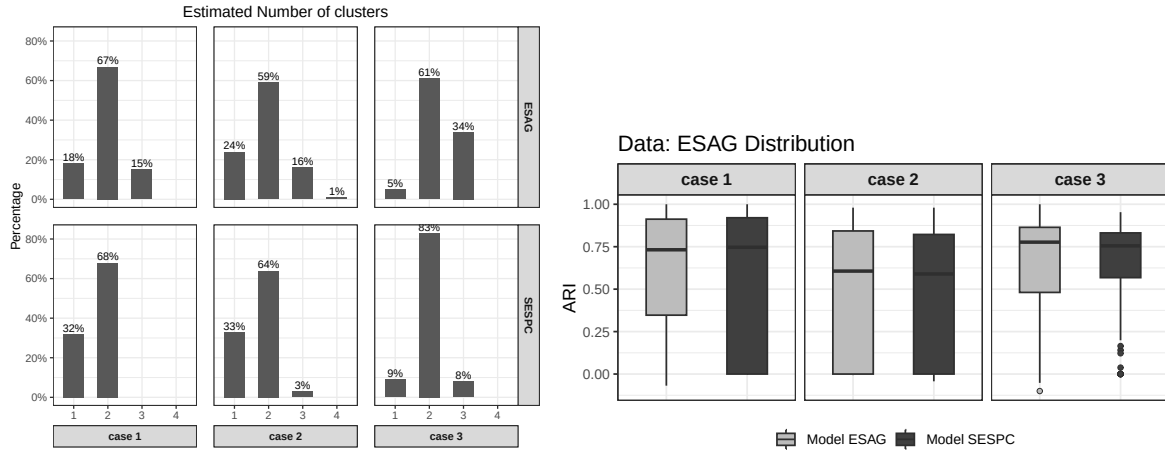


Figure 9: Hyper-spherical ESAG generated data with unequal clusters: $n = 200$, $\omega = 45^\circ$, The values of γ_1, γ_2 and μ_1, μ_2 are presented in the Appendix.

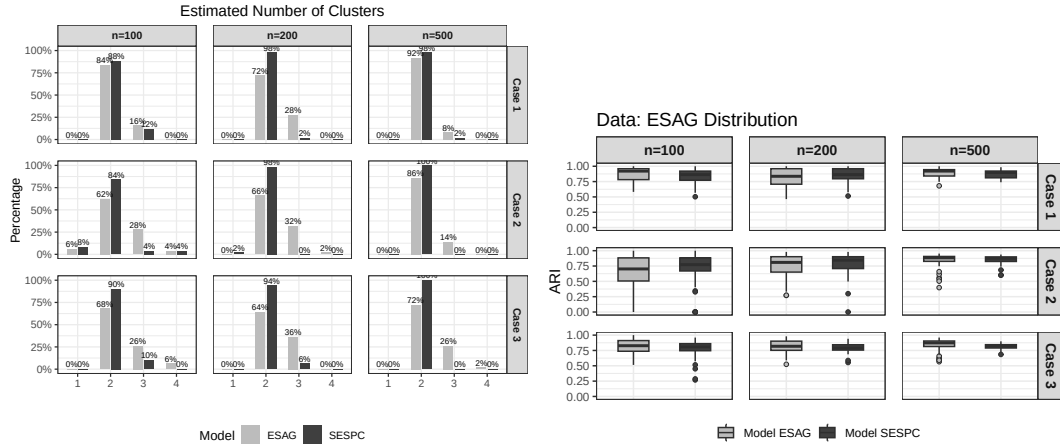


Figure 10: Hyper-spherical ESAG generated data with equal clusters: effect of n . $\omega = 45^\circ$, The values of γ_1, γ_2 and μ_1, μ_2 are presented in the Appendix.

4.3 Computational cost

Figure 11 presents the computational cost for each model as a function of the sample size $n = 200, 500, 1000$. In particular, for each model, 10 iterations were considered and the required time (in minutes) to obtain the results was stored, by using a computer with Intel(R) Core(TM) i7-8550U CPU with 16 GB RAM without performing parallel computing (an option which is available for these computations). For each iteration, each algorithm searches for $K = 1, \dots, 4$ clusters. The box-plots of the needed time for each model for $n = 200, 500, 1000$ are presented in Figure 11. It can be seen that, as expected, as the sample size increases the computational time also increases in both cases. Comparing the two models, it is evident that, regardless of the sample size, the SESPC model is faster.

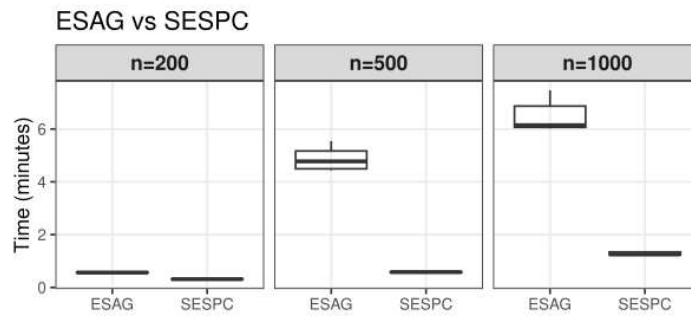


Figure 11: Box-plots of ESAG vs SESPC computational cost of searching for the optimal number of clusters.

5 Real data analysis

Based on the analysis presented in the previous sections, the performance of the two models was evaluated across a range of scenarios involving simulated spherical and hyper-spherical datasets. The findings consistently demonstrate that the SESPC model exhibits strong performance when the underlying data arise from either the ESAG or the SESPC distribution, highlighting its adaptability across different frameworks. Having established the models' behaviour under controlled simulation settings, it would be interesting to assess the performance of both models on real data arising from various directional statistics application domains, including seismological recordings, as well as to explore their potential in clustering settings involving multivariate non-directional data. The first two datasets are located on the sphere, whereas the other two are Euclidean data that were normalized to become unit-vectors and reside on the hyper-sphere. The rest of this Section is outlined as follows: For each example, the true labelled data are visualized on the sphere, providing a reference for the underlying grouping structure. The clustering solutions produced by each model are then examined and compared, allowing us to evaluate which model successfully recovers the true cluster configuration.

5.1 Seismological data

The first two datasets contain locations of the earthquakes that took place (a) in the North American continent during 1900 and 2025¹ and (b) in a cube near Fiji since 1964². For North America we collected the earthquakes whose magnitude was 5 or higher in the Richter scale, whereas for the Fiji region the magnitude of the earthquakes was of 4 Richter or higher. North America contributes with 366 locations from Canada, Dominican Republic, Guatemala, Haiti, Jamaica, Mexico, and USA, with the last two countries containing the 87.43% of the observations. The Fiji region consists of 1,000 locations from New Zealand, Solomon Islands and Vanuatu, but the vast majority (98.3%) refers to earthquakes that happened offshore. Figure 12 shows the locations of the earthquakes.

5.1.1 North America

As it was previously stated our goal is to investigate the performance of each model in real life dataset. Regarding the first dataset to be investigated, for both models, we searched for the optimal number of components K , using $K = 1, \dots, 6$. Regarding the computational cost between the two approaches no significant difference occurred. In particular, fitting the SESPC, the search required 4.75 seconds, whereas the ESAG required 6.51 seconds. Figures 13(a)–13(b) present the ICL values for the mixtures of both distributions using the selected range of K , indicating that both mixtures agree that 4 clusters are enough. Figures 13(c)–13(d) visualize, on the sphere, the 4 clusters identified by each mixture of distributions. Evidently, the two distributions agree in the blue and green clusters, but disagree in the

¹This dataset was downloaded from <https://earthquake.usgs.gov>.

²This dataset constitutes a subsample from a larger dataset of containing 5,000 observations and is available by default in R by the names *quakes*.

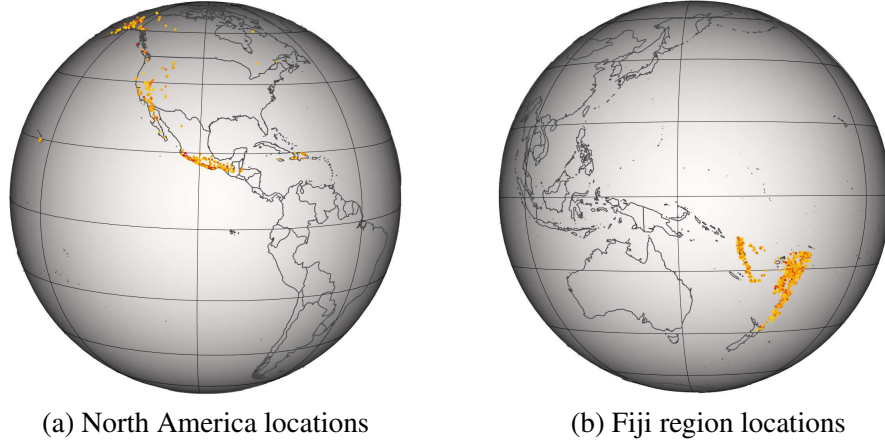


Figure 12: The locations of the earthquakes in North America and Fiji region. Different colors correspond to different levels of magnitude in Richter scale.

red and orange clusters. Hawaii is clustered as green color in both models. Notable differences also occur with some western and eastern USA locations, as indicated by the different colours. It should be highlighted the fact that the SESPC mixtures have discriminated the locations almost perfectly, they are linearly separable, whereas the ESAG mixtures have created overlapping clusters that are not linearly separable.

An alternative visualization of the results is illustrated in Figures 13(e)–13(f), where the locations (latitude and longitude) are placed in a 2D plot. Note that, distances in the 2D plot are not geodesic distances, two points that look far apart may be close on the sphere and vice versa. Since the clustering was done in 3D Cartesian space (or on the sphere directly), a visually comparison of the latitude/longitude coordinates can make a geometrically good clustering look bad and vice versa. For this reason we also projected the data onto the first two principal components of the Cartesian coordinates. Figure 13(g)–13(h) shows the PCA projected data and the results are nearly the same, the 4 groups are nearly linearly separable using the SESPC mixtures.

We repeated the analysis creating a concomitant model, using the magnitude of the earthquakes as a covariate to link to the mixing probabilities. We did not attempt the mixture of regressions model simply because the magnitude is affected by the locations and not vice versa, but it seemed natural to affect the mixing probabilities. The results were nearly the same, the effect of the magnitude was shown to be rather weak.

5.1.2 Fiji region

Since Fiji region contains more locations, we searched for the optimal number of components K , using $K = 1, \dots, 10$. When fitting the SESPC, with parallel processing using 9 cores, the search required 69.98 seconds, whereas the ESAG required 110.83 seconds. Figures 14(a)–14(b) present the ICL values for the mixtures of both distributions using the selected range of K . The SESPC mixtures

selected 4 clusters, whereas the ESAG mixtures were less parsimonious, selecting 7 clusters. Figures 14(c)–14(d) visualize, on the sphere, the clusters identified by each mixture of distributions. Figure 14(e)–14(f) present the PCA projected data with colours identifying the different clusters. Looking at Figures 14(c)–14(f) we can observe that the SESPC mixtures created clusters that are easier to separate in comparison to the clusters created using the ESAG mixtures.

We used again the magnitude of the earthquakes as a covariate in the concomitant model. The BIC and ICL of the SESPC mixtures reduced significantly, yet the concomitant model selected 4 clusters and the groups were affected, but to a small degree. The magnitude had a significant impact on the ESAG mixtures, where this time 6 clusters were selected. Figure 15 displays the ICL for the ESAG concomitant model and the spherical plot of the data with the clusters in different colours.

5.2 Wine quality data

The third dataset is the wine quality dataset³. The dataset contains 12 physicochemical features of red (1,599 observations) and white (4,898 observations) variants of the Portuguese "Vinho Verde" wine. For more details see Cortez et al. (2009).

Figure 16(a) displays the data projected onto the sphere with the 2 groups in different colours. The SESPC and ESAG mixture models both correctly identified 2 clusters and their estimated clusters are displayed in Figures 16(c)–(d). The ESAG is shown to have identified the two clusters more properly and this is evident from the ARI values that were 0.546 and 0.700 for the SESPC and ESAG mixtures, respectively. The SC mixture model identified 13 clusters in the original data and 2 in the projected spherical data. However, the estimated clusters differ significantly as seen in Figure 16(b) and this was evident in the ARI value that was the lowest, only 0.476.

5.3 Wholesale data

The fourth dataset refers to clients of a wholesale distributor⁴. It includes the annual spending in monetary units (m.u.) on 6 diverse product categories. There are two channels of distribution, either hotels/restaurants/catering or retail sales.

Figure 17(a) displays the data projected onto the sphere with the 2 groups in different colours. The SC mixture model identified 14 clusters in the original data and 2 in the projected spherical data. Interestingly, the ESAG mixture model detected 3 clusters, whereas the SESPC model detected 2 clusters, as seen in Figures 17(c)–(d).

6 Conclusions

We proposed mixture models based on two elliptically symmetric distributions defined on the sphere for clustering spherical data. The benefit of these mixture models over rotationally symmetric distri-

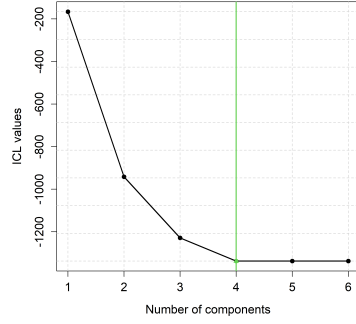
³This dataset was downloaded from <https://archive.ics.uci.edu/dataset/186/wine+quality>.

⁴This dataset was downloaded from <https://archive.ics.uci.edu/dataset/292/wholesale+customers>.

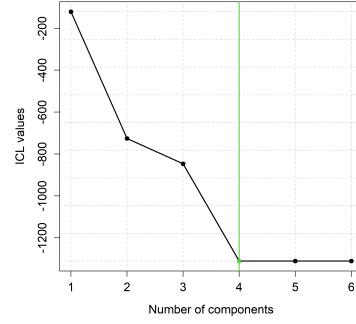
butions is that they escape the isotropic covariance matrix assumption and allow for a more general covariance matrix that produces ellipsoidal contours. To address the dimensionality issues when it comes to fitting we proposed to project the hyper-spherical data onto the sphere and fit the proposed mixture models there.

The simulation studies revealed that when the ground-truth model is the ESAG mixtures, both distributions perform equally well, but if the data are generated from a mixture of SESPC distributions, the SESPC mixture model outperforms the ESAG mixture model. Similar conclusions were derived for the hyper-spherical data cases.

When applied to two real spherical data containing the locations of earthquakes the two mixture models identified the same number of components in North America, albeit not clustering the same locations. The clusters identified by the SESPC mixtures are easier to linearly separate than the ESAG mixtures. With Fiji region though, the SESPC mixtures identified 4 clusters, whereas the ESAG mixtures identified 7 clusters. The concomitant model made no impact on the North America dataset, whereas for the Fiji region it reduced the number of clusters identified by ESAG. When applied to two real hyper-spherical data the proposed mixtures applied to the projected data on the sphere performed better than mixtures of rotationally symmetric distributions applied either on the hyper-sphere or on the sphere.



(a) ICL for SESPC mixtures



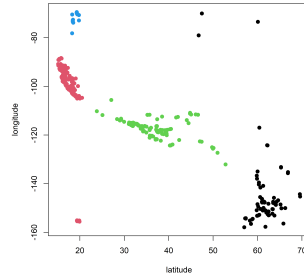
(b) ICL for ESAG mixtures



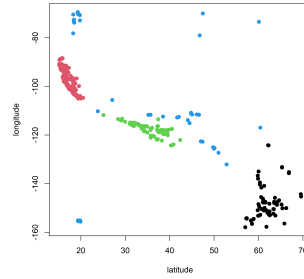
(c) Spherical plot of SESPC mixtures



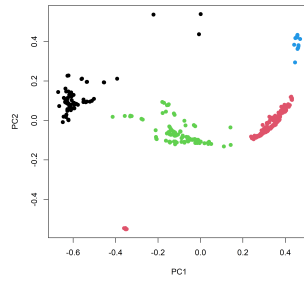
(d) Spherical plot of ESAG mixtures



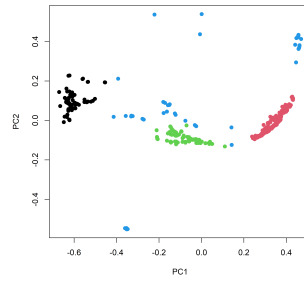
(e) 2D plot of SESPC mixtures



(f) 2D plot of ESAG mixtures

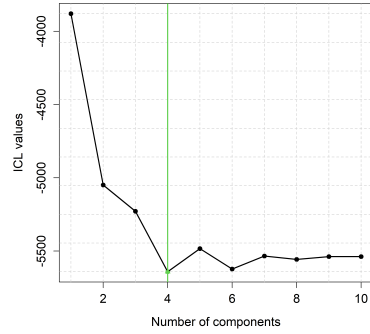


(g) PCA plot of SESPC mixtures

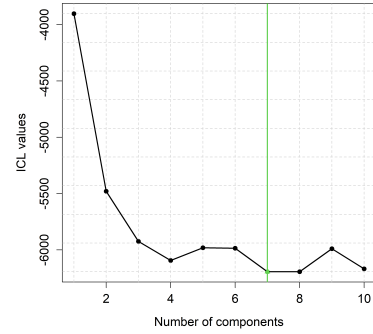


(h) PCA plot of ESAG mixtures

Figure 13: North America: ICL plots for the two mixtures and locations of the earthquakes coloured according to their estimated cluster.



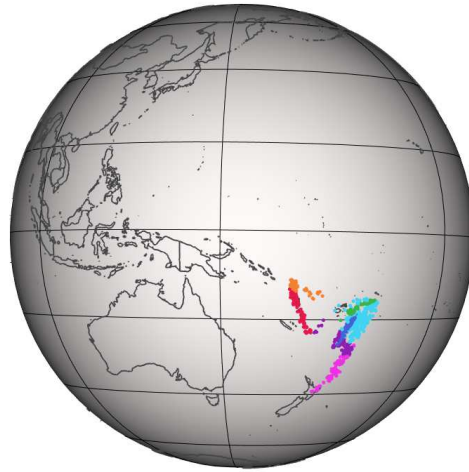
(a) ICL for SESPC mixtures



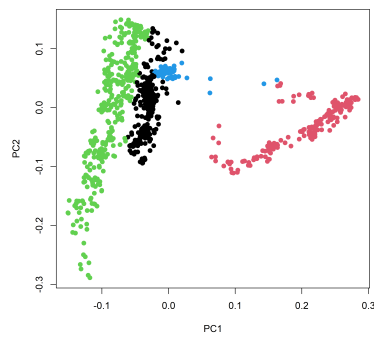
(b) ICL for ESAG mixtures



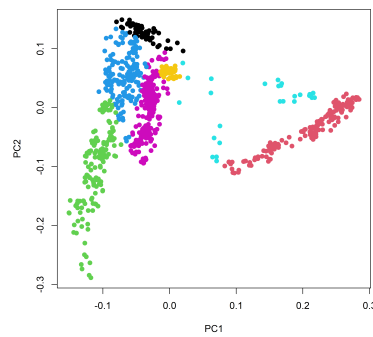
(c) Spherical plot of SESPC mixtures



(d) Spherical plot of ESAG mixtures



(e) PCA plot of SESPC mixtures



(f) PCA plot of ESAG mixtures

Figure 14: Fiji region: ICL plots for the two mixtures and locations of the earthquakes coloured according to their estimated cluster.

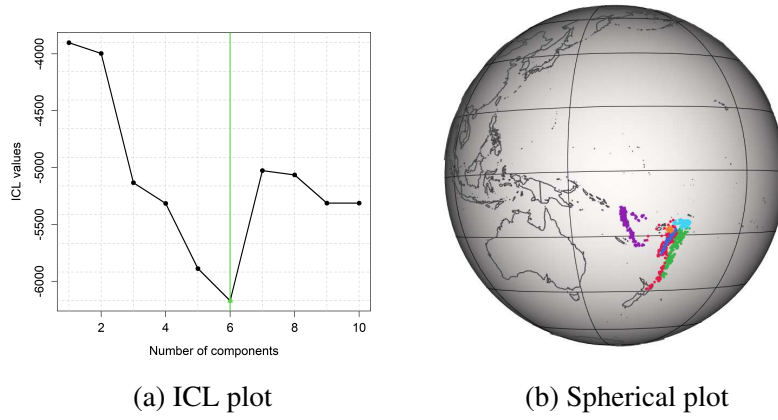


Figure 15: Fiji region: ICL and spherical plot for the ESAG concomitant model.

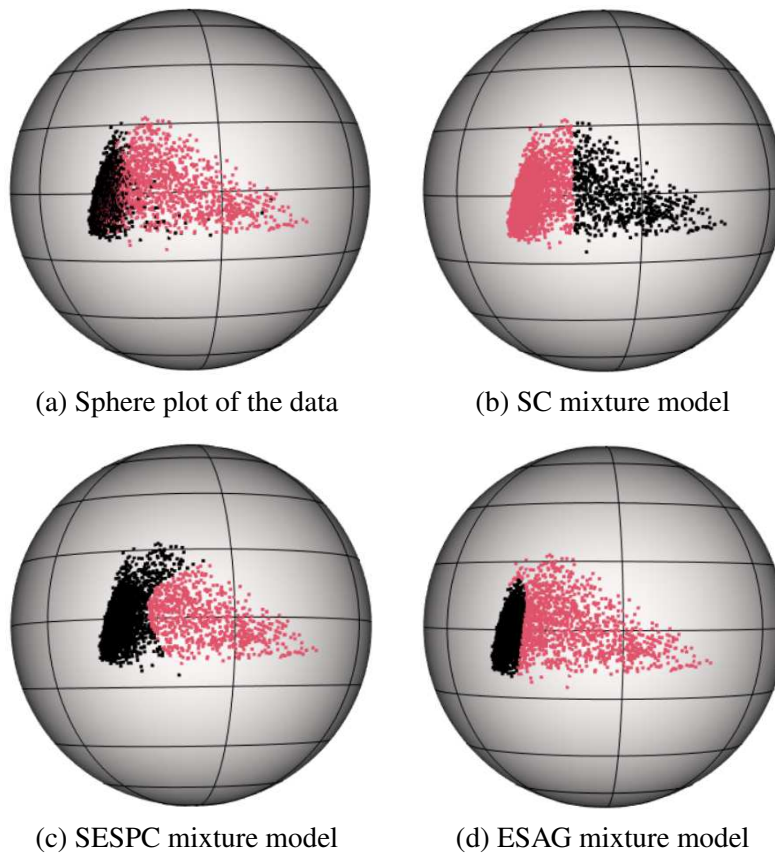
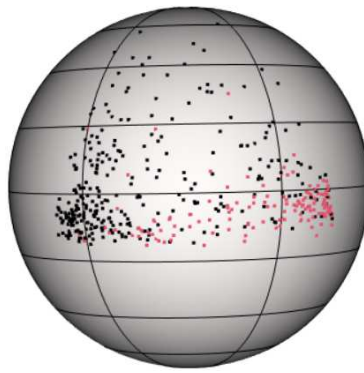
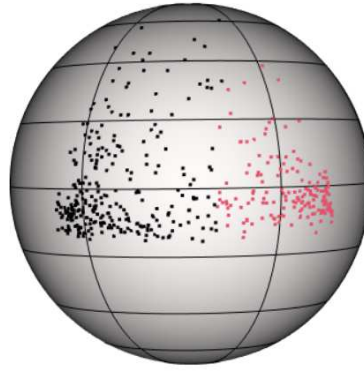


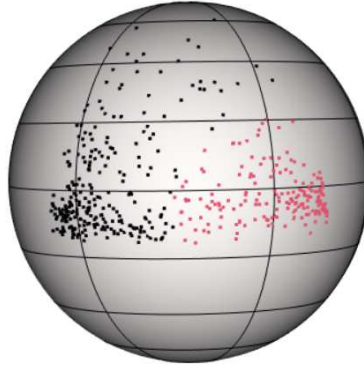
Figure 16: Wine quality data: Sphere plot of the data with the true groups in different colours and the results of the SESPC and ESAG mixture models.



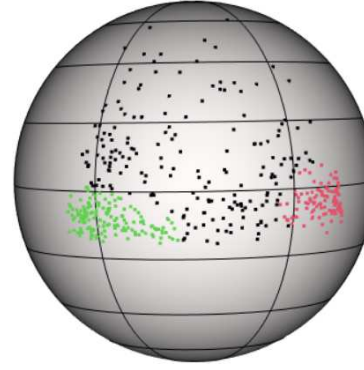
(a) Sphere plot of the data



(b) SC mixture model



(c) SESPC mixture model



(d) ESAG mixture model

Figure 17: Wholesale data: Sphere plot of the data with the true groups in different colours and the results of the SESPC and ESAG mixture models.

References

- Banerjee, A., Dhillon, I. S., Ghosh, J., and Sra, S. (2005). Clustering on the unit hypersphere using von Mises–Fisher distributions. *Journal of Machine Learning Research*, 6:1345–1382.
- Barbaro, F. and Rossi, F. (2021). Sparse mixture of von Mises-Fisher distribution. In *ESANN 2021-European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pages 263–268. Ciaco-idoc.com.
- Baudry, J.-P. and Celeux, G. (2015). EM for mixtures: Initialization requires special care. *Statistics and Computing*, 25:713–726.
- Biernacki, C., Celeux, G., and Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7):719–725.
- Biernacki, C., Celeux, G., and Govaert, G. (2003). Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statistics & Data Analysis*, 41(3-4):561–575.
- Bullock, I. M., Feix, T., and Dollar, A. M. (2014). Analyzing human fingertip usage in dexterous precision manipulation: Implications for robotic finger design. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1622–1628. IEEE.
- Chang, T. (1986). Spherical regression. *The Annals of Statistics*, 14(3):907–924.
- Cortez, P., Cerdeira, A., Almeida, F., Matos, T., and Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, 47(4):547–553.
- Dempster, A., Laird, N., and Rubin, D. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38.
- Dong, A. and Melnykov, V. (2024). Contaminated Kent mixture model for clustering non-spherical directional data with heavy tails or scatter. *Statistics & Probability Letters*, 208:110058.
- Fisher, R. A. (1953). Dispersion on a sphere. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 217(1130):295–305.
- Fletcher, P. T., Lu, C., Pizer, S. M., and Joshi, S. (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23(8):995–1005.
- Fraley, C., Raftery, A., and Wehrens, R. (2005). Incremental model-based clustering for large datasets with small clusters. *Journal of Computational and Graphical Statistics*, 14(3):529–546.

- Fruhworth-Schnatter, S., Celeux, G., and Robert, C. P. (2019). *Handbook of mixture analysis*. Boca Raton: CRC press.
- Golzy, M. and Markatou, M. (2020). Poisson kernel-based clustering on the sphere: convergence properties, identifiability, and a method of sampling. *Journal of Computational and Graphical Statistics*, 29(4):758–770.
- Gopal, S. and Yang, Y. (2014). Von Mises–Fisher Clustering Models. In *International Conference on Machine Learning*, pages 154–162. PMLR.
- Heaton, M., Katzfuss, M., Berrett, C., and Nychka, D. (2014). Constructing valid spatial processes on the sphere using kernel convolutions. *Environmetrics*, 25(1):2–15.
- Hornik, K. and Grün, B. (2014). movMF: An R package for fitting mixtures of von Mises-Fisher distributions. *Journal of Statistical Software*, 58(10):1–31.
- Jung, S., Dryden, I. L., and Marron, J. S. (2012). Analysis of principal nested spheres. *Biometrika*, 99(3):551–568.
- Kagan, Y. Y. (2013). Double-couple earthquake source: symmetry and rotation. *Geophysical Journal International*, 194(2):1167–1179.
- Kagan, Y. Y. and Jackson, D. D. (2015). Likelihood analysis of earthquake focal mechanism distributions. *Geophysical Journal International*, 201(3):1409–1415.
- Karlis, D. and Xekalaki, E. (2003). Choosing initial values for the EM algorithm for finite mixtures. *Computational Statistics & Data Analysis*, 41(3-4):577–590.
- Kato, S. and McCullagh, P. (2020). Some properties of a Cauchy family on the sphere derived from the Möbius transformations. *Bernoulli*, 26(4):3224–3248.
- Kendall, D. G. (1974). Pole-seeking Brownian motion and bird navigation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(3):365–402.
- Kent, J. T. (1982). The Fisher-Bingham distribution on the sphere. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(1):71–80.
- Kent, J. T., Hussein, I., and Jah, M. K. (2016). Directional distributions in tracking of space debris. In *2016 19th International Conference on Information Fusion (FUSION)*, pages 2081–2086. IEEE.
- Mardia, K. V. (1975). Statistics of directional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 37(3):349–371.
- McGraw, T., Vemuri, B. C., Yezierski, B., and Mareci, T. (2006). Von Mises-Fisher mixture model of the diffusion ODF. In *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006*, pages 65–68. IEEE.

- McLachlan, G. (2000). Finite mixture models. *New York: John Wiley & Sons.*
- McNicholas, P. D. (2016). *Mixture model-based classification*. Boca Raton: Chapman and Hall/CRC.
- Michael, S. and Melnykov, V. (2016). An effective strategy for initializing the EM algorithm in finite mixture models. *Advances in Data Analysis and Classification*, 10:563–583.
- Paine, P., Preston, S. P., Tsagris, M., and Wood, A. T. (2018). An elliptically symmetric angular Gaussian distribution. *Statistics and Computing*, 28(3):689–697.
- Papastamoulis, P., Martin-Magniette, M.-L., and Maugis-Rabusseau, C. (2016). On the estimation of mixtures of Poisson regression models with large number of components. *Computational Statistics & Data Analysis*, 93:97–106.
- Peel, D., Whiten, W. J., and McLachlan, G. J. (2001). Fitting mixtures of Kent distributions to aid in joint set identification. *Journal of the American Statistical Association*, 96(453):56–63.
- Røge, R. E., Madsen, K. H., Schmidt, M. N., and Mørup, M. (2017). Infinite von Mises–Fisher mixture modeling of whole brain fMRI data. *Neural Computation*, 29(10):2712–2741.
- Scealy, J. and Wood, A. T. (2019). Scaled von Mises–Fisher distributions and regression models for paleomagnetic directional data. *Journal of the American Statistical Association*.
- Straub, J., Chang, J., Freifeld, O., and Fisher III, J. (2015). A Dirichlet process mixture model for spherical data. In *Artificial Intelligence and Statistics*, pages 930–938. PMLR.
- Taghia, J., Ma, Z., and Leijon, A. (2014). Bayesian estimation of the von-Mises Fisher mixture model with variational inference. *IEEE transactions on pattern analysis and machine intelligence*, 36(9):1701–1715.
- Tsagris, M. and Alzeley, O. (2025). Circular and Spherical Projected Cauchy Distributions: A Novel Framework for Circular and Directional Data Modeling. *Australian & New Zealand Journal of Statistics*, 67(1):77–103.
- Tsagris, M., Papastamoulis, P., and Kato, S. (2025). Directional data analysis: spherical Cauchy or Poisson kernel-based distribution? *Statistics and Computing*, 35(2):51.
- Walsh, D., Arnold, R., and Townend, J. (2009). A Bayesian approach to determining and parametrizing earthquake focal mechanisms. *Geophysical Journal International*, 176(1):235–255.
- Yao, W. and Xiang, S. (2024). *Mixture Models: Parametric, Semiparametric, and New Directions*. Boca Raton: CRC Press.
- Zehao, Y. and Xianzheng, H. (2024). A new parameterization for elliptically symmetric angular Gaussian distributions of arbitrary dimensions. *Electronic Journal of Statistics*, 18(1):301–334.

Appendix

Choice of the parameters regarding the hyper-sphere ESAG distribution

Since we are investigating a 10-dimensional ESAG distribution the vector γ consists of 44 values ($44 = 0.5(d^2 + d - 2)$). As a result, for each cluster, we consider the vectors γ_1, γ_2 by generating 44 random number from $N(0, 1)$ for each case. In Figure 19, the corresponding vectors γ_1, γ_2 are illustrated, for the three investigated cases presented in Figure 3. As for the choice of the mean vectors μ'_1, μ'_2 , they were derived similarly to the numerical analysis for Spherical data and are presented in Figure 18 using $(\tau_1, \tau_2) = (20, 20)$.

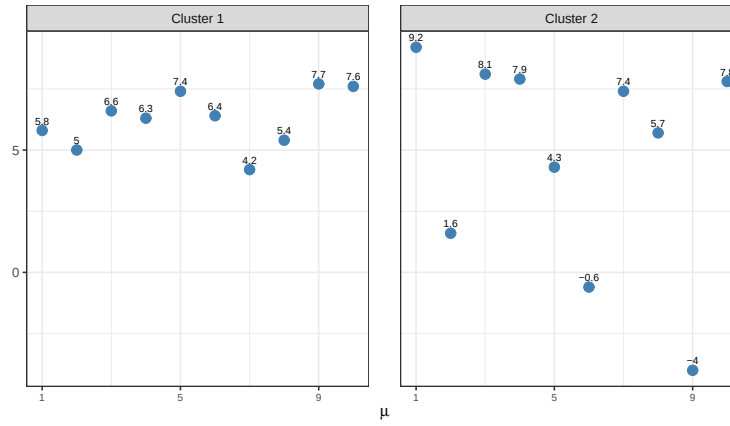
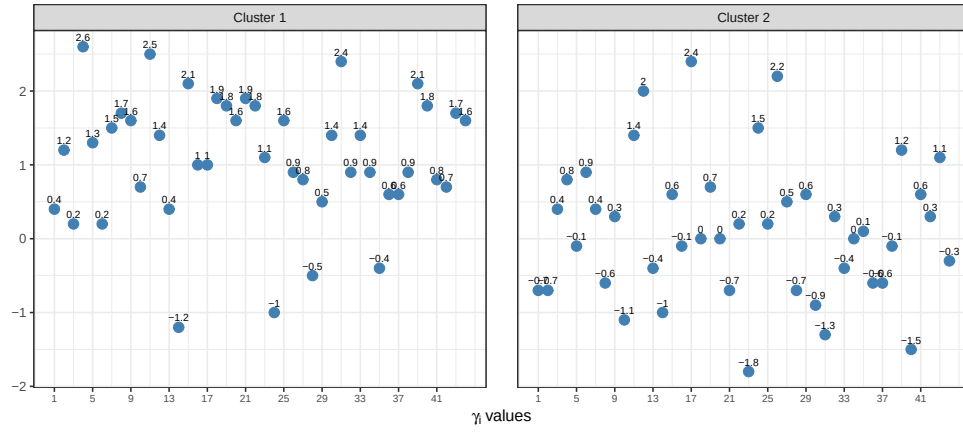
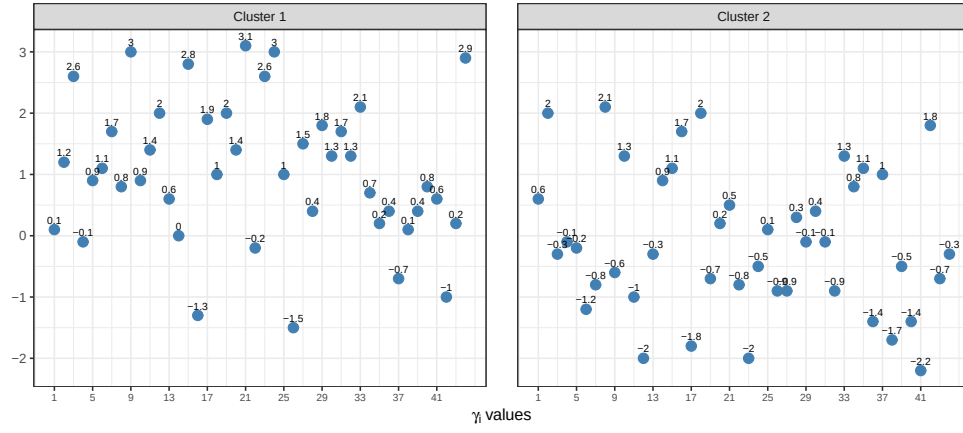


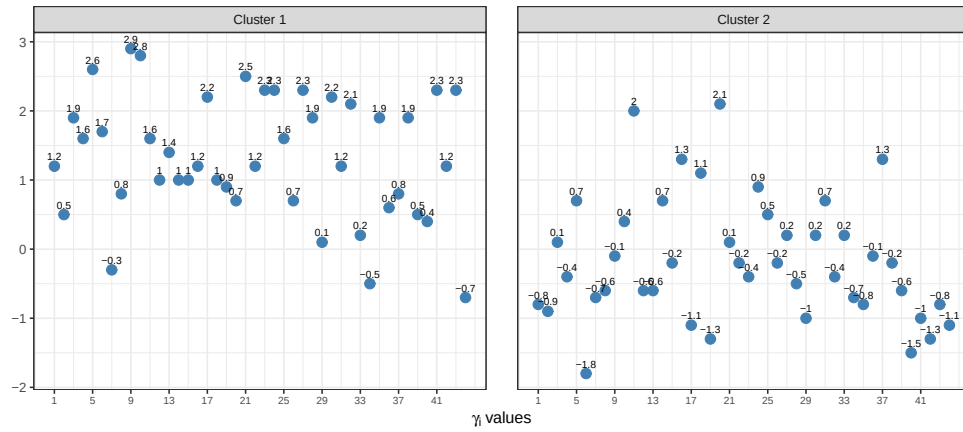
Figure 18: Values of μ'_1, μ'_2 using $(\tau_1, \tau_2) = (20, 20)$ for the 3 cases presented in Figure 3.



(a) Case 1



(b) Case 2



(c) Case 3

Figure 19: Values of γ_1, γ_2 for the 3 cases presented in Figure 3.