



Mixtures of peaked power Batschelet distributions for circular data with application to saccade directions

Kees Mulder^{a,*}, Irene Klugkist^a, Daan van Renswoude^b, Ingmar Visser^b

^a Department of Methodology and Statistics, Faculty of Social and Behavioral Sciences, Utrecht University, Utrecht, Netherlands

^b Department of Psychology, University of Amsterdam, Amsterdam, Netherlands

ARTICLE INFO

Article history:

Received 16 January 2019

Received in revised form 10 October 2019

Accepted 26 November 2019

Available online 1 February 2020

Keywords:

Circular statistics
Peaked distributions
Metropolis–Hastings

ABSTRACT

Circular data are encountered throughout a variety of scientific disciplines, such as in eye movement research as the direction of saccades. Motivated by such applications, mixtures of peaked circular distributions are developed. The peaked distributions are a novel family of Batschelet-type distributions, where the shape of the distribution is warped by means of a transformation function. Because the Inverse Batschelet distribution features an implicit inverse that is not computationally feasible for large or complex data, an alternative called the Power Batschelet distribution is introduced. This distribution is easy to compute and mimics the behavior of the Inverse Batschelet distribution. Inference is performed in both the frequentist framework, through Expectation–Maximization (EM) and the bootstrap, and the Bayesian framework, through MCMC. All parameters can be fixed, which may be done by assumption to reduce the number of parameters. Model comparison can be performed through information criteria or through bridge sampling in the Bayesian framework, which allows performing a wealth of hypothesis tests through the Bayes factor. An R package, `flexcircmix`, is available to perform these analyses.

© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Eye movements are commonly used to study aspects of cognition and its development (Henderson, 2003; Itti & Koch, 2001). Eye movements consist of point fixations and movements between fixations called saccades. In particular, eye movements are of paramount importance in studying the top-down division of attention. For a review, see Rayner (2009).

In eye movement research, a major quantity of interest is the *saccade direction*, the angle between two consecutive fixations. For example, one topic of interest using saccade directions investigates the existence of general directional biases (Tatler & Vincent, 2009), such as a preference for saccades along the horizontal axis (Foulsham, Kingstone, & Underwood, 2008) or a preference for leftward saccades (Foulsham, Gray, Nasiopoulos, & Kingstone, 2013). Another topic of interest is eye movement behavior when reading (Rayner, 2009). Furthermore, distributions of eye movement directions are used to assess the closeness of algorithms to human performance on a variety of eye movement tasks, such as visual search (Najemnik & Geisler, 2008) and saccadic decision making (Engbert, Trukenbrod, Barthelmé,

& Wichmann, 2015; Le Meur & Coutrot, 2016; Tatler, Brockmole, & Carpenter, 2017).

Previously, it has been difficult to directly analyze a sample of saccade directions. One pragmatic solution to the difficulty of analyzing saccade directions is categorizing the angles in a number of general directions, such as in Foulsham et al. (2008). However, such analyses have reduced power, provide less precise interpretation, and require arbitrary selection of a method of categorization.

A natural model of saccade directions can be obtained by viewing saccade directions as circular data, that is, data measured in angles. Circular data differ from linear data in the sense that circular data are measured in a periodical sample space. For example, an angle of 1° is quite close to an angle 359° , although linear intuition suggests otherwise. Circular data are frequently encountered in scientific fields as diverse as life sciences (Mardia, 2011), behavioral biology (Bulbert, Page, & Bernal, 2015), cognitive psychology (Kaas & Van Mier, 2006), bioinformatics (Mardia, Hughes, Taylor, & Singh, 2008), political sciences (Gill & Hangartner, 2010) and environmental sciences (Arnold & SenGupta, 2006). In this study, a model will be developed that leans on the field of circular statistics (Fisher, 1995; Mardia & Jupp, 2000; Pewsey, Neuhäuser, & Ruxton, 2013) to provide satisfying inference for saccade direction data.

Previously, Van Renswoude, Johnson, Raijmakers, and Visser (2016) used mixtures of von Mises distributions to model saccade

* Corresponding author.

E-mail addresses: kt.mulder@uu.nl (K. Mulder), i.klugkist@uu.nl (I. Klugkist), D.R.vanRenswoude@uva.nl (D. van Renswoude), I.Visser@uva.nl (I. Visser).

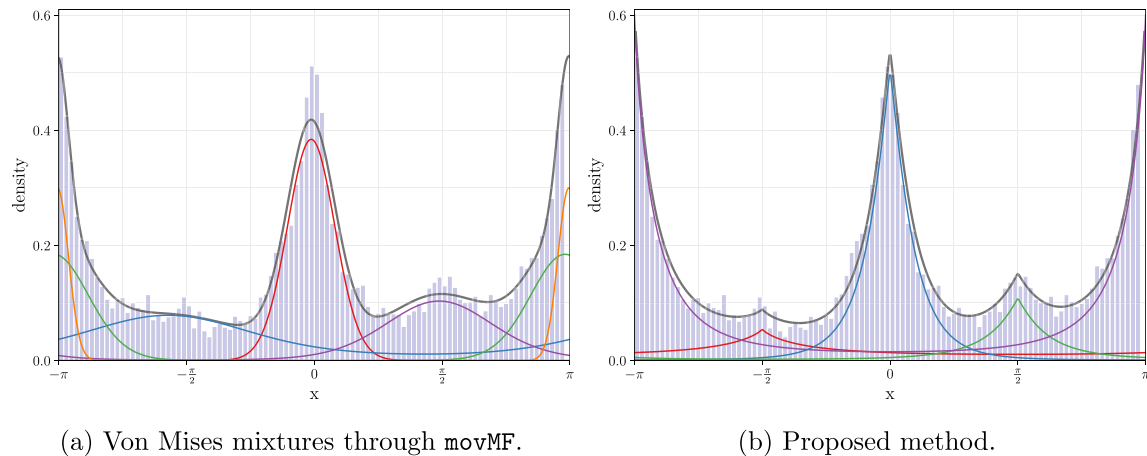


Fig. 1. A comparison of two approaches of analyzing saccade direction. Each colored distribution represents a single component of the mixture model.

direction data, using the R package `movMF` (Hornik & Grün, 2014). The `movMF` package was developed in the more general case for von Mises–Fisher mixtures of distributions on p -dimensional hyperspheres, with circular mixtures resulting as a special case. The van Renswoude saccade direction data and the von Mises mixture fit are displayed in Fig. 1a. It can be seen that the peakedness of the data is not captured well. The peaked mixture model in Fig. 1b is the final model to be introduced in this work. It naturally incorporates peakedness and requires fewer parameters. It can clearly be seen that using the von Mises mixture approach for saccade direction data is not a natural fit, and as such has major drawbacks.

In this paper, four major drawbacks of analyzing saccade directions using the von Mises mixture method of Hornik and Grün (2014) will be addressed. First, the `movMF` approach provides estimates for the parameters of the model by using the Expectation-Maximization (EM) algorithm, but no measure of their uncertainty, such as confidence intervals or standard errors. Second, it can be seen that saccade direction distributions are very often sharp-peaked or flat-topped distributions, which are not directly modeled by this approach. Instead, the mixture model will deal with peaked data by fitting multiple components on a single mode, which precludes interpretation of the component parameters. Third, because the mixture model deals with peakedness by fitting multiple components for a single mode, it is impossible to compare variances of components, which is something of interest in many saccade direction studies, such as in Van Renswoude et al. (2016). Fourth, there is often a desire to fix component means (or other parameters) to pre-specified values, in order to improve power, which is not possible currently.

The model that will be developed in this paper for saccade direction data has two main characteristics. First, it will be a mixture of circular distributions. Second, it will employ flexible distributions in order to naturally model sharp-peaked and flat-topped components. Inference will be developed in a frequentist framework through an EM-algorithm and the bootstrap, and in a Bayesian framework through MCMC sampling. In order to speed up the required computations, a new distribution will be introduced that mimics the behavior of the symmetric density introduced in Jones and Pewsey (2012).

As a motivating example, this study will rely on saccade direction data which was previously published on in Van Renswoude et al. (2016). The main interest in this work is in describing behavioral differences in free-viewing between adults and infants (see also Aslin (2007)). The data consists of 12 367 saccades from adults and 4832 saccades from infants. For details on data

collection, see Van Renswoude et al. (2016). This data is plotted in Fig. 2. Because the hypotheses of interest inform the development of the model, they will be revisited here. First, the researchers are interested in reaffirming a horizontal bias, that is, there are more saccade directions along the horizontal axis than the vertical axis. Second, infants are expected to have larger variance in their saccade directions. Third, the researchers are interested in the difference in the horizontal bias of infants and adults. For all of these hypotheses, currently one would be limited to descriptive analyses. The methods developed in this paper will allow full statistical inference. The methods are available in the R package `flexcircmix`, freely available on GitHub.

The structure of this paper will proceed as follows. In Section 2, the base distribution of the mixture model will be discussed, and the Power Batschelet distribution will be introduced. Inference for the resulting mixture model will be discussed in Section 3. The method will be illustrated on both synthetic data and the van Renswoude data in Section 4. Finally, some concluding remarks will be given in Section 5.

2. Family of Batschelet distributions

In this section, we will introduce Batschelet-type distributions. First, the Inverse Batschelet distribution of Jones and Pewsey (2012) will be recapped. Then, this approach will be adapted into the Power Batschelet distribution. Lastly, a note on computing the circular variance for such distributions will be given.

2.1. Inverse Batschelet distribution

The Inverse Batschelet distribution is a peaked or flat-topped circular distribution. It is constructed by modifying a base distribution, for which the von Mises distribution will be used. The von Mises distribution can be seen as a circular analogue to the normal distribution. In the following, it will be introduced shortly.

Denote the unit circle by \mathbb{S}^1 , and the set of observed angles (in radians) by $\theta = \theta_1, \dots, \theta_n$, with $\theta_i \in \mathbb{S}^1$. For notational simplicity, we assume $\theta \in [-\pi, \pi)$. The von Mises distribution is given by

$$\mathcal{M}(\theta \mid \mu, \kappa) = [2\pi I_0(\kappa)]^{-1} \exp \{ \kappa \cos(\theta - \mu) \}, \quad (1)$$

where θ is the observed angle, $\mu \in [-\pi, \pi)$ is the mean direction, $\kappa \in \mathbb{R}^+$ is a concentration parameter and $I_0(\cdot)$ is the modified Bessel function of the first kind and order zero. Note that this density is periodic, so $\mathcal{M}(\theta \mid \mu, \kappa) = \mathcal{M}(\theta + 2k\pi \mid \mu, \kappa)$, $\forall k \in \mathbb{Z}$. Various von Mises densities are displayed in Fig. 1a as the separate components of the mixture.

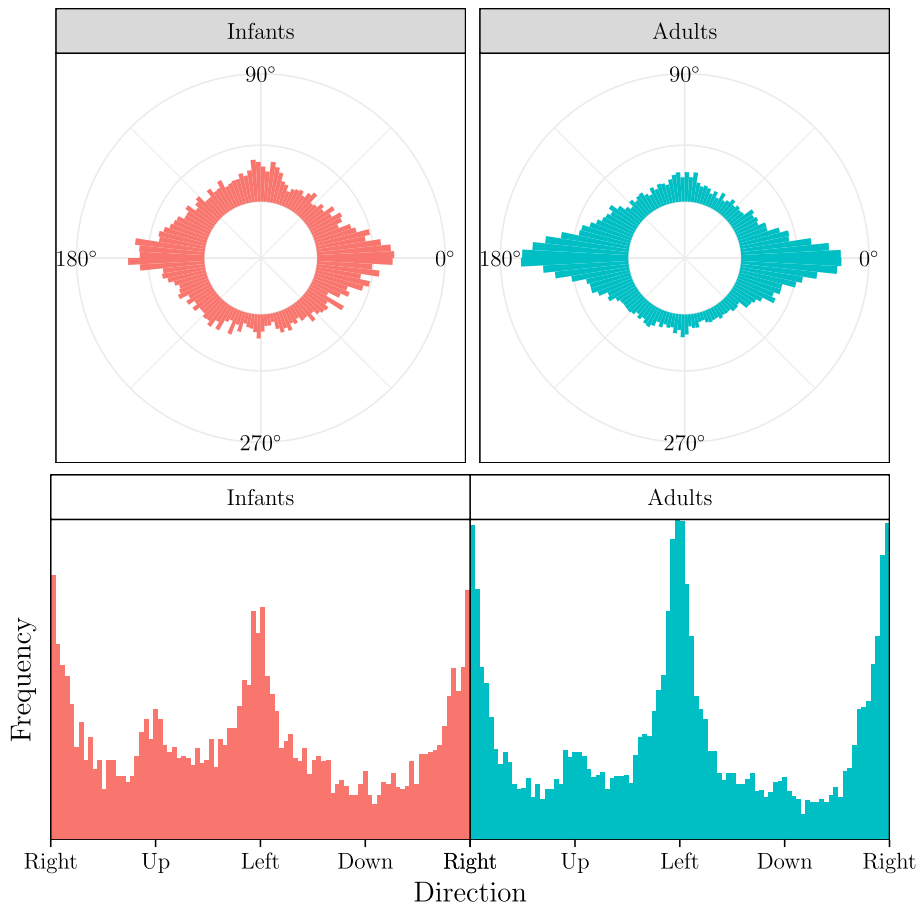


Fig. 2. Plots of the example data. The top plot provides the data in polar coordinates. The bottom plot shows the data on the real line, where the left and right sides of the plot represent the same point on the circle.

Clearly, saccade directions tend to follow more peaked densities than the von Mises density. Two approaches to incorporate peakedness in the model are the Jones–Pewsey distribution (Jones & Pewsey, 2005) and the Inverse Batschelet distribution (Jones & Pewsey, 2012). Both options have the von Mises distribution as a special case. However, the latter is of somewhat simpler form and allows for more peaked distributions, so it will be employed here.

The core idea of the peaked densities developed in Batschelet (1981) is that given a circular density $f(\theta)$, a new distribution emerges if we take $f(\tau(\theta))$ for some bijective function τ which maps the circle onto itself. We will refer to all distributions obtained by this construction as **Batschelet distributions**, possibly with a prefix relating to the specific bijective function τ used. Attention will be limited to using the von Mises density as the base distribution f , such that we will work in practice with $f_\kappa(\tau(\theta - \mu))$, with κ the concentration parameter of the von Mises distribution.

The function originally used by Batschelet (1981) is given by

$$\tau(\theta) = \theta + \lambda \sin \theta, \tag{2}$$

where the peakedness parameter $\lambda \in [-1, 1]$ can be used to obtain a family of flat-topped densities. Using the inverse of $\tau(\theta)$ instead results in a family of peaked densities (Abe, Shimizu, & Pewsey, 2010; Pewsey, Shimizu, & de la Cruz, 2011). A family of densities incorporating both flat-topped and peaked members was developed in Jones and Pewsey (2012) and will be employed here.

The von Mises based symmetric Inverse Batschelet density is given by

$$f(\theta \mid \mu, \kappa, \lambda) = [2\pi I_0(\kappa) K_{\kappa, \lambda}]^{-1} \exp\{\kappa \cos t_\lambda(\theta - \mu)\} \tag{3}$$

where

$$t_\lambda(\theta) = \frac{1 - \lambda}{1 + \lambda} \theta + \frac{2\lambda}{1 + \lambda} s_\lambda^{-1}(\theta) \tag{4}$$

with $s_\lambda^{-1}(\theta)$ being the inverse of $s_\lambda(\theta) = \theta - \frac{1}{2}(1 + \lambda) \sin(\theta)$, and

$$K_{\kappa, \lambda} = \frac{1 + \lambda}{1 - \lambda} - \frac{2\lambda}{1 - \lambda} \int_{-\pi}^{\pi} [2\pi I_0(\kappa)]^{-1} \times \exp\{\kappa \cos(\theta - (1 - \lambda) \sin(\theta)/2)\} d\theta. \tag{5}$$

Note that $t_\lambda(\theta)$ is not available analytically because $s_\lambda^{-1}(\theta)$ is not. Therefore, evaluation of the density requires both numerical integration and numerical inversion. The density is plotted with various values of λ in Fig. 3a. It can be seen that the peaked distribution observed for the saccade data can be obtained from this distribution when $0 < \lambda \leq 1$.

Although for many applications the computational burden of numerically inverting a function for each density evaluation is acceptable, such computations quickly become burdensome upon incorporation of the density into a larger model, such as a mixture model. This can similarly occur when using certain methods for uncertainty quantification, such as MCMC or the bootstrap. Therefore, in order to be able to employ Batschelet distributions in a broader context of models, an alternative to $t_\lambda(\theta)$ will be introduced in the following section. The major advantage will be that the alternative will not require numerical inversion.

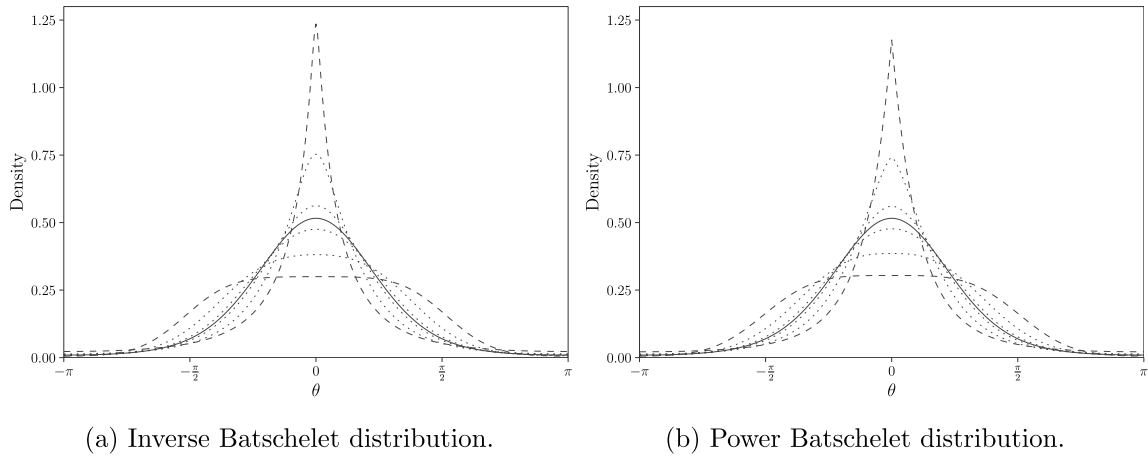


Fig. 3. Two types of Batschelet distributions, based on the von Mises distribution with $\mu = 0$, $\kappa = 2$. In order of increasing height at $\theta = 0$, the peakedness parameter $\lambda = \{-.8, -.4, -.1, 0, .1, .4, .8\}$. In each figure, $\lambda = \{-.8, .8\}$ are dashed, while $\lambda = 0$ is a solid black line, with all others dotted. It can be seen that the densities are extraordinarily similar.

2.2. Power Batschelet distribution

In order to improve computational efficiency in more complex models, $t_\lambda(\theta)$ can be replaced by a function of similar shape, but more appealing computational properties. We propose

$$t_\lambda^*(\theta) = \text{sign}(\theta)\pi \left(\frac{|\theta|}{\pi} \right)^{\gamma(\lambda)}, \tag{6}$$

with $\gamma(\lambda) \in \mathbb{R}^+$, which has the basic properties required of it, namely to be a mapping of the circle onto itself, so long as $-\pi \leq \theta \leq \pi$ is assumed. In practice, this property is generally forced upon the original data θ_o by taking $\theta = [(\theta_o + \pi) \bmod 2\pi] - \pi$. Note this does not change the angle, merely its numerical representation.

Next, the function $\gamma(\lambda)$ should be chosen such that changing λ mimics the behavior of parameter λ of the Inverse Batschelet distribution in Eq. (3). First, in order to keep the parametrization where $-1 \leq \lambda \leq 1$ with negative values corresponding to flat-topped densities, take

$$\gamma(\lambda) = \frac{1 - c\lambda}{1 + c\lambda}, \tag{7}$$

where c is some fixed constant, chosen such that t_λ^* closely approximates t_λ . In order to choose c , the difference between t_λ^* and t_λ was numerically minimized over values of c ,¹ resulting in $c = 0.4052284$. The two functions t_λ^* and t_λ , are plotted together in Fig. 4. It is clear that the functions, although they do not exactly coincide, are strongly comparable. In practical use, the resulting density of the Power Batschelet distribution was found to be evaluated more than several hundred times faster than the Inverse Batschelet distribution. The resulting density is shown in Fig. 3b, where again, we conclude that the densities are strongly similar.

The new continuous function $t_\lambda^*(\theta)$ is trivial to compute and has several attractive properties. For example, note that the we simply have $t_\lambda^{*-1}(\theta) = t_{-\lambda}^*(\theta)$.

Using this function, the Power Batschelet distribution is then defined as

$$f_{PB}(\theta | \mu, \kappa, \lambda) = [K_{\kappa,\lambda}^*]^{-1} \exp\{\kappa \cos t_\lambda^*(\theta - \mu)\}, \tag{8}$$

¹ To be precise, c was chosen such that for a specific λ , the mean absolute difference between $t_\lambda(\theta)$ and $t_\lambda^*(\theta)$ evaluated at 100 points evenly spread on the circle was minimized. The final $c = 0.4052284$ is the average between the optimal c for $\lambda = 1$ and $\lambda = -1$. The two functions do not have to coincide exactly, as being able to directly compare values for λ is only used for interpretability.

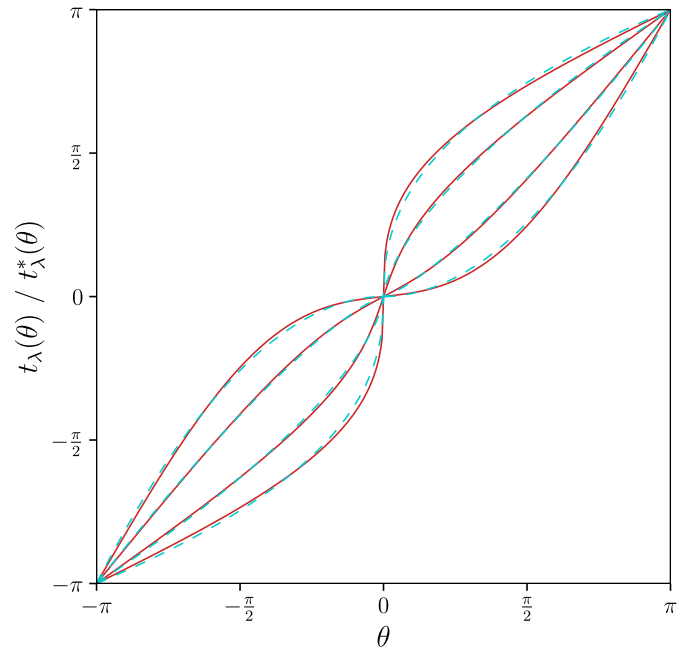


Fig. 4. Comparison of t_λ^* (blue, dashed) used for the Power Batschelet distribution and t_λ (red, solid) used for the Inverse Batschelet distribution, with, in order of increasing height at $\pi/2$, peakedness parameter $\lambda = \{-.8, -.3, .5, 1\}$.

where

$$t_\lambda^*(\theta) = \text{sign}(\theta)\pi \left(\frac{|\theta|}{\pi} \right)^{\frac{1-0.4052284\lambda}{1+0.4052284\lambda}}, \tag{9}$$

and the inverse of the normalizing constant is

$$K_{\kappa,\lambda}^* = \int_{-\pi}^{\pi} \exp\{\kappa \cos t_\lambda^*(\theta - \mu)\} d\theta, \tag{10}$$

which must still be numerically integrated. The Power Batschelet distribution generally shares the properties of the Inverse Batschelet distribution, in that it is symmetric around μ and unimodal. Several further properties of this distribution are discussed in Appendix.

A possible problem is that if $0 < \lambda \leq 1$, we have $0 < \gamma(\lambda) < 1$, and thus $\frac{dt_\lambda^*(\theta)}{d\theta} \Big|_{\theta=0} = \infty$, so the function is not twice differentiable for that range of λ , nor smooth. That is, the probability

density is continuous, and so is $t_\lambda^*(\cdot)$, but its derivative is not, nor is $\frac{df_{PB}(\theta|\mu,\kappa,\lambda)}{d\theta}$. As a result, not all regularity conditions for maximum likelihood estimation are met, in very similar fashion to the commonly used Laplace (double exponential) distribution. In addition, due to the role of this distribution as a close approximation to the Inverse Batschelet, results for the Power Batschelet distribution can be seen as an approximation to the results of the Inverse Batschelet distribution.

If one is concerned about the regularity conditions for the Power Batschelet distribution, the Inverse Batschelet distribution is an alternative which is also implemented in the package `flexcircmix`. However, any analysis with that method may take several orders of magnitude longer. In practice, we have not run into any issues related to this, so we prefer the computational efficiency of the Power Batschelet distribution.

2.3. Measures of circular dispersion

While for the von Mises distribution the circular variance is known to decrease monotonically with increasing κ regardless of the other parameters, this does not hold true for Batschelet distributions, because the peakedness parameter λ also exerts strong influence on the circular variance. However, it is desirable to compare the circular variance across components in the mixture model discussed in the following sections. Therefore, we compute the circular variance v , given by $v = 1 - \rho$, where ρ is the population resultant length associated with the circular density $f(\theta | \phi)$, where ϕ denotes a vector of parameters. In the general case, it is given by

$$\rho = E[\cos \Theta] = \int_{-\pi}^{\pi} \cos \theta p(\theta | \phi) d\theta. \tag{11}$$

If the data has the von Mises distribution, it is known that $\rho = \frac{I_1(\kappa)}{I_0(\kappa)}$ (Mardia & Jupp, 2000), but in general, computing ρ will require numerical integration. Denoting the normalizing constant by $C(\kappa, \lambda) = [\int_{-\pi}^{\pi} \exp\{\kappa \cos t_\lambda(\theta)\} d\theta]^{-1}$, we have

$$\begin{aligned} \rho(\kappa, \lambda) &= E[\cos \Theta] = \int_{-\pi}^{\pi} \cos \theta p(\theta | \mu, \kappa, \lambda) d\theta \tag{12} \\ &= C(\kappa, \lambda) \int_{-\pi}^{\pi} \cos \theta \exp\{\kappa \cos t_\lambda(\theta)\} d\theta. \tag{13} \end{aligned}$$

This means we should need at most two numerical integrations. Lastly, the circular standard deviation can be computed by $\sigma_c = \sqrt{-2 \log \rho}$ (Fisher, 1995).

3. Inference for Batschelet mixtures

The mixture of Batschelet distributions is given by

$$f(\theta | \mu, \kappa, \lambda, \alpha) = \sum_{j=1}^J \alpha_j f_B(\theta | \mu_j, \kappa_j, \lambda_j), \tag{14}$$

where j indexes the J components in the mixture, α_j are component weights, and $f_B(\cdot)$ is the chosen density, either Inverse Batschelet or Power Batschelet.

First, in Section 3.1, an EM algorithm will be presented. Second, in Section 3.2, a method for inference through MCMC is presented. A note on identifiability is given in 3.3. Note that the number of components will be assumed to be known initially. In Section 3.4, model selection and hypothesis testing will be discussed, which can be used to select the number of components as well evaluate many types of hypotheses. For a discussion of direct inference on mixtures with an unknown number of components, see Richardson and Green (1997).

3.1. EM algorithm

Directly maximizing the observed data log-likelihood of a mixture model is generally difficult. Therefore, the EM-algorithm will be employed, which exploits the fact that the complete data maximum likelihood, that is, with observed labels, is easier to maximize.

The EM-algorithm consists of the following steps:

(Initialize) Define an $n \times J$ matrix $\mathbf{W} = \{w_1, \dots, w_J\}^T$, where w_j are n -vectors. Initialize the parameters $\mu, \kappa, \lambda, \alpha$ at some user-specified values.

(E-step) Compute, for all i, j , the elements of \mathbf{W} as

$$w_{i,j} = \frac{\alpha_j f_B(\theta_i | \mu_j, \kappa_j, \lambda_j)}{\sum_{s=1}^J \alpha_s f_B(\theta_i | \mu_s, \kappa_s, \lambda_s)}. \tag{15}$$

(M-step) For each component k , maximize

$$\ell(\mu_j, \kappa_j, \lambda_j, w_j | \theta) = \sum_{i=1}^n w_{i,j} \log f_B(\theta_i | \mu_j, \kappa_j, \lambda_j). \tag{16}$$

The maximization of this log-likelihood of the parameters of a Batschelet distribution may proceed through the Nelder–Mead simplex (Nelder & Mead, 1965), as was done in Jones and Pewsey (2012).

Note that the EM algorithm is not guaranteed to find a global maximum. Therefore, reasonable (or multiple) starting values should be used, in order to assess the validity of the final results. If avoiding convergence to local maxima is of particular importance, one could consider implementing a stochastic version of the EM algorithm (Diebolt & Ip, 1996; Nielsen et al., 2000). However, the Bayesian MCMC approach described in Section 3.2 shares the advantages of such methods.

In addition to obtaining estimates of the mixture model, it is essential to infer the uncertainty around these estimates. For mixture models, asymptotic standard errors obtained from inverting the Fisher Information generally require very large datasets in order to have desirable properties (McLachlan & Peel, 2004). Therefore, parameter uncertainty will need to be assessed either through bootstrapping, or through MCMC.

Bootstrapping (Efron & Tibshirani, 1994) was implemented through a nonparametric bootstrap. In order to reduce computational burden, the EM algorithm of each bootstrap sample was given the full data estimates as starting values.

3.2. Bayesian inference

A Bayesian analysis of the finite mixture of von Mises-based Batschelet distributions is available through MCMC sampling (Chib & Greenberg, 1995; Gilks, Richardson, & Spiegelhalter, 1995). For an introduction focused on mixture models, see Frühwirth-Schnatter (2006). Besides providing uncertainty quantification naturally by performing inference on the posterior distribution rather than a set of estimates, the Bayesian paradigm also provides computational advantages in this case. In particular, the MCMC algorithm is less likely to converge to local maxima.

As is common for Bayesian sampling for mixture models, the parameter space is augmented by a vector of latent variables $\mathbf{z} \in \{1, \dots, J\}^n$ which contains a group label for each observation. After randomly assigning each observation to a group during every iteration, the remaining problem simplifies to MCMC sampling for each component separately. First, in Section 3.2.1, priors for this model will be discussed. Then, the MCMC algorithm will be provided in Section 3.2.2.

3.2.1. Priors

Although subjective priors can be chosen in practical use, attention here will be restricted to (somewhat) non-informative priors. Priors are required for α_j , μ_j , κ_j and λ_j , either jointly or separately. In principle, priors could even be set for the group assignments.

The component weights α_j are given the conjugate Dirichlet prior distribution, with vector prior parameter $\mathbf{n}_0 \in [\mathbb{R}^+]^J$. If $\mathbf{n}_0 = \mathbf{1}_J$, this prior is uninformative.

The mean directions $\mu_j \in [-\pi, \pi)$ are given a circular uniform prior, $p(\mu_j) = [2\pi]^{-1}$ which is proper.

The concentration parameters κ_j are given a constant prior $p(\kappa_j) \propto 1$, which is improper. In principle, the Jeffreys prior for the von Mises distribution could also be used. The Jeffreys prior is proportional to the square root of the determinant of the Fisher Information Matrix $\mathcal{I}(\phi)$, so that for the von Mises distribution it is given by

$$p(\phi) \propto \sqrt{\det[\mathcal{I}(\phi)]} = \sqrt{\kappa A(\kappa) A'(\kappa)}, \quad (17)$$

where $A(\kappa) = I_1(\kappa)/I_0(\kappa)$ and $A'(\kappa) = \frac{d}{d\kappa}A(\kappa)$. However, note that this is *not* the Jeffreys prior for the Inverse Batschelet distribution (as the given $\mathcal{I}(\phi)$ is the Fisher information of the von Mises), nor for a mixture of any circular distributions, nor proper. However, it can be used as a relatively diffuse default prior for cases in which very large values of κ are deemed unlikely. For the case of the von Mises distribution, [Hornik and Grün \(2013\)](#) show that the resulting posterior is almost surely proper if $n \geq 2$. A final alternative is to use a relatively diffuse non-conjugate proper prior, such as one from the gamma family of distributions.

The peakedness parameter λ can be given a proper uniform prior $p(\lambda_j) = 1/2$, $\lambda_j \in [-1, 1)$. However, [Jones and Pewsey \(2012\)](#) note that in maximum likelihood estimation of the Inverse Batschelet model, estimates often fall on the boundary of the parameter space. Boundary avoiding priors can be used here to prevent this behavior of the estimates. In particular, one may posit that large values of $|\lambda_j|$ are a priori unlikely. This belief can be captured in a rescaled Beta(a, b) prior, so that $p(\lambda_j) \propto f_{\text{Beta}}\left(\frac{\lambda+1}{2} \mid a, b\right)$. If $a = 1, b = 1$, this results in the uniform prior on $[-1, 1]$, while $1 < a, b \leq 2$ gives a range of priors which favor smaller values for $|\lambda_j|$, and thus less peaked and less flat-topped densities.

3.2.2. MCMC algorithm

In the application of MCMC sampling, only the group assignments in latent variable \mathbf{z} and the mixture weights α have known full conditional distributions. All other parameters are updated using the Metropolis–Hastings algorithm ([Hastings, 1970](#); [Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953](#)). After selecting starting values, the algorithm will be performed for $m = 1, \dots, M$ iterations, which will constitute a sample from the posterior distribution. One iteration m of the algorithm proceeds as follows:

(z_i) For each observation θ_i , sample $z_i \in 1, \dots, J$ with group probabilities

$$P(z_i = j) = \frac{\alpha_j f_B(\theta_i \mid \mu_j, \kappa_j, \lambda_j)}{\sum_{s=1}^J \alpha_s f_B(\theta_i \mid \mu_s, \kappa_s, \lambda_s)}.$$

This represents assigning this observation to one of the possible mixture components.

(α) Sample the vector of mixture weights

$$\alpha = \alpha_1, \dots, \alpha_j \sim \text{Dirichlet}(\mathbf{n} + \mathbf{n}_0),$$

where $\mathbf{n} = \left\{ \sum_{i=1}^n I(z_i = 1), \dots, \sum_{i=1}^n I(z_i = J) \right\}^T$, with $I(\cdot)$ the indicator function, and \mathbf{n}_0 the vector prior parameter

for the Dirichlet, which is set to $\mathbf{n}_0 = \mathbf{1}_n$ by default for an uninformative prior. Note that the definition of \mathbf{n} means that all we need to do to sample the mixture weights is counting the number of observations assigned to each group.

($\mu_j, \kappa_j, \lambda_j$) For each mixture component $j \in 1, \dots, J$, sample parameters $\mu_j, \kappa_j, \lambda_j$. Note that none of these parameters have known distributions, so we resort to Metropolis–Hastings throughout.

- Sample μ_j using an MH-step. As for the proposal μ_j^* , we make use of the fact that the distribution reduces to the von Mises distribution if $\lambda = 0$. Therefore, we can draw from the known distribution of the mean of the von Mises distribution, because it will be somewhat close to the desired distribution. We can take either the previously sampled μ , by sampling from $\mu_j^* \sim \mathcal{M}\left(\mu_j^{(m-1)}, R_j \kappa\right)$, or use the current sample mean direction $\bar{\theta}$, by sampling from $\mu_j^* \sim \mathcal{M}\left(\bar{\theta}_j, R_j \kappa\right)$, where $\bar{\theta}_j$ and R_j are computed from the sample assigned to component j .
- Sample κ_j using a MH-step, using a gamma distribution with mean $\kappa^{(m-1)}$ as the proposal distribution. The variance of this gamma distribution is a tuning parameter, which can be changed to improve computational efficiency of the algorithm. If the variance is $\kappa^{(m-1)}$, the proposal is the χ^2 -distribution with $\kappa^{(m-1)}$ degrees of freedom. In practice, setting the variance to $.05\kappa^{(m-1)}$ seems to work well.
- Sample λ_j using a MH-step, using a uniform proposal

$$U \left[\max(-1, \lambda_j^{(m-1)} - \varepsilon), \min(1, \lambda_j^{(m-1)} + \varepsilon) \right]. \quad (18)$$

Note that although the proposal distribution seems symmetric, this is not the case if the current value is less than ε from the boundary. Because the proposal is not symmetric, the proposal distribution must be included in the MH ratio. Again, ε is a tuning parameter, and we will set $\varepsilon = 0.01$.

It is well known that if parameters are strongly correlated, MCMC sampling can benefit from joint proposals for the correlated parameters. The parameters κ and λ are correlated, although not in an extreme fashion. Therefore, κ and λ may be sampled jointly, although this did not always prove beneficial in practice.

3.3. Model identifiability

In general, mixture models may not be identifiable ([Teicher, 1963](#)), a property which manifests itself most often through label switching, where component k represents a different unobserved subpopulation in different bootstrap or MCMC samples from the model. However, forcing an ordering on the means may be sufficient for identification ([Everitt, 2004](#)).

For the case of Batschelet mixtures on the circle, means are sometimes fixed by design, because this can be a reasonable assumption for saccade data. If the means are fixed, the model is identifiable as long as $\{\kappa_j = 0, \lambda = 0\}$ in no more than one component. If this assumption is violated, any convex combination of mixture weights α_j of components where $\{\kappa_j = 0, \lambda = 0\}$ gives the same probability density, so the model is not identified. This is unlikely in practice and can simply be checked in the output.

If the means are not fixed but estimated, label switching may occur. In practical inference, if label switching has occurred, this would be evident in bootstrap or MCMC samples. If label switching has occurred, a post-processing step may be used to solve this issue (Jasra, Holmes, & Stephens, 2005; Stephens, 2000). However, care must be taken in ensuring a circular ordering rather than a linear ordering.

3.4. Model selection and hypothesis testing

It is often relevant to compare several models and select the best among them, for example to select the required number of mixture components. The most common approach to model selection is through information criteria such as AIC (Akaike, 1987) and BIC (Schwarz et al., 1978) in frequentist settings, and DIC (Spiegelhalter, Best, Carlin, & Van Der Linde, 2002) and WAIC (Watanabe, 2010) in Bayesian settings (for an overview, see Wagenmakers and Waldorp (2006)). Such tools are provided in the R package `flexcircmix` accompanying this paper and provide an approximate comparison of the fit of various models.

However, the Bayesian approach also allows us to perform more sophisticated model comparisons naturally by comparing the models on their posterior model probability. Consider a set of Q models $\mathcal{M}_1, \dots, \mathcal{M}_Q$ each indexed by a set of free parameters ϕ , that are to be compared on their probability after observing data. This set of models can also take the form of hypotheses to be compared. For example, one could compare a 3-component model with a 4-component model, a model with mean directions fixed at the cardinal directions versus a model where mean directions can vary freely, or a model that allows peaked distributions (i.e. $\lambda \in (-1, 1)$) versus a von Mises mixture model (i.e. $\lambda = 0$).

In order to obtain the posterior model probability, the prior probability of the models under consideration must first be assessed. Here, and throughout the rest of this work, the models will be assumed to have equal prior probability, so $p(\mathcal{M}_s) = 1/Q$. As a result, the prior probability drops out of the rest of the formulae.

Then, regardless of the set of models under consideration, we can compute the posterior model probability

$$\text{pmp}(\mathcal{M}_s) = \frac{p(\mathcal{M}_s | \theta)}{\sum_{q=1}^Q p(\mathcal{M}_q | \theta)} \quad (19)$$

where $p(\mathcal{M}_s | \theta)$ is the *marginal likelihood*, given by

$$p(\mathcal{M}_s | \theta) = \int_{\Omega_\phi} p(\theta | \phi) d\phi, \quad (20)$$

where Ω_ϕ is the sample space of the parameter vector for the model \mathcal{M}_s . This integral is in general not easy to compute and has sparked a wealth of methods for computing it (for an overview, see Ardia, Baştürk, Hoogerheide, and Van Dijk (2012) and Friel and Wyse (2012)). Perhaps the most promising and stable sampling-based solution is found in bridge sampling (Meng & Wong, 1996), which was recently made more easily applicable as a post-processing step on MCMC output through the R package `bridgesampling` (Gronau et al., 2017). Broadly speaking, bridge sampling produces an estimate of the marginal likelihood by evaluating additional samples from a known density that approximates the posterior. For details, see Gronau et al. (2017) and Meng and Wong (1996).

Two issues arise for this specific application. First, the sample of mean direction parameters μ_j lies on a circular parameter space. Bridge sampling will find a known density that approximates the posterior by using the linear mean and the covariance matrix of the MCMC samples, for example by using the multivariate normal density with the same mean vector and covariance matrix. The approximation need only be roughly correct, which

is not necessarily the case for our model. For example, if we have a circular parameter with a mean direction near zero, some sampled values will lie in both intervals $[0, .1]$ and $[2\pi - .1, 2\pi]$. The linear mean will then incorrectly lie near π , and the linear variance will be far too large. To solve this, we will change the numerical representation of the mean direction sample of μ_j such that it lends itself better to the linear approximation. To do this, first the posterior mean direction $\bar{\mu}_j$ is computed from the sample of mean directions μ_j . Then, by taking $\mu_j^* = [(\mu_j - \bar{\mu}_j + \pi) \bmod 2\pi] - \pi + \bar{\mu}_j$, a numerical representation is obtained that does not have any 'gaps' on the real line, but corresponds to the same set of angles μ_j .

The second issue is that the sample of component weight parameters $\alpha_j \in [0, 1]$ lie on a simplex, that is, they are constrained to sum to one. The bridge sampling usually lies on the real line, such as the aforementioned multivariate normal distribution, which means the constrained parameter space is ignored, so that almost surely invalid proposals are sampled. Therefore, these parameters are given a stick-breaking representation and are then logit-transformed, in a similar manner as in Stan (Carpenter et al., 2017). For details on the transformation, its inverse and associated Jacobian, see the Stan reference manual (Stan Development Team, 2017). The solutions for both circular and simplex parameters were contributed to the latest version of the `bridgesampling` package.

4. Illustration

In order to illustrate the methods presented in this work, they will be applied to two examples.

First, the method is applied to a synthetic dataset in Section 4.1, where it is shown that the true parameters of a data generating process can be recovered. Then, in Section 4.2, the method is shown to provide new insights in the saccade direction data from Van Renswoude et al. (2016).

4.1. Synthetic data

Here, the methods developed in this paper will be applied to a synthetic data set for which parameter values are known. In order to sample from the Inverse Batschelet distribution, the sampling algorithm from Jones and Pewsey (2012) was applied. A data set consisting of 1000 angles was sampled with parameters $\mu = \{-1, 1, 2\}$, $\kappa = \{20, 4, 15\}$, $\lambda = \{-.7, 0, .7\}$, and $\alpha = \{.25, .25, .5\}$.

The results are shown in Table 1 and Fig. 5. First, it is clear that the method is able to recover mean direction. Also, it can be seen that both the bootstrapped confidence intervals and the credible intervals generally include the true value, and cases in which this is not true can be attributed to sampling error.

As mentioned previously, the joint likelihood of $\{\kappa, \lambda\}$ is correlated. Because of this, it can be seen in Table 1 that neither κ nor λ can be estimated precisely, with both having confidence intervals that are quite wide. A remarkable property of this method is that while neither of the variance-related parameters is estimated very precisely, the circular standard deviation, computed as in Section 2.3, has tighter confidence intervals and is estimated more precisely, so inference on it will be more powerful than inference directly on κ or λ .

The components weights are estimated adequately in all cases.

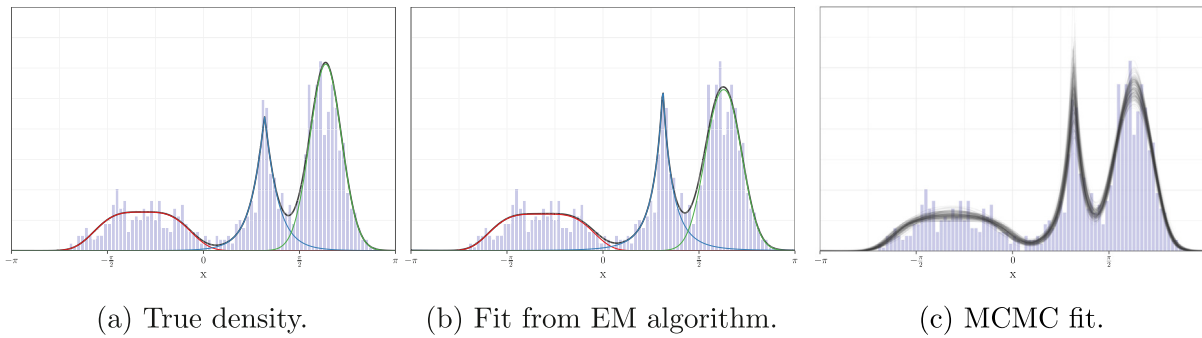


Fig. 5. Synthetic data plots using the power Batschelet distribution. The sample of synthetic data, plotted along with a sample of 200 densities of which the parameters were sampled in the MCMC. That is, the spread of probability density functions provides a rough uncertainty bound for the true probability at each point.

Table 1
Synthetic data fits using the Power Batschelet distribution.

	Truth	EM	Boot. CI		Bayes (MCMC)		
		Est.	2.5%	97.5%	Median	2.5%	97.5%
μ_1	-1.00	-1.00	-1.09	-0.91	-1.00	-1.04	-0.96
κ_1	20.00	24.92	8.67	63.37	27.67	8.91	59.99
λ_1	-0.70	-0.81	-1.00	-0.47	-0.85	-0.99	-0.51
α_1	0.25	0.26	0.23	0.28	0.26	0.23	0.29
σ_{c1}	0.52	0.55	0.51	0.60	0.56	0.51	0.60
μ_2	1.00	0.98	0.95	1.01	1.00	0.97	1.03
κ_2	4.00	3.26	3.17	3.94	3.52	2.80	5.17
λ_2	0.70	0.89	0.66	1.00	0.81	0.54	0.98
α_2	0.25	0.26	0.23	0.29	0.27	0.23	0.31
σ_{c2}	0.34	0.42	0.33	0.44	0.38	0.27	0.53
μ_3	2.00	1.97	1.95	2.00	1.98	1.96	2.00
κ_3	15.00	16.29	10.87	44.57	25.08	11.37	67.42
λ_3	0.00	-0.09	-0.36	0.08	-0.22	-0.47	0.06
α_3	0.50	0.48	0.45	0.51	0.48	0.44	0.51
σ_{c3}	0.26	0.28	0.26	0.30	0.28	0.25	0.30

4.2. Free-viewing data

Here, the method will be applied to a real world example, the free-viewing dataset that was originally published in Van Renswoude et al. (2016), shown again in Fig. 6. As discussed in Section 1, there are several hypotheses one might wish to learn about using this dataset. One hypothesis of interest is whether

infants have a larger circular variance for each of their mixture components. Another is whether the horizontal bias, that is, the preference for left-right movements, is weaker for infants than for adults. These hypotheses will be assessed here.

Because the mixture model has a fairly large number of parameters, it can be fruitful to consider fixing parameters about which we do not need to learn. For free-viewing data, modes are always observed oriented exactly in the cardinal directions, which will simplify our modeling problem somewhat. The mean directions can be chosen to be fixed at $\mu_1 = -\pi/2$ (upward), $\mu_2 = 0$ (rightward), $\mu_3 = \pi/2$ (downward), $\mu_4 = \pi$ (leftward). Because fewer components are needed in the Batschelet mixture and because the means are fixed, the model actually has fewer parameters than the von Mises mixture model. It is also possible to loosen this assumption slightly by placing a strong prior on the mean directions centered on the aforementioned cardinal directions.

For Bayesian inference, the priors were chosen according to the considerations in Section 3.2.1. To be specific, the prior for μ_j was circular uniform, for κ_j the Jeffreys prior of the von Mises distribution, for λ_j the prior was the rescaled beta distribution proportional to $f_{\text{Beta}}\left(\frac{\lambda+1}{2} \mid \sqrt{2}, \sqrt{2}\right)$, and finally the prior for α_j was Dirichlet($\alpha = \{\sqrt{2}, \dots, \sqrt{2}\}^T$). The MCMC algorithm was run for 46 000 iterations, split into 46 parallel chains each having a burn in of 1000.

A bootstrap was run with 10 000 bootstrap replications. For both adults and infants, the results from the EM algorithm with

Table 2

Free-viewing fit using the Power Batschelet mixture model, for adults (left) and infants (right). Mean directions were fixed at $\mu_1 = -\pi/2$ (upward), $\mu_2 = 0$ (rightward), $\mu_3 = \pi/2$ (downward), $\mu_4 = \pi$ (leftward).

		Adults						Infants					
		EM		Boot. CI		Bayes (MCMC)		EM		Boot. CI		Bayes (MCMC)	
		Est.	2.5%	97.5%	Median	2.5%	97.5%	Est.	2.5%	97.5%	Median	2.5%	97.5%
Up	κ_1	0.81	0.69	1.03	5.94	0.85	666.66	0.38	0.25	0.91	6.76	0.14	635.15
	λ_1	1.00	1.00	1.00	0.41	-0.53	0.97	-1.00	-1.00	0.17	-0.56	-0.97	0.55
	α_1	0.11	0.09	0.12	0.02	0.01	0.12	0.15	0.11	0.18	0.05	0.00	0.28
	σ_{c1}	1.54	1.37	1.64	0.32	0.09	1.46	1.96	1.46	2.13	0.68	0.03	2.33
Right	κ_2	3.05	2.61	3.50	2.46	2.09	3.39	2.34	1.90	3.75	2.22	1.49	3.78
	λ_2	0.67	0.56	0.78	0.77	0.58	0.91	0.72	0.40	0.90	0.67	0.38	0.93
	α_2	0.32	0.31	0.35	0.36	0.31	0.40	0.27	0.23	0.32	0.32	0.23	0.43
	σ_{c2}	0.48	0.42	0.58	0.62	0.44	0.76	0.65	0.43	0.84	0.69	0.43	1.05
Down	κ_3	1.93	1.65	3.34	4.33	1.51	30.56	1.41	1.21	1.94	2.14	0.90	18.24
	λ_3	0.76	0.44	0.90	0.41	-0.20	0.85	0.99	0.61	1.00	0.65	-0.09	0.96
	α_3	0.10	0.07	0.12	0.06	0.05	0.14	0.18	0.15	0.20	0.13	0.06	0.34
	σ_{c3}	0.80	0.47	0.95	0.38	0.23	1.01	1.12	0.80	1.23	0.73	0.25	1.42
Left	κ_4	1.84	1.70	1.97	1.62	1.50	2.02	1.34	1.18	1.49	1.28	0.98	2.06
	λ_4	0.96	0.88	1.00	0.97	0.87	1.00	0.98	0.82	1.00	0.88	0.61	0.99
	α_4	0.47	0.44	0.50	0.54	0.43	0.58	0.41	0.37	0.46	0.47	0.28	0.61
	σ_{c4}	0.89	0.81	0.96	1.00	0.79	1.07	1.16	1.04	1.26	1.16	0.75	1.39

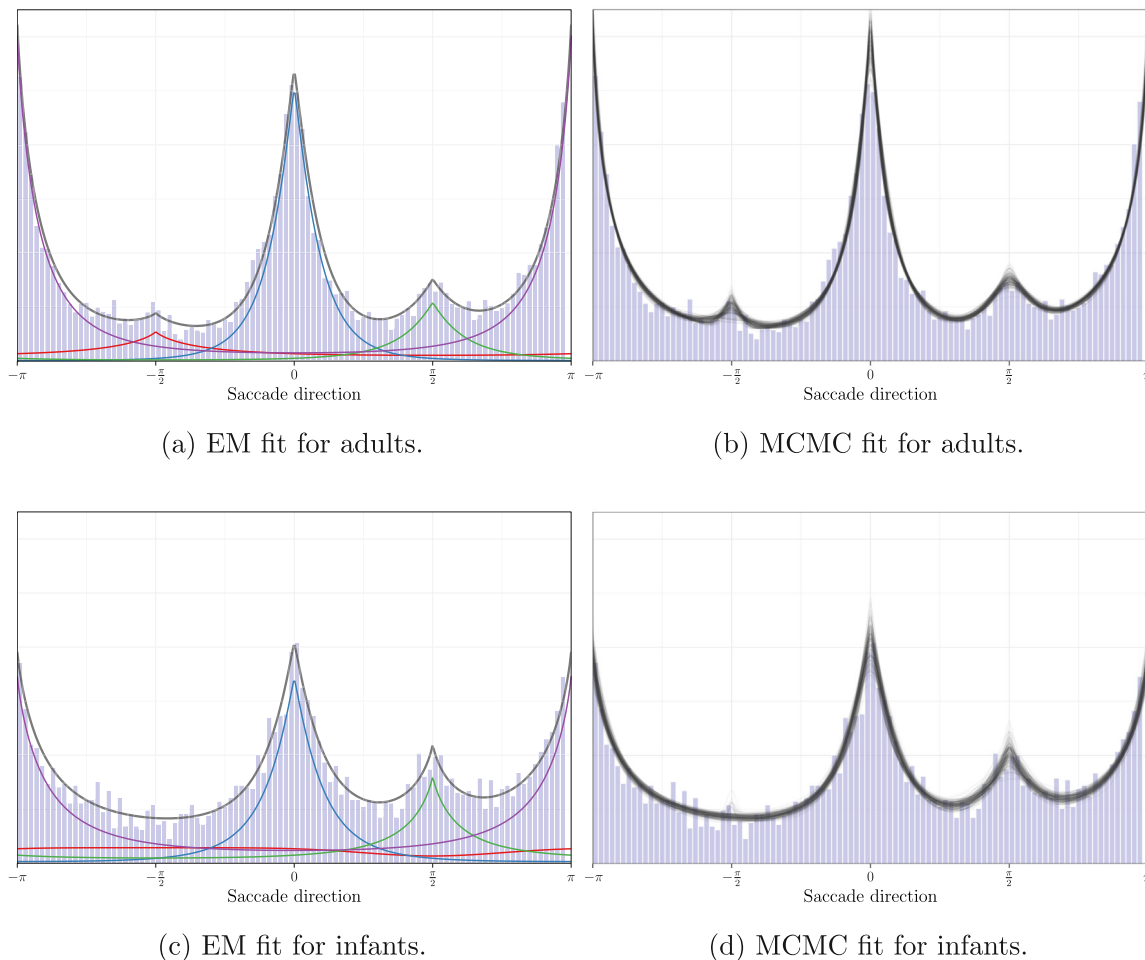


Fig. 6. Free-viewing data fit for adults (top) and infants (bottom) using the power Batschelet distribution. The x -axis is given in radians, where 0 corresponds to the rightward direction, while $\pi/2$ corresponds to the downward direction. In the left plot, it can be seen that a third component is estimated for adults, while this is not the case for infants. In the right plot, the densities are plotted that result from sampled parameter sets from the MCMC.

bootstrapped standard errors will be displayed together with the Bayesian approach in Table 2.

4.2.1. Adults

Results for the adult sample are displayed in Figs. 6a and 6b. Visually, the model fit seems excellent, and it can be seen that observed distributions are generally quite peaked. For adults, it can be seen that all four components contribute to the overall shape of the model.

4.2.2. Infants

For infants, judging from the convergence plots in Fig. 7, there might be grounds to assume that fewer than four components may suffice. Specifically, we can see that component 1 (upward) has component weight α that tends to zero. This can also be seen in Fig. 6c, where the red (upward) component 1 is taken as almost completely flat.

4.2.3. Horizontal bias comparison

The main question of whether the horizontal bias of adults and infants differ can be addressed by comparing the circular standard deviation in Table 2, as well as compare the component weights α .

The horizontal components are component 2 and 4. For the rightward component 2, with ($\mu_2 = 0$), the estimates are generally somewhat similar between adults and infants, as the confidence and credible intervals overlap.

For the leftward component 4 ($\mu_4 = \pi$), the confidence and credible intervals of adults and infants do not overlap, which can also be observed in Fig. 6 by noting that this component has a different shape between Figs. 6c and 6a. For the component weight α_4 , adults have EM-estimate and 95% bootstrap confidence interval $\hat{\alpha}^{(EM)} = 0.465$ (0.443, 0.502) and posterior median and credible interval $\hat{\alpha}^{(MCMC)} = 0.542$ (0.431, 0.58), compared to infants which have $\hat{\alpha}^{(EM)} = 0.405$ (0.37, 0.455) and $\hat{\alpha}^{(MCMC)} = 0.472$ (0.28, 0.614).

For the circular standard deviation σ_{c4} , adults have $\hat{\sigma}_c^{(EM)} = 0.886$ (0.812, 0.964) and $\hat{\sigma}_c^{(MCMC)} = 0.995$ (0.786, 1.07), compared to infants which have $\hat{\sigma}_c^{(EM)} = 1.16$ (1.036, 1.258) and $\hat{\sigma}_c^{(MCMC)} = 1.161$ (0.749, 1.391). Therefore, infants seem to have a larger variance on this component than adults. From this, it can be concluded that the horizontal bias exists, and differs between adults and infants.

4.2.4. Hypothesis testing

Using the model comparison methods discussed in Section 3.4, several models of interest can be compared using Bayesian hypothesis tests.

First, it is worthwhile to investigate whether in general infants and adults differ. This can be done by comparing the model that is discussed in Section 4.2.3, which allows separate parameters for infants and adults, to a model where both are given the same parameters (for which parameter estimates are not shown). For this model, the log Bayes Factor in favor of the model that has

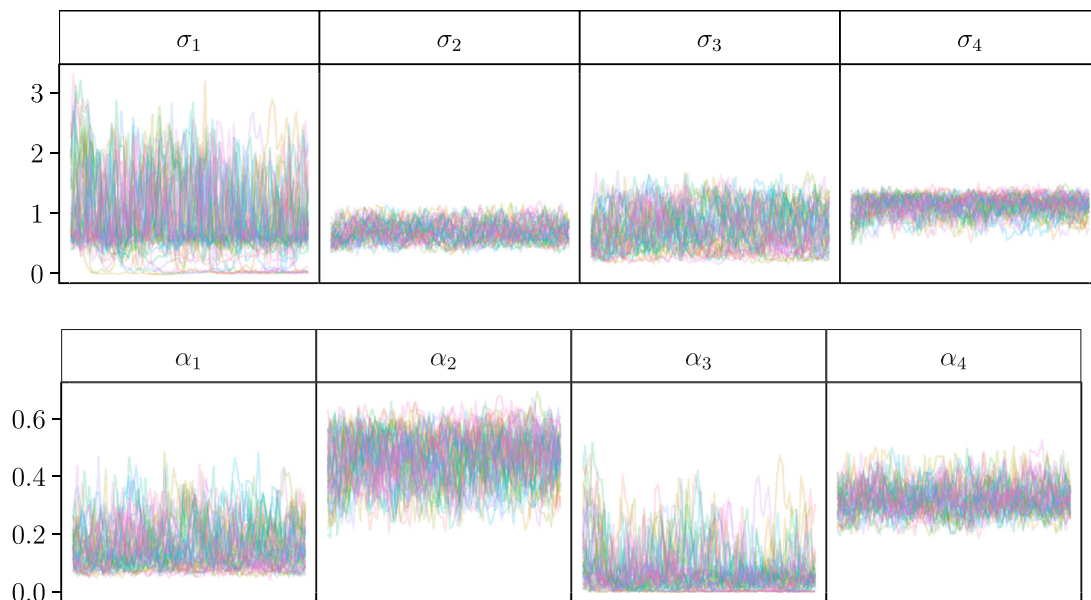


Fig. 7. Convergence plot for the infant data. The plot shows 46 chains of 1000 MCMC iterations, with a thinning factor of 50.

separate parameters is 52.7, which is associated with a posterior model probability (assuming equal prior odds) in favor of separate parameters close to 100%. Therefore, we have separated the groups throughout.

Second, as can be seen in Fig. 6c, one may be unsure about the number of required components for infants. Although four components were assumed so far, three may suffice. Again, we find almost certain evidence, log Bayes Factor 70.7, posterior probability $\approx 100\%$, that the model with four components fits better than the model with only three components.

Finally, the assumption that the means can be fixed to the cardinal directions can be checked. To test whether this was a valid assumption, a model with free means is run for both adults and infants (parameter estimates not shown). This model is then compared to the model with fixed means from before, which gives a log Bayes factor of 61 in favor of free means. This suggests that the data can be fit better by allowing the means to be freely estimated. This could be understood as a systematic bias of the means away from the cardinal directions contained in the stimuli used. However, for the sake of simplicity, the means were kept fixed throughout this paper.

5. Discussion

In this paper, a new mixture model for flexible circular distributions was developed. It can be used to distinguish clusters in samples of directions, for example those obtained from saccade directions. The main contribution is the development of mixture models for circular data that allow for peaked and flat-topped shapes. In order to do this, a new family of distributions was introduced, the Power Batschelet distributions, that mimic the distributions developed in Jones and Pewsey (2012), but that enjoy more appealing computational properties.

The method developed here can be used as a method to investigate whether two sets of saccade directions differ from each other, as shown in Section 4.2. This allows eye-tracking researchers to answer more complex questions about their saccade data, and draw inference where this was previously not possible.

In this paper, a Bayes Factor was computed comparing a model where groups have equal parameters with a model where groups each have their own parameters. We have allowed all parameters to differ between groups, while it is also common to vary only

a set of parameters of interest. A typical example is ANOVA-type models where means are different between groups but variances are set equal. In the current study all parameters were estimated separately because all free parameters were assumed to differ between infants and adults. We investigated whether the component variances differed between groups, which are jointly determined by κ and λ . In addition, we assumed that the mixture weights would vary along with the variances. However, in some cases ANOVA-type models may be required, but they would only require minor adaptation of our methods, for example by adapting the M-step of the EM algorithm or jointly evaluating the component posteriors in the MCMC.

Although developed in the context of saccade directions, the method has potentially much broader applications. For example, wind directions are commonly modeled with circular distributions (Bao, Gneiting, Gritter, Guttorp, & Raftery, 2010; Bowers, Morton, & Mould, 2000; Holzmann, Munk, Suster, & Zucchini, 2006), and sometimes feature peaked distributions. Also, observed arrival times can sometimes be governed by an event occurring at a single time point, causing strongly peaked distributions as well. Also, the method is a strong contender for any form of nonparametric fit on a set of univariate circular data. In that case, the current method functions as a flexible parametric alternative to a fully nonparametric analysis. The mixture of Batschelet distributions then allows much more extensive interpretation and inference than a nonparametric analysis typically would, at a minor cost of flexibility.

It should be noted that in general saccade data is time-series data where each observation is correlated with the previous observations. In the free-viewing paradigm, this autocorrelation is not of core interest, as the time series consist of only 3–5 saccades per viewed image. In general, fitting the saccade directions with a flexible model such as the one provided here while ignoring the time-series structure violates an assumption of the model. However, the time series structure of saccade directions usually does not follow a traditional autocorrelation structure, because, for example, reaching the end of a page results in negative autocorrelation. Therefore, this issue should be addressed separately. The current approach does allow a powerful parametric comparison of different groups of saccade direction data, even when ignoring the autocorrelation.

For some applications, interest might specifically lie in the autocorrelation structure, while the circular distributions still exhibit peakedness. One approach to extend the current work is to employ a Hidden Markov model (HMM) structure, where the components are Batschelet-type or Batschelet mixtures. HMM's have been successfully applied in the context of eye-tracking data (Stuijzand, 2016). Another approach to obtain longitudinally correlated circular observations may be through the covariance structure of the multivariate von Mises distribution (Lagona, 2016). In this approach, peaked marginals such as the Power Batschelet distribution may be obtained through copula-based methods, which have been applied in the circular context (Lagona, 2019).

In future studies, it may be fruitful to model the saccade direction jointly with saccade length, instead of the saccade direction separately. One promising model that has not been applied to the field of eye-tracking is the Abe-Ley model for cylindrical data (Abe & Ley, 2017). In this model, saccades with larger length naturally have a higher (circular) concentration, a property which is commonly observed in vision research. However, the complex form of such models means that extensions such as mixtures are not available, although a hidden Markov model has been developed (Lagona & Picone, 2016; Lagona, Picone, & Maruotti, 2015; Ranalli, Lagona, Picone, & Zambianchi, 2018).

Finally, it should be noted that methods developed here are made accessible to eye-tracking researchers through the easy-to-use R package `flexcircmix`. This means that the methods can be readily applied to new data, without requiring extensive technical knowledge. Hopefully, the methods developed in this paper will provide a valuable new direction for eye-tracking researchers to perform more valid parametric inference on the angles of saccades throughout a broad range of applications.

Acknowledgments

This work was supported by a Vidi grant awarded to I. Klugkist from NWO, the Dutch Organization for Scientific Research (NWO 452-12-010).

The authors are grateful to two reviewers for helpful comments.

Appendix. Properties of the Power Batschelet distribution

The probability density function of the Power Batschelet distribution is defined as

$$f_{PB}(\theta | \mu, \kappa, \lambda) = [K_{\kappa, \lambda}^*]^{-1} \exp\{\kappa \cos t_{\lambda}^*(\theta - \mu)\}, \tag{21}$$

where

$$t_{\lambda}^*(\theta) = \text{sign}(\theta)\pi \left(\frac{|\theta|}{\pi} \right)^{\gamma(\lambda)}, \tag{22}$$

with $\gamma(\lambda) = \frac{1-c\lambda}{1+c\lambda}$, where $c = 0.4052284$ and the inverse of the normalizing constant being

$$K_{\kappa, \lambda}^* = \int_{-\pi}^{\pi} \exp\{\kappa \cos t_{\lambda}^*(\theta - \mu)\} d\theta, \tag{23}$$

which is usually numerically integrated.

Note that for any symmetric base density, such as the von Mises used in this case, the $\text{sign}(\theta)$ in Eq. (22) is optional, because $f_{PB}(\theta - \mu | \mu, \kappa, \lambda) = f_{PB}(-(\theta - \mu) | \mu, \kappa, \lambda)$.

The log-likelihood is

$$\ell(\mu, \kappa, \lambda | \theta) = -\log[K_{\kappa, \lambda}^*] + \kappa \sum_{i=1}^n \cos t_{\lambda}^*(\theta_i - \mu), \tag{24}$$

so that, assuming $\theta_i \neq \mu \forall i \in 1, \dots, n$, the score functions are

$$\frac{\partial \ell(\mu, \kappa, \lambda | \theta)}{\partial \mu} = \kappa \pi^{1-\gamma(\lambda)} \gamma(\lambda) \sum_{i=1}^n |\theta_i - \mu|^{\gamma(\lambda)-1} \sin t_{\lambda}^*(\theta_i - \mu) \tag{25}$$

$$\begin{aligned} \frac{\partial \ell(\mu, \kappa, \lambda | \theta)}{\partial \kappa} &= -[K_{\kappa, \lambda}^*]^{-1} \int_{-\pi}^{\pi} \cos t_{\lambda}^*(\theta) e^{\kappa \cos t_{\lambda}^*(\theta)} d\theta \\ &+ \sum_{i=1}^n \cos t_{\lambda}^*(\theta_i - \mu) \end{aligned} \tag{26}$$

$$\begin{aligned} \frac{\partial \ell(\mu, \kappa, \lambda | \theta)}{\partial \lambda} &= -[K_{\kappa, \lambda}^*]^{-1} \int_{-\pi}^{\pi} h(\theta, \mu, \kappa, \lambda) e^{\kappa \cos t_{\lambda}^*(\theta)} d\theta \\ &+ \sum_{i=1}^n h(\theta_i, \mu, \kappa, \lambda) \end{aligned} \tag{27}$$

where

$$h(\theta, \mu, \kappa, \lambda) = \frac{\partial \kappa \cos t_{\lambda}^*(\theta - \mu)}{\partial \lambda} \tag{28}$$

$$\begin{aligned} &= \frac{\kappa \gamma'(\lambda) |\theta - \mu|^{\gamma(\lambda)} \sin \left(\frac{|\theta - \mu|^{\gamma(\lambda)}}{\pi^{\gamma(\lambda)-1}} \right)}{\pi^{\gamma(\lambda)-1}} \\ &\times \log \left(\frac{\pi}{|\theta - \mu|} \right), \end{aligned} \tag{29}$$

with $\gamma'(\lambda) = \frac{2c}{(1+c\lambda)^2}$. Due to the form of the score functions, as well as the condition that $\theta_i \neq \mu \forall i \in 1, \dots, n$, it is clear that the Hessian and Fisher Information will not be easy to work with. Therefore, it is preferred to work directly with the log-likelihood.

References

Abe, T., & Ley, C. (2017). A tractable, parsimonious and flexible model for cylindrical data, with applications. *Econometrics and Statistics*, 4, 91–104.

Abe, T., Shimizu, K., & Pewsey, A. (2010). Symmetric unimodal models for directional data motivated by inverse stereographic projection. *Journal of the Japan Statistical Society*, 40(1), 45–61.

Akaike, H. (1987). Factor analysis and aic. In *Selected papers of hirotugu akaike* (pp. 371–386). Springer.

Ardia, D., Baştürk, N., Hoogerheide, L., & Van Dijk, H. K. (2012). A comparative study of Monte Carlo methods for efficient evaluation of marginal likelihood. *Computational Statistics & Data Analysis*, 56(11), 3398–3414.

Arnold, B. C., & SenGupta, A. (2006). Recent advances in the analyses of directional data in ecological and environmental sciences. *Environmental and Ecological Statistics*, 13(3), 253–256.

Aslin, R. N. (2007). What's in a look? *Developmental Science*, 10(1), 48–53.

Bao, L., Gneiting, T., Grit, E. P., Guttorp, P., & Raftery, A. E. (2010). Bias correction and Bayesian model averaging for ensemble forecasts of surface wind direction. *Monthly Weather Review*, 138(5), 1811–1821.

Batschelet, E. (1981). *Circular statistics in biology*, Vol. 111 (p. 388). London: Academic Press.

Bowers, J., Morton, I., & Mould, G. (2000). Directional statistics of the wind and waves. *Applied Ocean Research*, 22(1), 13–30.

Bulbert, M. W., Page, R. A., & Bernal, X. E. (2015). Danger comes from all fronts: predator-dependent escape tactics of túngara frogs. *PloS One*, 10(4), e0120546.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., et al. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1).

Chib, S., & Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4), 327–335.

Diebolt, J., & Ip, E. H. S. (1996). Stochastic EM: method and application. In *Markov chain Monte Carlo in practice* (pp. 259–273). Springer.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

Engbert, R., Trukenbrod, H. A., Barthelmé, S., & Wichmann, F. A. (2015). Spatial statistics and attentional dynamics in scene viewing. *Journal of Vision*, 15(1), 14.

Everitt, B. S. (2004). Mixture distributions—I. *Encyclopedia of Statistical Sciences*, 7.

Fisher, N. I. (1995). *Statistical analysis of circular data*. Cambridge: Cambridge University Press.

- Foulsham, T., Gray, A., Nasiopoulos, E., & Kingstone, A. (2013). Leftward biases in picture scanning and line bisection: A gaze-contingent window study. *Vision Research*, 78, 14–25.
- Foulsham, T., Kingstone, A., & Underwood, G. (2008). Turning the world around: Patterns in saccade direction vary with picture orientation. *Vision Research*, 48(17), 1777–1790.
- Friel, N., & Wyse, J. (2012). Estimating the evidence—a review. *Statistica Neerlandica*, 66(3), 288–308.
- Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer Science & Business Media.
- Gilks, W. R., Richardson, S., & Spiegelhalter, D. (1995). *Markov chain Monte Carlo in practice*. CRC press.
- Gill, J., & Hangartner, D. (2010). Circular data in political science and how to handle it. *Political Analysis*, 18(3), 316–336.
- Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., et al. (2017). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, 81, 80–97.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97–109.
- Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7(11), 498–504.
- Holzmann, H., Munk, A., Suster, M., & Zucchini, W. (2006). Hidden Markov models for circular and linear-circular time series. *Environmental and Ecological Statistics*, 13(3), 325–347.
- Hornik, K., & Grün, B. (2013). On conjugate families and Jeffreys priors for von Mises–fisher distributions. *Journal of Statistical Planning and Inference*, 143(5), 992–999.
- Hornik, K., & Grün, B. (2014). Movmf: An R package for fitting mixtures of von Mises–fisher distributions. *Journal of Statistical Software*, 58(10), 1–31.
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3), 194–203.
- Jasra, A., Holmes, C. C., & Stephens, D. A. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science*, 50–67.
- Jones, M. C., & Pewsey, A. (2005). A family of symmetric distributions on the circle. *Journal of the American Statistical Association*, 100(472), 1422–1428.
- Jones, M. C., & Pewsey, A. (2012). Inverse Batschelet distributions for circular data. *Biometrics*, 68(1), 183–193.
- Kaas, A. L., & Van Mier, H. I. (2006). Haptic spatial matching in near peripersonal space. *Experimental Brain Research*, 170(3), 403–413.
- Lagona, F. (2016). Regression analysis of correlated circular data based on the multivariate von Mises distribution. *Environmental and Ecological Statistics*, 23(1), 89–113.
- Lagona, F. (2019). Copula-based segmentation of cylindrical time series. *Statistics & Probability Letters*, 144, 16–22.
- Lagona, F., & Picone, M. (2016). Model-based segmentation of spatial cylindrical data. *Journal of Statistical Computation and Simulation*, 86(13), 2598–2610.
- Lagona, F., Picone, M., & Maruotti, A. (2015). A hidden Markov model for the analysis of cylindrical time series. *Environmetrics*, 26(8), 534–544.
- Le Meur, O., & Coutrot, A. (2016). Introducing context-dependent and spatially-variant viewing biases in saccadic models. *Vision Research*, 121, 72–84.
- Mardia, K. V. (2011). How new shape analysis and directional statistics are advancing modern life-sciences. In *Int. statistical Inst.: Proc. 58th world statistical congress*.
- Mardia, K. V., Hughes, G., Taylor, C. C., & Singh, H. (2008). A multivariate von Mises distribution with applications to bioinformatics. *The Canadian Journal of Statistics*, 36(1), 99–109.
- Mardia, K. V., & Jupp, P. E. (2000). *Directional statistics*, Vol. 494. John Wiley & Sons.
- McLachlan, G., & Peel, D. (2004). *Finite mixture models*. John Wiley & Sons.
- Meng, X.-L., & Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica*, 831–860.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6), 1087–1092.
- Najemnik, J., & Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision*, 8(3), 4.
- Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7(4), 308–313.
- Nielsen, S. F., et al. (2000). The stochastic EM algorithm: estimation and asymptotic results. *Bernoulli*, 6(3), 457–489.
- Pewsey, A., Neuhäuser, M., & Ruxton, G. (2013). *Circular statistics in R*. OUP Oxford.
- Pewsey, A., Shimizu, K., & de la Cruz, R. (2011). On an extension of the von Mises distribution due to Batschelet. *Journal of Applied Statistics*, 38(5), 1073–1085.
- Ranalli, M., Lagona, F., Picone, M., & Zambianchi, E. (2018). Segmentation of sea current fields by cylindrical hidden Markov models: a composite likelihood approach. *Journal of the Royal Statistical Society. Series C. Applied Statistics*, 67(3), 575–598.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506.
- Richardson, S., & Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: series B (statistical methodology)*, 59(4), 731–792.
- Schwarz, G., et al. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 64(4), 583–639.
- Stan Development Team (2017). *Stan modeling language users guide and reference manual*. Retrieved from <http://mc-stan.org/>, (Version 2.17.0).
- Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 62, (4), 795–809.
- Stuijzand, B. G. (2016). *Advanced statistical methods to interpret eye movements: on time-series and individual differences* (Unpublished doctoral dissertation), University of Bristol.
- Tatler, B. W., Brockmole, J. R., & Carpenter, R. H. S. (2017). Latest: A model of saccadic decisions in space and time. *Psychological Review*, 124(3), 267.
- Tatler, B. W., & Vincent, B. T. (2009). The prominence of behavioural biases in eye guidance. *Visual Cognition*, 17(6–7), 1029–1054.
- Teicher, H. (1963). Identifiability of finite mixtures. *The Annals of Mathematical Statistics*, 1265–1269.
- Van Renswoude, D. R., Johnson, S. P., Raijmakers, M. E. J., & Visser, I. (2016). Do infants have the horizontal bias? *Infant Behavior and Development*, 44, 38–48.
- Wagenmakers, E.-J., & Waldorp, L. (2006). Editors introduction. *Journal of Mathematical Psychology*, 50(2), 99–100. <http://dx.doi.org/10.1016/j.jmp.2005.01.005>, Retrieved from <http://www.sciencedirect.com/science/article/pii/S0022249606000022>, (Special Issue on Model Selection: Theoretical Developments and Applications).
- Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research (JMLR)*, 11(Dec), 3571–3594.