Publié sous
L.-P. Rivest & S. Kato (2019). A random-effects model for clustered circular data.Canadian Journal of Statistics, 47, 712-728
https://doi.org/10.1002/cjs.11520

The Canadian Journal of Statistics
Vol. xx, No. yy, 2019, Pages 1–25
La revue canadienne de statistique

A random-effects model for clustered circular data

Louis-Paul Rivest[1]* and Shogo Kato[2]

[1]*Département de mathématiques et de statistique, Université Laval, 1045 Avenue de la médecine, Québec, Qc, G1V 0A6*
[2]*Institute of Statistical Mathematics, Tachikawa,Tokyo 190-8562, Japan E-mail: skato@ism.ac.jp*

Key words and phrases: Angular regression ; Circula ; intra-cluster correlation ; Uniform distribution ; Random effects ; von Mises distribution.

MSC 2010: Primary 62H11 ; secondary 62P10

Abstract: This article considers a circular regression model for clustered data, where both the cluster effects and the regression errors have von Mises distributions. It involves β, a vector of parameters for the fixed effects, and two concentration parameters for the error distribution. A measure of intra-cluster circular correlation and a predictor for an unobserved cluster random effect are studied. Preliminary estimators for the vector β and the two concentration parameters are proposed, and their performance is compared with that of the maximum likelihood estimators in a simulation study. A numerical example investigating the factors impacting the orientation taken by a sand hopper when released is presented.
Résumé: Ce travail considère un modèle de régression circulaire pour des données en grappe dans lequel l'effet de grappe et les erreurs expérimentales ont une loi de von Mises. Il est indexé par β, un vecteur de paramètres pour les effets fixes, et deux paramètres de concentration pour la distribution des erreurs. Une mesure de corrélation intra grappe et un prédicteur pour un effet aléatoire grappe non observé sont étudiés. Des estimateurs préliminaires pour le vecteur β et les deux paramètres pour la concentration des erreurs sont proposés et leur performance est comparée à celle des estimateurs du maximum de vraisemblance dans une étude par simulation. Un exemple numérique portant sur la direction prise par une puce de mer à sa libération est présentée. *La revue canadienne de statistique* xx: 1–25; 2019 © 2019 Société statistique du Canada

1. INTRODUCTION

The modelling of dependent angles has become an important topic in Statistics for the reason that complex natural phenomena involve angular measurements. See, for instance, Mastrantonio, Lasinio, & Gelfand (2016) and Hernandez-Stumpfhauser, Breidt, & van der Woerd (2017). Standard methods for linear variables do not apply, as angles of 0 and 360 degrees are not really different (Mardia & Jupp, 2000). Circular statistics (Pewsey, Neuhäuser, & Ruxton, 2013) attempt to develop statistical methodology for the analysis of these types of data. The construction of probability models for circular variables is now receiving more attention.

Accounting for the dependence between angular variables is an important problem, and there is a rich literature on the topic. Models for bivariate linear data can be used in this context: It suffices to project the bivariate random vectors onto the circumference of the unit circle. This applies in a complex regression setting featuring latent random effects for a within-cluster dependency. In the first step, a normal mixed model for bivariate linear variables is constructed; the bivariate random vectors are projected onto the circumference of the unit circle in the second step. The latent random effects can be discrete (Maruotti, 2016) and dependent (Maruotti et al., 2016). Rather than projecting normal variables, one can wrap them around the circle; this leads to a class of models considered by Jona-Lasinio, Gelfand, & Jona-Lasinio (2012). Multivariate extensions of the von Mises distribution (Mardia et al., 2008; Lagona, 2016) that belong to the exponential family are also available for dependent random angles. Model-free estimation of regression parameters can also be constructed using estimating functions (Artes, Paula, & Ranvaud, 2000). This article suggests another route, namely circular random effects, to model dependencies in circular data.

A multivariate angular density constructed with cluster-level random effects having a von Mises distribution is proposed by Holmquist & Gustafsson (2017). The present article considers a hierarchical angular regression model in which the errors follow Holmquist and Gustafsson's distribution. Its mean direction depends on dependent variables and a vector β of unknown parameters, whereas the mean direction is held constant in Holmquist & Gustafsson (2017). Section 2 presents the evaluation of the angular intra-cluster correlation and the prediction of the circular random cluster effects under the proposed model. Section 3 discusses maximum likelihood estimation and inference procedures for the regression parameters β that are robust to certain misspecification of the random effects. Section 4 compares the proposed model with a projected normal regression model. Section 5 presents a simulation study designed to investigate the sampling properties of estimators of the model parameters and their robustness to a misspecification of the distributions of the random effects and of the errors. The proposed model is easy to use: The log-likelihood and the score function have simple closed form expressions and its parameters have a meaningful interpretation. This is exemplified, in Section 6, by a relatively complex example featuring many explanatory variables.

2. A HIERARCHICAL REGRESSION MODEL FOR CIRCULAR VARIABLES

This article considers angular data having a hierarchical structure. Index $i,\ i = 1, \ldots, m$, represents the cluster and index $j,\ j = 1, \ldots, n_i$, denotes the individual within cluster i. The data set is represented as $\{(y_{ij}, x_{ij}, z_{ij}) : i = 1, \ldots, m,\ j = 1, \ldots, n_i\}$ where $y_{ij} \in [-\pi, \pi)$ is the dependent angle, and x_{ij} and z_{ij} are vectors of angular and linear explanatory variables, respectively. The proposed model for y_{ij} is

$$y_{ij} = \mu_{ij} + a_i + e_{ij},\ i = 1, \ldots, m,\ j = 1, \ldots, n_i, \tag{1}$$

where $\mu_{ij} = \mu_{ij}(\beta)$, the predicted angle for y_{ij}, is the fixed part of the model that depends on a vector of unknown parameters β, and $\{a_i\}$ and $\{e_{ij}\}$ are independent sets of random angles having von Mises distributions centred at 0 with respective concentration parameters $\kappa_a \geq 0$ and $\kappa_e \geq 0$. Note that the von Mises density with mean direction μ and concentration parameter κ, $VM(\mu, \kappa)$, is

$$f(u) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(u - \mu)\}, \quad u, \mu \in [-\pi, \pi),\ \kappa \geq 0,$$

where I_0 denotes the modified Bessel function of the first kind of order 0. In Equation (1) the a_i model the dependence between the y_{ij} in the same cluster. When the mean direction μ_{ij} is constant and equal to μ, the model of Holmquist & Gustafsson (2017) is obtained. In general, the mean direction μ_{ij} is a parametric function of some explanatory variables as discussed in Section 2.1.

Model (1) has interesting special cases. If $\kappa_a = \infty$, then the model has independent identically distributed errors with a von Mises distribution. When $\kappa_a = 0$, the marginal distribution of y_{ij} is uniform and the joint distribution within a cluster is a multivariate circula, using the terminology introduced by Jones, Pewsey, & Kato (2015). When μ_{ij} does not depend on i or j, a circula model specifies that the clusters' mean directions are uniformly distributed. Holmquist &

Gustafsson (2017) investigate tests for this hypothesis. When $\kappa_e = 0$, then the y_{ij} are independent uniformly distributed random angles and model (1) depends neither on μ_{ij} nor on κ_a.

2.1. Modelling fixed effects

Several link functions can be used to express the mean direction μ_{ij} in terms of explanatory variables, see, for instance, Fisher & Lee (1992), Presnell, Morrison, & Littell (1998), Downs & Mardia (2002), and Kato, Shimizu, & Shieh (2008). This article uses the multivariate predictor introduced in Rivest et al. (2016). The mean direction μ_{ij} is determined as the projection onto the circumference of the unit circle of a mean vector defined in R^2, whose components are linear functions of the explanatory variables. This approach proves useful in modelling animal movement in Rivest et al. (2016) and Nicosia et al. (2017). This section investigates the specification of μ_{ij} for a single unit and it is convenient to drop the two subscripts i and j The explanatory variables come in pairs, $(x_0, z_0), \dots, (x_p, z_p)$, where $x_k \in [-\pi, \pi)$ is an angle, $z_k > 0$ is a positive variable and subscript $k = 0, \dots, p$ is used to denote a particular pair. The proposed mean vector, a function of the vectors $x = (x_0, \dots, x_p)^\top$ and $z = (z_0, \dots, z_p)^\top$, is given by

$$\gamma(x, z) = \begin{pmatrix} \gamma_1(x, z) \\ \gamma_2(x, z) \end{pmatrix} = \sum_{k=0}^{p} \beta_k z_k \begin{pmatrix} \cos x_k \\ \sin x_k \end{pmatrix}, \tag{2}$$

and the associated mean direction is $\mu(\beta) = \arctan\{\gamma_2(x, z), \gamma_1(x, z)\}$, where $\arctan(a, b)$ gives an angle in $[-\pi, \pi)$ whose sine and cosine are respectively given by $a/\sqrt{a^2 + b^2}$ and $b/\sqrt{a^2 + b^2}$. Under this model, an angular covariate x_k can be seen as a target and $\beta_k z_k$ reflects its relative importance in the determination of $\mu(\beta)$. When all the targets have the same importance, $\mu(\beta)$ is the mean direction of the covariates $\{x_k\}$. Note also that a model where the coefficient of (x_k, z_k) is negative is the same as one where $(x_k + \pi, z_k)$ has a positive coefficient. Thus attractive and repulsive targets can enter into the vector (2). An alternative link function with unrelated expressions for the sine and cosine components of $\gamma(x, z)$ is investigated by Cremers, Mulder, & Klugkist (2018).

The parameters in Equation (2) have interesting interpretations. Consider an explanatory angle x^*, and a model with two explanatory pairs $(x_1, z_1) = (x^*, 1)$ and $(x_2, z_2) = (x^* + \pi/2, 1)$ with coefficients β_1 and β_2, respectively. This model is identical with one with a single covariate $(x^* + \theta, 1)$ with $\theta = \arctan(\beta_2, \beta_1)$, on account of the identity

$$\beta_1 \begin{pmatrix} \cos x^* \\ \sin x^* \end{pmatrix} + \beta_2 \begin{pmatrix} \cos(x^* + \pi/2) \\ \sin(x^* + \pi/2) \end{pmatrix} = \sqrt{\beta_1^2 + \beta_2^2} \begin{pmatrix} \cos(x^* + \theta) \\ \sin(x^* + \theta) \end{pmatrix}.$$

In this case $x^* + \theta$, rather than x^*, is impacting the mean direction $\mu(\beta)$. A model with $(x_1, z_1) = (x^*, 1)$ and $(x_2, z_2) = (x^*, z^*)$, with respective coefficients β_1 and β_2 in Equation (2), has an interaction between the angular covariate x^* and the linear covariate z^*: The impact of x^* on the vector (2) depends on z^*. A linear covariate z^* is added to Equation (2) through the pairs $(0, z^*)$ and $(\pi/2, z^*)$. This brings, in Equation (2), a term proportional to $z^*(\cos\theta, \sin\theta)^\top$ where θ depends on the β-coefficients for $(0, z^*)$ and $(\pi/2, z^*)$. Large values of z^* drag the mean direction $\mu(\beta)$ toward θ.

Since multiplying all the parameters β by a constant does not change the mean direction $\mu(\beta)$, the parameters in Equation (2) are not all identifiable. To reduce the size of the parameter space we identify an angular covariate, say x_0, that is dominant in $\mu(\beta)$ and set the coefficient of the first pair $(x_0, 1)$ equal to $\beta_0 = 1$. The other parameters then give the relative importance of the covariate pairs $(x_k, z_k),\ k = 1, \dots, p$ relative to $(x_0, 1)$.

Parameter identifiability is still an issue once a reference covariate has been identified. Some conditions for identifiability are provided in Proposition 1 of Rivest et al. (2016) as illustrated by a model for a single linear covariate z^*. Such a model introduces, in Equation (2), four pairs $(0,1), (\pi/2,1), (0,z^*), (\pi/2,z^*)$ depending on a linear covariate z^*. Taking $(x_0, z_0) = (0,1)$ as the reference gives a model where the vector of parameters β has dimension 3,

$$\mu(\beta) = \arctan(\beta_1 + \beta_3 z^*, 1 + \beta_2 z^*).$$

In this expression β_1, β_2, and β_3 are identifiable except at the independence model $\mu(\beta) = \arctan(\beta_1, 1)$ that is obtained with $\beta_3 = \beta_2 \times \beta_1$. The parameter β_2 is then redundant. We return to this issue in Section 4. In general, the fixed-effects mean direction $\mu(\beta)$ depends on a reference pair, $(x_0, 1)$ and on p covariate pairs $(x_k, z_k),\ k = 1, \ldots, p$. Thus the vector of unknown regression parameters β has length p.

2.2. The joint density of the proposed model

The marginal distribution of y_{ij} in model (1) is the convolution of two von Mises distributions, respectively, *VM*(μ_{ij}, κ_a) and *VM*$(0, \kappa_e)$. Its density is given by

$$f_y(u) = \frac{I_0\left(\{\kappa_a^2 + \kappa_e^2 + 2\kappa_a\kappa_e \cos(u - \mu_{ij})\}^{1/2}\right)}{2\pi I_0(\kappa_a) I_0(\kappa_e)} \qquad \text{for } u \in [-\pi, \pi). \tag{3}$$

The shape of this density is discussed in Mardia & Jupp (2000, pp. 38, 45). It is unimodal and can be closely approximated by a von Mises density, especially when $\kappa_a^2 + \kappa_e^2$ is large. The joint density, given by Holmquist & Gustafsson (2017, Eq. 11), is presented in the next proposition.

Proposition 1 *The joint density of* $(y_{i1}, \ldots, y_{in_i})$ *is, for* $i = 1, \ldots, m$, *given by*

$$f_{n_i}(y_{i1}, \ldots, y_{in_i}) = \frac{I_0(\sigma_i)}{(2\pi)^{n_i} I_0(\kappa_e)^{n_i} I_0(\kappa_a)}, \quad y_{i1}, \ldots, y_{in_i} \in [-\pi, \pi), \tag{4}$$

where σ_i^2 *is the squared length of the vector*

$$\begin{pmatrix} \kappa_a \\ 0 \end{pmatrix} + \kappa_e \sum_{j=1}^{n_i} \begin{pmatrix} \cos(y_{ij} - \mu_{ij}) \\ \sin(y_{ij} - \mu_{ij}) \end{pmatrix}. \tag{5}$$

The joint density (4) is easily evaluated as it depends on $\sum \cos(y_{ij} - \mu_{ij})$, its square, and $\{\sum \sin(y_{ij} - \mu_{ij})\}^2$. The density (4) takes its maximum value at $y_{ij} = \mu_{ij}$. When both κ_e and κ_a are large, its contours are very close to those of an equicorrelated normal model.

By construction, the family of angular densities $\{f_n(y_1, \ldots, y_n) :\ n = 2, 3, \ldots\}$ is closed under marginalization. This means that if $f_n(y_1, \ldots, y_n)$ is the density of angles $(y_1, \ldots, y_n)$ then the marginal density of $(y_1, \ldots, y_d)$ can, for $d < n$, also be expressed in the form (4). Alternatives, such as the multivariate von Mises distribution of Mardia et al. (2008), do not satisfy this property. Indeed, the marginal density of a subset of d variables in a vector of size n with a multivariate von Mises distribution does not have a d-dimensional von Mises distribution. This makes the multivariate von Mises inappropriate when the size of the clusters varies.

2.3. Evaluation of the intra-cluster correlation

To investigate how the concentration parameters κ_a and κ_e in Equation (4) impact the association between the y's in the same cluster, several circular correlation coefficients are available. These correlations are expressed in terms of trigonometric moments of the joint distribution of the random angles (y_{i1}, y_{i2}), defined in Equation (1) with $n_i = 2$. For simplicity we drop the subscript i and assume $\mu_1 = \mu_2 = 0$; the joint density of (y_1, y_2) is given by Equation (4) with $n = 2$.

Proposition 2 *The trigonometric moments of the random angles (y_1, y_2), whose density is given by (4) with $n_i = 2$ and $\mu_{ij} = 0$, are given by*

$$E[\cos\{p_1 y_1 + p_2 y_2\}] = A_{p_1}(\kappa_e) A_{p_2}(\kappa_e) A_{p_1+p_2}(\kappa_a),$$
$$E[\sin\{p_1 y_1 + p_2 y_2\}] = 0,$$

where $p_1, p_2 \in \mathbb{Z}$, $A_p(\kappa) = I_p(\kappa)/I_0(\kappa)$, and I_p denotes the modified Bessel function of the first kind of order p.

From this proposition, we deduce that

(a) $E\{\cos y_1\} = A_1(\kappa_e) A_1(\kappa_a)$,
(b) $E\{\sin y_1 \sin y_2\} = A_1(\kappa_e)^2 \{1 - A_2(\kappa_a)\}/2$,
(c) $E\{\sin^2 y_1\} = \{1 - A_2(\kappa_e) A_2(\kappa_a)\}/2$,

where $A_2(\kappa) = 1 - 2A_1(\kappa)/\kappa$ (Mardia & Jupp, 2000, p. 41). We propose using the sine-sine correlation introduced by Rivest (1982) and Jammalamadaka & Sarma (1988) to evaluate the within-cluster correlation. Using Proposition 2 this correlation is equal to

$$\rho_{SS} \equiv \frac{E\{\sin y_1 \sin y_2\}}{E\{\sin^2 y_1\}} = \frac{A_1^2(\kappa_e)\{1 - A_2(\kappa_a)\}}{1 - A_2(\kappa_a) A_2(\kappa_e)}. \tag{6}$$

It is clear that ρ_{SS} takes only positive values. It is equal to 0 if and only if y_1 and y_2 are independent; this occurs: (i) when $\kappa_e = 0$ or (ii) when κ_a goes to ∞ for a fixed $\kappa_e > 0$. This coefficient is monotonically increasing with respect to κ_e and monotonically decreasing with respect to κ_a. When κ_e is fixed, the maximum value or ρ_{SS} is $A_1^2(\kappa_e)$ which is achieved at $\kappa_a = 0$. For any $\kappa_a \geq 0$, ρ_{SS} tends to 1 as $\kappa_e \to \infty$.

When both κ_a and κ_e are large, the approximate distributions of a_i and e_{ij} in model (1) are normal, centred at 0, with respective variances κ_a^{-1} and κ_e^{-1} (Mardia & Jupp, 2000, Ch. 4). Then the model (1) is equivalent to the normal mixed model considered by Battese, Harter, & Fuller (1988). For this model, the association is measured by the intra-cluster correlation $\kappa_e(\kappa_a + \kappa_e)^{-1}$ and ρ_{SS} converges to that value when κ_a and κ_e converge to ∞.

An important peculiarity of the angular model (4) is that, for a fixed level of residual variability indexed by κ_e, ρ_{SS} is bounded away from 1 as it reaches its maximum value at $\kappa_a = 0$. Thus the magnitude of κ_e drives the intra-cluster correlation: A small κ_e only allows a small level of intra-cluster correlation. Indeed, as noted by Rivest et al. (2016), it is nearly impossible to fit angular regression models when the concentration parameter of the errors is small. This contrasts with the linear case where intra-cluster correlations between 0 and 1 are possible when one of the two variance components is fixed.

2.4. Prediction of the random effect a_i and of a new value y_{i0} in cluster i

In Equation (1), the random angles a_i are not observed. This section considers methods to predict a_i using $\tilde{a}_i$, a function of $\{y_{ij} - \mu_{ij} : j = 1, \ldots, n_i\}$. It assumes that the parameters $(\beta, \kappa_a, \kappa_e)$ are known. The best predictor (BP) $\tilde{a}_i$ is the function of $\{y_{ij} - \mu_{ij}\}$ that minimizes the circular variance $1 - E\{\cos(\tilde{a}_i - a_i)\}$. The solution, given in the next proposition, is to take $\tilde{a}_i$ equal to the conditional mean direction of a_i given $\{y_{ij} - \mu_{ij}\}$. The proof of Proposition 3 is easily deduced from that of Proposition 1.

Proposition 3 *The conditional distribution of a_i given $\{y_{ij} : j = 1, \ldots, n_i\}$ is von Mises with mean direction $\tilde{a}_i$, given by*

$$\tilde{a}_i = \arctan\left(\kappa_e \sum_j \sin(y_{ij} - \mu_{ij}), \kappa_a + \kappa_e \sum_j \cos(y_{ij} - \mu_{ij})\right), \tag{7}$$

and concentration parameter σ_i, where $\tilde{a}_i$ and σ_i are respectively the direction and the length of the vector (5).

When $\kappa_a = 0$, $\tilde{a}_i$ is the mean direction of the fixed-effects residuals within cluster i. A positive value of κ_a shrinks $\tilde{a}_i$ toward 0. The extent of the shrinking decreases as the cluster sample size n_i increases. This shrinkage is similar to what is obtained when estimating a random cluster effect in models for linear variables (Battese, Harter, & Fuller, 1988).

The prediction for an unobserved y_{i0} is $\mu_{i0} + \tilde{a}_i$ where μ_{i0}, the mean direction for the unobserved y, depends on the covariates (x_{i0}, z_{i0}) for this unit. The prediction error is

$$y_{i0} - \mu_{i0} - \tilde{a}_i = e_{i0} + a_i - \tilde{a}_i.$$

As e_{i0} and $a_i - \tilde{a}_i$ are independent, its circular variance is

$$\begin{aligned} 1 - E\{\cos(e_{i0} + a_i - \tilde{a}_i)\} &= 1 - E\{\cos e_{i0}\} E[E\{\cos(a_i - \tilde{a}_i) | y_{i1}, \ldots, y_{in_i}\}] \\ &= 1 - A_1(\kappa_e) E\{A_1(\sigma_i)\}. \end{aligned} \tag{8}$$

The conditional distribution of y_{i0} given $\{y_{ij} - \mu_{ij}\}$ is a sum of two independent von Mises random variables, a $VM(0, \kappa_e)$ plus a $VM(\mu_{i0} + \tilde{a}_i, \sigma_i)$. Its density is given in Equation (3). When both κ_e and κ_a are large, or $O(\kappa)$, $y_{ij} - \mu_{ij}$ is $O_p(\kappa^{-1/2})$ and one has the following approximation (where $O(\kappa)$ is used for non-random values, that is parameters, while $O_p(\kappa)$ is used for random variables).

$$\begin{aligned} \tilde{a}_i &\approx \frac{\kappa_e \sum_j \sin(y_{ij} - \mu_{ij})}{\kappa_a + n_i \kappa_e} \\ &= \frac{\rho}{1 + (n_i - 1)\rho} \sum_j \sin(y_{ij} - \mu_{ij}), \end{aligned}$$

where $\rho = \kappa_e(\kappa_a + \kappa_e)^{-1}$ is the intra-cluster correlation. Thus $\tilde{a}_i$ is approximately equal to the best linear unbiased predictor (BLUP) for the effect of cluster i in the approximate mixed linear model for y (Rao & Molina, 2015, p. 174).

In general, parameters are estimated and the empirical best predictor (EBP) $\hat{a}_i$ is obtained from Equation (7) by replacing the unknown parameters by their estimators. Formula (8) then underestimates the circular variance of the prediction error of the EBP since it does not account for the estimation of the parameters. Techniques to correct this bias are available in the linear case (Rao & Molina, 2015) and they could be adapted to angular variables.

3. PARAMETER ESTIMATION

This section proposes a strategy to estimate the parameters of the density (4). Simple consistent preliminary estimators are first proposed. They can be used as initial parameter values to calculate the maximum likelihood estimates. The sampling properties of the maximum likelihood estimators under a misspecification of the error distribution are also investigated.

3.1. Preliminary estimators

The parameter vector β can be estimated using an algorithm for fitting the regression model $y_{ij} = \mu_{ij} + \varepsilon_{ij}$ where the errors ε_{ij} are assumed to be independent and identically distributed with a density that is symmetric with respect to 0. Such algorithms are available for wrapped Cauchy and von Mises errors, see Kato, Shimizu, & Shieh (2008) and Rivest et al. (2016). Let $\hat{\mu}_{ij}$ be the estimated mean direction for y_{ij}. Since $E\{\cos(y_{ij} - \mu_{ij} - y_{ik} + \mu_{ik})\} = A_1^2(\kappa_e)$, a simple moment estimator for κ_e is

$$\hat{\kappa}_e = A_1^{-1}\left\{\sqrt{\left|\sum_i \sum_{j>k} \frac{2\cos(y_{ij} - \hat{\mu}_{ij} - y_{ik} + \hat{\mu}_{ik})}{mn_i(n_i - 1)}\right|}\right\}.$$

In addition, note that $E\{\cos(y_{ij} - \mu_{ij})\} = A_1(\kappa_e)A_1(\kappa_a)$. This suggests the following moment estimator for κ_a,

$$\hat{\kappa}_a = \begin{cases} A_1^{-1}(S), & S < 1, \\ \infty, & S \geq 1, \end{cases}$$

where

$$S = \frac{1}{mA_1(\hat{\kappa}_e)}\left|\sum_{i=1}^{m} n_i^{-1} \sum_{j=1}^{n_i} \cos(y_{ij} - \hat{\mu}_{ij})\right|.$$

These preliminary estimators are consistent. The preliminary estimator for β is obtained by maximizing a misspecified log-likelihood that assumes independent errors with a density that is symmetric with respect to 0. Since according to Equation (4) the marginal density of the errors is also symmetric with respect to 0, the expectation of the misspecified score function at the true value for β is 0. Thus its preliminary estimator converges in probability to the true value. Applying the weak law of large numbers, the moments entering in the definition of $\hat{\kappa}_e$ and $\hat{\kappa}_a$ converge in probability to the corresponding theoretical moments. Therefore these two estimators are consistent.

3.2. Score function and Fisher information

This section derives the score function defining the maximum likelihood estimators of the parameters. It shows that the Fisher information matrix is block-diagonal, one block for β and one for (κ_e, κ_a). This pattern mimics the well-known result for normal linear mixed models in which the regression and variance parameters are orthogonal.

The log-likelihood to be maximized is given by

$$L(\beta, \kappa) = \sum_{i=1}^{n} [\log\{I_0(\sigma_i)\} - n_i \log\{I_0(\kappa_e)\} - \log\{I_0(\kappa_a)\}], \tag{9}$$

where σ_i is the length of the vector (5). To derive the score function, the following results concerning the partial derivatives of σ_i are needed.

Proposition 4 *Let $\dot{\mu}_{ij}$ denote the vector of partial derivatives of μ_{ij} with respect to the vector β. Then the partial derivatives of σ_i are given by*

$$\frac{\partial \sigma_i}{\partial \beta} = \kappa_e \sum_j \dot{\mu}_{ij} \sin(y_{ij} - \mu_{ij} - \tilde{a}_i),$$

$$\frac{\partial \sigma_i}{\partial \kappa_e} = \sum_j \cos(y_{ij} - \mu_{ij} - \tilde{a}_i), \quad \textit{and}$$

$$\frac{\partial \sigma_i}{\partial \kappa_a} = \cos \tilde{a}_i,$$

where $\tilde{a}_i$ is defined in Equation (7).

The score function is easily evaluated using Proposition 4,

$$s(\beta, \kappa_e, \kappa_a) = \sum_{i=1}^{m} \begin{pmatrix} \kappa_e A_1(\sigma_i) \sum_j \dot{\mu}_{ij} \sin(y_{ij} - \mu_{ij} - \tilde{a}_i) \\ A_1(\sigma_i) \sum_j \cos(y_{ij} - \mu_{ij} - \tilde{a}_i) - n_i A_1(\kappa_e) \\ A_1(\sigma_i) \cos \tilde{a}_i - A_1(\kappa_a) \end{pmatrix}.$$

The Fisher information matrix is the covariance matrix of the above. It consists of two diagonal blocks, for β and (κ_e, κ_a) respectively, and a covariance block. We next show that the covariance entries are equal to 0.

From Equation (7), note that $\tilde{a}_i$ is an odd function of $\{y_{ij} - \mu_{ij}\}$: Replacing $\{y_{ij} - \mu_{ij}\}$ by $\{-y_{ij} + \mu_{ij}\}$ changes $\tilde{a}_i$ into $-\tilde{a}_i$ while σ_i is left unchanged. Thus the score function for each entry of the vector β is an odd function of $\{y_{ij} - \mu_{ij}\}$ while the score functions for the two κ'parameters are even. Since $\{y_{ij} - \mu_{ij}\}$ has a symmetric distribution, the expectation of an odd function of these random variables is 0. Thus the covariances between the score functions of the β's and the κ's are null and the Fisher information matrix is block-diagonal: One block for β and one for (κ_e, κ_a).

It does not seem possible to get a simple expression for the covariance matrix of the score vector. General-purpose optimization programs that maximize the log-likelihood can provide a numerical approximation of the Hessian at the maximum likelihood estimates. The negative of this Hessian can be used to estimate $\hat{I}(\beta, \kappa_e, \kappa_a)$, the observed Fisher information matrix, and to obtain approximations to the sampling variances of the maximum likelihood estimators.

3.3. Misspecified error models

Model (1) can be misspecified in several ways. The distribution of the random effects, a_i and e_{ij}, might differ from the von Mises distribution. For example, they might have a symmetric, non-von-Mises distribution, such as a wrapped Cauchy or a wrapped normal. They might also be heterogeneous and κ_e could, for instance, vary with the explanatory variables. Finally the equicorrelation assumption of the errors within a cluster might fail and a wrapped autoregressive normal model might provide a better description. In this section we show that $\hat{\beta}$, the estimator of the fixed-effects parameter obtained by maximizing (9), is consistent under these three symmetric alternatives.

Even when the model is misspecified, $\hat{\beta}$ is consistent provided that the expectation of the score function for β, evaluated at the true β value, is zero. The argument presented in Section 3.2, showing that the score function of β is an odd function of $\{y_{ij} - \mu_{ij}\}$, remains valid for any values of κ_a and κ_e, provided that the misspecified model has symmetric error distributions. Since the score function at the true value of β is an odd function of random variables symmetric with respect to 0, its expectation is 0. The von Mises maximum likelihood estimator of β is therefore robust to a misspecification of the error distribution. This discussion does not apply to random effects with an asymmetric distribution, such as the family of Kato & Jones (2015) or the sine skewed von Mises distribution of Abe & Pewsey (2016). The sampling properties of $\hat{\beta}$ in models with asymmetric random effects are studied in Section 5, using Monte Carlo methods.

For misspecified models, the covariance matrix of $\hat{\beta}$ can be evaluated as the (β, β) component of a standard sandwich estimator,

$$v(\hat{\beta}, \hat{\kappa}_e, \hat{\kappa}_a) = \hat{I}(\beta, \kappa_e, \kappa_a)^{-1} \left(\sum_i s_i s_i^\top \right) \hat{I}(\beta, \kappa_e, \kappa_a)^{-1},$$

where s_i is an estimator of the score vector for cluster i,

$$s_i = \begin{pmatrix} \hat{\kappa}_e A_1(\hat{\sigma}_i) \sum_j \widehat{\dot{\mu}}_{ij} \sin(y_{ij} - \hat{\mu}_{ij} - \hat{a}_i) \\ A_1(\hat{\sigma}_i) \sum_j \cos(y_{ij} - \hat{\mu}_{ij} - \hat{a}_i) - n_i A_1(\hat{\kappa}_e) \\ A_1(\hat{\sigma}_i) \cos(\hat{a}_i) - A_1(\hat{\kappa}_a) \end{pmatrix}$$

and $\hat{a}_i$ is given by Equation (7) with the parameters replaced by their estimators.

4. COMPARISONS WITH PROJECTED NORMAL MODELS

A projected normal model is obtained by adding independent normal deviates to the two components of γ in the model (2). A random angle with a projected normal distribution is constructed as $y = \arctan\{\gamma_2(x, z) + u_2, \gamma_1(x, z) + u_1\}$ where u_1 and u_2 have independent $N(0, 1)$ distributions. The mean direction of y is $\mu(\beta) = \arctan\{\gamma_2(x, z), \gamma_1(x, z)\}$ (Presnell, Morrison, & Littell, 1998). Assuming that u_1 and u_2 are small with respect to the entries of $\gamma(x, z)$, a first-order Taylor series approximation to y gives

$$y = \arctan\left\{ \frac{\gamma_2(x, z) + u_2}{\gamma_1(x, z) + u_1} \right\} \approx \mu(\beta) + \frac{\gamma_1(x, z)u_2 - \gamma_2(x, z)u_1}{\gamma_1^2(x, z) + \gamma_2^2(x, z)}, \tag{10}$$

and the angular error has approximately a $N\left(0, \{\gamma_1^2(x, z) + \gamma_2^2(x, z)\}^{-1}\right)$ distribution. This highlights the fact that in a projected normal model, in contrast to model (1), the regression coefficients determine both the mean direction and the error concentration of y. With this joint specification, the identifiability constraints discussed in Section 2.1 disappear: The $p + 1$ β-parameters in Equation (2) are estimable. This estimability comes at the cost of a strong assumption: The same linear combination (2) of the explanatory variables determines both the mean direction and the error concentration of y. An interpretation of the β-parameters in terms of only the mean direction can be misleading (Rivest et al., 2016). (See also Cremers, Mulder, & Klugkist (2018) for a discussion of parameter interpretation in this context.) Considering equation (3.5.22) in Mardia & Jupp (2000), a close approximation to the distribution of y in Equation (10) is then $VM\{\mu(\beta), \gamma_1^2(x, z) + \gamma_2^2(x, z)\}$. Rivest et al. (2016) refer to it as a consensus von Mises model because the error concentration is proportional to the level of agreement between the p targets appearing in Equation (2). They also suggest fitting a consensus model, which is not plagued by identifiability problems, as a first step in fitting a general regression model with homogeneous errors. It helps to identify the important covariates.

A cluster effect can be introduced in a projected normal model by expressing $u_j = b_j + e_j,\ j = 1, 2$ where (b_1, b_2) are independent $N(0, \psi)$ variables representing a cluster effect while the unit errors (e_1, e_2) have independent $N(0, 1-\psi)$ distributions (Nuñez-Antonio & Gutiérrez-Peña, 2014; Hall & Shen, 2015). Considering Equation (10), y is approximately equal to a sum featuring a fixed effect, a cluster effect, and an experimental error, with the β-parameters appearing in all three terms. This makes their interpretation challenging. Other random-effects specifications in a projected normal model are discussed by Maruotti (2016). The projected normal construction is unable to model the fixed and the random effects separately or to distinguish parameters for the mean direction from those associated with the error concentration. For our model, the parameter vector β controls only the mean direction of y. Section 3.3 emphasizes that β can be consistently estimated even if von Mises random effects do not provide an accurate description of the errors within a cluster. Also the random-effects model allows for uniformly distributed cluster effects which do not seem possible with the projected normal construction presented in this section. Thus the proposed model should be considered in situations where expressing the mean direction of y in terms of explanatory variables is an important objective. Projected normal models might be best suited for complex dependency structures encountered with circular time series and spatial variables in which the mean direction is a fixed constant.

5. MONTE CARLO SIMULATIONS

In this section we present the details of, and results from, a simulation performed to investigate the precision of the estimators obtained by maximizing the log-likelihood (9), as compared to the simple preliminary estimators of Section 3.1. We considered an angular regression model where the fixed part involves two pairs $(x_0, 1)$ and $(x_1, 1)$ and a single β parameter. The mean direction for unit (i, j) was given by

$$\mu_{ij} = \arctan\{\sin x_{0ij} + \beta \sin x_{1ij}, \cos x_{0ij} + \beta \cos x_{1ij}\}$$

where the two explanatory angles for unit (i, j), x_{0ij} and x_{1ij}, were simulated from centred von Mises distributions with concentration parameter values of 2 and 4 respectively. The dependent angle y is driven by x_0 and x_1 and β gives the relative importance of x_1 with respect to x_0. We used $\beta = 0.5$ in the simulations. For this model, $\dot{\mu}_{ij} = \sin(x_{1ij} - \mu_{ij})/\ell_{ij}$, where $\ell_{ij} = [\{\sin(x_{0ij}) + \beta \sin(x_{1ij})\}^2 + \{\cos(x_{0ij}) + \beta \cos(x_{1ij}\}^2]^{1/2}$. See Rivest et al. (2016) for details. The simulations used values of m and n_i equal to 5, 20, and 50, for a total of nine scenarios.

The preliminary estimates (Pre in Table 1) of β were obtained by fitting a model with independent and identically distributed von Mises errors following Rivest & al. (2016). Those for the κ-parameters were calculated as proposed in Section 3.1. The maximum likelihood estimates (MLE in Table 1) were calculated by maximizing (9) using the quasi-Newton BFGS algorithm implemented in the R function `optim`, that uses the score function derived in Section 3.2 as an input. The observed Fisher information matrix $\hat{I}(\beta, \kappa_e, \kappa_a)$ was estimated as minus the numerical approximation of the Hessian provided by `optim` and the covariance matrix of the maximum likelihood estimates was estimated by $\hat{I}(\beta, \kappa_e, \kappa_a)^{-1}$.

The Monte Carlo study used 10,000 replications for each scenario. Biases, B, and standard deviations, Sd, of an estimator $\hat{\theta}$, equal to one of $(\hat{\beta}, \hat{\kappa}_a, \hat{\kappa}_e)$, were calculated using the formulas

$$B(\hat{\theta}) = \bar{\hat{\theta}} - \theta \text{ and } Sd(\hat{\theta}) = \sqrt{\frac{1}{N-1}\sum_{b=1}^{N}(\hat{\theta}_b - \bar{\hat{\theta}})^2}, \text{ where } \bar{\hat{\theta}} = \frac{1}{N}\sum_{b=1}^{N}\hat{\theta}_b,$$

and N is 10,000 minus the number of runs for which the numerical algorithm to maximize the log-likelihood (9) failed to converge. Convergence failure was not an issue for any scenario as it occurred in fewer than 0.7% of the Monte Carlo repetitions. See the supplementary material for more information on these convergence failures. Biases, as an estimator of $Sd(\hat{\beta})$, of the standard error estimators derived from the observed Fisher information matrix were also evaluated using

$$\frac{1}{N}\sum_{b=1}^{N}\sqrt{\hat{I}(\hat{\beta}, \hat{\kappa}_e, \hat{\kappa}_a)_b^{(\beta,\beta)}} - Sd(\hat{\beta}),$$

where $\hat{I}(\hat{\beta}, \hat{\kappa}_e, \hat{\kappa}_a)^{(\beta,\beta)}$ is the (β, β) entry of the inverse of the observed Fisher information matrix. Simulation results for von Mises errors with sample sizes $m = 50$ and $n_i = 5$ are presented in Table 1 and the complete results for the nine scenarios are presented in Table S1 of the supplementary material.

TABLE 1: Bias, and between brackets, standard deviation of preliminary (Pre) and maximum likelihood (MLE) estimators for the mixed angular model under four scenarios with $m = 50$ and $n_i = 5$. The s.e. rows present the biases of the standard error estimators of $\hat{\beta}$ derived from the inverse of the observed Fisher information matrix.

		Simulation scenario, $(\kappa_e; \kappa_a)$			
Parameter	Estimator	(2; 2)	(1.6; 0.5)	(3.6; 2)	(2.6; 0.5)
β	Pre	0.007 (0.184)	0.049 (0.902)	0.001 (0.152)	0.038 (0.737)
	MLE	0.003 (0.125)	0.007 (0.160)	$< 10^{-3}$(0.085)	0.001 (0.104)
	s.e.	0.004	0.008	0.001	0.002
κ_e	Pre	0.007 (0.175)	−0.124 (0.278)	0.009 (0.321)	−0.206 (0.420)
	MLE	0.022 (0.174)	0.015 (0.149)	0.046 (0.322)	0.029 (0.224)
	s.e	0.001	$< 10^{-3}$	0.004	0.004
κ_a	Pre	0.117 (0.483)	0.206 (2.40)	0.087 (0.418)	0.093 (1.42)
	MLE	0.093 (0.471)	0.016 (0.244)	0.073 (0.415)	0.009 (0.227)
	s.e.	0.030	0.004	0.022	0.004

Considering the theoretical values of the intra-cluster correlation ρ_{SS} given by Equation (6), the first two scenarios, $(\kappa_e = 2; \kappa_a = 2)$ and $(\kappa_e = 1.6; \kappa_a = 0.5)$, of Table 1 have moderate intra-cluster correlation of 0.37 and 0.38, respectively. It is 0.59 for the last two scenarios. Parameter κ_a impacts the moment estimator for β. When $\kappa_a = 0.5$, this estimator is much worse than the maximum likelihood estimator. In the (1.6; 0.5) scenario, the error mean resultant length is $A_1(1.6) \times A_1(0.5) = 0.15$. These nearly uniformly distributed errors make the preliminary estimator of β highly unstable. Indeed, this estimator is not well defined if $\kappa_a = 0$ since the marginal error distribution is then uniform. The maximum likelihood estimator circumvents this problem by incorporating, in its estimating equations, estimates of the random effects a_i. The properties of the preliminary estimator for β impact the moment estimators for κ_e and κ_a negatively when $\kappa_a = 0.5$. Table 1 also reveals that the observed Fisher information matrix provides reliable estimators of the standard errors as the biases are all positive and smaller than 10% of the target values.

The results of additional simulations for m and n_i equal to $5, 20, 50$ and the same parameter values as in Table 1 are presented in the Table S1 of the supplementary material. In most cases, the maximum likelihood estimators slightly overestimate their target parameters and the observed Fisher information matrix gives a conservative estimator of the standard error as it has a small positive bias. With only $m = 5$ clusters, κ_a is very difficult to estimate and $Sd(\hat{\kappa}_a) > 2\kappa_a$ for the three values of n_i considered. Indeed, in all simulations, $\hat{\kappa}_a$ has the largest standard errors and it is the most difficult parameter to estimate. The maximum likelihood estimators of β and κ_e have good sampling properties except when $m = n_i = 5$.

Additional simulations with a_i and e_{ij} having either a Kato–Jones distribution (Kato & Jones, 2015) or a von Mises distribution, were carried out to investigate the impact of misspecified random effects on the inference for β (see Table S2 of the supplementary material). The Kato–Jones family was selected because it can be asymmetric, with tails heavier than the von Mises as it generalizes the wrapped Cauchy distribution. The results, presented in the supplementary material, show that the inference on β is robust to a misspecification of either of the two distributions. However, when the errors e_{ij} have a Kato–Jones distribution, the von Mises observed Fisher information matrix leads to negatively biased estimators for the standard error of $\hat{\beta}$. Then, the sandwich variance estimator of Section 3.3 should be used instead. The findings of this Monte Carlo study support the conclusion of Section 3.3 that the inference on the regression parameter β, using the proposed model, is robust to a misspecification of the errors.

6. EXAMPLE: AN ANGULAR REGRESSION MODEL WITH REPEATED MEASURES

Here we present the analysis of the spring subset of the data set considered by D'Elia (2001) containing repeated measurements on the escape orientation, with respect to north, of 59 sand hoppers (*Talitrus saltator*). Each sand hopper was released five times over a short time period at the centre of a circular arena with a 120 cm diameter. On each release, its direction with respect to north was recorded when it left the arena. Several explanatory variables were measured at the time of the experiment: the sun azimuth (*Azi*), the wind speed (*SpeeW*) and its direction (*DirW*), and a measure of the sand hopper eye asymmetry (*Eye*). In addition, as explained in D'Elia (2001), there is a counterclockwise time trend ($T = j - 1$) over the 5 releases that has to be included as a within-sand-hopper covariate. Analyses of these data are presented in D'Elia (2001), Lagona (2016), and Maruotti (2016). Lagona (2016) used the multivariate von Mises distribution of Mardia et al. (2008) and the link function of Fisher & Lee (1992), while Maruotti (2016) used a mixture of projected normal models. The data set and the R-code for the analyses reported in this section are available in the supplementary material.

6.1. Modelling fixed effects

For the fixed part μ_{ij}, the following pairs (x_k, z_k) of covariates were considered in the model selection phase: $(\mathit{Azi}, 1)$, $(\mathit{Azi} + \pi/2, 1)$, $(\mathit{DirW}, 1)$, $(\mathit{DirW} + \pi/2, 1)$, $(\mathit{DirW}, \mathit{SpeeW})$, and $(\mathit{DirW} + \pi/2, \mathit{SpeeW})$. The variable *Azi* was selected as the reference explanatory angle x_0. The two linear covariates, *Eye* and T, were entered in the model through the pairs $(0, \mathit{Eye})$, $(\pi/2, \mathit{Eye})$ and $(0, T)$, $(\pi/2, T)$. The full model had 12 explanatory pairs (x_k, z_k) and variable selection was done using a backward elimination procedure with a 5% significance level in a regression that assumed independent identically distributed von Mises errors (Rivest et al., 2016). The selected

TABLE 2: Preliminary (Pre) and maximum likelihood estimates, with von Mises (s.e.) and sandwich (s.e.R) standard errors, of the parameters of model (1) fitted to the sand hopper data, with the mean direction involving either six parameters (MLE, see model (11)) or four parameters (MLE2).

Parameter	β_1	β_2	β_3	β_4	β_5	β_6	κ_e	κ_a
Pre	0.186	-2.893	-0.005	1.910	1.304	-0.170	4.368	4.731
MLE	0.062	-0.632	-0.004	0.538	2.597	-0.168	4.170	4.559
s.e.	0.242	0.155	0.001	1.007	0.853	0.027	0.348	1.010
MLE2	-	-0.619	-0.004	-	2.742	-0.170	4.184	4.338
s.e.2	-	0.155	0.001	-	0.806	0.026	0.347	0.907
s.e.R	-	0.185	0.001	-	0.676	0.029	0.603	0.809

model involves seven predictors and $\mu_{ij}(\beta)$ has the following form,

$$\arctan\{\sin(\mathit{Azi}) + \beta_1 \sin(\mathit{DirW}) + \beta_3\, \mathit{SpeeW} \sin(\mathit{DirW}) + \beta_5\, \mathit{Eye} + \beta_6 T,$$
$$\cos(\mathit{Azi}) + \beta_1 \cos(\mathit{DirW}) + \beta_2 + \beta_3\, \mathit{SpeeW} \cos(\mathit{DirW}) + \beta_4\, \mathit{Eye}\}\,. \tag{11}$$

Preliminary estimates (Pre) of the six β-parameters and of (κ_e, κ_a) are given in Table 2. The circular variance of the dependent angle is 0.40, while it is 0.23 for the fixed-effects residuals. The seven explanatory variables allow a reduction of nearly 50% in the circular variance of y.

6.2. The random-effects model

Final parameter estimates were obtained by maximizing the log-likelihood (9) using the quasi-Newton algorithm discussed in Section 5. The maximum likelihood estimates and their standard errors are reported in Table 2.

As the Wald statistics for the null hypotheses $H_0: \ \beta_1 = 0$ and $H_0: \ \beta_4 = 0$ are small, a restricted model, with $\beta_1 = \beta_4 = 0$ was fitted. This results in a small drop in the maximum log-likelihood, from 260.06 to 259.78, and leads to the estimates MLE2 in Table 2. Robust standard errors, s.e.R, calculated with the sandwich covariance estimate $v(\hat\beta, \hat\kappa_e, \hat\kappa_a)$ defined in Section 3.3, are also provided. There are no systematic differences between the two sets of standard errors, either robust or model-based. The sine-sine correlation calculated with the final estimates of κ is 0.467, so the within-sand-hopper dependency is important. As $\hat\kappa_a = 4.37$ is large, the preliminary estimates calculated without accounting for the within-sand-hopper dependency are fairly accurate.

Circular plots of the conditional residuals $y_{ij} - \hat\mu_{ij} - \hat a_i$ and of the estimates of the random effects $\hat a_i$ are provided in Figures 1 and 2, respectively. They were produced using `circular`, the R-package of Agostinelli & Lund (2013). Both are approximately unimodal and symmetric. The circular variance of the conditional residuals is 0.11, as compared to 0.23 for the fixed-effects residuals $y_{ij} - \hat\mu_{ij}$. Thus the between-sand-hopper variability is important. Figure 1 shows some large residuals that are not really compatible with a von Mises distribution with parameter $\hat\kappa_e = 4.179$. This was ascertained with the sand-hopper residual mean resultant lengths $\bar R_i, \ i = 1, \ldots, 59$, whose range, (0.40, 0.99), is much larger than what is expected for von Mises samples of size 5 with concentration $\hat\kappa_e$, considering the approximate distribution given by Mardia & Jupp (2000, p. 80). A heavy-tailed distribution, such as a wrapped Cauchy, for the errors e_{ij} in model (1) might provide a better fit. The correlation matrix of the vectors $\{[\sin(y_{i1} - \hat\mu_{i1} - \hat a, \ldots, \sin(y_{i5} - \hat\mu_{i5} - \hat a_i)] : \ i = 1, \ldots, 59\}$ still shows a time trend that was identified in D'Elia (2001) and Lagona (2016), with larger correlations between residuals at con-

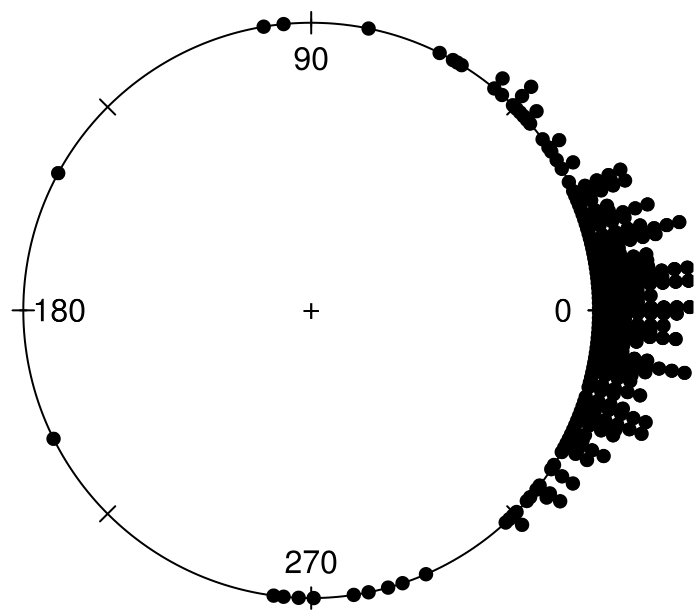

FIGURE 1: Raw circular data plot of the conditional residuals.

secutive times. Thus it would appear that the proposed model does not fully describe the error structure underlying the sand hopper movements. However, as proved in Section 3.3, inference for the vector β, and hence on the fixed part of the model, is unaffected by this misspecification of the error distribution.

We now provide a brief interpretation of the estimates of the elements of the vector β. The most important dependent angle is the sun azimuth: Sand hoppers move in the direction of the sun on their release. The variable *Eye* measures eye asymmetry; it is positive when the right eye is dominant. As $\hat{\beta}_5 > 0$, a dominant right eye increases the sine component of the model and suggests that a sand hopper then has a tendency to go west (or left). In a similar way, as $\hat{\beta}_6 < 0$ the later trials for a sand hopper have a smaller sine value suggesting an eastward time trend. The interaction between wind direction and wind speed gives a coefficient of $-0.004 \times$ *SpeeW* to the wind direction in Equation (2). As *SpeeW* ranges between 3 and 340, this coefficient varies between -0.01 and -1.36. When *SpeeW* is large, the coefficient of wind direction in Equation (2) is negative with large magnitude. Thus a strong wind impacts a sand hopper's orientation as it then tends to go upwind when it is released; a mild wind has almost no impact on its orientation.

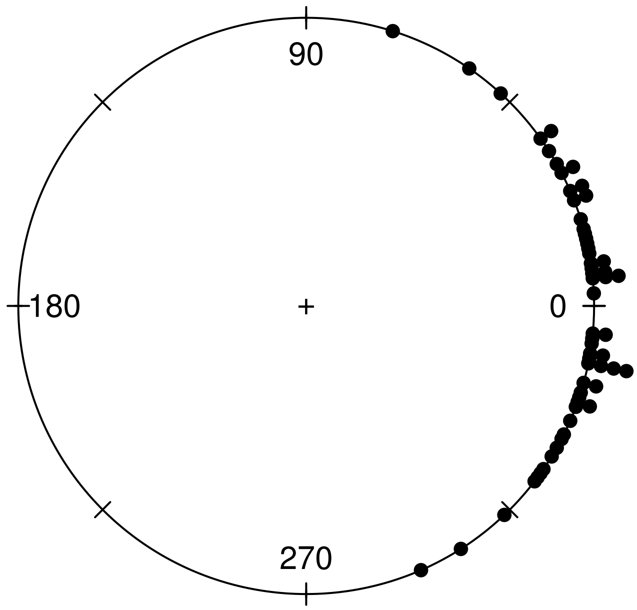

FIGURE 2: Raw circular data plot of the predictors of the cluster effects.

This interaction has not been noticed in previous analyses of this data set.

The analysis based on the random-effects model provides a parsimonious description of a complex data set using only six parameters. An alternative analysis using a projected normal density with latent classes presented in Maruotti (2016) uses 26 parameters to fit these data (see also Maruotti et al., 2016). A projected normal analysis uses explanatory variables to explain both the mean direction and the error heterogeneity noted in the discussion of Figure 1. It does not distinguish parameters associated with the mean direction from those driven by the error concentration. This makes their interpretation more challenging, as noted by Maruotti (2016), and a comparison with the analysis based on a random-effects model is not really feasible.

7. DISCUSSION

Holmquist & Gustafsson (2017) use model (1) to carry out tests for the null hypothesis of uniformity of the random effects, $H_0:\ \kappa_a = 0$. This article uses model (1) to account for a within-cluster dependency in an angular regression model. The implementation of the proposed model in a regression context has been discussed, including the evaluation of the within-cluster dependency, the prediction of angular random effects, and parameter estimation. The fixed part of the proposed model determines the mean direction of y. It can be modelled independently of the errors, and the inference derived from the proposed random-effects model is robust to a misspecification of the error distribution. This contrasts with a projected normal analysis whose parameters are more difficult to interpret as they are driven by the within-cluster dependence, the mean direction, and the concentration of the dependent angle. The example of Section 6 shows that the two models can lead to analyses that are drastically different. As pointed out by a referee, it would be interesting to extend the model proposed in this article by allowing the parameters β in μ_{ij} to be random variables that vary from one cluster to the next. For such a model the log-likelihood does not have a closed form expression. First it would have to be approximated using quadrature methods, then an optimizer such as R's `optim` would have to be used to maximize this approximate log-likelihood. This will be future work.

ACKNOWLEDGEMENTS

The authors are grateful to Antonello Maruotti for providing the sand hopper data set analyzed in section 6. This article benefitted from the financial support of the Institute of Statistical Mathematics of Japan, the Fonds de Recherche du Québec, Nature et Technologies, of the Canada Research Chair in Statistical Sampling and Data Analysis, and of the Natural Sciences and Engineering Research Council of Canada. The article of the second author was supported by JSPS KAKENHI Grant Number JP25400218.

BIBLIOGRAPHY

Abe, T. & Pewsey, A. (2011). Sine-skewed circular distributions. *Statistical Papers*, 52, 683–707.

Agostinelli, C. & Lund, U. (2013). R package *circular*: Circular Statistics (version 0.4-7). https://r-forge.r-project.org/projects/circular/.

Artes, R., Paula, G. A., & Ranvaud, R. (2000). Analysis of circular longitudinal data based on generalized estimating equations. *Australian and New Zealand Journal of Statistics*, 42, 347–358.

Battese, G. E., Harter, R. M., & Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83, 28–36.

Cremers, J., Mulder, K. E., & Klugkist, I. (2018). Circular interpretation of regression coefficients, *British Journal of Mathematical and Statistical Psychology*, 71, 75–95

D'Elia, A. (2001). A statistical model for orientation mechanism. *Statistical Methods and Applications*, 10, 157–174.

Downs, T. D. & Mardia, K. (2002). Circular regression. *Biometrika*, 89, 683–697.

Fisher, N. I. & Lee, A. J. (1992). Regression models for an angular response. *Biometrics*, 48, 665–677.

Hall, D. B. & Shen, J. (2015). Marginal projected multivariate linear models for clustered angular data. *Australian and New Zealand Journal of Statistics*, 57, 241–257.

Hernandez-Stumpfhauser, D., Breidt, F. J., & van der Woerd, M. J. (2017). The general projected normal distribution of arbitrary dimension: modeling and Bayesian inference. *Bayesian Analysis*, 12, 113–133.

Holmquist, B. & Gustafsson, P. M. J. (2017). A two-level directional model for dependence in circular data. *Canadian Journal of Statistics*, 45, 461–478.

Jammalamadaka, S. R. & Sarma, Y. R. (1988). A correlation coefficient for angular variables. In K. Matusita K. (Ed.) *Statistical Theory and Data Analysis II*, 349–364.

Jona-Lasinio, G., Gelfand, A. F., & Jona-Lasinio, M. (2012). Spatial analysis of wave direction data using wrapped Gaussian processes. *Annals of Applied Statistics*, 6, 1478–1498.

Jones, M. C., Pewsey, A., & Kato, S. (2015). On a class of circulas: copulas for circular distributions. *Annals of the Institute of Statistical Mathematics*, 67, 843–862.

Kato, S., Shimizu, K., & Shieh, G. S. (2008). A circular-circular regression model. *Statistica Sinica*, 18, 633–646.

Kato, S. & Jones, M. C. (2015). A tractable and interpretable four-parameter family of unimodal distributions on the circle. *Biometrika*, 102, 181–190.

Lagona, F. (2016). Regression analysis of correlated circular data based on the multivariate von Mises distribution. *Environmental and Ecological Statistics*, 23, 89–113.

Mardia, K. V., Hughes, G., Taylor, C. C., & Singh, H. (2008). A multivariate von Mises distribution with applications to bioinformatics. *Canadian Journal of Statistics*, 36, 99–109.

Mardia, K. V. & Jupp, P. E. (2000). *Directional Statistics*. Wiley: New York.

Maruotti, A. (2016). Analyzing longitudinal circular data by projected normal models: a semi-parametric approach based on finite mixture models. *Environmental and Ecological Statistics*, 23, 257–277.

Maruotti, A., Punzo, A., Mastrantonio, G., & Lagona, F. (2016). A time-dependent extension of the projected normal regression model for longitudinal circular data based on a hidden Markov heterogeneity structure. *Stochastic Environmental Research and Risk Assessment*, 30, 1725–1740.

Mastrantonio, G., Lasinio, G. J., & Gelfand, A. E. (2016). Spatio-temporal circular models with non-separable covariance structure. *Test*, 25(2), 331–350.

Nicosia, A., Duchesne, T., Rivest, L.-P., & Fortin, D. (2017). A general hidden state random walk model for animal movement. *Computational Statistics and Data Analysis*, 105, 76–95.

Nuñez-Antonio, G. & Gutiérrez-Peña, E. (2014). A Bayesian model for longitudinal circular data based on the projected normal distribution. *Computational Statistics and Data Analysis*, 71, 506–519.

Pewsey, A., Neuhäuser, M., & Ruxton, G.D. (2013). *Circular Statistics in R*, Oxford University Press: Oxford.

Presnell, B., Morrison, S. P., & Littell, R. (1998). Projected multivariate linear models for directional data. *Journal of the American Statistical Association*, 93, 1068–1077.

Rao, J. N. K. & Molina, I. (2015). *Small Area Estimation,* 2nd ed., Wiley: New York.

Rivest, L.-P. (1982). Some statistical methods for bivariate circular data. *Journal of the Royal Statistical Society Series B*, 44, 81–90.

Rivest, L.-P., Duchesne, T., Nicosia, A., & Fortin, D. (2016). A general angular regression model for the analysis of data on animal movement in Ecology, *Journal of the Royal Statistical Society, Series C*, 66, 445–463.

APPENDIX

Proof of Proposition 2. The joint distribution of (y_1, y_2) is given in Equation (4); for $-\pi \leq y_1, y_2 < \pi$ it can be expressed as

$$f(y_1, y_2, a) = C \int_{-\pi}^{\pi} \exp\left\{ \kappa_a \cos a + \kappa_e \sum_{j=1}^{2} \cos(y_j - a) \right\} da.$$

where $C^{-1} = (2\pi)^3 I_0(\kappa_a)\{I_0(\kappa_e)\}^2$. Then the trigonometric moments of (y_1, y_2) can be calculated as

$$\begin{aligned}
&E\left\{e^{i(p_1y_1+p_2y_2)}\right\} \\
&= C\int_{[-\pi,\pi)^3} e^{i(p_1y_1+p_2y_2)} \exp\left\{\kappa_a \cos a + \kappa_e \sum_{j=1}^{2} \cos(y_j - a)\right\} da\, dy_1\, dy_2 \\
&= C\int \exp(\kappa_a \cos a)\left(\int e^{ip_1y_1} \exp\left\{\kappa_e \cos(y_1 - a)\right\}\, dy_1\right) \\
&\qquad \times \left(\int e^{ip_2y_2} \exp\left\{\kappa_e \cos(y_2 - a)\right\}\, dy_2\right) da \\
&= \frac{A_{p_1}(\kappa_e)A_{p_2}(\kappa_e)}{2\pi I_0(\kappa_a)} \int e^{i\{(p_1+p_2)a\}} \exp(\kappa_a \cos a)\, da \\
&= A_{p_1}(\kappa_e)A_{p_2}(\kappa_e)A_{p_1+p_2}(\kappa_a).
\end{aligned}$$

The second equality follows from Fubini's theorem. ■

Proof of Proposition 4. It is convenient to let $\tilde{y}_{ij} = y_{ij} - \mu_{ij}$. One has

$$\begin{aligned}
\frac{\partial \sigma_i}{\partial \beta} &= \frac{1}{2\sigma_i}\frac{\partial \sigma_i^2}{\partial \beta} \\
&= \kappa_e \frac{2\{\kappa_a + \kappa_e \sum \cos \tilde{y}_{ij}\} \sum \dot{\mu}_{ij} \sin \tilde{y}_{ij} - 2\kappa_e \sum \sin \tilde{y}_{ij} \sum \dot{\mu}_{ij} \cos \tilde{y}_{ij}}{2\sigma_i} \\
&= \kappa_e \sum \dot{\mu}_{ij} \sin(\tilde{y}_{ij} - \tilde{a}_i),
\end{aligned}$$

where $\tilde{a}_i$ is defined in Equation (7). In a similar way

$$\begin{aligned}
\frac{\partial \sigma_i}{\partial \kappa_e} &= \frac{1}{2\sigma_i}\frac{\partial \sigma_i^2}{\partial \kappa_e} = \frac{2\{\kappa_a + \kappa_e \sum \cos \tilde{y}_{ij}\} \sum \cos \tilde{y}_{ij} + 2\kappa_e \{\sum \sin \tilde{y}_{ij}\}^2}{2\sigma_i} \\
&= \sum \cos(\tilde{y}_{ij} - \tilde{a}_i).
\end{aligned}$$

The partial derivative with respect to κ_a follows similarly. ■

Received 9 December 2017
Accepted 8 March 2019