Environmental and Ecological Statistics (2021) 28:21–34
https://doi.org/10.1007/s10651-020-00471-3

Bayesian predictive model selection in circular random effects models with applications in ecological and environmental studies

Onur Camli[1] · Zeynep Kalaylioglu[1]

Received: 24 November 2019 / Revised: 21 September 2020 / Accepted: 30 September 2020 /
Published online: 26 October 2020

Abstract

In this paper we present a detailed comparison of the prediction error based model selection criteria in circular random effects models. The study is primarily motivated by the need for an understanding of their performance in real life ecological and environmental applications. Prediction errors are based on posterior predictive distributions and the model selection methods are adjusted for the circular manifold. Plug-in estimators of the circular distance parameters are also considered. A Monte Carlo experiment scheme taking the account of various realistic ecological and biological scenarios is designed. We introduced a coefficient that is based on conditional expectations to examine how the deviation from von Mises (vM) distribution, the standard choice in applications, effects the performances. Our results show that the performances of widely used circular predictive model selection criteria mostly depend on the sample size as well as within-sample-correlation. The approaches and selection strategies are then applied to investigate orientational behaviour of *Talitrus saltator* under the risk of dehydration and direction of wind with respect to associated atmoshperic variables.

Keywords Directional statistics · Model comparison · Sandhoppers · von Mises distribution · Wind direction

Handling Editor: Pierre Dutilleul.

✉ Zeynep Kalaylioglu
kzeynep@metu.edu.tr

Onur Camli
camli@metu.edu.tr

[1] Department of Statistics, Middle East Technical University, Ankara, Turkey

1 Introduction

Circular random effects models are used in various different environmental and ecological studies in which response data are angular or directional and observed for each subject multiple times (D'Elia 2001; Hall and Shen 2015; Maruotti 2016; Maruotti et al. 2016; McMillan et al. 2013; Nunez-Antonio and Gutierrez-Pena 2014). For instance, as seen in our Sect. 5.1, *Talitrus saltator*, a beach amphipod, are observed in many ecological studies to investigate their orientation with respect to important environmental and ecological factors (Scapini 1997). In these studies, the response data consist of directional recordings of their consecutive jumps whereas covariate data consist of several environmental factors including wind direction and sun azimuth. As in such experiments on animal orientations, the aim is to understand the orientation of a particular species under given and advancing ecological conditions as well as to determine the environmental factors that play significant role in their orientation. In this regard, the focus is on determining the most significant factors that should take place in a model with good predictive abilities. This can be accomplished by using a predictive model comparison/selection method applied over a set of candidate nested circular random effect models. Circular prediction errors are based on circular distance between the observed and predicted.

Describing a distance function on the circular space requires adjusting the standard formulas for the circle. Ravindran and Ghosh (2011) considered a model selection criteria based on minimizing a predictive loss defined on a circle and conducted a simulation study to investigate the performance of this method in a non-regression setting in terms of selecting the most suitable distribution for its predictive abilities. In the regression setting, Maruotti (2016) employed the trigonometric distance formula suggested by Jammalamadaka and SenGupta (2001) to define prediction error and used it to make a selection over the set consisting of models with different random effect distributions. These criteria are used by others in the circular literature to evaluate the predictive performances of models (Mastrantonio et al. 2016).

However, to the best of our knowledge, there is no literature describing the performances of existent predictive model selection methods for circular data. Results of such an investigation would provide the researchers with a guideline regarding the performances of these predictive model selection methods with respect to the characteristics of their study design such as number of subjects, number of replications from each subject, and correlation between the observations obtained from the same subject. It is important to interpret the result of a model selection operation in the light of such information. Therefore, the aim of this paper is to provide the performances of predictive model selection criteria in most common practical scenarios whose aim is to obtain a parsimonious predictive model for repetitively occurring circular data and to develop guidelines for the researchers.

The rest of the paper is hence organized as follows. Section 2 previews Bayesian longitudinal circular-linear random effects models. In this section, Bayesian analysis of these models is also given. We present the predictive loss based model selection criteria and some comments in Sect. 3. Section 4 presents several simulation examples elaborating on the performance of the predictive model selection approaches. These approaches are then applied in Sect. 5 on *T. saltator* and wind direction datasets to illustrate the use of circular predictive model selection discussed here to identify predictive environmental factors on the orientation of these amphipods. Section 6 concludes the article with a critical discussion of our findings and lays out a general guideline for researchers including environment, ecology, and biology scientists.

2 Model preview

2.1 Description of the model

Let $\theta_{ij} \in [-\pi, \pi]$, $i = 1, ..., n,\ j = 1, ..., m_i$ be a circular random variable in a longitudinal study. θ_{ij} denotes the circular response for subject i on the jth measurement time. Also let $\mathbf{X}_{ij}$ be the vector of P linear covariates for subject i observed at jth time point. Let $\mathbf{Z}_{ij}$ be a known subset of $\mathbf{X}_{ij}$ of dimension q that may include 1 for a random intercept. We consider vM distribution as this is the distribution of choice for circular regression in most applied problems. Random effects vM model (which we will denote hereafter by LCREM standing for longitudinal circular random effects model) is given by the following hierarchical framework:

$$\begin{aligned} \theta_{ij}|\mathbf{b}_i &\sim vM(\mu_{ij}, \kappa), \\ \mu_{ij} &= \mu + g(\mathbf{Z}_{ij}\mathbf{b}_i + \mathbf{X}_{ij}\beta) \\ \mathbf{b}_i &\sim N_q(\mathbf{0}, \Sigma). \end{aligned} \tag{1}$$

for $i = 1, ..., n$ and $j = 1, ..., m_i$ where μ_{ij} and κ are the conditional mean function and concentration parameter given $\mathbf{b}_i$, $\mathbf{b}_i$ a $q \times 1$ vector of unobserved subject-specific random effects for subject i which is usually assumed to follow a multivariate normal distribution (N_q) with zero mean for identifiability and variance-covariance matrix Σ. Also, $\mu \in [-\pi, \pi]$ is an offset parameter and β is a $P \times 1$ vector of regression coefficients (fixed effects), g is a link function such as $g(u) = 2\arctan(u)$ in which case inverse tangent is the link, $\mathbf{b}_i$ and circular residuals are assumed independent. It is also assumed that θ_{ij} and $\theta_{ij'}$ are conditionally independent given subject-specific random effects. A simpler form of the model, which includes only a subject-specific random intercept, was methodologically undertaken earlier by D'Elia (2001) for a vM distribution and by Nunez-Antonio and Gutierrez-Pena (2014), and Maruotti (2016) for a projected normal distribution. Here**,** we consider the general version of these models where there are both random intercepts and random coefficients. If there is a belief that, in the underlying natural phenomena, covariates such as environmental factors affect subjects differently, then random coefficient model can be used. Otherwise, random intercept can be used to model dependent angular data.

Interpretation of the regression coefficients in a circular model is different than those in a regression model defined on Euclidean space and thus requires special attention in practice. The correct interpretation requires the origin and the direction of rotation to be explicitly stated per data application. To illustrate, consider that the circular variable is direction of wind, which is a common variable in air pollution studies. Suppose that the origin is North, the direction of rotation is clockwise and the inverse tangent link is used. Then, a positive β_p refers to a clockwise advance from North in the response with a one unit increase in the pth covariate. In other words, when there is one unit change in the pth covariate, there will be a magnitude of β_p change in $\tan(\frac{\theta_{ij}-\mu}{2})$.

2.2 Bayesian analysis

Bayesian analysis of the model is straightforward. We can use the prior distributions that are used for standard Bayesian analyses of vM distribution and random effects models. Namely, $\mu \sim G_{[-\pi,\pi]}$, $\beta \sim N_Q(\mu_\beta, \Sigma_\beta)$, $\kappa \sim Ga(a_\kappa, b_\kappa)$ and $\Sigma^{-1} \sim Wishart(R, d)$, where $G_{[-\pi,\pi]}$ is a distribution defined on $[-\pi, \pi]$ such as circular uniform distribution, μ_β, Σ_β, a_κ, b_κ, μ_0, σ_0^2, R and d are fixed hyper-parameters.

Letting $D_{obs} = \{\boldsymbol{\theta}, \boldsymbol{X}\}$ and $D_{comp} = \{\boldsymbol{\theta}, \boldsymbol{X}, \mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be the observed and complete data respectively, joint posterior distribution of all unknown quantities is given by

$$f(\mu, \boldsymbol{\beta}, \kappa, \Sigma, \boldsymbol{b_1}, \ldots, \boldsymbol{b_n}|D_{obs}) \propto L(\mu, \boldsymbol{\beta}, \kappa, \Sigma|D_{comp}) f(\mu) f(\boldsymbol{\beta}) f(\kappa) f(\Sigma)$$

where $L(\mu, \boldsymbol{\beta}, \kappa, \Sigma|D_{comp})$ is the complete data likelihood function and the rest are the prior probability density functions (pdf's). Complete data likelihood function is given as follows

$$L(\mu, \boldsymbol{\beta}, \kappa, \Sigma|D_{comp}) = \prod_{i=1}^{n} \left(\prod_{j=1}^{m} f(\theta_{ij}|\mathbf{b}_i, \mu, \boldsymbol{\beta}, \kappa) \right) f(\mathbf{b}_i|\Sigma), \tag{2}$$

where $f(\theta_{ij}|\mathbf{b}_i, \mu, \boldsymbol{\beta}, \kappa)$ denotes the conditional circular pdf (vM) and $f(\mathbf{b}_i|\Sigma)$ is the prior distribution for $\mathbf{b}_i$ for $i = 1, \ldots, n$. Hence,

$$\begin{aligned} &L(\mu, \boldsymbol{\beta}, \kappa, \Sigma|D_{comp}) \\ &\quad = \prod_{i=1}^{n} \left(\prod_{j=1}^{m} [2\pi I_0(\kappa)]^{-1} \exp\{\kappa \cos(\theta_{ij} - \mu - 2\arctan(\mathbf{b}_{0i} + \boldsymbol{\beta X}_{ij}))\} \right) \\ &\qquad |\Sigma|^{-0.5} exp\{-0.5\mathbf{b}_i' \Sigma^{-1} \mathbf{b}_i\} \end{aligned} \tag{3}$$

In order to obtain the predicted data, a posterior predictive distribution is used as follows. Let $\theta^{pred} = (\theta_1^{pred}, ..., \theta_n^{pred})$ and $\theta^{obs} = (\theta_1^{obs}, ..., \theta_n^{obs})$ be predicted and observed data, respectively. Predicted data drawn from posterior predictive distribution is given below

$$f(\theta_i^{pred}|\theta^{obs}) = \int f(\theta_i^{pred}|\mu, \boldsymbol{\beta}, \kappa, \Sigma, \mathbf{b}) f(\mu, \boldsymbol{\beta}, \kappa, \Sigma, \mathbf{b}|\theta^{obs}) d\mu, d\boldsymbol{\beta}, d\kappa, d\Sigma, d\mathbf{b}, \quad for\ i = 1, ..., n, \tag{4}$$

where $f(\theta^{pred}|\mu, \boldsymbol{\beta}, \kappa, \Sigma, \mathbf{b})$ is the posterior predictive density of the data which is the density evaluated at θ^{pred} given the parameter vector $(\mu, \boldsymbol{\beta}, \kappa, \Sigma, \mathbf{b})$ and $f(\mu, \boldsymbol{\beta}, \kappa, \Sigma, \mathbf{b}|\theta^{obs})$ is the joint posterior density.

Posterior inference of parameters and predictions are obtained using MCMC algorithms that can be easily implemented in in OpenBUGS (see Appendix 1). Under ergodicity conditions, sample moments of random samples from full conditional densities converge in probability to posterior moments.

3 Predictive density based model assessment, comparison, and selection

Usual prediction error that is defined for linear response variables is not applicable for predictive models defined on the circles. To define circular prediction error, one can consider the trigonometric distance (Jammalamadaka and SenGupta 2001; Maruotti 2016; Maruotti et al. 2016). Letting K and L be two distinct points on the circle with angles from the origin symbolized by α and β respectively, trigonometric distance function denoted by d(. , .) is given by $d(\alpha, \beta) = 1 - cos(\alpha - \beta)$ and $d(\alpha, \beta) \in [0, 2]$ which is a monotone increasing function of the angle between the points K and L. Denoting the predicted and observed angular longitudinal data by θ_{ij}^{pred} and θ_{ij}^{obs} respectively, total circular prediction error (CPE) and the predictive model selector (CPD_1) are given by

$$\begin{aligned} CPE &= \sum_{i=1}^{n} \sum_{j=1}^{m_i} d(\theta_{ij}^{pred}, \theta_{ij}^{obs}) \\ CPD_1 &= E[CPE|\theta^{obs}] \end{aligned} \tag{5}$$

where expectation is over the posterior predictive density.

Predictive loss function can also be defined as a circular distance, which is the absolute difference between the predicted and the observed responses adjusted for the circular sample space (Ravindran and Ghosh 2011). Then the total absolute prediction error (APE) and the corresponding model selector, denoted here by CPD_2, are given by

$$\begin{aligned} APE &= \sum_{i=1}^{n} \sum_{j=1}^{mm_i} min\left(|\theta_{ij}^{pred} - \theta_{ij}^{obs}|, 2\pi - |\theta_{ij}^{pred} - \theta_{ij}^{obs}|\right) \\ CPD_2 &= E\left[APE|\theta^{obs}\right] \end{aligned} \tag{6}$$

The models with lower CPD values have a better predictive ability. Note that, unlike the loss functions for linear data, CPD_2 for circular data can not be decomposed into two terms as the sum of penalty and a term for goodness of fit.

One can also consider the *plug-in* estimators of CPE and APE where predictions are plugged in by their posterior estimators. The resulting plugged-in CPE and plugged-in APE are denoted by PCPE and PAPE and presented below.

$$PCPE = \sum_{i=1}^{n}\sum_{j=1}^{mm_i}\left(1 - \cos(E[\theta_{ij}^{pred}|\theta_{ij}^{obs}] - \theta_{ij}^{obs}\right) \tag{7}$$

$$PAPE = \sum_{i=1}^{n}\sum_{j=1}^{mm_i} min\left(|E[\theta_{ij}^{pred}|\theta_{ij}^{obs}] - \theta_{ij}^{obs}|, 2\pi - |E[\theta_{ij}^{pred}|\theta_{ij}^{obs}] - \theta_{ij}^{obs}|\right) \tag{8}$$

It must be noted that lower PCPE and PAPE indicate better predictive performances.

OpenBUGS sofware was used to perform the calculations of CPD_1 and CPD_2. An examplary code provided in Fig. 1 in Appendix shows how CPD_1 and CPD_2 were calculated. PCPE and PAPE are computed in R using the output delivered by OpenBUGS.

4 Simulation study

In this section performances of these model selection methods are investigated from different aspects under several scenarios that represent studies conducted in practice. In that sense, we consider unbalanced longitudinal studies that are common especially in ecological researches. Average number of observations per subject is considered to be seven, which can be found in many studies in practice on animal orientations. The number of observations per subject hence is generated from a discrete uniform distribution with a mean equals seven. Our Monte Carlo experiment is controlled for sample size, effect size and latent within cluster variation enabling us to evaluate the performances for different underlying settings. We begin with describing the specific aims of the two separate simulations studies and give their layout, which is followed by the specifics of the Bayesian computations. In the first simulation study, the aim is to evaluate and compare the suggested circular model selection methods in their finite sample behaviors as well as in their consistency property, i.e. whether the frequency of selecting the true model converges to unity with increasing sample size. For this simulation study, data are generated as follows. Circular data are generated from $\theta_{ij} \sim vM(\mu_{ij}, \kappa = 2)$ following the true conditional mean model TM: $\mu_{ij} = \mu + 2\arctan(b_{0i} + (b_{1i} + \beta_1)x_{ij} + \beta_2 x_{ij}^2)$. Here two different true settings are considered for the fixed effects. These are $(\beta_1, \beta_2)^T = (2.5, 1.5)$ and $(\beta_1, \beta_2)^T = (2.5, 0.3)$ representing an emphasized quadratic effect and a relatively moderate one respectively. Random effects are generated following $N_2(0, \Sigma)$ where variance of random slope and covariance are set at $\Sigma_{22} = 4$ and $\Sigma_{12} = 0.25$ respectively. Three different settings are considered for variance of random intercept, Σ_{11}, as 0.18, 0.72 and 5.28 that lead to different within cluster variation. Time-dependent linear explanatory variable x_{ij} are generated from N(0,1) without loss of generality. Each simulated data set is fitted by the following circular random effects models denoted by M1 and M2 and predictive selectors are calculated for each model.

M1: $\mu_{ij} = \mu + 2\arctan(b_{0i} + (b_{1i} + \beta_1)x_{ij} + \beta_2 x_{ij}^2)$
M2: $\mu_{ij} = \mu + 2\arctan(b_{0i} + (b_{1i} + \beta_1)x_{ij})$.

In the second simulation study, the aim is to investigate the strength of the methods in their decision for choosing a vM model. The strength of this decision is defined as the ratio of expected value of the criterion under a non-vM assumption to its expected value under vM assumption. The criterion with a ratio further away from unity is more decisive in its identification of vM. For this study, data are generated from the following models denoted by TM1 and TM2 where the underlying distributions are very similar to each other.

$$\begin{aligned}&\text{TM1: } \theta_{ij} \sim vM(\mu_{ij}, \kappa = 2)\\ &\mu_{ij} = \mu + 2\arctan(b_{0i} + (b_{1i} + \beta_1)x_{ij}),\end{aligned}$$

$$\begin{aligned}&\text{TM2: } \theta_{ij} \sim WC(\mu_{ij}, \rho = 0.62)\\ &\mu_{ij} = (\mu + b_{0i} + (b_{1i} + \beta_1)x_{ij})\ [mod\ 2\pi]\end{aligned}$$

where WC is Wrapped Cauchy distribution, μ_{ij}s denote the mean direction while κ and ρ denote the concentration parameters and the true settings for other parameters are as before. For each simulated dataset, vM regression (TM1) is fitted.

Bayesian analysis of the models in these studies is carried out in OpenBUGS. We used the noninformative priors given in Sect. 2.2. Trace plots and Brooks–Gelman–Rubin statistic are used for convergence diagnostics and determining the burn-in period. MCMC iterations were run until Monte Carlo errors based on the Markov chain were less than 5% of the posterior standard deviations. Finally, posterior means (i.e. expectation of the posterior distributions) are used for estimating the model parameters and the predictions. All Monte Carlo scenarios were repeated 100 times.

The results of the investigation are given in Tables 1 and 2. Table 1 gives the frequency of selecting the true model for each criterion. The results show that model selection methods based on CPD_1 and CPD_2 favor the true model more often compared to the other methods implying overall superiority in identification of the true mean model. However the methods perform equally well if the underlying mean model has strong non-linearity. The results also show that each criterion is consistent as the frequency of true selection converges to one with increasing sample size, fastest for CPD_1 and CPD_2. Furthermore, Table 2 in particular presents that the strength ratio for CPD_1 is generally greater than the others implying that CPD_1 can identify the true vM more decisively.

5 Applications

In this section we employ two datasets to illustrate the use of the methods. The first dataset is a standard dataset used in circular literature (Sect. 5.1) whereas the second dataset comes from an atmospheric study (Sect. 5.2). Aim of Sect. 5.1 is to provide a detailed analysis showing the contribution of our model selection methods with respect to the current circular literature available on this dataset. Aim of Sect. 5.2 is to illustrate an application of the model selection methods in practice.

Table 1 Frequency of selecting the true mean model

Σ_{11}	Criterion	$\beta_2 = 0.3$					$\beta_2 = 1.5$				
		n = 20	50	100	250	500	n = 20	50	100	250	500
0.18	CPD_1	49	60	76	96	98	100	100	100	100	100
	CPD_2	49	61	77	96	98	100	100	100	100	100
	$PCPE$	35	56	66	87	92	97	100	100	100	100
	$PAPE$	32	55	60	86	90	97	100	100	100	100
0.72	CPD_1	49	68	76	91	98	100	100	100	100	100
	CPD_2	49	68	76	90	98	100	100	100	100	100
	$PCPE$	34	56	57	79	83	99	100	100	100	100
	$PAPE$	36	56	58	79	83	99	100	100	100	100
5.28	CPD_1	36	73	83	84	99	98	100	100	100	100
	CPD_2	36	71	83	84	99	98	100	100	100	100
	$PCPE$	23	52	67	75	97	98	100	100	100	100
	$PAPE$	23	50	70	75	97	98	100	100	100	100

Table 2 Monte Carlo approximation of strength ratio

Σ_{11}	Criterion	n			
		50	100	250	500
0.18	CPD_1	1.518	1.544	1.547	1.538
	CPD_2	1.358	1.375	1.377	1.372
	$PCPE$	1.328	1.334	1.348	1.344
	$PAPE$	1.230	1.233	1.244	1.242
0.72	CPD_1	1.598	1.688	1.622	1.626
	CPD_2	1.414	1.428	1.430	1.432
	$PCPE$	1.368	1.387	1.397	1.391
	$PAPE$	1.268	1.281	1.287	1.285
5.28	CPD_1	1.715	1.741	1.705	1.710
	CPD_2	1.494	1.512	1.487	1.491
	$PCPE$	1.233	1.235	1.220	1.233
	$PAPE$	1.182	1.183	1.171	1.181

5.1 Orientation of Talitrus saltator

In this section we use the sandhopper dataset to assess the performance of predictive model selection criteria in a real data application. The particular sandhopper species in consideration is *Talitrus saltator* and the main question is the factors effecting their orientation towards the sea under the risk of high dehydration. The dataset is taken from Nunez-Antonio and Gutierrez-Pena (2014) and originally given by Borgioli et al. (1999a, 1999b). The data have previously been analyzed by Borgioli et al. (1999a, 1999b), D'Elia (2001), Lagona (2016), Maruotti (2016), Maruotti et al. (2016), Nunez-Antonio and Gutierrez-Pena (2014), and Song (2007). The dataset includes their escape directions in angles with respect to North that are measured every ten minutes resulting in five recordings. The study also includes data on environmental factors such as direction of the wind and sun azimuth as well as biological characteristics of the sandhoppers such as their left and right ocular diameters. Wind direction and sun azimuth are angular and categorized prior to modeling as in the previous circular analysis of the data (Nunez-Antonio and Gutierrez-Pena 2014; Lagona 2016; Maruotti 2016; Rivest and Kato 2019). Accordingly, there are four categories for wind direction and two categories for sun azimuth. For wind: wind from land [337°, 66°] (reference); wind from longshore-east (LSE) [67°, 156°]; wind from sea (SEA) [157°, 246°]; wind from longshore-west (LSW) [247°, 336°]. For sun azimuth: morning (MOR) [124°, 149°]; afternoon [240°, 269°] (reference). Finally, an eye symmetry index is constructed, as in Borgioli et al. (1999a, 1999b), and D'Elia (2001), using the ocular diameters and used in the preceding analyses: Eye = log(max diameter of right eye × min diameter of right eye) − log(max diameter of left eye × min diameter of left eye) which measures the difference between the sizes of the right and the left eye.

5.1.1 Exploratory analysis

According to the circular histograms previously given in D'Elia (2001), marginal distribution of the escape directions for each release is a symmetric and unimodal circular distribution. Circular summary statistics, mean direction ($\bar{\theta}$), mean resultant length ($\bar{R}$), circular variance (V), circular symmetry coefficient (s) and p-values of large-sample test for circular symmetry, that are originally given in Pewsey (2002) are reproduced in Table 3. These results show that the mean direction ($\bar{\theta}$) at each release is close to 201° [which is the theoretical escape direction (TED)] and there is a gradual approach to TED after each jump. Also notice the increase and decrease over the releases in terms of $\bar{R}$ and V respectively which indicates that their orientations tend to the same direction as they get closer to the sea. Circular symmetry coefficient (s) being around zero verifies that each marginal distribution is a symmetric and unimodal distribution. We performed a large-sample test for reflective symmetry using the method by Pewsey (2002) to investigate whether the hypothesis of circular reflective symmetry is supported by the data. Since all p-values are greater than 0.05 as seen in the last column of Table 3, it is clear that each marginal distribution is a symmetric distribution.

We tested the plausability of vM assumption with data on five consecutive releases collapsed together. Watson's goodness of fit test (Jammalamadaka and SenGupta 2001) and its p-value are 0.0883 and 0.092 (at significance level 0.05), respectively, implying that vM seems to be a plausible distribution for the data.

Circular autocorrelation coefficients for escape directions are given in Table 4. Clearly, there is a noticeable circular autocorrelation between successive releases and as seen in the results autocorrelation within same animal decreases as the lag between the two jump points increases. The angles observed on the last two jumps are the most correlated (0.87). This within-correlation needs to be accounted for in the regression analysis.

Table 3 Circular summary statistics for each marginal distribution

Release	$\bar{\theta}$	$\bar{R}$	V	s	p-value
1st	167.088	0.523	0.477	−0.244	0.283
2nd	171.401	0.528	0.472	0.026	0.912
3rd	193.242	0.576	0.424	0.098	0.733
4th	190.887	0.627	0.373	0.170	0.582
5th	194.585	0.667	0.333	0.204	0.543

Table 4 Autocorrelation coefficient for escape directions

Release	1st	2nd	3rd	4th	5th
1st	1	0.76	0.58	0.61	0.56
2nd		1	0.70	0.67	0.68
3rd			1	0.77	0.69
4th				1	0.87
5th					1

5.1.2 Modeling using LCREM

Previously Borgioli et al. (1999a, 1999b), D'Elia (2001), and Song (2007) used a variance component model under a vM distribution assumption while later on Nunez-Antonio and Gutierrez-Pena (2014), Maruotti (2016) and Maruotti (2016) considered random effects model with projected normal distribution. On the multivariate end, Lagona (2016) employed a fixed effects regression with multivariate vM distribution. Here we use LCREM and consider the models shown in Table 5 that will compete based on the circular predictive model selection criteria. Time variable in the models is the order of the jump taking the numeric values from one to five.

As can be seen in Table 6, CPD_1 and CPD_2 tend to select "Sun+Eye" or "Sun+Wind+Eye" while PAPE and PCPE tend to select "Sun+Wind" or "Sun+Wind+Eye". The full model is selected by all four criteria. This is in line with the previous findings in the literature. Our results further show that "Sun+Eye" model is equally adequate in terms of controlled prediction errors which indicates in turn that sun azimuth and eye ocular structure may be the main driving forces behind the orientation of sandhoppers towards the sea under dehydration threat. Further inference based on "Sun+Eye" model is given in Table 7. Estimate and Std in the table correspond to posterior means and posterior standard deviations. Accordingly, for Eye=0 and Time=1 (first jump), mean direction in the morning is $\mu_{ij} = 2.871 + 2\arctan(0.239 \times 1 - 1.101 \times 0 + 0.054 \times 1) \approx 197.097^\circ$ while in the afternoon it is $\mu_{ij} = 2.871 + 2\arctan(0.239 \times 0 - 1.101 \times 0 + 0.054 \times 1) \approx 170.168^\circ$. Other parameters can be interpreted in the similar fashion.

5.2 Eddy covariance flux data set: wind direction

The second dataset considered in this paper is Coastal Biodiversity and Ecosystem Service Sustainability eddy covariance flux data for Abbotts Hall, Essex (Hill and Chocholek 2016). Data collection was carried out at Abbotts Hall marsh from 15

Table 5 Nested models

Covariates	Mean models
Sun	$\mu_{ij} = \mu + 2\arctan(b_{0i} + \beta_1 MOR + \beta_2 Time)$
Sun + eye	$\mu_{ij} = \mu + 2\arctan(b_{0i} + \beta_1 MOR + \beta_2 Eye + \beta_3 Time)$
Sun + wind	$\mu_{ij} = \mu + 2\arctan(b_{0i} + \beta_1 MOR + \beta_2 LSE + \beta_3 SEA + \beta_4 LSW + \beta_5 Time)$
Sun + wind + eye	$\mu_{ij} = \mu + 2\arctan(b_{0i} + \beta_1 SUN + \beta_2 LSE + \beta_3 SEA + \beta_4 LSW + \beta_5 Eye + \beta_6 Time)$

Table 6 Model comparison for orientation of *Talitrus saltator*

Tools	Models			
	Sun	Sun + eye	Sun + wind	Sun + wind + eye
CPD_1	90.520	89.870	90.490	89.910
CPD_2	206.900	206.000	206.800	206.100
$PCPE$	231.055	231.110	230.184	230.329
$PAPE$	390.004	390.237	389.2160	389.153

Table 7 Parameter estimates

	Par. Est.	Std.	MC error	95% Credible Int.
μ	2.871	0.083	0.002	(2.701, 3.031)
Sun	0.239	0.096	0.003	(0.052, 0.428)
Eye	−1.101	0.101	0.002	(−1.300, −0.904)
Time	0.054	0.013	0.0003	(0.028, 0.080)
κ	3.655	0.278	0.005	(3.125, 4.213)
$\sigma^2_{b_0}$	0.463	0.124	0.002	(0.270, 0.755)

Table 8 Nested models for wind direction

Covariates	Mean models
Random intercept models	
Air Temp	$\mu_{ij} = \mu + 2\arctan(b_{0i} + \beta_1 AirTemp)$
Net Rad	$\mu_{ij} = \mu + 2\arctan(b_{0i} + \beta_1 NetRad)$
Air Temp + Net Rad	$\mu_{ij} = \mu + 2\arctan(b_{0i} + \beta_1 AirTemp + \beta_2 NetRad)$
Random intercept and slope models	
Air Temp	$\mu_{ij} = \mu + 2\arctan(b_{0i} + (\beta_1 + b_{1i}) AirTemp)$
Net Rad	$\mu_{ij} = \mu + 2\arctan(b_{0i} + (\beta_1 + b_{1i}) NetRad)$
Air Temp + Net Rad	$\mu_{ij} = \mu + 2\arctan(b_{0i} + (\beta_1 + b_{1i}) AirTemp + (\beta_2 + b_{2i}) NetRad)$

December 2012 till 27 January 2015 in which a total of 34 variables including soil variables and atmospheric variables were measured repeatedly on each day. Sample dataset considered here consists of atmospheric variables recorded seven times a day during the first 4 months of the year 2013 resulting in $m_i = 7$ for $i = 1, \ldots, n = 75$.

Table 9 Model comparison for wind direction

Tools	Models Random intercept			Random intercept and slope		
	Air Temp	Net Rad	Air Temp + Net Rad	Air Temp	Net Rad	Air Temp + Net Rad
CPD_1	523.40	524.80	524.10	522.70	524.00	524.20
CPD_2	822.20	824.10	823.10	821.20	822.90	823.20
$PCPE$	537.65	539.75	539.29	536.76	539.05	539.82
$PAPE$	841.49	844.01	843.56	840.86	843.78	844.20

Our aim is to illustrate the use of our model selection methods for determining some of the variables that can be used to predict the direction of the wind. Three of the models given in Table 8 are random intercept models and the others are both random intercept and slope models.

Table 9 gives the comparison results of these models. Accordingly, all four criteria tend to select “Air Temp” model with random intercept and slope. This means that the relationship between air temperature and wind direction can change on different days. The results also imply that, when air temperature and net radiation are compared, the former is more predictive for wind direction.

6 Conclusion

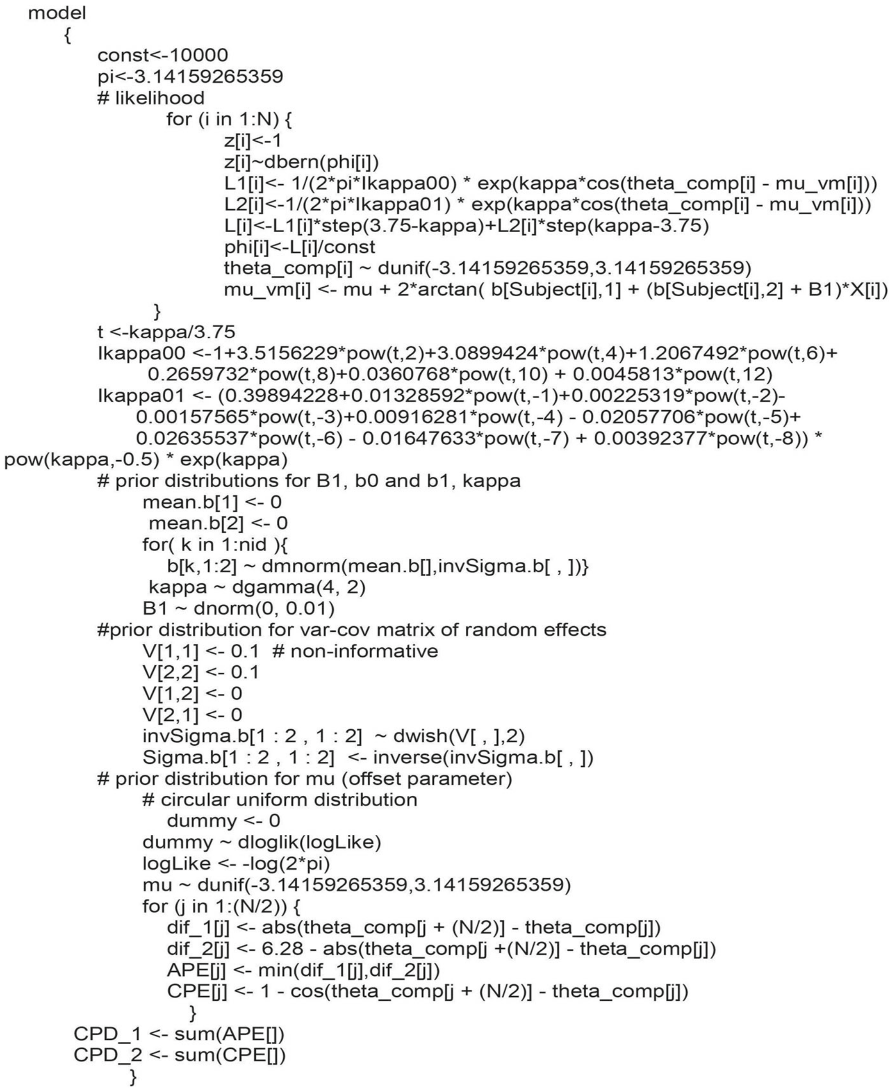
Model comparison and selection criteria based on circular prediction errors have been widely used for circular model applications. However, the performances of these methods in the circular regression analyses were unknown. In this paper we investigated the performance of these model selection criteria in circular vM based random effects models. vM distribution is the standard distribution of choice in most applied circular regression problems and thus the paper addresses model selection strategies in most of the applications.

We used extensive simulation studies illustrating the performances to identify the model with true mean function. We also examined the strength of the decisions of the methods on selecting the vM distribution when vM is indeed the true underlying process. The circular model selection approach based on CPD_1 or CPD_2 seems to have overall better performance. The results also show that the performance is primarily a function of sample size and within-sample-correlation. When uncertainty in the estimator is large and the function to be plugged-in is nonlinear (like APE and CPE), errors from using plug-in (in our case PAPE and PCPE) may be large (Rossi et al. 2005). This means that the size of uncertainty in the estimator affects the performance of plug-in estimators. (For instance, in another context, when estimating MSE, Maity and Sherman (2008) showed that plug-in estimators perform inferior compared to bootstrap method in adaptive linear regression). For the correct evaluation of performances of these estimators (PCPE and PAPE), uncertainty in the prediction, $\hat{\theta}^{pred}$, and theoretical properties of these tools should be further investigated.

Acknowledgements We sincerely thank the two anonymous reviewers for their critical reading, comments and suggestions that helped improve and clarify this manuscript.

A Appendix

See Fig. 1.



```
    model
        {
            const<-10000
            pi<-3.14159265359
            # likelihood
                    for (i in 1:N) {
                            z[i]<-1
                            z[i]~dbern(phi[i])
                            L1[i]<- 1/(2*pi*Ikappa00) * exp(kappa*cos(theta_comp[i] - mu_vm[i]))
                            L2[i]<-1/(2*pi*Ikappa01) * exp(kappa*cos(theta_comp[i] - mu_vm[i]))
                            L[i]<-L1[i]*step(3.75-kappa)+L2[i]*step(kappa-3.75)
                            phi[i]<-L[i]/const
                            theta_comp[i] ~ dunif(-3.14159265359,3.14159265359)
                            mu_vm[i] <- mu + 2*arctan( b[Subject[i],1] + (b[Subject[i],2] + B1)*X[i])
                    }
            t <-kappa/3.75
            Ikappa00 <-1+3.5156229*pow(t,2)+3.0899424*pow(t,4)+1.2067492*pow(t,6)+
                   0.2659732*pow(t,8)+0.0360768*pow(t,10) + 0.0045813*pow(t,12)
            Ikappa01 <- (0.39894228+0.01328592*pow(t,-1)+0.00225319*pow(t,-2)-
                 0.00157565*pow(t,-3)+0.00916281*pow(t,-4) - 0.02057706*pow(t,-5)+
                 0.02635537*pow(t,-6) - 0.01647633*pow(t,-7) + 0.00392377*pow(t,-8)) *
pow(kappa,-0.5) * exp(kappa)
            # prior distributions for B1, b0 and b1, kappa
                  mean.b[1] <- 0
                   mean.b[2] <- 0
                  for( k in 1:nid ){
                     b[k,1:2] ~ dmnorm(mean.b[],invSigma.b[ , ])}
                   kappa ~ dgamma(4, 2)
                  B1 ~ dnorm(0, 0.01)
            #prior distribution for var-cov matrix of random effects
                  V[1,1] <- 0.1  # non-informative
                  V[2,2] <- 0.1
                  V[1,2] <- 0
                  V[2,1] <- 0
                  invSigma.b[1 : 2 , 1 : 2]  ~ dwish(V[ , ],2)
                  Sigma.b[1 : 2 , 1 : 2]  <- inverse(invSigma.b[ , ])
            # prior distribution for mu (offset parameter)
                  # circular uniform distribution
                     dummy <- 0
                  dummy ~ dloglik(logLike)
                  logLike <- -log(2*pi)
                  mu ~ dunif(-3.14159265359,3.14159265359)
                  for (j in 1:(N/2)) {
                     dif_1[j] <- abs(theta_comp[j + (N/2)] - theta_comp[j])
                     dif_2[j] <- 6.28 - abs(theta_comp[j +(N/2)] - theta_comp[j])
                     APE[j] <- min(dif_1[j],dif_2[j])
                     CPE[j] <- 1 - cos(theta_comp[j + (N/2)] - theta_comp[j])
                        }
         CPD_1 <- sum(APE[])
         CPD_2 <- sum(CPE[])
               }
```

Fig. 1 OpenBUGS code for LCREM with $P = 1$, $Q = 2$

References

Borgioli C, Marchetti GM, Scapini F (1999a) Variation in zonal recovery in four *Talitrus saltator* populatıons from dıfferent coastlınes: a comparison of orientations in the field and in an experimental arena. Behav Ecol Sociobiol 45:7–85

Borgioli C, Martelli M, Porri F et al (1999b) Orientation in *Talitrus saltator* (montagu): trends in intrapopulations variability related to environmental and intrinsic factors. J Exp Mar Biol Ecol 238:29–47

D'Elia A (2001) A statistical model for orientation mechanism. Stat Methods Appl 10:157–174

Hall DB, Shen J (2015) Marginal projected multivariate linear models for clustered angular data. Aust N Z J Stat 57(2):241–257

Hill T, Chocholek M (2016) Coastal biodiversity and ecosystem service sustainability (CBESS) eddy covariance flux data for Abbotts Hall. NERC Environmental Information Data Centre, Bailrigg. https://doi.org/10.5285/8cfd9a2a-8b68-40c6-94a1-be8e02e869c1

Jammalamadaka RA, SenGupta A (2001) Topics in circular statistics. World Scientific Inc., New York

Lagona F (2016) Regression analysis of correlated circular data based on the multivariate von Mises distribution. Environ Ecol Stat 23(1):89–113

Maity A, Sherman M (2008) On adaptive linear regression. J Appl Stat 35(12):1409–1422

Maruotti A (2016) Analyzing longitudinal circular data by projected normal models: a semi-parametric approach based on finite mixture models. Stoch Environ Res Risk Assess 23:257–277

Maruotti A, Punzo A, Mastrantonio G et al (2016) A time-dependent extension of the projected normal regression model for longitudinal circular data based on a hidden Markov heterogeneity structure. Stoch Environ Res Risk Assess 30:1725–1740

Mastrantonio G, Lasinio GJ, Gelfand AE (2016) Spatio-temporal circular models with non-separable covariance structure. Test 25(2):331–350

McMillan GP, Hanson TE, Saunders G et al (2013) A two-component circular regression model for repeated measures auditory localization data. J R Stat Soc Ser C Appl Stat 62(4):515–534

Nunez-Antonio G, Gutierrez-Pena E (2014) A Bayesian model for longitudinal circular data based on the projected normal distribution. Comput Stat Data Anal 71:506–519

Pewsey A (2002) Testing circular symmetry. Can J Stat 30:591–600

Ravindran PK, Ghosh SK (2011) Bayesian analysis of circular data using wrapped distributions. J Stat Theory Pract 5:547–561

Rivest LP, Kato S (2019) A random-effects model for clustered circular data. Can J Stat. https://doi.org/10.1002/cjs.11520

Rossi PE, Allenby GM, McCulloch R (2005) Bayesian statistics and marketing. Wiley, Chichester

Scapini F (1997) Variation in scototaxis and orientation adaptation of *Talitrus saltator* populations subjected to different ecological constraints. Estuar Coast Shelf Sci 44:139–146

Song XKP (2007) Correlated data analysis: modeling analytics, and applications. Springer, Berlin

Onur Camli is currently a doctoral student in the Department of Statistics at the Middle East Technical University, Ankara, Turkey. He holds a M.Sc. degree in Statistics from METU. His research interests are related with Bayesian statistics, directional statistics, model selection and computational statistics.

Zeynep Kalaylioglu is currently Associate Professor in the Department of Statistics at the Middle East Technical University, Ankara, Turkey. Previously, she worked as a biostatistics scientist at an informatic company in the USA collaborating with NIH/DCEG researchers in developing methodologies for the analysis of genetic associations, gene–environment interactions, etiological heterogeneity by cancer subtypes and risk prediction for important cancer types. She holds a Ph.D. degree in Statistics from North Carolina State University. Her research focuses on developing Bayesian methodologies for environmental health and biological data, in particular model selection, directional data analysis, and nonrandom missingness.