Environ Ecol Stat (2016) 23:89–113
DOI 10.1007/s10651-015-0330-y

Regression analysis of correlated circular data based on the multivariate von Mises distribution

Francesco Lagona[1]

Received: 31 October 2014 / Revised: 12 October 2015 / Published online: 31 October 2015

Abstract A regression model for correlated circular data is proposed by assuming that samples of angular measurements are drawn from a multivariate von Mises distribution with mean and concentration parameters that depend on covariates through link functions. The model can flexibly accommodate heteroscedasticity, unstructured correlation, and specific autoregressive correlation structures. Because the computation of the normalizing constant of the multivariate von Mises distribution is unfeasible, inference is based on a computationally tractable Monte Carlo approximation of the log-likelihood. These methods are illustrated by fitting a number of regression models in two case studies: a longitudinal study of animal orientation, involving multiple time series of directional observations, and a study of marine currents, involving a spatial series of sea current directions.

Keywords Animal orientation · Circular data · Monte Carlo maximum likelihood · Multivariate von Mises · Pseudo-likleihood · Sea currents

1 Introduction

Correlated circular data arise when multiple angular measurements are taken on the same sample unit and are encountered across several research areas. Recent examples in the literature include bioinformatical studies of the dihedral angles that summarizes the structure of a protein (Mardia et al. 2007; Rueda et al. 2009), studies of radiation

Handling Editor: Bryan F. J. Manly.

✉ Francesco Lagona
francesco.lagona@uniroma3.it

[1] University Roma Tre, Via G. chiabrera 199, 00145 Rome, Italy

therapy that involve the positions of multiple X-ray beams on a circle (Abraham et al. 2013), environmental studies of wind and wave directions (Lagona and Picone 2012, 2013) and pattern recognition studies of online handwriting (Bahlmann 2006). Special cases of correlated circular data arise when one angular variable is repeatedly observed over time or across space. Examples include the analysis of time series of wind directions (Fisher and Lee 1994), longitudinal studies of animal orientation (Artes and Jørgensen 2000; D'Elia 2001; Nunez-Antonio and Gutiérrez-Peña 2012) and environmental studies of spatial series of wind or wave directions (Modlin et al. 2012; Jona-Lasinio et al. 2012; Bulla et al. 2012).

Regression models for correlated circular data are needed when the interest is on estimating the influence of covariates on circular responses, simultaneously accounting for the data correlation structure (Arnold and SenGupta 2006). The specification of a regression model in this context is however challenging because most of the circular distribution theory has been so far restricted to one-dimensional densities on the circle and two-dimensional densities on the torus (see Pewsey et al. 2013, Chapter 4, for a recent review). The difficulties of working with multivariate circular models stem primarily from the disparate topologies of the multi-dimensional torus and the Euclidean space; the former being concerned with multivariate circular data and the latter associated with multivariate linear data. As a result, most of the univariate circular densities proposed in the literature have not been extended to the multivariate setting. Noticeable exceptions include the multivariate wrapped normal distribution (Fisher 1993) and the multivariate von Mises (MVM) distribution (Mardia et al. 2008).

This paper introduces a multivariate circular regression model by assuming that vectors of multiple angular measurements are samples drawn from a MVM distribution. The MVM distribution is a natural extension of the univariate von Mises distribution to a multivariate setting and can be viewed as the circular counterpart of a multivariate Gaussian distribution. It is fully specified by a vector of mean direction parameters, a vector of concentration parameters and a symmetric matrix of conditional dependence parameters that capture the direct link between two variables when the other variables are conditioned on. The MVM regression model is specified by assuming that both the mean directions and the concentrations depend on covariates through link functions and allowing the matrix of the conditional dependence parameters to be fully unstructured. It can be viewed as a multivariate extension of the generalized linear model developed by Fisher and Lee (1992) for independent angular measurements.

Maximum likelihood estimation of this model requires special care (Presnell et al. 1998). The normalizing constant of the MVM distribution is numerically intractable and, as a result, inference must rely on a suitable approximation of the likelihood function. Following Geyer (1992), this paper focuses on a Monte Carlo approximation of the likelihood function. It is obtained by sampling from a MVM distribution, evaluated at a preliminary pseudo-likelihood estimate of the parameters. Pseudo-likelihood estimates are obtained by maximizing a computationally tractable pseudo-likelihood function, which is however typically multimodal. This requires a multi-start strategy where the optimization routine is initialized at several starting points of the parameter space.

The rest of the paper is organized as follows. The need of regression models for correlated circular data is motivated by two examples, namely a longitudinal study

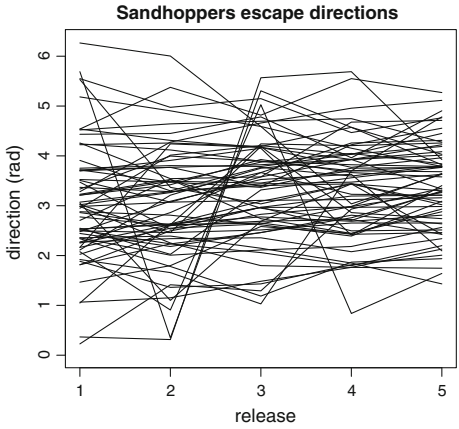

Fig. 1 Escape directions for sandhoppers over five consecutive releases

of animal orientation and an environmental study that involves a spatial series of sea currents directions (Sect. 2). Then, Sect. 3 introduces the proposed regression models, focusing on the properties that facilitate Monte Carlo maximum likelihood. The proposed strategy for parameter estimation is discussed in Sect. 4. Section 5 describes the application of these methods to the two motivating examples of Sect. 2. Major discussion points are finally summarized in Sect. 6.

2 Motivating examples

2.1 Animal orientation

Song (2007) reports a longitudinal study whose aim is to help understanding the escaping mechanism of sandhoppers (*Talitrus saltator*). The data concern 65 sandhoppers which were released sequentially on five occasions, observing escape directions along with a number of time-constant covariates. The covariates include wind direction, azimuth direction for the sun (Sun), and a eye asymmetry index (Eye). Wind directions are split into four categories: wind from land, wind from sea, wind from longshore-west and wind from longshore-east, with wind from the land taken as the reference category. Figure 1 displays the 65 short time series of the observed escape directions. The interest in this study is in estimating the influence of the covariates on the response (escape direction), simultaneously accounting for the dependence between observations taken on the same sandhopper. These data were previously examined by D'Elia (2001) and Song (2007), who estimated a von Mises mixed-effects regression model, assuming that escape directions are conditionally independent given a sandhopper-specific random intercept. Under a Bayesian setting, Nunez-Antonio and Gutiérrez-Peña (2012) re-examined these data by exploiting a bivariate mixed-effects model, based on the projected bivariate normal distribution, and assuming escape directions as conditionally independent given a subject-specific random intercept.

Mixed-effects models account for longitudinal dependence by introducing random effects and assuming that responses are conditionally independent given the random

effects. In the case of normal longitudinal data, more general dependence structures can be examined by removing the conditional independence assumption and assuming that each longitudinal trajectory is drawn from a multivariate normal distribution with an unstructured variance-covariance matrix (Diggle et al. 2002). In the sandhoppers orientation study, where we observe trajectories of escape directions at five occasions, the dependence structure of multiple releases can be investigated by assuming that each trajectory is a sample from a 5-dimensions MVM distribution and comparing unstructured and structured matrices of dependence parameters.

2.2 Marine currents

Correlated circular data may arise not only in the form of time series, but also in the form of spatial series. The second motivating example considered in this paper is a case study of spatial angular measurements. Figure 2 displays the directions of surface sea currents obtained by high-frequency (HF) radars installed in the eastern part of the northern Adriatic. These data are part of the NASCUM (North Adriatic Surface Current Mapping) project (Mihanovic et al. 2011; Cosoli et al. 2012). The figure includes the observations taken on May 4th at 11 a.m., across a grid of 245 points having a horizontal resolution of 2 km $\times$ 2 km.

These data are normally used to estimate spatial gradients that summarize the large-scale variation of the data, along with sources of small-scale variation due to residual spatial auto-correlation (Lagona et al. 2015). A statistical approach to estimate variation at multiple spatial scales plays a crucial role in studies that involve semi-enclosed basins, such as the Adriatic Sea, where currents are influenced by the local orographic conditions and numerical models, well suited for oceans, may provide inaccurate results (Bertotti and Cavalieri 2009). Orography effects clearly appear in Fig. 2, where the outcome variability increases as the grid points approach the coast.

The estimation of a spatial gradient in this context requires models that simultaneously include both heteroscedasticy and spatial dependence, in addition to a spatial trend. In the case of normal spatial series, these models are specified by assuming

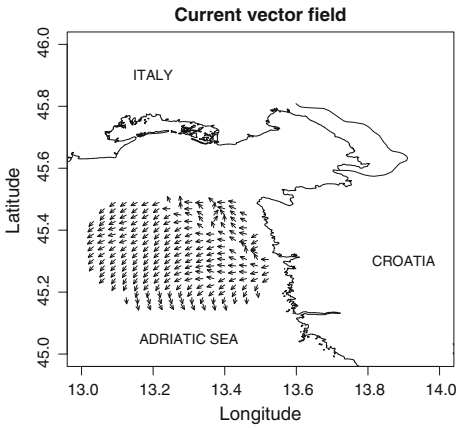

Fig. 2 Sea current directions in the Northen Adriatic sea on May 4th, 11 a.m., measured across a grid of 245 point. The grid is located at the head of the Adriatic Sea, between the Gulf of Trieste and the Bay of Kvarner. The coastlines in the right side of the picture are the Croatian coasts of the Istrian peninsula. The Lagoon of Grado-Marano appears at the top of the picture

that the spatial series is a single sample drawn from a multivariate normal distribution. Because the dimension of the normal distribution is equal to the sample size, the variance-covariance matrix needs to be structured, i.e. known up to a small number of unknown parameters. In the case of normal spatial data, the literature is rich of different proposals to include heteroscedasticity and spatial autocorrelation in a variance-covariance matrix (Gaetan and Guyon 2010). Assuming that a spatial series of circular data is drawn from a multivariate von Mises distribution allows us to mimic these proposals in a circular context.

3 A multivariate von Mises regression model

A generalized linear model (GLM) for independent angular responses can be specified by assuming that a vector of angles $\boldsymbol{y} = (y_1, \ldots, y_j, \ldots, y_J)$, $y_j \in [-\pi, \pi)$, is drawn from J independent von Mises distributions with location parameters $\mu_1, \ldots, \mu_J$ and concentrations $\kappa_1, \ldots, \kappa_J$. The influence of the covariates on these parameters can be conveniently studied by specifying link functions that map linear predictors to the appropriate parameter space. Following Fisher and Lee (1992), the location parameters can be for example modeled as

$$\mu_j(\boldsymbol{\beta}) = \beta_0 + g(\boldsymbol{x}_j^{\mathsf{T}}\boldsymbol{\beta}), \tag{1}$$

where $\beta_0 \in [-\pi, \pi)$ is an offset parameter, $\boldsymbol{x}_j^{\mathsf{T}}$ is a row-vector of K covariates, associated with observation y_j, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)$ is a vector of regression coefficients and, finally, g is a known monotone function that maps the real line to the circle $[-\pi, \pi)$. In the case of a single continuous covariate x, g is a curve on the surface of an infinite cylinder $[-\pi, \pi) \times (-\infty, \infty)$, which rotates once around the cylinder as x varies from $-\infty$ to $+\infty$. For example, a negative value of a single regression coefficient β_k indicates that a clock-wise rotation of the response is associated with an increase of the k-th covariate. If the link function g is such that $g(0) = 0$, then the offset β_0 can be conveniently interpreted as an origin.

Similarly, the concentration parameters can be modeled as follows

$$\kappa_j(\boldsymbol{\gamma}) = h(\boldsymbol{w}_j^T\boldsymbol{\gamma}), \tag{2}$$

where $\boldsymbol{w}_j^T$ is a row-vector of H covariates, not necessarily equal to $\boldsymbol{x}_j^{\mathsf{T}}$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_H)$ is a vector of parameters and h is a known link function that maps the real line to $(0, +\infty)$.

Under this setting, the data joint distribution is given by the product of J von Mises densities, i.e.,

$$f(\boldsymbol{y}; \boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{j=1}^{J} f_{\mathrm{vm}}(y_j; \mu_j(\boldsymbol{\beta}), \kappa_j(\boldsymbol{\gamma})), \tag{3}$$

where

$$f_{\text{vm}}(y;\mu,\kappa)=\frac{\exp(\kappa\cos(y-\mu))}{2\pi I_0(\kappa)}$$

is the von Mises density of mean μ and concentration κ, and

$$I_m(\kappa)=\frac{1}{\pi}\int_0^{\pi} e^{\kappa\cos t}\cos(mt)dt$$

is the modified Bessel function of order m.

The independence assumption can be relaxed by assuming that the response vector $\boldsymbol{y}$ is a sample drawn from a multivariate von Mises distribution. It is specified by a multivariate density, known up to the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ and a $J\times J$ symmetric association matrix Λ with zero diagonal entries ($\lambda_{jj}=0,\lambda_{jh}=\lambda_{hj},j,h=1,\ldots,J$), namely

$$f_{\text{MVM}}(\boldsymbol{y};\boldsymbol{\beta},\boldsymbol{\gamma},\Lambda)=\frac{\exp\left(\boldsymbol{\kappa}(\boldsymbol{\gamma})^{\mathsf{T}}\boldsymbol{c}(\boldsymbol{y}-\boldsymbol{\mu}(\boldsymbol{\beta}))+\frac{1}{2}\boldsymbol{s}(\boldsymbol{y}-\boldsymbol{\mu}(\boldsymbol{\beta}))^{\mathsf{T}}\Lambda\boldsymbol{s}(\boldsymbol{y}-\boldsymbol{\mu}(\boldsymbol{\beta}))\right)}{C(\boldsymbol{\gamma},\Lambda)} \tag{4}$$

where $C(\boldsymbol{\gamma},\Lambda)$ is the normalizing constant, and

$$\begin{aligned}\boldsymbol{\mu}(\boldsymbol{\beta})&=(\mu_1(\boldsymbol{\beta}),\ldots,\mu_J(\boldsymbol{\beta}))\\ \boldsymbol{\kappa}(\boldsymbol{\gamma})&=(\kappa_1(\boldsymbol{\gamma}),\ldots,\kappa_J(\boldsymbol{\gamma}))\\ \boldsymbol{c}(\boldsymbol{y}-\boldsymbol{\mu}(\boldsymbol{\beta}))&=(\cos(y_1-\mu_1(\boldsymbol{\beta})),\ldots,\cos(y_J-\mu_J(\boldsymbol{\beta})))\\ \boldsymbol{s}(\boldsymbol{y}-\boldsymbol{\mu}(\boldsymbol{\beta}))&=(\sin(y_1-\mu_1(\boldsymbol{\beta})),\ldots,\sin(y_J-\mu_J(\boldsymbol{\beta}))).\end{aligned}$$

Density (4) extends the MVM density proposed by Mardia et al. (2008) to the GLM setting, by assuming that both the location and the concentration parameters depend on covariates.

Most of the MVM properties have been discussed recently (Mardia et al. 2008; Mardia and Voss 2014). Among these properties, of interest in this paper is the convenient von Mises form taken by the univariate conditional distributions. Formally, under (4), the conditional distribution of each angle y_j given the rest of the sample $\boldsymbol{y}^{(j)}=(y_1,\ldots,y_{j-1},y_{j+1},\ldots,y_J)$ is von Mises. Precisely,

$$\begin{aligned}f(y_j|\boldsymbol{y}^{(j)};\boldsymbol{\beta},\boldsymbol{\gamma},\Lambda)&=f_{\text{VM}}\left(y_j;\nu_j(\boldsymbol{y}^{(j)};\boldsymbol{\beta},\boldsymbol{\gamma},\Lambda),\frac{1}{\sigma_j^2(\boldsymbol{y}^{(j)};\boldsymbol{\beta},\boldsymbol{\gamma},\Lambda)}\right)\\ &=\frac{\exp\left(\frac{1}{\sigma_j^2(\boldsymbol{y}^{(j)};\boldsymbol{\beta},\boldsymbol{\gamma},\Lambda)}\cos(y_j-\nu_j(\boldsymbol{y}^{(j)};\boldsymbol{\beta},\boldsymbol{\gamma},\Lambda))\right)}{2\pi I_0\left(\frac{1}{\sigma_j^2(\boldsymbol{y}^{(j)};\boldsymbol{\beta},\boldsymbol{\gamma},\Lambda)}\right)},\end{aligned} \tag{5}$$

where

$$\nu_j(\boldsymbol{y}^{(j)};\boldsymbol{\beta},\boldsymbol{\gamma},\Lambda)=\mu_j(\boldsymbol{\beta})+\tan^{-1}\frac{\sum_{h\neq j}\lambda_{jh}\sin(y_h-\mu_h(\boldsymbol{\beta}))}{\kappa_j(\boldsymbol{\gamma})}$$

is the conditional mean parameter and, finally,

Table 1 Examples of dependence structures for longitudinal circular data

Model	Concentration parameters	Dependence parameters	Number of parameters
Unstructured	κ_j	λ_{jk}	$J + \frac{J(J-1)}{2}$
	$j = 1, \ldots, J$	$j = 1, \ldots, J-1, k = j+1, \ldots, J$	
Heteroscedastic AR(1)	$\kappa_j = h(\boldsymbol{w}_j^\top \boldsymbol{\gamma})$	$\lambda_{jk} = \begin{cases} \lambda & \text{if } \mid j-k \mid = 1 \\ 0 & \text{otherwise} \end{cases}$	$H + 1$
	$\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_H)$		
Homoscedastic AR(1)	$\kappa_j = \kappa$	$\lambda_{jk} = \begin{cases} \lambda & \text{if } \mid j-k \mid = 1 \\ 0 & \text{otherwise} \end{cases}$	$1 + 1$

$$\frac{1}{\sigma_j^2(\boldsymbol{y}^{(j)}); \boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda)} = \sqrt{\kappa_j^2(\boldsymbol{\gamma}) + \left(\sum_{h \neq j} \lambda_{jh} \sin(y_h - \mu_h(\boldsymbol{\beta})) \right)^2}$$

is the conditional concentration parameter. If all the entries of Λ are equal to zero, the density (4) reduces to the product (3) of J univariate von Mises densities and the proposed model reduces to the Fisher-Lee regression model. Otherwise, the univariate marginal distributions are in general not von Mises with a normalizing constant that can be accurately approximated only when $J = 2$.

The MVM density (4) reflects a model with covariate-specific location parameters and concentrations, and an unstructured association matrix Λ. An unstructured association matrix can be of interest for exploratory purposes, when little is a priori known on the association structure of the responses. Other cases may suggest a specific structure of matrix Λ. Under (4), the conditional distribution of each angular measurement y_j given the rest of the sample depends on the observations only through the entries of the jth row of Λ. In other words, $\lambda_{jk} = 0$ implies that the responses y_j and y_k are conditionally independent given the rest of the sample. A variety of auto-regressive structures can be therefore specified by suitably modeling the entries of Λ.

Table 1, for example, list three different models can can be specified in the case of longitudinal circular data with covariates.

The first model features an unstructured association matrix Λ and unstructured concentrations. It can be simplified by modeling the concentration parameters through a suitable design matrix and equating the dependence parameters of non-adjacent observations to zero. These constraints lead to the second model of Table 1, namely a heteroscedastic first-order autoregressive [AR(1)] model. It can be further simplified by assuming a single concentration parameter, leading to the third model of Table 1, namely a homoscedastic AR(1) model.

Table 2 displays three examples of dependence structures under a spatial setting. In this case, matrix Λ can be modeled by specifying a spatial neighborhood structure on the set of the observation sites $1, \ldots, n$. A spatial neighborhood structure is defined by associating each site i with a set of sites $N(i) \subset \{1, \ldots, n\}$. A first-order autoregressive (AR(1)) spatial structure can be defined by assuming that

Table 2 Examples of dependence structures for spatial circular data

Model	Concentration parameters	Dependence parameters	Number of parameters
Heteroscedastic spatial AR(1)	$\kappa_j = h(\boldsymbol{w}_j^\top \boldsymbol{\gamma})$	$\lambda_{jk} = \begin{cases} \lambda & \text{if } j \in N(i) \\ 0 & \text{otherwise} \end{cases}$	$H+1$
	$\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_H)$		
Homoscedastic spatial AR(1)	$\kappa_j = \kappa$	$\lambda_{jk} = \begin{cases} \lambda & \text{if } j \in N(i) \\ 0 & \text{otherwise} \end{cases}$	$1+1$
Heteoscedastic spatial AR(2)	$\kappa_j = h(\boldsymbol{w}_j^\top \boldsymbol{\gamma})$	$\lambda_{jk} = \begin{cases} \lambda_1 & \text{if } j \in N_1(i) \\ \lambda_2 & \text{if } j \in N_2(i) \\ 0 & \text{otherwise} \end{cases}$	$H+1$
	$\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_H)$		

$$\lambda_{jk} = \begin{cases} \lambda & \text{if } k \in N(j) \\ 0 & \text{otherwise} \end{cases}.$$

The spatial structure above can be associated with either heteroscedastic or homoscedastic concentrations, leading to the first two models displayed in Table 2. Higher-order spatial auto-regressive models can be specified by defining non-overlapping neighborhood structures (Lagona 2002). For example, a spatial AR(2) model can be specified by associating each site i with two neighborhoods, say $N_1(i)$ and $N_2(i)$, $N_1(i) \cap N_2(i) = \emptyset$, and associating each neighborhood with a specific parameter, namely

$$\lambda_{jk} = \begin{cases} \lambda_1 & \text{if } j \in N_1(i) \\ \lambda_2 & \text{if } j \in N_2(i) \\ 0 & \text{otherwise.} \end{cases}$$

Such constraints lead to the third model of Table 2, namely a spatial AR(2) with heteroscedastic concentrations.

4 Markov chain Monte Carlo maximum likelihood

Let $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_i, \ldots, \boldsymbol{y}_n$ be a sample of n vectors $\boldsymbol{y}_i = (y_{i1}, \ldots y_{ij}, \ldots, y_{iJ})$, independently drawn from the MVM distribution (4). The maximum likelihood estimates of the parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda)$, can in principle be obtained by maximizing the log-likelihood function

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f_{\mathrm{MVM}}(\boldsymbol{y}_i; \boldsymbol{\theta}). \tag{6}$$

Maximum likelihood estimation is however complicated by the intractability of the normalizing constant of the MVM density. More precisely, in the bivariate case, the normalizing constant of this density can be numerically approximated by a sum of Bessel functions (Singh et al. 2002). As a result, traditional Bayesian and maximum-likelihood methods can be exploited to estimate the parameters (Singh et al. 2002; Bhattacharya and SenGupta 2009; Mardia 2010). When the dimension is larger than two, however, the analytical form of the normalizing constant is not known and numerical integration is not feasible. In this case, pseudo-likelihood or composite-likelihood methods have been suggested (Mardia et al. 2008). These procedures avoid the computation of intractable normalizing constants at the price of estimates that are in general less efficient than the maximum likelihood estimates (Geyer 1992; Varin et al. 2011). In the particular case of the MVM distribution, the quality of pseudo-likelihood estimates has been investigated by Mardia et al. (2009).

The estimation strategy proposed in this paper is based on a Monte Carlo (MC) approximation of the log-likelihood. MC maximum likelihood estimation (Geyer 1992) is based on the maximization of a stochastic approximation of the log-likelihood, obtained by drawing a sample of M J-dimensional vectors $\boldsymbol{y}_1^\star, \ldots, \boldsymbol{y}_M^\star$ from the model under scrutiny, evaluated at a preliminary estimate $\tilde{\boldsymbol{\theta}}$. This approach is attractive when (1) a computationally efficient procedure is available to obtain pseudo-likelihood estimates $\tilde{\boldsymbol{\theta}}$, (2) samples from the assumed model can be easily drawn without knowledge of the normalizing constant and (3) the stochastic approximation of the log-likelihood takes a simple form that makes maximization straightforward. The MVM density fulfills these three requirements. Under a MVM, the univariate conditional distribution of each component given the others is von Mises. The Besag's (1975) pseudo-likelihood function, defined as a product of univariate conditional densities, is therefore available in closed form. Maximization of such function yields therefore maximum pseudo-likelihood estimates $\tilde{\boldsymbol{\theta}}$ with little computational effort, even in the presence of covariates that modulate location parameters and concentrations. Second, a simple Gibbs sampling scheme can be implemented to obtain samples from the MVM, relying on the efficient algorithms that are available for sampling from the univariate von Mises distribution. Third, the centered MVM is a canonical exponential family distribution and, as a result, conventional Newton-type procedures can be exploited to maximize the Monte Carlo approximation of the log-likelihood. These points are discussed in the following sections.

4.1 Simulating from a multivariate von Mises distribution

In the bivariate case ($J = 2$), the MVM normalizing constant can be numerically approximated and the univariate marginal distributions can be efficiently computed. As a result, bivariate samples can be obtained by a simple two-step procedure (Mardia et al. 2007). First, a conventional rejection sampling procedure is exploited to draw a sample from a univariate marginal distribution, by taking a univariate von Mises

density as proposal. Second, a sample is drawn from the conditional von Mises density (5), exploiting one of the available routines for simulating a univariate von Mises distribution.

When the dimension J is larger than 2, this procedure is not longer practical. Alternative simulation strategies are therefore necessary. One approach relies on drawing multivariate samples via rejection sampling, exploiting the product of von Mises univariate densities as proposal distribution. This method is restricted to small or moderate J, and works well for the case of high concentrations (Mardia and Voss 2014). An alternative method, exploited in this paper, relies on Gibbs sampling (GS). The GS algorithm is initializated at an arbitrary configuration, say $\boldsymbol{y}_0$, and a sequence of vectors $(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_k, \ldots)$ is generated, according to a sweeping strategy. A sweeping strategy is an infinite sequence of indexes $(i_1, i_2, \ldots, i_k, \ldots)$, $i_k \in \{1, 2, \ldots, J\}$, so that at the k-th iteration, $\boldsymbol{y}_{k-1}$ is updated to a new vector $\boldsymbol{y}_k$ by sampling a new coordinate y_{jk}, drawn from the conditional von Mises distribution $f(y_j|\boldsymbol{y}_{k-1}^{(j)})$. If every index in $\{1, 2, \ldots, J\}$ is visited infinitely often, then process $(\boldsymbol{y}_k, k > 0)$ is Markov and it converges weakly to the MVM distribution $f(\boldsymbol{y})$, as $k \to \infty$ (Geman and Geman 1984).

In this paper, a centered MVM distribution with concentrations $\kappa_1, \ldots, \kappa_J$ and association matrix Λ is simulated by initializing the algorithm at a configuration $\boldsymbol{y}_0$, obtained by taking a sample of J independent values from J univariate von Mises distributions with means 0 and concentrations $\kappa_1, \ldots, \kappa_J$. A sequence of vectors $\boldsymbol{y}_k$ is then generated by randomly selecting an updating scheme $(i_1, \ldots, i_J)$ for each cycle k.

4.2 Maximum pseudo-likelihood estimation

A convenient pseudo-likelihood function was defined by Besag (1975) as the product of univariate conditional densities. According to this proposal, the maximum pseudo-likelihood estimates (MPLE) $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}, \tilde{\Lambda})$ can be obtained by maximizing the following product of conditional von Mises densities

$$\mathrm{PL}(\boldsymbol{\theta}) = \prod_{i=1}^{n}\prod_{j=1}^{J} f_{\mathrm{VM}}\left(y_{ij};\, \nu_j(\boldsymbol{y}_i^{(j)};\boldsymbol{\theta}),\, \frac{1}{\sigma_i^2(\boldsymbol{y}_i^{(j)};\boldsymbol{\theta})}\right). \tag{7}$$

Ignoring the constant terms 2π, the log-pseudo-likelihood function takes the following form

$$\begin{aligned}\mathrm{pl}(\boldsymbol{\theta}) &= \log \mathrm{PL}(\boldsymbol{\theta}) \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{J} \log I_0\left(\frac{1}{\sigma_j^2(\boldsymbol{y}_i^{(j)};\boldsymbol{\theta})}\right) + \sum_{i=1}^{n}\sum_{j=1}^{J} \frac{\cos\left(y_{ij} - \nu_j(\boldsymbol{y}_i^{(j)};\boldsymbol{\theta})\right)}{\sigma_j^2(\boldsymbol{y}_i^{(j)};\boldsymbol{\theta})}.\end{aligned} \tag{8}$$

When all the entries of Λ are equal to zero, $\tilde{\boldsymbol{\theta}}$ is a maximum likelihood estimate. Otherwise, the maximum pseudo-likelihood estimates can be exploited as starting points of an algorithm that converges to the maximum likelihood estimate.

The pseudo-likelihood function (8) can be maximized by traditional Newton-type algorithms, such as those available in the R command `optim`. However, (8) is in general not globally concave and may include a number of local maxima. Care should be hence taken in the selection of the starting point that initializes the algorithm. Following Fisher and Lee (1992), the specific starting point exploited in this paper was specified by equating a number of parameters with zero, namely $\gamma_h^{(0)} = 0, h > 1$, $\lambda_{ji}^{(0)} = 0, \quad j, i = 1, \ldots, J$ and initializing the regression coefficient $\boldsymbol{\beta}$, the offset β_0 and the concentration parameter γ_0 as follows. Given the two functions

$$S(\boldsymbol{\beta}) = \frac{1}{nJ} \sum_{i=1}^{n} \sum_{j=1}^{J} \sin(y_{ij} - g(\boldsymbol{x}_{ij}^{\mathsf{T}} \boldsymbol{\beta}))$$

$$C(\boldsymbol{\beta}) = \frac{1}{nJ} \sum_{i=1}^{n} \sum_{j=1}^{J} \cos(y_{ij} - g(\boldsymbol{x}_{ij}^{\mathsf{T}} \boldsymbol{\beta})),$$

we define the function

$$R(\boldsymbol{\beta}) = \left(S^2(\boldsymbol{\beta}) + C^2(\boldsymbol{\beta}) \right)^{1/2}.$$

First, maximization of $R(\boldsymbol{\beta})$ gives a starting point $\boldsymbol{\beta}^{(0)}$ for the the regression $\boldsymbol{\beta}$; then the offset β_0 can be initialized at

$$\beta_0^{(0)} = \begin{cases} \tan^{-1} \frac{S(\boldsymbol{\beta}^{(0)})}{C(\boldsymbol{\beta}^{(0)})} & \text{if } C(\boldsymbol{\beta}^{(0)}) \geq 0 \\ \pi + \tan^{-1} \frac{S(\boldsymbol{\beta}^{(0)})}{C(\boldsymbol{\beta}^{(0)})} & \text{if } C(\boldsymbol{\beta}^{(0)}) < 0, \end{cases}.$$

Finally, γ_0 can be initialized at the value

$$\gamma_0 = A_1^{-1}(R(\boldsymbol{\beta}^{(0)})),$$

where $A_1(x) = \frac{I_1(x)}{I_0(x)}$ is the ratio of the first and zeroth order Bessel functions of the first kind, and $A_1^{-1}(x)$, can be numerically approximated, for example using the `A1inv(x)` function of the R library `circular`. Although this strategy worked well with the applications of this paper, it does not guarantee the convergence to a global maximum. Care should be taken when local maxima are close to each other and, in these cases, the pseudo-likelihood surface must be carefully explored before choosing suitable starting points.

4.3 Markov chain Monte Carlo approximation of the log-likelihood

With a slight abuse of notation, let vech$(\cdot)$ denote the matrix operator that arranges the $J(J-1)/2$ supradiagonal entries of a $J \times J$ symmetric matrix into a column vector. The centered MVM can be conveniently expressed as

$$\begin{aligned} f_{MVM}(y; \boldsymbol{\gamma}, \Lambda) &= \frac{\exp\left(\boldsymbol{t}_1(\boldsymbol{y})^\top \boldsymbol{\kappa}(\boldsymbol{\gamma}) + \boldsymbol{t}_2(\boldsymbol{y})^\top \text{vech}(\Lambda)\right)}{C(\boldsymbol{\gamma}, \boldsymbol{\Lambda})} \\ &= \frac{\varphi(\boldsymbol{y}; \boldsymbol{\gamma}, \Lambda)}{C(\boldsymbol{\gamma}, \Lambda)}, \end{aligned} \tag{9}$$

where

$$\begin{aligned} \boldsymbol{t}_1(\boldsymbol{y}) &= \boldsymbol{c}(\boldsymbol{y}) \\ \boldsymbol{t}_2(\boldsymbol{y}) &= \text{vech}\left(\boldsymbol{s}(\boldsymbol{y})\boldsymbol{s}(\boldsymbol{y})^\top\right) \end{aligned}$$

are two sufficient statistics, respectively associated with the parameters $\boldsymbol{\kappa}(\boldsymbol{\gamma})$ and $\text{vech}(\Lambda) = (\lambda_{12}, \lambda_{13}, \ldots, \lambda_{1J}, \lambda_{23}, \lambda_{24}, \ldots, \lambda_{2J}, \ldots, \lambda_{J-1,J})$.

Let $\boldsymbol{y}^\star_m$, $m = 1, \ldots, M$ be a sample of M vectors, drawn from the centered MVM distribution (9), evaluated at the preliminary estimates (e.g., the MPLE) $\tilde{\boldsymbol{\gamma}}$ and $\tilde{\Lambda}$, by using one of the two methods discussed in Sect. 4.1. On the basis of this sample, the ratio $C(\boldsymbol{\kappa}, \Lambda)/C(\tilde{\boldsymbol{\kappa}}, \tilde{\Lambda})$ can be approximated by a Monte Carlo mean, as follows

$$\begin{aligned} \frac{C(\boldsymbol{\kappa}, \Lambda)}{C(\tilde{\boldsymbol{\kappa}}, \tilde{\Lambda})} &= \int_{(-\pi,\pi)^J} \frac{\varphi(\boldsymbol{y}; \boldsymbol{\kappa}, \Lambda)}{\varphi(\boldsymbol{y}; \tilde{\boldsymbol{\kappa}}, \tilde{\Lambda})} f_{MVM}(\boldsymbol{y}; \tilde{\boldsymbol{\kappa}}, \tilde{\lambda}) d\boldsymbol{y} = \mathbb{E}\left(\frac{\varphi(\boldsymbol{y}; \boldsymbol{\kappa}, \Lambda)}{\varphi(\boldsymbol{y}; \tilde{\boldsymbol{\kappa}}, \tilde{\Lambda})}\right) \\ &\approx \frac{1}{M} \sum_{m=1}^{M} \exp\left(\boldsymbol{t}_1^\top(\boldsymbol{y}^\star_m)(\boldsymbol{\kappa}(\boldsymbol{\gamma}) - \boldsymbol{\kappa}(\tilde{\gamma})) + \boldsymbol{t}_2^\top(\boldsymbol{y}^\star_m)(\text{vech}(\Lambda) - \text{vech}(\tilde{\Lambda}))\right), \end{aligned} \tag{10}$$

where the expected value above is with respect to the centered distribution (9) and the approximation holds due the MC-likelihood approximation theory (Geyer 1992). The stochastic approximation above allows approximating the log-likelihood function

$$l(\boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda) = \sum_{i=1}^{n} \boldsymbol{t}_1^\top(\boldsymbol{y}_i - \boldsymbol{\mu}(\boldsymbol{\beta}))\boldsymbol{\kappa}(\boldsymbol{\gamma}) + \boldsymbol{t}_2^\top(\boldsymbol{y}_i - \boldsymbol{\mu}(\boldsymbol{\beta}))\text{vech}(\Lambda) - n \log \frac{C(\boldsymbol{\kappa}, \Lambda)}{C(\tilde{\boldsymbol{\kappa}}, \tilde{\Lambda})}$$

by the Monte Carlo log-likelihood

$$\begin{aligned} l_{\text{MC}}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda) = &\sum_{i=1}^{n} \boldsymbol{t}_1^\top(\boldsymbol{y}_i - \boldsymbol{\mu}(\boldsymbol{\beta}))\boldsymbol{\kappa}(\boldsymbol{\gamma}) + \boldsymbol{t}_2^\top(\boldsymbol{y}_i - \boldsymbol{\mu}(\boldsymbol{\beta}))\text{vech}(\Lambda) \\ &- n \log \frac{1}{M} \sum_{m=1}^{M} \exp A_m\left(\boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}, \Lambda, \tilde{\Lambda}\right) \end{aligned} \tag{11}$$

where

$$A_m\left(\boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}, \Lambda, \tilde{\Lambda}\right) = \left(\boldsymbol{t}_1^{\mathsf{T}}(\boldsymbol{y}_m^{\star})(\boldsymbol{\kappa}(\boldsymbol{\gamma}) - \boldsymbol{\kappa}(\tilde{\gamma})) + \boldsymbol{t}_2^{\mathsf{T}}(\boldsymbol{y}_m^{\star})(\mathrm{vech}(\Lambda) - \mathrm{vech}(\tilde{\Lambda}))\right),$$

and obtaining the maximum likelihood estimate $\hat{\boldsymbol{\theta}}_{\mathrm{MC}}$ as the maximum point of l_{MC}.

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda)$ be the vector of the unknown parameters. According to [18], the information matrix

$$\hat{I}_M = -\frac{\partial^2 l_{\mathrm{MC}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\mathrm{MC}}}$$

is a consistent estimate of

$$\hat{I}(\hat{\boldsymbol{\theta}}) = -\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

where $l(\boldsymbol{\theta})$ is the exact likelihood function and $\hat{\boldsymbol{\theta}}$ is the exact maximum likelihood estimate. Hence the diagonal elements of $\hat{I}_M^{-1}$ can be exploited to obtain the estimated variances of the MCMC parameter estimates. The accuracy of these variance estimates increases with the number M. The general theory developed by Geyer (1994) does not give information on the rate of convergence. Hence a suitably large number M must be chosen on empirical grounds. A possible strategy is to choose M in a way that the Monte Carlo variability, i.e. the covariance matrix of $(\hat{\boldsymbol{\theta}}_{\mathrm{MC}} - \hat{\boldsymbol{\theta}})$, is small with respect to the 'true' variability, i.e. the covariance matrix of $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$, estimated by $\hat{I}_M^{-1}$. The Monte Carlo variability can be estimated by standard time series method (Geyer and Thompson 1992).

The quality of the Monte Carlo samples can be assessed by examining the time series of the sufficient statistics $\boldsymbol{t}(\boldsymbol{y}_m^{\star})$. Time series of good quality should be stationary and the corresponding autocorrelation functions should decay rapidly to zero. This is usually accomplished by taking only the vectors $\boldsymbol{y}_m^{\star} = \boldsymbol{y}_{B+m\Delta}^{\star}$ from the whole time series $\boldsymbol{y}_1^{\star}, \boldsymbol{y}_2^{\star}, \ldots$, where B is the so-called burn-in period and Δ is the so-called spacing.

The Monte Carlo likelihood (11) can be maximized by traditional Newton-type algorithms, such as those available in the R command `optim`. Occasional overflows and numerical instabilities may occur during the optimization process, if the quantity

$$\omega_m(\boldsymbol{\gamma}, \Lambda) = \boldsymbol{t}_1^{\mathsf{T}}(\boldsymbol{y}_m^{\star})(\boldsymbol{\kappa}(\boldsymbol{\gamma}) - \boldsymbol{\kappa}(\tilde{\gamma})) + \boldsymbol{t}_2^{\mathsf{T}}(\boldsymbol{y}_m^{\star})(\mathrm{vech}(\Lambda) - \mathrm{vech}(\tilde{\Lambda})),$$

is too large for some $\boldsymbol{y}_m^{\star}$. In these cases, a scaled version of the MC log-likelihood can be specified by defining

$$\omega_{\max} = \max_{m=1,\ldots,M} \omega_m(\boldsymbol{\gamma}, \Lambda),$$

and setting

$$l_{\mathrm{MC}}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda) = \sum_{i=1}^{n} \boldsymbol{t}_1(\boldsymbol{y}_i - \boldsymbol{\mu}(\boldsymbol{\beta}))^{\mathsf{T}} \boldsymbol{\kappa}(\boldsymbol{\gamma}) + \boldsymbol{t}_2(\boldsymbol{y}_i - \boldsymbol{\mu}(\boldsymbol{\beta}))^{\mathsf{T}} \mathrm{vech}(\Lambda) - n\omega_{\max} - n \log \frac{1}{M} \sum_{m=1}^{M} \exp B_m \left(\boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}, \Lambda, \tilde{\Lambda} \right) \tag{12}$$

where

$$B_m \left(\boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}, \Lambda, \tilde{\Lambda} \right) = \boldsymbol{t}_1^{\mathsf{T}}(\boldsymbol{y}_m^{\star})(\boldsymbol{\kappa}(\boldsymbol{\gamma}) - \boldsymbol{\kappa}(\tilde{\gamma})) + \boldsymbol{t}_2^{\mathsf{T}}(\boldsymbol{y}_m^{\star})(\mathrm{vech}(\Lambda) - \mathrm{vech}(\tilde{\Lambda})) - \omega_{\max}.$$

4.4 Simulation studies

MCMC maximum likelihood estimates are asymptotically normally distributed under general conditions (Geyer 1994). Little is however known about the distributional properties of these estimates under specific models and moderate sample sizes. Ad-hoc simulation studies are therefore advisable. This Section examines the behavior of the MCMC-MLEs under two settings, inspired by the applications described in Sect. 2.

First, the case of a homoscedastic AR(1) model with one covariate, in a longitudinal setting, was examined. Precisely, $n = 65$ values $x_1, \ldots, x_n$ were drawn from a uniform distribution with support $(-1, 1)$. Second, a sample of size $n = 65$ is drawn from a 3-parameters multivariate ($J = 5$) von Mises distribution, obtained from (4) by assuming that $\mu_{ij} = 2\tan^{-1}(\beta x_i)$, $\kappa_{ij} = \exp(\gamma)$ and, finally,

$$\lambda_{jk} = \begin{cases} \lambda & \text{if } \mid j - k \mid = 1 \\ 0 & \text{otherwise} \end{cases} \qquad j, k = 1, \ldots, 5$$

For each triplet of true parameter values (β, γ, λ), 500 sets of $n \times J$ pseudo-observations were generated by a Gibbs sampler; the 500 data sets were obtained from 500 independent Markov chains. For each pseudo-observed data set, MCMC estimates of the parameters were calculated. These MCMC estimates were computed using a sample size of $M = 2000$ generated from a Gibbs sampler started from the zero-state, that is, the initial value is the $y_{ij} = 0$ for each i and j. The burn-in period and the spacing were respectively $B = 200$ and $\Delta = 2$. Under this setting, the time series of the sufficient statistics appeared to be stationary and with an exponentially decreasing autocorrelation function.

The distribution of the estimates of the regression coefficient $\hat{\beta}_{\mathrm{MC}}$, the log-concentration parameter $\hat{\gamma}_{\mathrm{MC}}$ and the dependence parameter $\hat{\lambda}_{\mathrm{MC}}$, respectively, was examined by computing the bias, the variance, the skewness (measured by the index

$m_3/m_2^{3/2}$) and the kurtosis (measured by the excess index $(m_4/m_2^2) - 3$). The parameter values were inspired by the applications of Sect. 5 and chosen to reflect weak and strong covariate effects ($\beta = 0.5, 2$), and different scenarios of concentration ($\gamma = 0, 1, 2$, corresponding to the concentrations $\kappa = \exp\gamma = 1, e, e^2$) and temporal dependence ($\lambda = 0, 2, 4, 6$). Notice that if $\lambda = 0$, then the data are independent and the MCMC maximum likelihood procedure is not necessary. However, the performance of the MCMC estimation procedure was reported for comparison purposes.

The results are summarized by the three pictures in the left column of Fig. 3, which shows that bias, variance and skewness increase as the ratio $\lambda/\exp(\gamma)$ increases. Because similar patterns were observed regardless of the value taken by β, only the cases with $\beta = 2$ were reported. The highest values of bias, skewness and kurtosis occur when $\gamma = 0$ (i.e., $\kappa = \exp(0) = 1$). This phenomenon is however alleviated when γ takes higher values. These results seem to suggest that, under the considered settings, the normal approximation to the distribution of the MCMC estimates appears reasonable so long as the temporal dependence between adjacent observations is not too large and the concentration is not too small.

Similar results were obtained in the spatial setting (right column of Fig. 3). The particular case of a homoscedastic spatial AR(1) model was considered. First, $n = 15\times15$ values $x_1, \ldots, x_n$ were drawn from a uniform distribution with support $(-1, 1)$. Then, each value was associated with a point of coordinates (x_i, y_i) in a regular 15×15 lattice, $x_i, y_i = 1, \ldots, 15$. Spatial series of circular observations were drawn from a 3-parameters multivariate ($J = 15 \times 15$) von Mises distribution, obtained from (4) by assuming that $\mu_i = 2\tan^{-1}(\beta x_i)$, $\kappa_i = \exp(\gamma)$ and, finally, $\Lambda = \lambda \boldsymbol{C}$, where $\boldsymbol{C}$ is a $15^2 \times 15^2$ connectivity matrix with entries $c_{ij} = 1$ if $j \in N(i)$, and 0 otherwise, where $N(i)$ includes the four nearest locations with coordinates $(x_i - 1, y_i)$, $(x_i + 1, y_i)$,$(x_i, y_i - 1)$, $(x_i, y_i + 1)$.

For each set of true parameter values, 500 sets of $n \times J$ pseudo-observations were generated by running 500 independent Markov chains. Figure 3 shows again that the normal approximation to the MCMC MLEs deteriorates as the ratio $\lambda/\exp(\gamma)$ increases. Under the considered spatial setting, the behavior of the estimates seems worse than that exhibited under a longitudinal setting. This was somehow expected, as in the spatial setting we are using a single spatial series of limited size, while in a longitudinal setting multiple observations were used.

5 Applications

5.1 Animal orientation

The effects of the covariates on escape direction can be estimated by using an inverse tangent link function. Precisely, we may assume that the mean direction of the ith animal at time j is given by

$$\mu_{ij}(\boldsymbol{\beta}) = \beta_0 + 2\tan^{-1}(\boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\beta}),$$

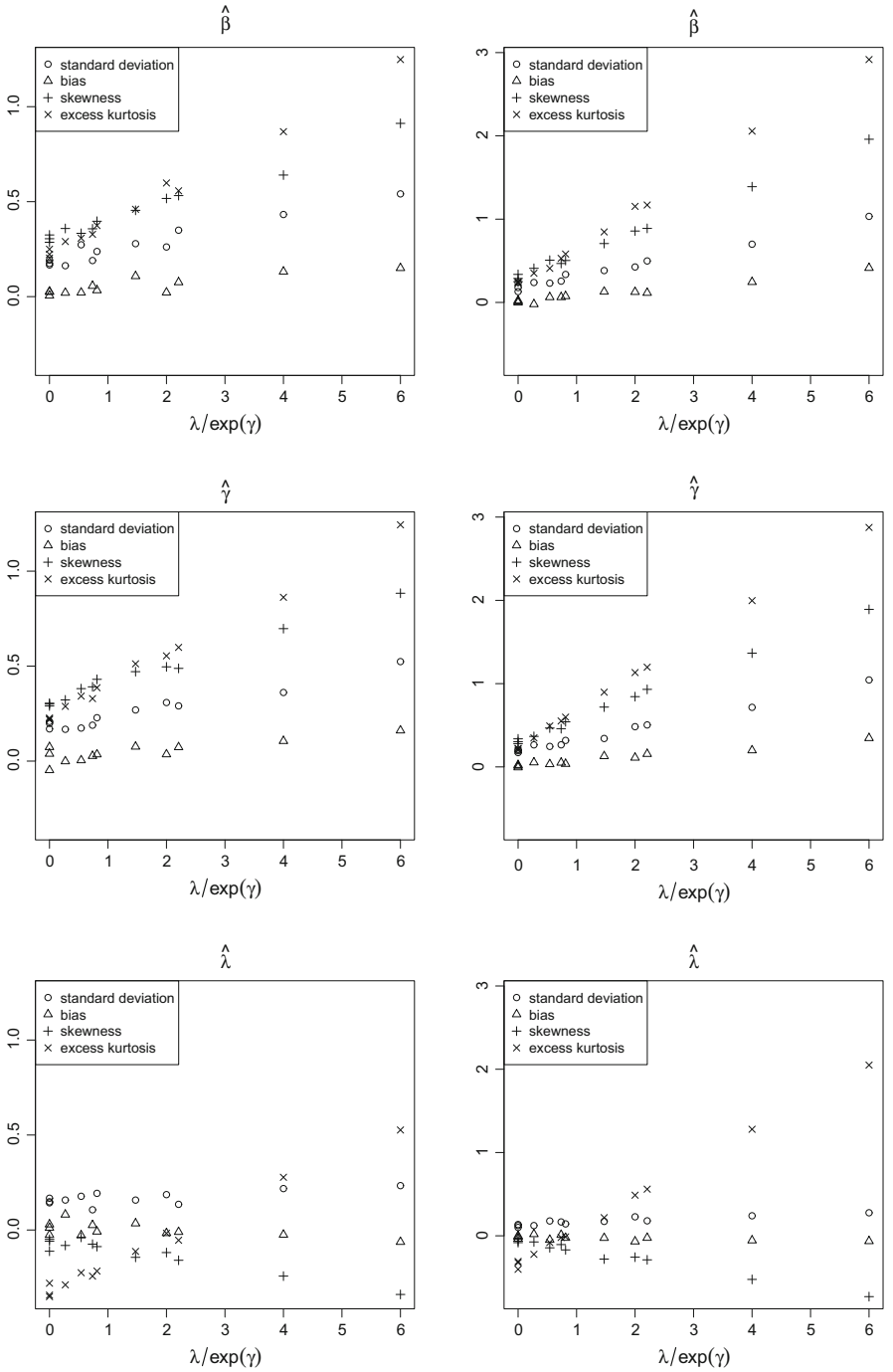

Fig. 3 Results of the simulation study. Summary statistics of the sampling distribution of $\hat{\beta}_{\mathrm{MC}}$ (*top*), $\hat{\gamma}_{\mathrm{MC}}$ (*middle*) and $\hat{\lambda}_{\mathrm{MC}}$ (*bottom*) under an autoregressive time series model (*left*) and a spatial autoregressive model (*right*)

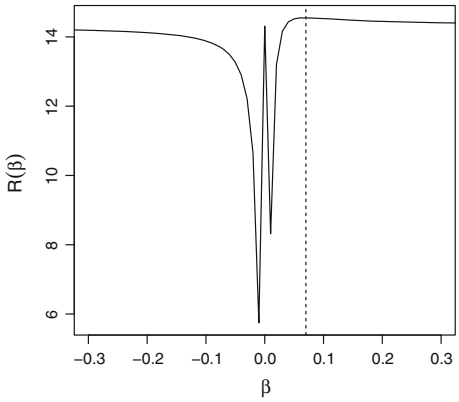

Fig. 4 The multimodal shape of an univariate $R(\beta)$ function, given the animal orientation data. The *dashed line* indicates the maximum of the function

where the vector $\boldsymbol{x}_i^{\mathsf{T}}$ includes wind direction, azimuth and eye asymmetry of the ith animal. In this case study, using the inverse tangent link function may lead to a multimodal pseudo-likelihood function. Figure 4 displays the shape of the univariate R function

$$R(\beta) = \left(S^2(\beta) + C^2(\beta)\right)^{1/2}$$

$$S(\beta) = \frac{1}{nJ}\sum_{i=1}^{n}\sum_{j=1}^{J}\sin(y_{ij} - g(x_i\beta))$$

$$C(\beta) = \frac{1}{nJ}\sum_{i=1}^{n}\sum_{j=1}^{J}\cos(y_{ij} - g(x_i\beta)),$$

computed by taking the covariate azimuth direction for the sun. The presence of two local maxima that are close to each other clearly illustrates the care that must be taken with starting values as discussed previously. In this case, the value 0.07 was taken as the starting point for the regression coefficient of the azimuth direction for the sun, to initialize the maximization of the pseudo-likelihood function.

Table 3 shows the results obtained by fitting the three models described in Table 1. Standard errors were computed by inverting the observed information matrix. The parameter estimates seem to indicate that the ratio $\lambda/\exp(\gamma)$ is small and, according to the simulation studies in Sect. 4.1, the normal approximation to the sampling distribution of the MCMC-MLEs should be reasonable. The first two columns of the table display the estimates and the standard errors obtained by an unstructured model where five time-specific concentrations $\kappa_t, t = 1, \ldots, 5$ and ten conditional dependence parameters $\lambda_{t\tau}$ $(t = 1, \ldots, 5, \tau > t)$ are separately estimated along with the fixed effects $\boldsymbol{\beta}$ of the available covariates. Under this model, sun azimuth, wind direction and eye symmetry appear to influence significantly a sandhopper's orientation. Only the difference between the effect of wind blowing from land and wind blowing from longshore-East seems not significant, using a 95 % confidence interval. The signs of the fixed effects are consistent with the previous analyses of these data (D'Elia 2001; Song 2007; Nunez-Antonio and Gutiérrez-Peña 2012). In these studies,

Table 3 Estimates of three models for longitudinal circular data

	Unstructured model			Heteroscedastic AR(1) model			Homoscedastic AR(1) model	
	Estimate	S.E.		Estimate	S.E.		Estimate	S.E.
Sun	0.050	0.001		0.051	0.001		0.050	0.001
Eye symmetry	−1.647	0.223		−1.676	0.307		−1.553	0.320
Longshore-East	−0.058	0.112		−0.098	0.123		−0.114	0.128
Sea	−0.553	0.081		−0.539	0.100		−0.617	0.108
Longshore-West	−0.592	0.128		−0.612	0.130		−0.690	0.135
κ_1	1.819	0.309	γ_0	0.719	0.146	κ	2.811	0.219
κ_2	2.763	0.457						
κ_3	3.329	0.537	γ_1	0.113	0.040			
κ_4	4.967	0.833						
κ_5	4.097	0.677						
λ_{12}	1.825	0.513	λ	2.465	0.213	λ	2.392	0.216
λ_{13}	0.425	0.581						
λ_{14}	1.236	0.706						
λ_{15}	−0.873	0.645						
λ_{23}	2.267	0.640						
λ_{24}	0.156	0.838						
λ_{25}	0.113	0.765						
λ_{34}	1.811	0.860						
λ_{35}	0.444	0.789						
λ_{45}	4.709	0.923						
AIC	401.661			156.543			174.870	

however, some of the covariates are not significant. These studies included time of release as an additional covariate, and it is therefore possible that this trend component, not included in our analysis, triggered the explanatory power of the remaining covariates. In addition, the increasing concentration parameters seem to indicate that the data are slightly heteroscedastic, a feature that was not considered in the previous analyses. More interestingly, the estimated conditional dependence parameters reveal a specific first-order autoregressive structure of the outcomes: using a 95 % confidence interval, only the parameters λ_{ij} with $\mid i - j \mid = 1$ are significant, while the remaining parameters are not significant. Such a dependence structure was mentioned by D'Elia (2001), but it was not incorporated in their analysis.

These results suggest that some of the parameters in the unstructured model are redundant and that some simplification can be applied. A more parsimonious model can be specified by assuming that the concentrations follow a parametric function of time of release, say $\kappa_t = \exp(\gamma_0 + \gamma_1 t)$, where $t = 1, \ldots, 5$ is time of release, and that the dependence parameters follow a first-order auto-regressive structure. Columns 3 and 4 of Table 3 display the estimates and the standard errors obtained by fitting this heteroscedastic AR(1) model. Estimates and standard errors of the fixed effects are similar to those estimated under the unstructured model. The significant slope γ_1 indi-

Table 4 Estimates of three models for spatial circular data

	Heteroscedastic spatial AR(2)			Heteroscedastic spatial AR(1)			Homoscedastic spatial AR(1)	
	Estimate	S.E.		Estimate	S.E.		Estimate	S.E.
Longitude	−1.435	0.238		−1.429	0.237		−1.681	0.246
Latitude	−3.630	0.278		−3.610	0.274		−3.881	0.310
γ_0	1.659	0.142	γ_0	1.635	0.141	γ_0	2.013	0.090
γ_1	2.545	0.506	γ_1	2.572	0.495			
λ_1	6.726	0.820	λ	6.504	0.612	λ	5.830	0.529
λ_2	−0.007	0.634						
AIC	259.210			251.341			269.015	

cates that escape directions tend to be more and more concentrated, perhaps reflecting a learning effect. Finally, the significant estimate of λ summarizes the positive association between subsequent releases. By comparing the AIC values of the unstructured model and this model, the latter seems a better choice. The last columns of Table 3 indicate the results obtained by assuming equal concentrations, $\kappa_t = \kappa$ along with a first-order auto-regressive structure, i.e. fitting a homoscedastic AR(1) model. This further constrain increases the AIC, favoring a model that allows for heteroscedasticity.

5.2 Marine currents

Table 4 displays the results obtained by assuming that the data are a sample drawn from a MVM distribution with location parameters

$$\mu_j(\beta_0, \beta_1, \beta_2) = \beta_0 + 2\tan^{-1}\left((x_{1j} - \bar{x}_1)\beta_1 + (x_{2j} - \bar{x}_2)\beta_2\right),$$

where x_{1j} and x_{2j} are respectively the longitude and the latitude of the jth observation point. The pseudo-likelihood maximization was initialized at point $(-0.93, 0.26)$, which is the maximum of the bivariate R function is this case study (Fig. 5).

This spatial gradient component was estimated along with the three different dependence structures displayed in Table 2. Standard errors were computed by inverting the observed information matrix. The parameter estimates seem to indicate that the ratio $\lambda/\exp(\gamma)$ is small and, according to the simulation studies in Sect. 4.1, the normal approximation to the sampling distribution of the MCMC-MLEs seems defensible.

The first two columns of Table 4 display the estimates under an heteroscedastic spatial AR(2) model. This model accounts for heteroscedasticity by assuming that the concentration at the jth observation site is a parametric function of the minimum distance d_j from the coast, say

$$\kappa_j(\gamma_0, \gamma_1) = \exp(\gamma_0 + \gamma_1 d_j).$$

It additionally assumes a second-order spatial auto-regressive structure, by associating two dependence parameters, λ_1 and λ_2, with two disjoint neighborhood structures,

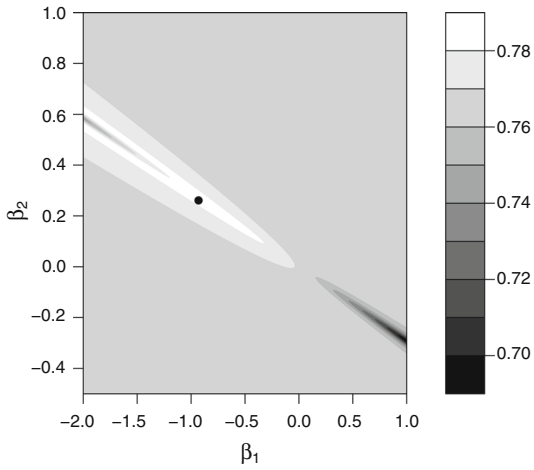

Fig. 5 The function R, as computed from the marine current data. The *black dot* indicates the maximum of the function

N_1 and N_2. For each site j, $N_1(j)$ include all the grid points k such that the distance between k and j is less or equal to 2 km. $N_2(j)$ includes instead all the points at a distance between 2 km and 4 km from j.

Under this model, the negative fixed effects reflect the clockwise rotation of currents as longitude and latitude increase. More interestingly, the slope γ_1 is significant at a 95 % confidence level, indicating that the concentration of current directions around the spatial gradient increases with the distance from the coast. Finally, the first-order auto-correlation parameter λ_1 is significant at a 95 % level, while the second-order parameter λ_2 seems not significant. This suggests that the residual of each outcome from the spatial gradient is conditionally independent on the second-order neighborhood, given the nearest first-order neighbors. In other words, ignoring a second-order neighborhood structure should lead to a better compromise between model parsimony and maximum likelihood.

Columns 3 and 4 of Table 4 display the estimates and the standard errors obtained by estimating an heteroscedastic spatial AR(1) model and a homoscedastic spatial AR(1) model. The resulting estimates are similar to those obtained under the previous model, with smaller standard errors. As expected, the AIC suggests that the more parsimonious first-order auto-regressive model is better than a model that assumes two spatial neighborhood structures.

The last estimated model (columns 5 and 6 of Table 4) is obtained by assuming equal concentrations, $\kappa_j = \kappa$ and a first-order auto-regressive structure, $\Lambda = \lambda C_1$. The AIC value ranks this model as the worst one, favoring the models that accounts for spatial heteroscedasticity.

6 Discussion

The multivariate von Mises density provides a flexible tool to specify regression models for correlated circular data. A distinct advantage of these models is the intuitively appealing interpretation of the parameters, which mimic the parameters of a multi-

variate normal distribution. In real case studies, such as those illustrated in this paper, the simple structure of a MVM regression model allows to account for unstructured dependence relationships, heteroscedastic concentrations and autoregressive structures in space and time.

The MVM density is however not closed under marginalization and its normalizing constant is intractable. These distributional features complicate the estimation of the parameters. On the other hand, the univariate conditional distributions are von Mises and the centered MVM is a canonical exponential family distribution. As a result, a straightforward Gibbs sampling procedure can be implemented to draw samples from the MVM distribution and specify a tractable Monte Carlo approximation of the log-likelihood. For large samples, the distribution of the MCMC-MLEs can be approximated by a normal distribution. However, simulation studies indicate that the quality of this approximation can deteriorate in small samples, especially when the concentration of the data is small and the dependence is strong. Further research is needed to propose suitable estimation strategies in these cases.

The MCMC approximation of the likelihood function depends on samples drawn from a MVM distribution, evaluated at preliminary pseudo-likelihood estimates. Pseudo-likelihood estimates are obtained by maximizing a computationally tractable pseudo-likelihood function, which is however typically multimodal. This requires a multi-start strategy where the optimization routine is initialized at several starting points of the parameter space (O'Hagan et al. 2012). Choosing these starting points according to a grid with a specific resolution is feasible only when the number of the parameters in a model is small. As the number of parameters increases, a grid-based search of the maximum is not longer feasible, due to the curse of dimensionality, and multi-start strategies rely on starting points that are randomly drawn from the parameter space.

Appendix

This Appendix displays the code to compute the principal functions, needed to fit an unstructured MVM regression model for longitudinal circular data. This is the most complex model of the paper. The code can be easily modified to fit the other simpler models considered in the paper.
The following function computes the sufficient statistics of a multivariate sample of circular observations (Sect. 4.3).

```
stat.suff <- function(Y) {
Y <- as.matrix(Y)
S <- matrix(0,nrow(Y),ncol(Y)+ncol(Y)*(ncol(Y)-1)/2)
for(i in 1:nrow(Y)){
    ss <- sin(Y[i,])%*%t(sin(Y[i,]))
    S[i,]<-c(cos(Y[i,]),ss[lower.tri(ss)==TRUE])
}
    return(S)
}
```

The following function draws a matrix of M samples for a centered MVM distribution (Sect. 4.1).

```
chain <- function(M,theta,y0){

# M: number of sweepings
# theta: parameters
# y0: initial configuration

    Sam <- matrix(0,nrow=M,ncol=J)
    K <- theta[1:J]
    L <- matrix(0,J,J)
    L[lower.tri(L)==TRUE] <- theta[(J+1):(J+(J*(J-1)/2))]
      L <- L + t(L)
    sam <- y0
    for(step in 1:M){
        sweeping <- sample(1:length(y0))
        for(j in sweeping){
        sin.tilde <- sum(L[j,]*sin(sam))
        tan.mu.cond <- sin.tilde/K[j]
        kappa.cond <- sqrt(K[j]^2+sin.tilde^2)
        sam[j] <- PI(rvm(1,atan(tan.mu.cond),kappa.cond))

    }
    Sam[step,] <- sam
    }

    return(Sam)
}
```

The following function is the pseudo-likelihood function, defined in Sect. 4.2.

```
# dat is a n x J matrix of outcomes

neg.pseudo.log.lik <- function(theta){

    K <- theta[1:J]
    L <- matrix(0,J,J)
    L[lower.tri(L)==TRUE] <- theta[(J+1):(J+(J*(J-1)/2))]
      L <- L + t(L)
    beta <- theta[ (J+(J*(J-1)/2)+1): (J+(J*(J-1)/2)
+ncol(cov))]
    nu.cond <- dat
    kappa.cond <- dat
        lin.pred <- rep(cov%*%beta[1:ncol(cov)],J)
        res <- dat - 2*atan(lin.pred)
```

```
    for(i in 1:nrow(dat)){
        for(j in 1:ncol(dat)){
        nu.cond[i,j] <- atan(sum(L[j,]*sin(res[i,]))
/K[j])
        kappa.cond[i,j] <- sqrt(K[j]^2+(sum(L[j,]*sin
(res[i,])))^2)
        }
    }
    fun.j <- matrix(0,J,1)
    for(j in 1:J){
    fun.j[j] <- sum(log(dvm(dat[,j],mu=nu.cond[,j],
kappa=kappa.cond[,j])))
    }
    fun <- sum(fun.j)
    return(-fun)
}
```

Preliminary MPLEs are obtained by a traditional quasi-Newton method, as follows

```
MPLE <- optim(fn = neg.pseudo.log.lik, par =
c(
apply(dat,2,function(sub) est.kappa(sub))
,rep(0,J*(J-1)/2),rep(0,ncol(cov))),hessian=TRUE,method=
"BFGS",control=list(trace=1))
```

The following function defines the Monte Carlo log-likelihood approximation, initialized at the value `theta.tilde`. The function `optim` is then used to obtain the final estimates (see Sect. 4.3).

```
theta.tilde <- MPLE$par[1:(J+J*(J-1)/2)]

y<-rvm(J,0,0.01)
sim <- chain(10000,theta.tilde,y)
sim <- sim[5000:10000,]

MC.neg.log.lik <- function(theta){

    THETA <- theta[1:(J+J*(J-1)/2)]
    BETA <- theta[ (J+(J*(J-1)/2)+1): (J+(J*(J-1)/2)
+ncol(cov))]

        lin.pred <- rep(cov%*%BETA[1:ncol(cov)],J)
        res <- dat - 2*atan(lin.pred)
    fun <-
    sum(stat.suff(res)%*%THETA) - nrow(dat)*log(
    sum(exp(stat.suff(sim)%*%(THETA-theta.tilde))))
```

```
return(-fun)
}

MC.mle <- optim(fn = MC.neg.log.lik,par = c(MPLE$par
[1:(J+J*(J-1)/2)],rep(0,5)),hessian=TRUE,
control=list(trace=1), method="BFGS")
MC.mle$par
```

References

Abraham C, Molinari N, Servien R (2013) Unsupervised clustering of multivariate circular data. Stat Med 32(8):1376–1382

Arnold B, SenGupta A (2006) Recent advances in the analyses of directional data in ecological and environmental sciences. Environ Ecol Stat 13(3):253–256

Artes R, Jørgensen B (2000) Longitudinal data estimating equations for dispersion models. Scand J Stat 27(2):321–334

Bahlmann C (2006) Directional features in online handwriting recognition. Pattern Recognit 39:115–125

Bertotti L, Cavalieri L (2009) Wind and wave predictions in the adriatic sea. J Mar Syst 78:S227–S234

Besag J (1975) Statistical analysis of non-lattice data. Statistician 24:179–195

Bhattacharya S, SenGupta A (2009) Bayesian inference for circular distributions with unknown normalising constants. J Stat Plan Inference 139:4179–4192

Bulla J, Lagona F, Maruotti A, Picone M (2012) A multivariate hidden markov model for the identification of sea regimes from incomplete skewed and circular time series. J Agricult Biol Environ Stat 17(4):544–567

Cosoli S, Gacic M, Mazzoldi A (2012) Surface current variability and wind influence in the northeastern adriatic sea as observed from high-frequency (hf) radar measurements. Cont Shelf Res 33:1–13

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

Diggle P, Heagerty P, Liang KY, Zeger S (2002) Analyis longitudinal data. Oxford University Press, Oxford

Fisher N (1993) The analysis of circular data. Cambridge University Press, Cambridge

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

Fisher N, Lee A (1994) Time series analysis of circular data. J R Stat Soc Ser B 56(2):327–339

Gaetan C, Guyon X (2010) Spatial statistics and modelling. Springer, New York

Geman S, Geman D (1984) Stochastic relaxation, gibbs distribution and the bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell 6:721–741

Geyer CJ (1992) Practical Markov chain Monte Carlo. Stat Sci 7:473–511

Geyer CJ (1994) On the convergence of Monte Carlo maximum likelihood calculations. J R Stat Soc Ser B 56:261–274

Geyer CJ, Thompson EA (1992) Constrained monte carlo maximum likelihood for dependent data. J R Stat Soc Ser B 54:657–699

Jona-Lasinio G, Gelfand A, Jona-Lasinio M (2012) Spatial analysis of wave direction data using wrapped gaussian processes. Ann Appl Stat 6:1478–1498

Lagona F (2002) Adjacency selection in Markov random fields for high spatial resolution hyperspectral data. J Geogr Syst 4(1):53–68

Lagona F, Picone M, Maruotti A, Cosoli S (2015) A hidden Markov approach to the analysis of space-time environmental data with linear and circular components. Stoch Environ Res Risk Assess 29:397–409

Lagona F, Picone M (2012) Model-based clustering of multivariate skew data with circular components and missing values. J Appl Stat 39(5):927–945

Lagona F, Picone M (2013) Maximum likelihood estimation of bivariate circular hidden markov models from incomplete data. J Stat Comput Simul 83:1223–1237

Mardia K, Taylor C, Subramaniam G (2007) Protein bioinformatics and mixtures of bivariate von mises distributions for angular data. Biometrics 63:505–512

Mardia KV (2010) Bayesian analysis for bivariate von Mises distributions. J Appl Stat 37(3):515–528

Mardia KV, Voss J (2014) Some fundamental properties of a multivariate von Mises distribution. Commun Stat-Theory Methods 43(6):1132–1144
Mardia KV, Hughes G, Taylor CC, Singh H (2008) A multivariate von Mises distribution with applications to bioinformatics. Can J Stat 36(1):99–109
Mardia KV, Kent JT, Hughes G, Taylor CC (2009) Maximum likelihood estimation using composite likelihoods for closed exponential families. Biometrika 96(4):975–982
Mihanovic H, Cosoli S, Vilibic I, Ivankovic D, Dadic V, Gacic M (2011) Surface current patterns in the northern adriatic extracted from high frequency radar data using self organizing map analysis. J Geophys Res 116(C08):033
Modlin D, Fuentes M, Reich B (2012) Circular conditional autoregressive modeling of vector fields. Environmetrics 23(1):46–53
Nunez-Antonio G, Gutiérrez-Peña E (2012) A bayesian model for longitudinal circular data based on the projected normal distribution. Comput Stat Data Anal 71:506–519
O'Hagan A, Brendan Murphy T, Gormley IC (2012) Computational aspects of fitting mixture models via the expectation-maximization algorithm. Comput Stat Data Anal 56:3843–3864
Pewsey A, Neuhauser M, Ruxton D (2013) Circular statistics in R. Oxford University Press, Oxford
Presnell B, Morrison SP, Littell RC (1998) Projected Multivariate Linear Models for Directional Data. J Am Stat Assoc 93:1068–1077
Rueda C, Fernandez M, Peddada S (2009) Estimation of parameters subject to order restrictions on a circle with application to estimation of phase angles of cell-cycle genes. J Am Stat Assoc 104:338–347
Singh H, Hnizdo V, Demchuk E (2002) Probabilistic model for two dependent circular variables. Biometrika 89:719–723
Song PXK (2007) Correlated data analysis. Springer, New York
Varin C, Reid N, Firth D (2011) An overview on composite likeihood methods. Stat Sin 21:5–42

Francesco Lagona is associate professor of Statistics at the University of Roma Tre, Rome (Italy) and Research Fellow at the Max Planck Institute for Demographic Research, Rostock (Germany). His research interests include the statistical analysis of spatial and temporal data with applications to environmental studies, cognitive impairment, image analysis and cellular biology.