

Regression analysis of correlated circular data

Modelli di regressione per dati circolari correlati

Francesco Lagona

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 suitable link functions. The model accommodates for heteroscedasticity, unstructured correlation, and specific autoregressive correlation structures. Inference is based on a Monte Carlo approximation of the log-likelihood, due to the intractability of the normalizing constant. The model is illustrated on two case studies: a longitudinal study of animal orientation and a study on the spatial distribution of sea current directions.

Riassunto Si propone un modello di regressione per dati circolari correlati e generati da una distribuzione di von Mises multivariata con medie e concentrazioni che dipendono da covariate. Specifiche forme di eteroschedasticità e correlazione strutturata e non strutturata possono essere facilmente incluse nella specificazione del modello. A causa della intrattabilità della costante di normalizzazione della distribuzione von Mises multivariata, la stima dei parametri si basa su un'approssimazione Monte Carlo della verosimiglianza. Le potenzialità del modello vengono illustrate usando un esempio di dati circolari longitudinali ed un esempio di dati spaziali circolari.

Key words: Circular data, Monte Carlo maximum likelihood, Multivariate von Mises, Animal orientation, Sea currents

Francesco Lagona
University Roma Tre, e-mail: francesco.lagona@uniroma3.it

1 Introduction

Correlated circular data arise when multiple angular measurements are taken on the same sample unit and are encountered across several research areas, including bioinformatical studies of the dihedral angles that summarizes the structure of a protein [7] and environmental studies of wind and wave directions [5]. 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. 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. Noticeable exceptions include the multivariate wrapped normal distribution [2] and the multivariate von Mises (MVM) distribution [7].

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 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 proposed MVM regression model is specified by assuming that both the mean directions and the concentrations depend on covariates through suitable link functions and allowing the matrix of the conditional dependence parameters to be fully unstructured, therefore extending the generalized linear model developed by [3] for independent angular measurements.

Maximum likelihood estimation (MLE) of the MVM regression model is complicated by the intractability of the normalizing constant of the MVM density. In the bivariate case, the normalizing constant of this density can be numerically approximated by a sum of Bessel functions. 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 [7]. 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 [4].

An alternative estimation strategy relies on a Monte Carlo (MC) approximation of the log-likelihood. MC maximum likelihood estimation [4] is based on the maximization of a stochastic approximation of the log-likelihood, obtained by drawing samples from the model of interest, evaluated at a pseudo-likelihood estimate. It is an attractive estimation method when the following conditions are fulfilled: a computationally efficient procedure is available to obtain pseudo-likelihood estimates (MPLEs); samples from the assumed model can be easily drawn without knowledge of the normalizing constant; and, 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 VM. As a result, first, finding MPLEs reduces essentially to a VM regression problem and requires a tiny computational effort, even in the presence of covariates that modulate directional

means 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 VM 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.

2 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 $\mathbf{y} = (y_1, \dots, y_J)$, $y_j \in [-\pi, \pi)$, is drawn from J independent VM distributions with means $\mu_1, \dots, \mu_J$ and concentrations $\kappa_1, \dots, \kappa_J$. The influence of the available covariates on these parameters can be conveniently studied by specifying suitable link functions that map linear predictors to the appropriate parameter space. Following [3], the means can be for example modeled as

$$\mu_j(\boldsymbol{\beta}) = \beta_0 + g(\mathbf{x}_j^\top \boldsymbol{\beta}), \quad (1)$$

where $\beta_0 \in [-\pi, \pi)$ is an offset parameter, $\mathbf{x}_j^\top$ is a row-vector of K covariates, associated with observation y_j , $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)$ is a vector of regression coefficients and, finally, g is a known monotone function that maps the real line to the circle $[-\pi, \pi)$. If the link function g is such that $g(0) = 0$, then the offset β_0 can be conveniently interpreted as an origin. A popular choice is given by $g(x) = 2 \tan^{-1}(x)$. Similarly, the concentration parameters can be modeled as follows

$$\kappa_j(\boldsymbol{\gamma}) = h(\mathbf{w}_j^\top \boldsymbol{\gamma}), \quad (2)$$

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

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

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

where

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

is the VM 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 $\mathbf{y}$ is a sample drawn from a MVM distribution, known up to the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ and a $J \times J$ symmetric association matrix Λ with zero diagonal entries ($\lambda_{jj} = 0$, $\lambda_{jh} = \lambda_{hj}$, $j, h = 1 \dots J$), namely

$$f_{\text{MVM}}(\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda) = \frac{\exp(\boldsymbol{\kappa}(\boldsymbol{\gamma})^\top \mathbf{c}(\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\beta})) + \frac{1}{2} \mathbf{s}(\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\beta}))^\top \Lambda \mathbf{s}(\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\beta})))}{C(\boldsymbol{\gamma}, \Lambda)} \quad (4)$$

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

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

Density (4) extends the MVM distribution proposed by [7] to the GLM setting, by assuming that both the means and the concentration parameters depend on covariates.

Most of the MVM properties have been discussed recently [7, 6]. Among these properties, of interest in this paper is the convenient VM form taken by the univariate conditional distributions. Formally, under (4), the conditional distribution of each angle y_j given the rest of the sample $\mathbf{y}^{(j)} = (y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_J)$ is a univariate VM density

$$f(y_j | \mathbf{y}^{(j)}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda) = f_{\text{VM}}\left(y_j; v_j(\mathbf{y}^{(j)}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda), \frac{1}{\sigma_j^2(\mathbf{y}^{(j)}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda)}\right)$$

where

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

is the conditional mean parameter and, finally,

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

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 VM densities and the proposed model reduces to the Fisher-Lee regression model. Otherwise, the univariate marginal distributions are in general not VM with a normalizing constant that can be accurately approximated only when $J = 2$.

The distribution (4) reflects a fully saturated model with covariate-specific means and concentrations, and an unstructured association matrix Λ . A saturated model 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 j th 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 . When $\mathbf{y} = (y_1 \dots y_T)$ is a time series of angular measurements, for example, a temporal first-order auto-regressive structure can be specified by setting

$$\lambda_{jk} = \begin{cases} \lambda & \text{if } |j - k| = 1 \\ 0 & \text{otherwise} \end{cases}. \quad (5)$$

When $\mathbf{y} = (y_j \dots y_J)$ is a spatial series, a first-order spatial auto-regressive structure can be specified by associating each site j with a neighborhood $N(j) \subset \{1 \dots J\}$ and setting

$$\lambda_{jk} = \begin{cases} \lambda & \text{if } k \in N(j) \\ 0 & \text{otherwise} \end{cases}. \quad (6)$$

3 Markov Chain Monte Carlo maximum likelihood

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

$$l(\boldsymbol{\theta}) = \sum_{i=1}^n \log f_{\text{MVM}}(\mathbf{y}_i; \boldsymbol{\theta}). \quad (7)$$

Because the normalizing constant C is unavailable in closed form, direct maximization of the log-likelihood is unfeasible. MLEs can be alternatively found by maximizing a stochastic approximation of the log-likelihood, obtained by drawing M samples $\mathbf{y}_1, \dots, \mathbf{y}_M$ from the model under scrutiny, evaluated at a preliminary estimate $\tilde{\boldsymbol{\theta}}$ [4].

Exploiting the matrix operator $\text{vech}(\cdot)$ 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

$$\begin{aligned} f_{\text{MVM}}(\mathbf{y}; \boldsymbol{\gamma}, \Lambda) &= \frac{\exp(\mathbf{t}_1(\mathbf{y})^\top \boldsymbol{\kappa}(\boldsymbol{\gamma}) + \mathbf{t}_2(\mathbf{y})^\top \text{vech}(\Lambda))}{C(\boldsymbol{\gamma}, \Lambda)} \\ &= \frac{\varphi(\mathbf{y}; \boldsymbol{\gamma}, \Lambda)}{C(\boldsymbol{\gamma}, \Lambda)}, \end{aligned} \quad (8)$$

where

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

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

Let $\mathbf{y}_m, m = 1 \dots M$ be M samples, drawn from the centered MVM distribution (8), evaluated at the preliminary estimates (e.g., the MPLE) $\tilde{\boldsymbol{\gamma}}$ and $\tilde{\Lambda}$. On the basis of these samples, the ratio $C(\boldsymbol{\kappa}, \Lambda)/C(\tilde{\boldsymbol{\kappa}}, \tilde{\Lambda})$ can be approximated by a Monte Carlo mean, as follows

$$\begin{aligned}\frac{C(\boldsymbol{\kappa}, \Lambda)}{C(\tilde{\boldsymbol{\kappa}}, \tilde{\Lambda})} &= \int_{(-\pi, \pi)^J} \frac{\varphi(\mathbf{y}; \boldsymbol{\kappa}, \Lambda)}{\varphi(\mathbf{y}; \tilde{\boldsymbol{\kappa}}, \tilde{\Lambda})} f_{MVM}(\mathbf{y}; \tilde{\boldsymbol{\kappa}}, \tilde{\Lambda}) d\mathbf{y} = \mathbb{E} \left(\frac{\varphi(\mathbf{y}; \boldsymbol{\kappa}, \Lambda)}{\varphi(\mathbf{y}; \tilde{\boldsymbol{\kappa}}, \tilde{\Lambda})} \right) \\ &\approx \frac{1}{M} \sum_{m=1}^M \exp \left(\mathbf{t}_1^\top(\mathbf{y}_m) (\boldsymbol{\kappa}(\boldsymbol{\gamma}) - \boldsymbol{\kappa}(\tilde{\boldsymbol{\gamma}})) + \mathbf{t}_2^\top(\mathbf{y}_m) (\text{vech}(\Lambda) - \text{vech}(\tilde{\Lambda})) \right),\end{aligned}\tag{9}$$

where the expected value above is with respect to the centered distribution (8) and the approximation holds due the MC-likelihood approximation theory [4]. The stochastic approximation above allows approximating the log-likelihood function by the Monte Carlo log-likelihood

$$\begin{aligned}l_{MC}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda) &= \sum_{i=1}^n \mathbf{t}_1(\mathbf{y}_i - \boldsymbol{\mu}(\boldsymbol{\beta}))^\top \boldsymbol{\kappa}(\boldsymbol{\gamma}) + \mathbf{t}_2(\mathbf{y}_i - \boldsymbol{\mu}(\boldsymbol{\beta}))^\top \text{vech}(\Lambda) \\ &\quad - n \log \frac{1}{M} \sum_{m=1}^M \exp \left(\mathbf{t}_1^\top(\mathbf{y}_m) (\boldsymbol{\kappa}(\boldsymbol{\gamma}) - \boldsymbol{\kappa}(\tilde{\boldsymbol{\gamma}})) + \mathbf{t}_2^\top(\mathbf{y}_m) (\text{vech}(\Lambda) - \text{vech}(\tilde{\Lambda})) \right),\end{aligned}\tag{10}$$

and obtaining the MC-MLEs as the maximum point of l_{MC} . In obtaining the MLEs in the two examples below, the algorithm was implemented as follows. The Gibbs sampler was started taking the observed data as the starting state. A sample size $M = 2000$ was used, discarding the first 100 samples (burning period). The final estimate was obtained by a process of successive modification beginning with the MPLE, that is, a sequence of Monte Carlo samples was generated and each sample was exploited to suggest a new estimate to be used for the next sample. This process was iterated up to convergence, i.e. when the increase of two successive log-likelihoods fell below $10^4\%$.

4 Illustrative examples

The first example concerns a longitudinal study of escape directions of 65 sand-hoppers which were released sequentially on five occasions [8]. The study include

wind direction, azimuth direction for the sun (Sun), and an 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.

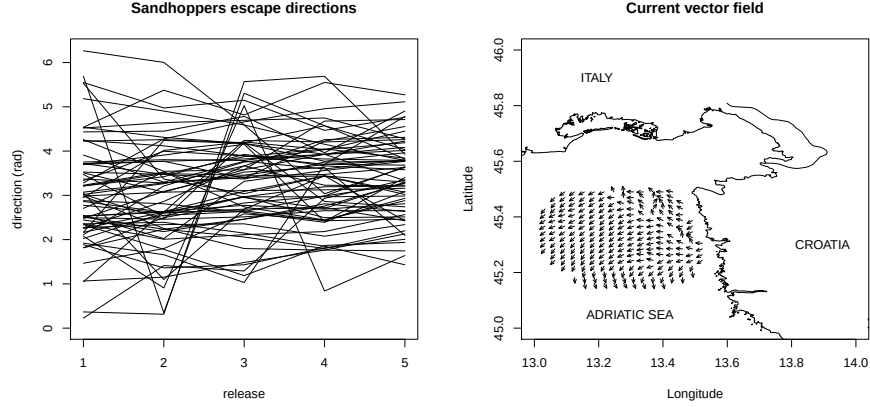


Fig. 1 Left: escape directions for sandhoppers over five consecutive releases. Right: Sea current directions in the Northern Adriatic sea on May 4th, 11 am, measured across a grid of 245 point.

Using the MVM distribution, these data can be easily explored without assuming a specific dependence structure on the data. The first two columns of Table 1 display the estimates and the standard errors obtained by a saturated model where 5 time-specific concentrations $\kappa_t, t = 1, \dots, 5$ and 10 conditional dependence parameters $\lambda_{t\tau}$ ($t = 1, \dots, 5, \tau > t$) are separately estimated along with the fixed effects β 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, at a 95% confidence level. In addition, the increasing concentration parameters seem indicating that the data are slightly heteroscedastic, a feature that was not considered in the previous analyses [8]. More interestingly, the estimated conditional dependence parameters reveal a specific first-order autoregressive structure of the outcomes: only the parameters λ_{ij} with $|i - j| = 1$ are significant, while the remaining parameters are not significant at the 95% level.

These results suggest a more parsimonious model with (1) concentrations that follow a parametric function of time of release, say $\kappa_t = \exp(\gamma_0 + \gamma_1 t)$, where $t = 1, \dots, 5$ is time of release, and (2) dependence parameters that follow the first-order auto-regressive structure (5). Columns 3 and 4 of Table 1 display the estimates and the standard errors obtained under these restrictions on the saturated model. Es-

Table 1 Estimates of three models for longitudinal circular data

	saturated model		heteroscedastic AR(1) model		homoscedastic AR(1) model	
	estimate	s.e.	estimate	s.e.	estimate	s.e.
Sun	0.003	0.001	0.003	0.001	0.003	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	

estimates and standard errors of the fixed effects are similar to those estimated under a saturated model. The significant slope γ_1 indicates 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 saturated model and this model, the latter seems a better compromise between parsimony and goodness of fit. The last columns of Table 1 indicate the results obtained by assuming equal concentrations, $\kappa_i = \kappa$ along with a first-order auto-regressive structure. This further constrain increases the AIC, favoring a model that allows for heteroscedasticity.

The second example is a case study of spatial angular measurements. Figure 1 displays the direction of surface 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 [1]. The figure includes the observations taken on May 4th at 11 am, across a grid of 245 points having a horizontal resolution of $2 \text{ km} \times 2 \text{ km}$. 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. A statistical approach to estimate variation at multiple spatial scale plays a crucial role in studies that involve semi-enclosed basins, such as the Adriatic Sea, where currents

Table 2 Estimates of three models for spatial circular data

	heteroscedastic circular AR(2) estimate s.e.		heteroscedastic circular AR(1) estimate s.e.		homoscedastic circular AR(1) 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			

are influenced by the local orographic conditions and numerical models, well suited for oceans, may provide inaccurate results. Orography effects clearly appear in Figure 1, where the outcome variability seems increasing as the grid points approach the coast.

Table 2 displays the results obtained by assuming that the data are a sample drawn from a MVM distribution with means

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

where x_{1j} and x_{2j} are respectively the longitude and the latitude of the j th observation point. This spatial gradient component was estimated along with three different dependence structures. Because the analysis is based on a single spatial series, the estimation of a saturated model is not possible and specific assumptions on the data dependence structure are unavoidable. The first model accounts for heteroscedasticity by assuming that the concentration at the j observation site is a parametric function of the minimum distance d_j from the coast, say

$$\kappa_j(\gamma_1, \gamma_2) = \exp(\gamma_0 + \gamma_1 d_j).$$

It additionally assumes a second-order spatial auto-regressive structure, by associating two auto-correlation parameters, λ_1 and λ_2 , with two disjoint neighborhood structures, 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. Matrix Λ is accordingly structured as follows:

$$\Lambda = \lambda_1 C_1 + \lambda_2 C_2,$$

where C_1 is a 245×245 connectivity matrix whose generic element $c_{jk}^{(1)}$ is equal to 1 if $j \in N_1(j)$ and 0 otherwise. Analogously, the generic element of C_2 , say $c_{jk}^{(2)}$, is equal to 1 if $j \in N_2(j)$ and 0 otherwise. The first two columns of Table 2 display the

estimates 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 is not. The residual of each outcome from the spatial gradient seems therefore conditionally independent on the second-order neighborhood, given the nearest neighbors.

These findings suggest that ignoring a second-order neighborhood structure should lead to a better compromise between model parsimony and goodness of fit. Columns 3 and 4 of Table 2 display the estimates and the standard errors obtained by estimating fixed effects and spatial concentration parameters along with the first-order matrix $\Lambda = \lambda C_1$. 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 2) 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.

References

1. Cosoli, S., Gacic, M., Mazzoldi, A.: Surface current variability and wind influence in the northeastern adriatic sea as observed from high-frequency (hf) radar measurements. *Continental Shelf Research*, **33**, 1-13 (2012)
2. Fisher, N.: *The Analysis of Circular Data*. Cambridge University Press, Cambridge (1993)
3. Fisher, N., Lee, A.J.: Regression models for an angular response. *Biometrics*, **48**, 665-677 (1992)
4. Geyer, C.J., Thompson, E.A.: Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society B*, **54**, 657-699 (1992)
5. Lagona F., Picone, M.: Maximum likelihood estimation of bivariate circular hidden Markov models from incomplete data. *Journal of Statistical Computation and Simulation*, **83**, 1223-1237 (2013)
6. Mardia, K.V., Voss, J.: Some fundamental properties of a multivariate von Mises distribution. *ArXiv e-prints* (2011)
7. Mardia, K.V., Hughes, G., Taylor, C.C., Singh, H.: A multivariate von Mises distribution with applications to bioinformatics. *Canadian Journal of Statistics*, **36**, 99-109 (2008)
8. Song, P.X.-K. *Correlated Data Analysis*. Springer, New York (2007)