Nonparametric estimation of circular trend surfaces with application to wave directions

Andrea Meilán-Vila · Rosa M. Crujeiras · Mario Francisco-Fernández

Abstract In oceanography, modeling wave fields requires the use of statistical tools capable of handling the circular nature of the data measurements. An important issue in ocean wave analysis is the study of height and direction waves, being direction values recorded as angles or, equivalently, as points on a unit circle. Hence, reconstruction of a wave direction field on the sea surface can be approached by the use of a linear-circular regression model, viewing wave directions as a realization of a circular spatial process whose trend should be estimated. In this paper, we consider a spatial regression model with a circular response and several real-valued predictors. Nonparametric estimators of the circular trend surface are proposed, accounting for the (unknown) spatial correlation. Some asymptotic results about these estimators as well as some guidelines for their practical implementation are also given. The performance of the proposed estimators is investigated in a simulation study. An application to wave directions in the Adriatic Sea is provided for illustration.

Andrea Meilán-Vila
Research group MODES, CITIC, Department of Mathematics, Faculty of Computer Science, Universidade da Coruña, Campus de Elviña s/n, 15071, A Coruña, Spain
Tel.: +34981167000 (ext 1301)
Fax: +34981167160
E-mail: andrea.meilan@udc.es

Rosa M. Crujeiras
Department of Statistics, Mathematical Analysis and Optimization, Faculty of Mathematics, Universidade de Santiago de Compostela, Rúa Lope Gómez de Marzoa s/n, 15782, Santiago de Compostela, Spain

Mario Francisco-Fernández
Research group MODES, CITIC, Department of Mathematics, Faculty of Computer Science, Universidade da Coruña, Campus de Elviña s/n, 15071, A Coruña, Spain

1 Introduction

In many scientific fields, such as oceanography, meteorology or biology, data are angular measurements (points on the circumference of the unit circle), exhibiting in some cases a spatial dependence structure which should be accounted for in any modeling approach. For instance, Casson and Coles (1998) provided a spatial analysis about the direction of maximum wind speed at locations on the Gulf and Atlantic coasts of the United States. On a series of simulated hurricane wind speeds, the authors aim to model the stochastic behavior of the extreme wind speeds jointly with their associated directions. In other scenarios, circular measurements are also accompanied by observations of real-valued random variables, as in García-Portugués et al. (2014), who analyzed the relation between orientation and size of wildfires in Portugal; or Mastrantonio et al. (2018), who proposed a Markov model for multivariate circular-linear data to forecast the wind speed and direction in the city of Taranto (Italy). Alternative approaches using copulas have been also considered in similar contexts. For instance, Carnicero et al. (2013), explored the relation between wind direction and rainfall amount in the North of Spain, as well as the dependence between the wind directions in two nearby buoys at the Atlantic ocean.

In certain situations, the circular data sample is geo-referenced, and the goal is to reconstruct the circular trend from a realization of a circular spatial process. This is the case in our motivating example, correspond-

Fig. 1 Random sample of 150 wave directions in the Adriatic Sea area on April 2, 2010 at 6am during a calm period.

ing to an application in oceanography. Wave directions are recorded in 1494 grid points on the Adriatic Sea area from a calm period transitioning to a storm period at different times. Fig. 1 shows a random sample of 150 observations during a calm period. As intuition suggests, these data seem to exhibit a spatial pattern. Considering other periods and moments, this dataset has been deeply studied by several authors using parametric methods. For example, Jona-Lasinio et al. (2012) analyzed outgoing wave directions from a storm period, formulating the wrapped Gaussian spatial process, as a spatial process for circular data. Mastrantonio et al. (2016) introduced the wrapped skew Gaussian process as an alternative to the wrapped Gaussian process which allows for asymmetric marginal distributions. This circular process was also used for analyzing wave directions. Wave directions from a calm period transitioning to a storm period were also modeled by Wang and Gelfand (2014). They developed the projected Gaussian spatial process, induced from a linear bivariate Gaussian spatial process. Motivated by the same real dataset, Lagona et al. (2015) introduced a hidden Markov model accounting for the correlation of spatio-temporal linear-circular data, providing an approach to identify regimes of marine currents.

An alternative to the previous approaches for modeling circular data at different spatial locations consists in the estimation of a circular spatial trend using smoothing methods. These techniques relax parametric assumptions of the generating process and, con-

sequently, provide a more flexible way to explore and model the data. This work aims to provide, using smoothing techniques, a circular trend surface estimation procedure. For this purpose, we consider a regression model with a circular response and an $\mathbb{R}^d$-valued covariate, assuming that the errors exhibit a spatial correlation structure. For a single real-valued covariate, Di Marzio et al. (2013) introduced a nonparametric estimator of the regression function when the errors are independent and also when the data come from mixing processes. The same approach has been also applied, with due modifications, in the context of time series by Di Marzio et al. (2012). The authors considered smoothing and prediction in the time domain for circular time-series data. Recently, Meilán-Vila et al. (2020b) proposed and studied nonparametric estimators of a circular regression function considering independent data and several real-valued covariates. In the present work, local polynomial-type regression estimators considering a model with a circular response and an $\mathbb{R}^d$-valued covariate, in the presence of (unknown) spatial correlation, are introduced and analyzed.

As in any smoothing approach, a crucial step in our proposal is the selection of an appropriate bandwidth or smoothing parameter (a symmetric $d \times d$ matrix, for an $\mathbb{R}^d$-valued covariate in our setting). This bandwidth matrix controls the shape and the size of the local neighborhood used for estimating the spatial trend, which directly impacts the smoothness of the estimator. In the Euclidean setting, considering a random design and independent data, Ruppert and Wand (1994) derived the asymptotic mean squared error (AMSE) for the multivariate local linear estimator. On the other hand, Opsomer et al. (2001) provided the corresponding results for the bidimensional ($d = 2$) case when the errors are correlated, while Liu (2001) generalized those results for an arbitrary dimension d. These error expressions allow to derive optimal bandwidths in the corresponding contexts. Following similar arguments, some guidelines to select locally optimal bandwidth matrices are given in this paper. Cross-validation (CV) bandwidth selection methods are also introduced and analyzed in practice. Different simulation scenarios are designed, considering circular spatial errors generated from wrapped and from projected Gaussian spatial processes.

This paper is organized as follows. Section 2 introduces the linear-circular regression model, with some highlights on the error process structure. Nonparametric estimators of the circular regression function are also proposed in this section. Section 3 contains some results about the asymptotic behavior of these estimators. Additionally, some proposals for bandwidth selection are introduced. A simulation study for assessing the perfor-

mance of the estimators and the bandwidth selectors is provided in Section 4. These simulations are carried out considering that the errors are drawn from wrapped (Section 4.2) and from projected (Section 4.3) Gaussian spatial processes. In Section 5, the application of the proposed approaches to estimate the wave direction trend surface in the Adriatic Sea is presented. Finally, Section 6 contains some conclusions, limitations of the proposal, and comments on further research.

2 Regression models and estimators

This section presents the regression models considered in this work and the proposed circular regression estimators. The linear-circular regression model for spatially correlated data is introduced in Section 2.1. For this model, nonparametric estimators of the circular regression function, based on considering two regression models for the sine and cosine components of the response variable, are proposed in Section 2.2. Specifically, Nadaraya–Watson (NW) and local linear (LL) type nonparametric estimators of the regression function are considered.

In what follows, $\boldsymbol{\nabla} g(\boldsymbol{x})$ and $\boldsymbol{\mathcal{H}}_g(\boldsymbol{x})$ will denote the vector of first-order partial derivatives and the Hessian matrix of a sufficiently smooth function g at $\boldsymbol{x}$, respectively. For a vector $\boldsymbol{u} = (u_1, \dots, u_d)^\top$ and an integrable function g, the integral $\int \cdots \int g(\boldsymbol{u}) \mathrm{d} u_1 \dots \mathrm{d} u_d$ will be simply denoted as $\int g(\boldsymbol{u}) \mathrm{d} \boldsymbol{u}$. Finally, for any matrix $\boldsymbol{A}$, $\boldsymbol{A}^\top$, $|\boldsymbol{A}|$, $\mathrm{tr}(\boldsymbol{A})$, $\lambda_{\max}(\boldsymbol{A})$ and $\lambda_{\min}(\boldsymbol{A})$ denote its transpose, determinant, trace, maximum eigenvalue and minimum eigenvalue, respectively.

2.1 A regression model with circular response

Let $\{(\boldsymbol{X}_i, \Theta_i)\}_{i=1}^n$ be a random sample from the $(d+1)$-valued random vector $(\boldsymbol{X}, \Theta)$, where Θ denotes a circular response, taking values on $\mathbb{T} = [0, 2\pi)$, which depends on a d-dimensional random variable $\boldsymbol{X}$, with density f and taking values in $D \subseteq \mathbb{R}^d$, through the following linear-circular regression model:

$$\Theta_i = [m(\boldsymbol{X}_i) + \varepsilon_i](\texttt{mod}\, 2\pi), \quad i = 1, \dots, n, \quad (1)$$

where m is a smooth trend or regression function, $\texttt{mod}$ stands for the modulo operation, and ε denotes a spatially correlated error process with zero mean direction. Taking into account the definition of the mean of a circular random variable, this condition is equivalent to $\mathbb{E}[\sin(\varepsilon) \mid \boldsymbol{X} = \boldsymbol{x}] = 0$. Additionally, we assume that

$$\mathbb{C}\mathrm{ov}[\sin(\varepsilon_i), \sin(\varepsilon_j) \mid \boldsymbol{X}_i, \boldsymbol{X}_j] = \sigma_1^2 \rho_{1,n}(\boldsymbol{X}_i - \boldsymbol{X}_j), \quad (2)$$

$$\mathbb{C}\mathrm{ov}[\cos(\varepsilon_i), \cos(\varepsilon_j) \mid \boldsymbol{X}_i, \boldsymbol{X}_j] = \sigma_2^2 \rho_{2,n}(\boldsymbol{X}_i - \boldsymbol{X}_j), \quad (3)$$

$$\mathbb{C}\mathrm{ov}[\sin(\varepsilon_i), \cos(\varepsilon_j) \mid \boldsymbol{X}_i, \boldsymbol{X}_j] = \sigma_{12} \rho_{3,n}(\boldsymbol{X}_i - \boldsymbol{X}_j), \quad (4)$$

with $\sigma_k^2 < \infty$, for $k = 1, 2$, and $\sigma_{12} < \infty$. The continuous stationary correlation functions $\rho_{k,n}$ satisfy $\rho_{k,n}(\boldsymbol{0}) = 1$, $\rho_{k,n}(\boldsymbol{x}) = \rho_{k,n}(-\boldsymbol{x})$, and $|\rho_{k,n}(\boldsymbol{x})| \leq 1$, for $\boldsymbol{x} \in D$, and $k = 1, 2, 3$. The subscript n in $\rho_{k,n}$ indicates that the correlation functions vary with n (specifically, the correlation functions shrink as n goes to infinity, as described below). Note also that the subscript k does not correspond to an integer sequence and it just indicates if the correlation corresponds to the sine process ($k = 1$), the cosine process ($k = 2$) or if it is the cross-correlation between them ($k = 3$).

2.2 Nonparametric regression estimators

The circular regression function m can be defined as the minimizer of the usual angular risk $\mathbb{E}\{1 - \cos[\Theta - m(\boldsymbol{x})] \mid \boldsymbol{X} = \boldsymbol{x}\}$. The solution of this optimization problem is given by:

$$m(\boldsymbol{x}) = \arctan2[m_1(\boldsymbol{x}), m_2(\boldsymbol{x})], \quad (5)$$

where $m_1(\boldsymbol{x}) = \mathbb{E}[\sin(\Theta) \mid \boldsymbol{X} = \boldsymbol{x}]$, $m_2(\boldsymbol{x}) = \mathbb{E}[\cos(\Theta) \mid \boldsymbol{X} = \boldsymbol{x}]$, and the function $\arctan2[y, x]$ returns the angle between the x-axis and the vector from the origin to (x, y). With this formulation, m_1 and m_2 can be regarded as the regression functions of two regression models having $\sin(\Theta)$ and $\cos(\Theta)$ as their responses, respectively. Specifically, we assume the models:

$$\sin(\Theta_i) = m_1(\boldsymbol{X}_i) + \xi_i \quad (6)$$

$$\cos(\Theta_i) = m_2(\boldsymbol{X}_i) + \zeta_i, \quad (7)$$

where the ξ_i and the ζ_i are error terms, absolutely bounded by 1, satisfying $\mathbb{E}(\xi \mid \boldsymbol{X} = \boldsymbol{x}) = \mathbb{E}(\zeta \mid \boldsymbol{X} = \boldsymbol{x}) = 0$. Additionally, for every $\boldsymbol{x} \in D$, set $s_1^2(\boldsymbol{x}) = \mathbb{V}\mathrm{ar}(\xi \mid \boldsymbol{X} = \boldsymbol{x})$, $s_2^2(\boldsymbol{x}) = \mathbb{V}\mathrm{ar}(\zeta \mid \boldsymbol{X} = \boldsymbol{x})$, $c(\boldsymbol{x}) = \mathbb{E}(\xi\zeta \mid \boldsymbol{X} = \boldsymbol{x})$, and taking into account that the errors in model (1) are spatially correlated, we use the notation $\mathbb{C}\mathrm{ov}(\xi_i, \xi_j \mid \boldsymbol{X}_i, \boldsymbol{X}_j) = C_{n,1}(\boldsymbol{X}_i, \boldsymbol{X}_j)$, $\mathbb{C}\mathrm{ov}(\zeta_i, \zeta_j \mid \boldsymbol{X}_i, \boldsymbol{X}_j) = C_{n,2}(\boldsymbol{X}_i, \boldsymbol{X}_j)$ and $\mathbb{C}\mathrm{ov}(\xi_i, \zeta_j \mid \boldsymbol{X}_i, \boldsymbol{X}_j) = C_{n,3}(\boldsymbol{X}_i, \boldsymbol{X}_j)$, for $i, j = 1, \dots, n$, and $i \neq j$.

A whole class of kernel-type estimators for $m(\boldsymbol{x})$ in (5) can be defined replacing in its expression the unknown functions $m_1(\boldsymbol{x})$ and $m_2(\boldsymbol{x})$ by suitable local polynomial estimators as follows:

$$\hat{m}_{\boldsymbol{H}}(\boldsymbol{x}; p) = \arctan2[\hat{m}_{1,\boldsymbol{H}}(\boldsymbol{x}; p), \hat{m}_{2,\boldsymbol{H}}(\boldsymbol{x}; p)], \quad (8)$$

where $\hat{m}_{1,\boldsymbol{H}}(\boldsymbol{x}; p)$ and $\hat{m}_{2,\boldsymbol{H}}(\boldsymbol{x}; p)$ denote the pth order local polynomial estimators (with bandwidth matrix $\boldsymbol{H}$) of $m_1(\boldsymbol{x})$ and $m_2(\boldsymbol{x})$, respectively (Ruppert and Wand, 1994; Liu, 2001).

Considering $p = 0$, the NW estimators of the regression functions m_j, $j = 1, 2$, at $\boldsymbol{x} \in D$, are respectively

defined as:

$$\hat{m}_{j,\boldsymbol{H}}(\boldsymbol{x};0) = \begin{cases} \dfrac{\sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{X}_i - \boldsymbol{x}) \sin(\Theta_i)}{\sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{X}_i - \boldsymbol{x})} \text{ if } j = 1, \\[4mm] \dfrac{\sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{X}_i - \boldsymbol{x}) \cos(\Theta_i)}{\sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{X}_i - \boldsymbol{x})} \text{ if } j = 2, \end{cases}$$

where, for $\boldsymbol{u} \in \mathbb{R}^d$, $K_{\boldsymbol{H}}(\boldsymbol{u}) = |\boldsymbol{H}|^{-1} K(\boldsymbol{H}^{-1}\boldsymbol{u})$ is the rescaled version of a d-variate kernel function K, and $\boldsymbol{H}$ is a $d \times d$ bandwidth matrix.

On the other hand, considering $p = 1$, the LL estimators for the regression functions m_j, $j = 1, 2$, at a given point $\boldsymbol{x} \in D$, are given by:

$$\hat{m}_{j,\boldsymbol{H}}(\boldsymbol{x};1) = \begin{cases} \boldsymbol{e}_1^\top (\boldsymbol{X}_{\boldsymbol{x}}^\top \boldsymbol{W}_{\boldsymbol{x}} \boldsymbol{X}_{\boldsymbol{x}})^{-1} \boldsymbol{X}_{\boldsymbol{x}}^\top \boldsymbol{W}_{\boldsymbol{x}} \boldsymbol{S} \text{ if } j = 1, \\[3mm] \boldsymbol{e}_1^\top (\boldsymbol{X}_{\boldsymbol{x}}^\top \boldsymbol{W}_{\boldsymbol{x}} \boldsymbol{X}_{\boldsymbol{x}})^{-1} \boldsymbol{X}_{\boldsymbol{x}}^\top \boldsymbol{W}_{\boldsymbol{x}} \boldsymbol{C} \text{ if } j = 2, \end{cases}$$

where $\boldsymbol{e}_1$ is a $(d+1) \times 1$ vector having 1 in the first entry and 0 in all other entries, $\boldsymbol{X}_{\boldsymbol{x}}$ is a $n \times (d+1)$ matrix having $(1, (\boldsymbol{X}_i - \boldsymbol{x})^\top)$ as its ith row, $\boldsymbol{W}_{\boldsymbol{x}} = \mathrm{diag}\{K_{\boldsymbol{H}}(\boldsymbol{X}_1 - \boldsymbol{x}), \ldots, K_{\boldsymbol{H}}(\boldsymbol{X}_n - \boldsymbol{x})\}$, $\boldsymbol{S} = (\sin(\Theta_1), \ldots, \sin(\Theta_n))^\top$ and $\boldsymbol{C} = (\cos(\Theta_1), \ldots, \cos(\Theta_n))^\top$.

Meilán-Vila et al. (2020b) derived the AMSE of the circular regression estimator given in (8), for independent data. In the present work, this estimator is studied for spatially correlated data, for $p = 0$ (corresponding to a NW-type estimator) and $p = 1$ (providing a LL-type estimator).

3 Theoretical results

Some asymptotic conditional properties of the estimator (8), with polynomial degrees $p = 0$ and $p = 1$, are derived in Section 3.1. An asymptotically optimal local bandwidth matrix for $\hat{m}_{\boldsymbol{H}}(\boldsymbol{x};p)$, with $p = 0$ and $p = 1$, is also provided in this section, whereas a suitably adapted cross-validation criterion is proposed in Section 3.2.

3.1 Asymptotic results

Asymptotic properties of $\hat{m}_{j,\boldsymbol{H}}(\boldsymbol{x};p)$, for $j = 1, 2$ and $p = 0, 1$, can be obtained using some results given in Liu (2001). The following assumptions on the design, the kernel function and the bandwidth matrix are needed to derive the asymptotic bias, the asymptotic variance of estimator $\hat{m}_{j,\boldsymbol{H}}(\boldsymbol{x};p)$, $j = 1, 2$, and the asymptotic covariance between $\hat{m}_{1,\boldsymbol{H}}(\boldsymbol{x};p)$ and $\hat{m}_{2,\boldsymbol{H}}(\boldsymbol{x};p)$, as well the asymptotic bias and variance of $\hat{m}_{\boldsymbol{H}}(\boldsymbol{x};p)$, $p = 0, 1$.

(A1) The design density f is continuously differentiable at $\boldsymbol{x} \in D$, and satisfies $f(\boldsymbol{x}) > 0$. Moreover, $s_j^2(\boldsymbol{x}) > 0$, and s_j^2 and all second-order derivatives of the regression functions m_j, for $j = 1, 2$, are continuous at $\boldsymbol{x}$.

(A2) The kernel K is a spherically symmetric density function, twice continuously differentiable, with compact support (for simplicity with a nonzero value only if $\|\boldsymbol{u}\| \leq 1$). Moreover, $\int \boldsymbol{u}\boldsymbol{u}^\top K(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = \mu_2 \boldsymbol{I}_d$, where $\mu_2 = \int u_i^2 K(\boldsymbol{u})\mathrm{d}\boldsymbol{u} \neq 0$, for all $i = 1, \ldots, d$, and $\boldsymbol{I}_d$ denotes the $d \times d$ identity matrix. Then, μ_2 is the second-order moment of the multivariate kernel K. It is also assumed that $\nu_0 = \int K^2(\boldsymbol{u})\mathrm{d}\boldsymbol{u} < \infty$.

(A3) K is Lipschitz continuous. That is, there exists a constant $\mathfrak{L} > 0$, such that,

$$|K(\boldsymbol{X}_1) - K(\boldsymbol{X}_2)| \leq \mathfrak{L}\|\boldsymbol{X}_1 - \boldsymbol{X}_2\|, \quad \forall \boldsymbol{X}_1, \boldsymbol{X}_2 \in D.$$

(A4) The bandwidth matrix $\boldsymbol{H}$ is symmetric and positive definite, with $\boldsymbol{H} \to 0$ and $n|\boldsymbol{H}|\lambda_{\min}^d(\boldsymbol{H}) \to \infty$, when $n \to \infty$. The ratio $\lambda_{\max}(\boldsymbol{H})/\lambda_{\min}(\boldsymbol{H})$ is bounded above

(A5) For the correlation functions $\rho_{k,n}$, $k = 1, 2, 3$, in (2), (3) and (4), respectively, there exist constants ρ_{M_k} and ρ_{c_k}, such that, $n \int |\rho_{k,n}(\boldsymbol{x})|\mathrm{d}\boldsymbol{x} < \rho_{\mathrm{M}_k}$ and $\lim_{n \to \infty} n \int \rho_{k,n}(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \rho_{\mathrm{c}_k}$. Moreover, for any sequence $\epsilon_n > 0$ satisfying $n^{1/2}\epsilon_n \to \infty$,

$$n \int_{\|\boldsymbol{x}\| \geq \epsilon_n} |\rho_{k,n}(\boldsymbol{x})|\mathrm{d}\boldsymbol{x} \to 0 \quad \text{as} \quad n \to \infty.$$

In assumption (A2), we are supposing that K is a kernel of order 2. That means that $\int \boldsymbol{u}K(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = 0$ (which is obviously satisfied by the symmetry of the kernel) and $\mu_2 \neq 0$ (see, for instance, Ruppert and Wand, 1994; Chacón and Duong, 2018). These conditions are satisfied by the most frequent kernels used in kernel-type estimators, such as the Gaussian or the multiplicative triweight kernel, which is the one used in our simulation and real data analysis in Sections 4 and 5, respectively. In Duong (2015), the values of μ_2 for different spherically symmetric kernels and for different dimensions d are provided. For example, in the case of $d = 2$, $\mu_2 = 1$ for the Gaussian kernel, and $\mu_2 = 1/10$ for the triweight kernel. In assumption (A4), $\boldsymbol{H} \to 0$ means that every entry of $\boldsymbol{H}$ goes to 0. This condition is equivalent to $\lambda_{\max}(\boldsymbol{H}) \to 0$ due to the symmetry and positive definiteness of $\boldsymbol{H}$. Further, $|\boldsymbol{H}| = O[\lambda_{\max}^d(\boldsymbol{H})]$, because $|\boldsymbol{H}|$ is equal to the product of all eigenvalues of $\boldsymbol{H}$. Assumption (A5) implies that the correlation functions $\rho_{k,n}$, for $k = 1, 2, 3$, depend on n, and the integrals $\int |\rho_{k,n}(\boldsymbol{x})|\mathrm{d}\boldsymbol{x}$, $k = 1, 2, 3$, should vanish as $n \to \infty$. The vanishing speed should not be slower than $O(n^{-1})$. For all $k = 1, 2, 3$, this assumption also entails that the

integrals of $|\rho_{k,n}(\boldsymbol{x})|$ are essentially dominated by the values of $\rho_{k,n}(\boldsymbol{x})$ near the origin $\boldsymbol{0}$. Hence, the correlations are assumed to be short-ranged. This means that they decrease rapidly when the distance between two observations increases, as $n \to \infty$. Two examples of correlation functions that satisfy the conditions of assumption (A5) are the exponential model

$$\rho_n(\boldsymbol{x}) = \exp(-an\|\boldsymbol{x}\|),$$

and the rational quadratic model

$$\rho_n(\boldsymbol{x}) = \frac{1}{1 + a(n\|\boldsymbol{x}\|)^2},$$

a being a positive constant in both cases (see Cressie, 1993).

The asymptotic conditional bias and variance of the circular regression estimator $\hat{m}_{\boldsymbol{H}}(\boldsymbol{x}; p)$, for $p = 0, 1$, given in (8), can be derived by using the asymptotic bias and variance of estimators $\hat{m}_{j,\boldsymbol{H}}(\boldsymbol{x}; p)$, $j = 1, 2$, as well as the asymptotic covariance between $\hat{m}_{1,\boldsymbol{H}}(\boldsymbol{x}; p)$ and $\hat{m}_{2,\boldsymbol{H}}(\boldsymbol{x}; p)$, for $p = 0, 1$. The asymptotic bias of estimators $\hat{m}_{j,\boldsymbol{H}}(\boldsymbol{x}; p)$, $j = 1, 2$, assuming models (6) and (7), is the same as that obtained considering independent data, derived in Härdle and Müller (2012) and Ruppert and Wand (1994), for $p = 0$ and $p = 1$, respectively. On the other hand, their asymptotic variances and the covariances between $\hat{m}_{1,\boldsymbol{H}}(\boldsymbol{x}; p)$ and $\hat{m}_{2,\boldsymbol{H}}(\boldsymbol{x}; p)$, for $p = 0, 1$, (considering assumption (A5)) can be deduced from the results obtained in Liu (2001). These expressions are given in the final Appendix. Note that although assumption (A5) establishes conditions on the correlations in (2), (3) and (4), for the sine and cosine process, and the cross-correlation, directly derived from model (1), using the sine and cosine addition formulas, it is straightforward to obtain equations relating these covariances with those coming from models (6) and (7). For further details on the relation between $C_{n,3}$ and $\rho_{k,n}$, $k = 1, 2, 3$, in (2), (3) and (4), see the final Appendix. Similar equations can be also obtained for $C_{n,1}$ and $C_{n,2}$. As in the Euclidean setting, the asymptotic bias of $\hat{m}_{\boldsymbol{H}}(\boldsymbol{x}; p)$, $p = 0, 1$, is the same for dependent and for independent data (Meilán-Vila et al., 2020b). However, the asymptotic conditional variance of these estimators depends on the spatial correlation. Considering an interior point in the support of f, the NW- and LL-type estimators of m have the same asymptotic variance. The following theorem provides this result. Its proof is included in the Appendix.

Theorem 1 *Given a sample $\{(\boldsymbol{X}_i, \Theta_i)\}_{i=1}^{n}$ on $D \times \mathbb{T}$ generated from model* (1) *and assuming that conditions* (A1)–(A5) *hold, the asymptotic conditional variance of*

the estimator $\hat{m}_{\boldsymbol{H}}(\boldsymbol{x}; p)$, for $p = 0, 1$, at a fixed interior point $\boldsymbol{x}$ in the support of f, is given by:

$$\mathbb{V}\mathrm{ar}[\hat{m}_{\boldsymbol{H}}(\boldsymbol{x}; p) \mid \boldsymbol{X}_1, \dots, \boldsymbol{X}_n] = \frac{\nu_0 \sigma_1^2 [1 + f(\boldsymbol{x})\rho_{c_1}]}{n|\boldsymbol{H}|\ell^2(\boldsymbol{x})f(\boldsymbol{x})}$$
$$+ o_{\mathbb{P}}\left(\frac{1}{n|\boldsymbol{H}|}\right), \quad (9)$$

where $\ell(\boldsymbol{x}) = \mathbb{E}[\cos(\varepsilon) \mid \boldsymbol{X} = \boldsymbol{x}]$.

Remark 1 Notice that the expression of the asymptotic conditional variance of estimator $\hat{m}_{\boldsymbol{H}}(\boldsymbol{x}; p)$, for $p = 0, 1$, has a similar structure to that obtained for the NW and LL estimators in a regression model with Euclidean response and spatially correlated errors. For independent data, it follows that $\rho_{c_1} = 0$ in Theorem 1, and, consequently, the asymptotic conditional variance of both estimators coincides with the expression obtained for independent data in Meilán-Vila et al. (2020b).

The AMSE of $\hat{m}_{\boldsymbol{H}}(\boldsymbol{x}; 0)$, defined as the sum of the square of the leading term of the bias (Meilán-Vila et al., 2020b) and the leading term of the variance (9), is given by:

$$\mathrm{AMSE}[\hat{m}_{\boldsymbol{H}}(\boldsymbol{x}; 0)] = \left\{\frac{1}{2}\mu_2 \mathrm{tr}[\boldsymbol{H}^2 \boldsymbol{\mathcal{H}}_m(\boldsymbol{x})]\right.$$
$$\left. + \frac{\mu_2}{\ell(\boldsymbol{x})f(\boldsymbol{x})}\boldsymbol{\nabla}^\top m(\boldsymbol{x})\boldsymbol{H}^2 \boldsymbol{\nabla}(\ell f)(\boldsymbol{x})\right\}^2$$
$$+ \frac{\nu_0 \sigma_1^2 [1 + f(\boldsymbol{x})\rho_{c_1}]}{n|\boldsymbol{H}|\ell^2(\boldsymbol{x})f(\boldsymbol{x})}$$
$$= \frac{1}{4}\mu_2^2 \mathrm{tr}^2[\boldsymbol{H}^2 \mathcal{B}(\boldsymbol{x})]$$
$$+ \frac{\nu_0 \sigma_1^2 [1 + f(\boldsymbol{x})\rho_{c_1}]}{n|\boldsymbol{H}|\ell^2(\boldsymbol{x})f(\boldsymbol{x})}, \quad (10)$$

with

$$\mathcal{B}(\boldsymbol{x}) = \frac{1}{\ell(\boldsymbol{x})f(\boldsymbol{x})}[\boldsymbol{\nabla}(\ell f)(\boldsymbol{x})\boldsymbol{\nabla}^\top m(\boldsymbol{x})$$
$$+ \boldsymbol{\nabla} m(\boldsymbol{x})\boldsymbol{\nabla}^\top(\ell f)(\boldsymbol{x})] + \boldsymbol{\mathcal{H}}_m(\boldsymbol{x}).$$

An asymptotically optimal local bandwidth matrix, $\boldsymbol{H}_{\mathrm{opt}}(\boldsymbol{x}; 0)$, for $\hat{m}_{\boldsymbol{H}}(\boldsymbol{x}; 0)$ can be directly derived minimizing equation (10) with respect to $\boldsymbol{H}$. Using Proposition 2.6 of Liu (2001), it can be obtained that this optimal local bandwidth is:

$$\boldsymbol{H}_{\mathrm{opt}}(\boldsymbol{x}; 0) = h^*(\boldsymbol{x}) \cdot [\tilde{\mathcal{B}}(\boldsymbol{x})]^{-1/2}, \quad (11)$$

where

$$h^*(\boldsymbol{x}) = \left\{\frac{\nu_0 |\tilde{\mathcal{B}}(\boldsymbol{x})|^{1/2}\sigma_1^2[1 + f(\boldsymbol{x})\rho_{c_1}]}{nd\mu_2^2 \ell^2(\boldsymbol{x})f(\boldsymbol{x})}\right\}^{1/d+4},$$

and

$$\tilde{\mathcal{B}}(\boldsymbol{x}) = \begin{cases} \mathcal{B}(\boldsymbol{x}) & \text{if } \mathcal{B}(\boldsymbol{x}) \text{ is positive definite,} \\ -\mathcal{B}(\boldsymbol{x}) & \text{if } \mathcal{B}(\boldsymbol{x}) \text{ is negative definite.} \end{cases}$$

The matrix $\tilde{\mathcal{B}}(\boldsymbol{x})$ determines the shape and the orientation in the d-dimensional space of the covariate region which is used to compute the local estimates. Such data regions are ellipsoids in $\mathbb{R}^d$, being the magnitude of the axes controlled by $\tilde{\mathcal{B}}(\boldsymbol{x})$. Similarly, an asymptotically optimal local bandwidth can be also obtained for the LL-type estimator. In this case, the AMSE of $\hat{m}_{\boldsymbol{H}}(\boldsymbol{x};1)$ is given by:

$$\text{AMSE}[\hat{m}_{\boldsymbol{H}}(\boldsymbol{x};1)] = \frac{1}{4}\mu_2^2\text{tr}^2[\boldsymbol{H}^2\mathcal{G}(\boldsymbol{x})]$$
$$+ \frac{\nu_0\sigma_1^2[1+f(\boldsymbol{x})\rho_{c_1}]}{n|\boldsymbol{H}|\ell^2(\boldsymbol{x})f(\boldsymbol{x})},$$

with $\mathcal{G}(\boldsymbol{x}) = \ell^{-1}(\boldsymbol{x})[\boldsymbol{\nabla}\ell(\boldsymbol{x})\boldsymbol{\nabla}^\top m(\boldsymbol{x}) + \boldsymbol{\nabla}m(\boldsymbol{x})\boldsymbol{\nabla}^\top\ell(\boldsymbol{x})] + \mathcal{H}_m(\boldsymbol{x})$. Consequently, the bandwidth matrix which minimizes this expression coincides with (11), but using $\mathcal{G}(\boldsymbol{x})$ instead of $\mathcal{B}(\boldsymbol{x})$.

Local bandwidth matrices may be useful for estimating the trend at a given point $\boldsymbol{x} \in D$, however, the nonparametric estimators computed with them may not be accurate enough for reconstructing the whole trend. An asymptotically optimal global bandwidth matrix can be selected minimizing the asymptotic mean integrated squared error (AMISE). Unfortunately, there is not a closed form solution for this optimization problem. Moreover, optimal bandwidth matrices, depending on unknown quantities, cannot be used for practical purposes. Practical bandwidth selection techniques, based on cross-validation methods, are considered in what follows.

3.2 Cross-validation bandwidth selection methods

A first approach to select the bandwidth $\boldsymbol{H}$ for $\hat{m}_{\boldsymbol{H}}(\boldsymbol{x};p)$, $p = 0,1$, consists on minimizing the cross-validation function:

$$\text{CV}(\boldsymbol{H}) = \sum_{i=1}^{n}\left\{1 - \cos\left[\Theta_i - \hat{m}_{\boldsymbol{H}}^{(i)}(\boldsymbol{X}_i;p)\right]\right\},$$

where $\hat{m}_{\boldsymbol{H}}^{(i)}(\boldsymbol{X}_i;p)$ is the estimator computed using all observations except $(\boldsymbol{X}_i,\Theta_i)$ and evaluated at $\boldsymbol{X}_i$. The CV criterion, as well as other smoothing parameter selection methods in nonparametric regression, should not be directly used for selecting the bandwidth when working with dependent data, given that its expectation is severely affected by the correlation (Opsomer et al., 2001; Liu, 2001).

In our setting, the CV criterion should be modified in order to account for the effect of the spatial correlation. With this issue in mind, we propose a modified cross-validation (MCV) criterion, which selects the bandwidth matrix $\boldsymbol{H}$ that minimizes the function:

$$\text{MCV}(\boldsymbol{H}) = \sum_{i=1}^{n}\left\{1 - \cos\left[\Theta_i - \hat{m}_{\boldsymbol{H}}^{N(i)}(\boldsymbol{X}_i;p)\right]\right\},$$

where $\hat{m}_{\boldsymbol{H}}^{N(i)}(\boldsymbol{X}_i;p)$ denotes the estimator computed using all observations except those located within a neighborhood of $\boldsymbol{X}_i$, $N(i)$, and evaluated at $\boldsymbol{X}_i$. For applying this criterion, the size of the neighborhood $N(i)$ must be selected. For simplicity, we consider the MCV criterion when $N(i) = \{\boldsymbol{X}_j : \|\boldsymbol{X}_j - \boldsymbol{X}_i\| \le l\}$. For $d = 2$, the neighborhood $N(i)$ consists of observations within the circle centered at $\boldsymbol{X}_i$ and radius l. If there is a strong spatial correlation, more observations should be omitted in the bandwidth selection procedure, and consequently, the value of l for constructing $N(i)$ should be larger. The use of the CV and MCV criteria to select the bandwidth matrix is explored through simulations in the following section.

4 Simulation study

In this section, the performance of the proposed estimators and the cross-validation bandwidth selection criteria are analyzed in a simulation study for $d = 2$. Section 4.1 contains a detailed description of the different simulation scenarios. In Sections 4.2 and 4.3, we briefly describe the procedures employed to generate wrapped and projected Gaussian spatial errors in model (1), respectively. The practical performance of the proposed circular regression estimator in (8) is analyzed considering these previous error simulation approaches.

4.1 General aspects

Assuming regression model (1), 500 samples of size n ($n = 100, 225, 400$) are generated, considering the spatial locations $\boldsymbol{X} = (X_1, X_2)$ on a bidimensional regular grid in the unit square. Two different regression functions (shown in the left panels of Fig. 4) are considered:

$$r_1 : m(\boldsymbol{X}) = \arctan2(6X_1^5 - 2X_1^3 - 1, -2X_2^5 - 3X_2 - 1),$$
$$r_2 : m(\boldsymbol{X}) = \arccos(X_1^5 - 1) + \frac{3}{2}\arcsin(X_2^3 - X_2 + 1).$$

Circular spatially correlated errors in model (1), for both regression functions, are generated from wrapped (Jona-Lasinio et al., 2012) and from projected Gaussian spatial processes (Wang and Gelfand, 2014). For each sample, the NW- and LL-type estimators of the circular regression function, given in (8), are computed. In both cases, a multiplicative triweight kernel is considered, while the bandwidth matrix $\boldsymbol{H}$ is selected by using the CV and MCV criteria. Different values of
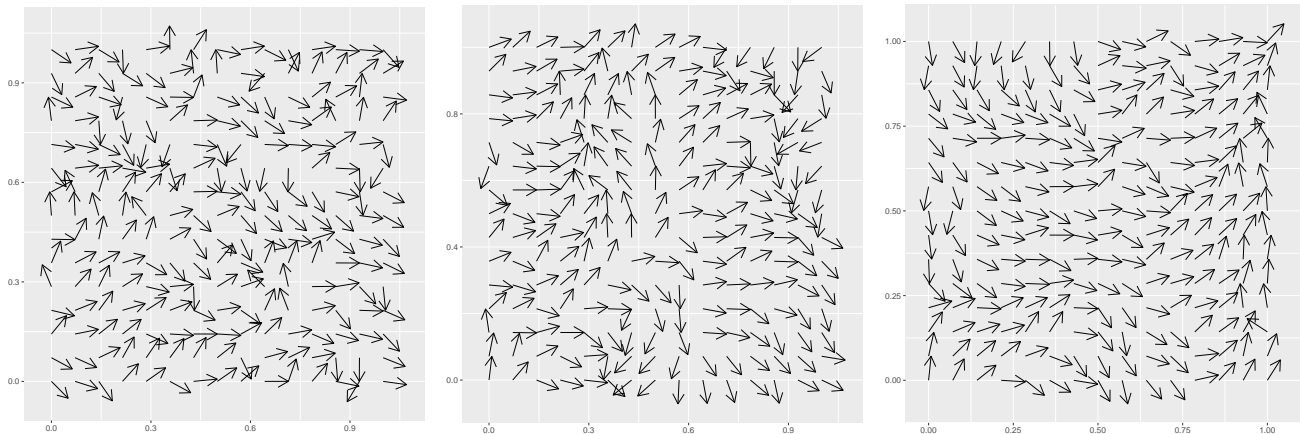
Fig. 2 Simulated samples of a wrapped Gaussian spatial process on a 15×15 grid with exponential correlation, being $a_{\mathrm{e}} = 0.1$ (left), $a_{\mathrm{e}} = 0.3$ (center) and $a_{\mathrm{e}} = 0.6$ (right), for $\mu = 0$ and $\sigma = 1$ in (13) and (14).

Table 1 Results obtained when the errors in model (1) are simulated from wrapped Gaussian spatial processes. Average (over 500 replicates) of the CASE given in (12), for the regression function r_1, using NW (left) and LL (right) type estimators. Bandwidth matrix is selected by minimizing CV ($\boldsymbol{H}_{\mathrm{CV}}$), MCV ($\boldsymbol{H}_{\mathrm{MCV}_1}$, $\boldsymbol{H}_{\mathrm{MCV}_2}$, $\boldsymbol{H}_{\mathrm{MCV}_3}$) and CASE ($\boldsymbol{H}_{\mathrm{CASE}}$) as a benchmark.

		NW					LL				
a_{e}	n	$\boldsymbol{H}_{\mathrm{CV}}$	$\boldsymbol{H}_{\mathrm{MCV}_1}$	$\boldsymbol{H}_{\mathrm{MCV}_2}$	$\boldsymbol{H}_{\mathrm{MCV}_3}$	$\boldsymbol{H}_{\mathrm{CASE}}$	$\boldsymbol{H}_{\mathrm{CV}}$	$\boldsymbol{H}_{\mathrm{MCV}_1}$	$\boldsymbol{H}_{\mathrm{MCV}_2}$	$\boldsymbol{H}_{\mathrm{MCV}_3}$	$\boldsymbol{H}_{\mathrm{CASE}}$
0.1	100	0.2087	0.0902	0.0721	0.0609	0.0387	0.1818	0.0933	0.0839	0.0672	0.0550
	225	0.2880	0.1291	0.0768	0.0602	0.0365	0.2649	0.1190	0.0782	0.0667	0.0532
	400	0.2932	0.1195	0.0702	0.0585	0.0359	0.2920	0.1114	0.0771	0.0667	0.0518
0.3	100	0.2852	0.1752	0.1342	0.0803	0.0529	0.2499	0.1651	0.1474	0.1143	0.1062
	225	0.3080	0.2054	0.1500	0.0788	0.0520	0.2979	0.2026	0.1546	0.1196	0.1053
	400	0.2764	0.1967	0.1351	0.0778	0.0497	0.2785	0.1965	0.1495	0.1168	0.1019
0.6	100	0.2316	0.1591	0.1316	0.0806	0.0677	0.2117	0.1520	0.1392	0.1146	0.1097
	225	0.2417	0.1775	0.1455	0.0798	0.0620	0.2361	0.1783	0.1488	0.1212	0.1093
	400	0.2177	0.1701	0.1325	0.0778	0.0569	0.2192	0.1725	0.1413	0.1171	0.1074

Table 2 Results obtained when the errors in model (1) are simulated from wrapped Gaussian spatial processes. Average (over 500 replicates) of the CASE given in (12), for the regression function r_2, using NW (left) and LL (right) type estimators. Bandwidth matrix is selected by minimizing CV ($\boldsymbol{H}_{\mathrm{CV}}$), MCV ($\boldsymbol{H}_{\mathrm{MCV}_1}$, $\boldsymbol{H}_{\mathrm{MCV}_2}$, $\boldsymbol{H}_{\mathrm{MCV}_3}$) and CASE ($\boldsymbol{H}_{\mathrm{CASE}}$) as a benchmark.

		NW					LL				
a_{e}	n	$\boldsymbol{H}_{\mathrm{CV}}$	$\boldsymbol{H}_{\mathrm{MCV}_1}$	$\boldsymbol{H}_{\mathrm{MCV}_2}$	$\boldsymbol{H}_{\mathrm{MCV}_3}$	$\boldsymbol{H}_{\mathrm{CASE}}$	$\boldsymbol{H}_{\mathrm{CV}}$	$\boldsymbol{H}_{\mathrm{MCV}_1}$	$\boldsymbol{H}_{\mathrm{MCV}_2}$	$\boldsymbol{H}_{\mathrm{MCV}_3}$	$\boldsymbol{H}_{\mathrm{CASE}}$
0.1	100	0.2231	0.1331	0.1233	0.1282	0.0871	0.1804	0.1163	0.1092	0.1114	0.0781
	225	0.3044	0.1475	0.1120	0.1122	0.0792	0.2574	0.1264	0.0976	0.0980	0.0704
	400	0.2941	0.1369	0.1008	0.1051	0.0748	0.2916	0.1160	0.0886	0.0973	0.0668
0.3	100	0.1927	0.1740	0.1622	0.1531	0.1277	0.1991	0.1688	0.1627	0.1526	0.1299
	225	0.2053	0.1879	0.1620	0.1519	0.1030	0.2090	0.1869	0.1563	0.1518	0.1277
	400	0.2154	0.1936	0.1529	0.1414	0.1017	0.2189	0.1875	0.1505	0.1432	0.1208
0.6	100	0.2392	0.1704	0.1605	0.1561	0.1440	0.2072	0.1605	0.1602	0.1597	0.1348
	225	0.1891	0.1765	0.1729	0.1553	0.1196	0.1903	0.1717	0.1669	0.1594	0.1343
	400	0.1947	0.1687	0.1624	0.1467	0.1063	0.1962	0.1653	0.1608	0.1509	0.1227

the radius l are considered in the MCV method. Given that the covariates are located in unit square, we set $l(b) = \sqrt{2}b/10$, where $b = 0$ would correspond with the CV method and $b = 10$ would provide the maximum distance between two points in the unit square. After some tests, only three values of b ($b = 1, 2, 3$) are considered. The corresponding CV and MCV bandwidths are denoted by $\boldsymbol{H}_{\mathrm{CV}}$ and $\boldsymbol{H}_{\mathrm{MCV}_b}$, $b = 1, 2, 3$, respec-

tively. Taking into account the structure of the regression functions and in order to save computational time, the bandwidth matrix is restricted to be diagonal with possibly different elements. The performance of the estimators and the bandwidth selectors is evaluated using the circular average squared error (CASE), defined as

$$\mathrm{CASE}(\boldsymbol{H}) = \frac{1}{n} \sum_{i=1}^{n} \left\{ 1 - \cos\left[m(\boldsymbol{X}_i) - \hat{m}_{\boldsymbol{H}}(\boldsymbol{X}_i; p)\right] \right\}, \quad (12)$$

for $p = 0, 1$, as a comparative error measure (Kim and SenGupta, 2017). Additionally, the diagonal optimal bandwidth matrix $\boldsymbol{H}_{\mathrm{CASE}}$ minimizing (12), obtained by intensive search, is also computed. Note that this bandwidth matrix can not be used in a practical situation where the true regression is unknown. For this reason, it can not be considered as a criterion to select the bandwidth, but it is used to get a benchmark value for comparison.

The computing time for running the whole procedure (simulate a sample, select the bandwidth matrix, compute the circular nonparametric estimator and evaluate the CASE) for just one of the 500 samples of size of 225 is around 2 seconds, no matter the bandwidth matrix selection method employed and regardless of the estimator (NW or LL) used. However, it should be noted that the computing time for obtaining a bandwidth matrix with the MCV criterion increases with b. In addition, the projection approach for the circular errors generation seems to be slightly more computationally expensive than the wrapping method.

4.2 The wrapping approach

Given a collection of spatial coordinates, $\boldsymbol{X}_i$, with $i = 1, \dots, n$, a realization of a spatial circular (error) process $\{\varepsilon_i,\ i = 1, \dots, n\}$ can be obtained using the wrapping method, introduced by Jona-Lasinio et al. (2012), for Gaussian spatial processes. In general, a wrapping approach consists on wrapping a linear variable around the unit circle. In this case, its circular density function is easily obtained by wrapping the density function of the linear random variable.

So, consider a realization $\{Y_i = Y(\boldsymbol{X}_i),\ i = 1, \dots, n\}$ from a real-valued Gaussian process, where each observation can be decomposed as:

$$Y_i = \mu + w_i, \quad i = 1, \dots, n, \tag{13}$$

being $\mu = \mu(\boldsymbol{X}_i)$ the mean and w a zero mean Gaussian spatial process with $\mathbb{Cov}(w_i, w_j \mid \boldsymbol{X}_i, \boldsymbol{X}_j) = \sigma^2 \rho_n(\boldsymbol{X}_i - \boldsymbol{X}_j)$. The variance of w is denoted by σ^2 and ρ_n is a continuous stationary correlation function satisfying $\rho_n(\boldsymbol{0}) = 1$, $\rho_n(\boldsymbol{x}) = \rho_n(-\boldsymbol{x})$, and $|\rho_n(\boldsymbol{x})| \leq 1$, $\forall \boldsymbol{x}$. Then, a realization of a wrapped Gaussian spatial process $\{\varepsilon_i,\ i = 1, \dots, n\}$, linked to the spatial coordinates $\boldsymbol{X}_i$, with $i = 1, \dots, n$, is obtained as:

$$\varepsilon_i = Y_i (\mathtt{mod}\, 2\pi), \quad i = 1, \dots, n.$$

Note that this realization can be written in vector form as $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^\top$, with mean direction vector $\mu \mathbf{1}_n$, being $\mathbf{1}_n$ a $n \times 1$ vector with every entry equal to 1, and covariance matrix $\sigma^2 \boldsymbol{R}_n$, where

$\boldsymbol{R}_n(i, j) = \rho_n(\boldsymbol{X}_i - \boldsymbol{X}_j)$ is the (i, j)-entry of the correlation matrix $\boldsymbol{R}_n$. In the simulation study, the unwrapped Gaussian spatial process to generate the errors is obtained assuming model (13), with constant mean, μ, equal to zero and exponential covariance structure

$$\mathbb{Cov}(w_i, w_j \mid \boldsymbol{X}_i, \boldsymbol{X}_j) = \sigma^2 [\exp(-\|\boldsymbol{X}_i - \boldsymbol{X}_j\|/a_{\mathrm{e}})], \tag{14}$$

where a_{e} is the practical range. This exponential spatial correlation function and the circular correlation of the corresponding wrapped Gaussian spatial process were compared by Jona-Lasinio et al. (2012, Fig. 4), obtaining very similar shapes for both correlations. Notice that although the vector of circular variables $\boldsymbol{\varepsilon}$ has almost zero mean direction, to properly apply the estimation procedure in practice, $\boldsymbol{\varepsilon}$ must be centered. In the simulation experiments, the value of the variance σ^2 in (14) is set equal to one, and different values of the parameter a_{e} are considered: $a_{\mathrm{e}} = 0.1$ (weak correlation), $a_{\mathrm{e}} = 0.3$ (medium correlation) and $a_{\mathrm{e}} = 0.6$ (strong correlation). Fixing the values of $\mu = 0$ and $\sigma = 1$, the effect of the range parameter a_{e} on a realization (on a 15×15 grid) of the wrapped circular spatial process can be seen in Fig. 2. Larger values of the range a_{e} yield a smoother pattern.

For the regression function r_1, Table 1 shows the average, over 500 replicates, of the CASE given in (12) considering the bandwidths selected by the CV and MCV methods, and the minimum value of CASE($\boldsymbol{H}$), i.e., CASE($\boldsymbol{H}_{\mathrm{CASE}}$), which can be viewed as a benchmark. Note that the optimal error increases as the dependence range becomes larger. It can be observed the poor behavior of the CV bandwidth, providing average values of the CASE far from the optimal value, not even decreasing for large sample sizes. In general, the MCV criterion appears to provide a significant improvement over the CV one when correlation is present. It can be observed that $\boldsymbol{H}_{\mathrm{MCV}_3}$ provides good results for all cases, decreasing the error as n gets larger. For stronger dependence (larger range values), this is the only selector that provides a reasonable behavior. Similar conclusions can be derived when considering the regression function r_2. The corresponding results are displayed in Table 2. Notice that when $a_{\mathrm{e}} = 0.1$ (weak spatial correlation), the best behavior is observed when $\boldsymbol{H}_{\mathrm{MCV}_2}$ is employed. As expected, for larger values of the practical range a_{e}, $\boldsymbol{H}_{\mathrm{MCV}_3}$ provides better results. Note that no major differences have been found if the NW- or LL-type estimators are employed.

4.3 The projection approach

An alternative way of generating a realization from a circular spatial process is by considering a projection
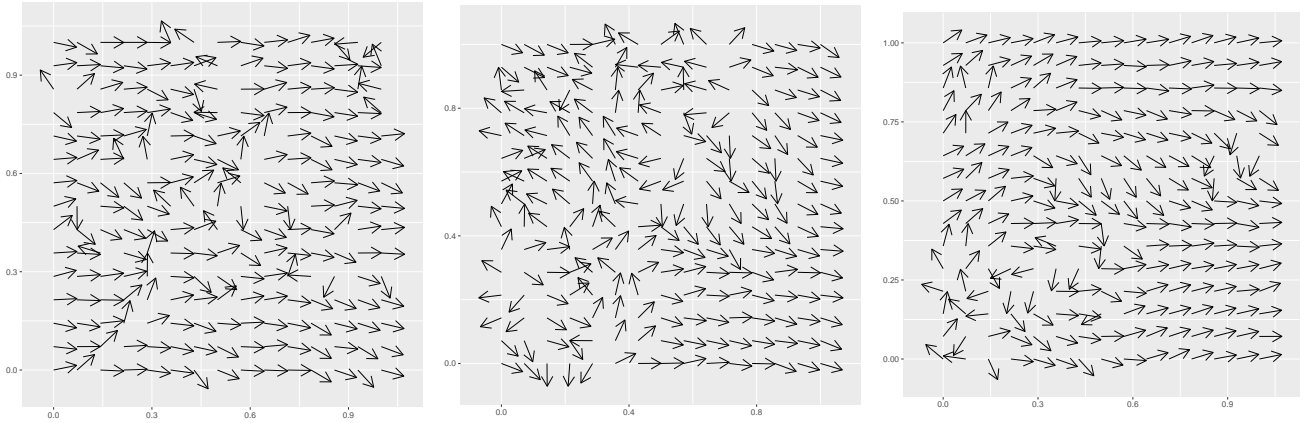
Fig. 3 Simulated samples of a projected Gaussian spatial process on a 15×15 grid with exponential correlation being $a_e = 0.1$ (left), $a_e = 0.3$ (center) and $a_e = 0.6$ (right), for $\boldsymbol{\mu} = (1,1)^\top$, $\sigma = 1$ and $\tau = 0.9$ in (15) and (16).

Table 3 Results obtained when the errors in model (1) are simulated from projected Gaussian spatial processes. Average (over 500 replicates) of the CASE given in (12), for the regression function r_1, using NW (left) and LL (right) type estimators. Bandwidth matrix is selected by minimizing CV ($\boldsymbol{H}_{\mathrm{CV}}$), MCV ($\boldsymbol{H}_{\mathrm{MCV}_1}$, $\boldsymbol{H}_{\mathrm{MCV}_2}$, $\boldsymbol{H}_{\mathrm{MCV}_3}$) and CASE ($\boldsymbol{H}_{\mathrm{CASE}}$) as a benchmark.

		NW					LL				
a_e	n	$\boldsymbol{H}_{\mathrm{CV}}$	$\boldsymbol{H}_{\mathrm{MCV}_1}$	$\boldsymbol{H}_{\mathrm{MCV}_2}$	$\boldsymbol{H}_{\mathrm{MCV}_3}$	$\boldsymbol{H}_{\mathrm{CASE}}$	$\boldsymbol{H}_{\mathrm{CV}}$	$\boldsymbol{H}_{\mathrm{MCV}_1}$	$\boldsymbol{H}_{\mathrm{MCV}_2}$	$\boldsymbol{H}_{\mathrm{MCV}_3}$	$\boldsymbol{H}_{\mathrm{CASE}}$
0.1	100	0.1465	0.0629	0.0482	0.0452	0.0249	0.1376	0.0735	0.0667	0.0596	0.0381
	225	0.2207	0.0874	0.0417	0.0386	0.0216	0.1976	0.0891	0.0552	0.0566	0.0321
	400	0.2639	0.0722	0.0380	0.0385	0.0209	0.2430	0.0745	0.0497	0.0535	0.0301
0.3	100	0.2562	0.1746	0.1382	0.1142	0.0564	0.2452	0.1823	0.1658	0.1534	0.1205
	225	0.2235	0.1851	0.1248	0.1123	0.0558	0.2300	0.1948	0.1555	0.1498	0.1203
	400	0.2498	0.1987	0.1279	0.1109	0.0474	0.2524	0.1974	0.1554	0.1478	0.1125
0.6	100	0.2368	0.1871	0.1567	0.1378	0.0620	0.1988	0.1772	0.1703	0.1622	0.1376
	225	0.2452	0.2126	0.1659	0.1371	0.0585	0.2072	0.1927	0.1710	0.1612	0.1361
	400	0.2424	0.1991	0.1600	0.1303	0.0525	0.2051	0.1868	0.1637	0.1578	0.1300

Table 4 Results obtained when the errors in model (1) are simulated from projected Gaussian spatial processes. Average (over 500 replicates) of the CASE given in (12), for the regression function r_2, using NW (left) and LL (right) type estimators. Bandwidth matrix is selected by minimizing CV ($\boldsymbol{H}_{\mathrm{CV}}$), MCV ($\boldsymbol{H}_{\mathrm{MCV}_1}$, $\boldsymbol{H}_{\mathrm{MCV}_2}$, $\boldsymbol{H}_{\mathrm{MCV}_3}$) and CASE ($\boldsymbol{H}_{\mathrm{CASE}}$) as a benchmark.

		NW					LL				
a_e	n	$\boldsymbol{H}_{\mathrm{CV}}$	$\boldsymbol{H}_{\mathrm{MCV}_1}$	$\boldsymbol{H}_{\mathrm{MCV}_2}$	$\boldsymbol{H}_{\mathrm{MCV}_3}$	$\boldsymbol{H}_{\mathrm{CASE}}$	$\boldsymbol{H}_{\mathrm{CV}}$	$\boldsymbol{H}_{\mathrm{MCV}_1}$	$\boldsymbol{H}_{\mathrm{MCV}_2}$	$\boldsymbol{H}_{\mathrm{MCV}_3}$	$\boldsymbol{H}_{\mathrm{CASE}}$
0.1	100	0.1764	0.1111	0.1055	0.1083	0.0813	0.1551	0.1023	0.0935	0.0957	0.0792
	225	0.2324	0.1177	0.0865	0.0901	0.0578	0.2023	0.1024	0.0769	0.0790	0.0513
	400	0.2689	0.1048	0.0772	0.0854	0.0516	0.2470	0.0909	0.0701	0.0717	0.0465
0.3	100	0.2701	0.2132	0.1822	0.1752	0.1457	0.2510	0.1998	0.1832	0.1796	0.1621
	225	0.2890	0.2354	0.1919	0.1723	0.1043	0.2708	0.2268	0.1913	0.1749	0.1412
	400	0.2850	0.2223	0.1734	0.1583	0.0940	0.2773	0.2134	0.1769	0.1687	0.1326
0.6	100	0.2458	0.2068	0.1934	0.1847	0.1472	0.2344	0.1995	0.1900	0.1877	0.1849
	225	0.2558	0.2211	0.1953	0.1777	0.1010	0.2441	0.2143	0.1923	0.1856	0.1517
	400	0.2444	0.2082	0.1806	0.1679	0.0960	0.2403	0.2017	0.1801	0.1752	0.1439

approach, as proposed by Wang and Gelfand (2014). A projected Gaussian spatial process is constructed from a bivariate Gaussian spatial process, and their correlation structures can be related, as shown in Wang and Gelfand (2014). This procedure allows to obtain samples of a projected spatial circular error as follows. First, a bivariate Gaussian spatial process, $\boldsymbol{Y}$, observed at a collection of spatial coordinates $\boldsymbol{X}_i$, with $i = 1, \ldots, n$,

is considered. The observations $\boldsymbol{Y}_i = (Y_{1i}, Y_{2i})$, with $\boldsymbol{Y}_i = \boldsymbol{Y}(\boldsymbol{X}_i)$ and $(Y_{1i}, Y_{2i}) = (Y_{1i}(\boldsymbol{X}_i), Y_{2i}(\boldsymbol{X}_i))$, can be decomposed as:

$$\boldsymbol{Y}_i = \boldsymbol{\mu} + \boldsymbol{w}_i, \quad i = 1, \ldots, n, \tag{15}$$

where $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{X}_i) \in \mathbb{R}^2$ is the mean vector and $\boldsymbol{w}$ is a zero mean bivariate Gaussian spatial process with cross covariance function $\rho_n(\boldsymbol{x}) \otimes \boldsymbol{T}$, being ρ_n a continuous
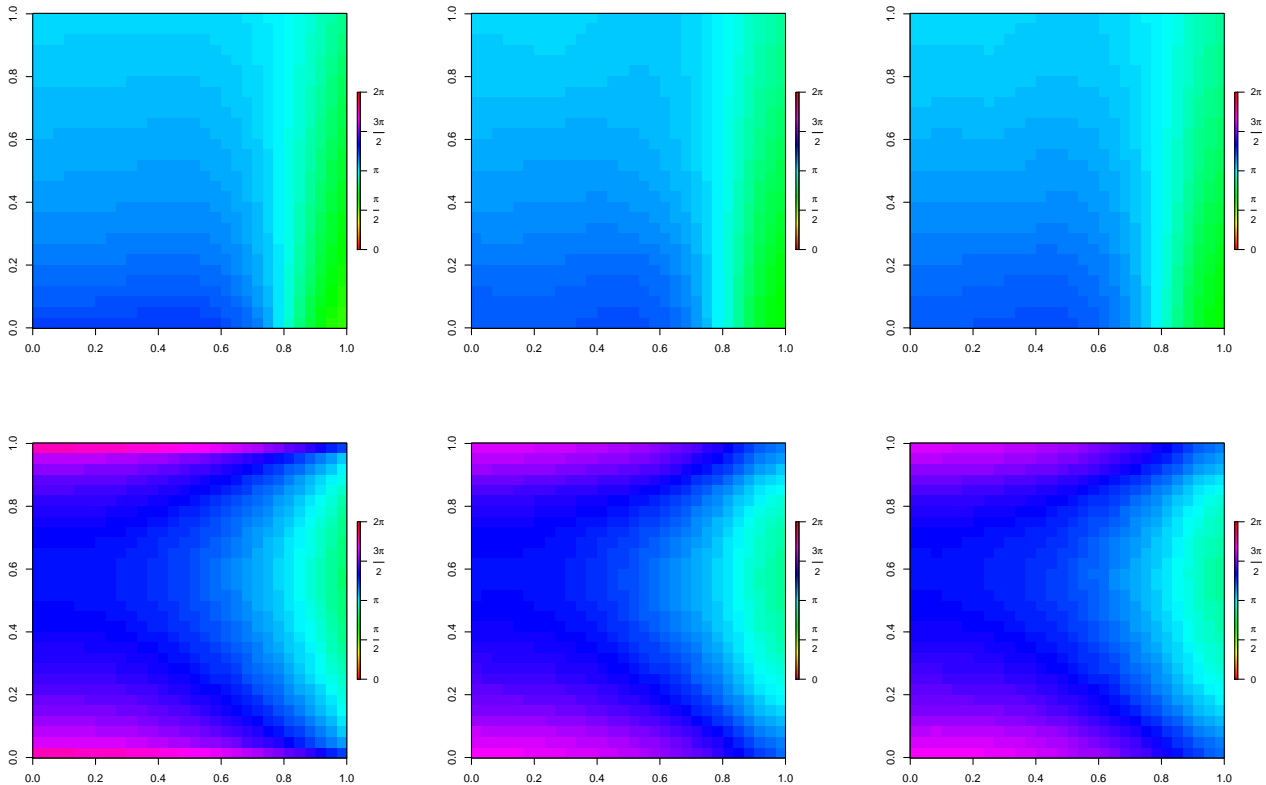
Fig. 4 Theoretical regression function (left), jointly with the NW-type estimators using wrapped (center) and projected Gaussian spatial processes (right) to generate the errors in model (1), for the regression function r_1 (top row) and r_2 (bottom row).

stationary correlation function and T a matrix, defined as:

$$T = \begin{pmatrix} \sigma^2 & \tau\sigma \\ \tau\sigma & 1 \end{pmatrix},$$

with $\sigma > 0$ and $\tau \in [-1, 1]$. The operator $\otimes$ denotes the Kronecker product. A realization of a circular spatial error process (in vector form), $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$ is obtained as:

$$\varepsilon_i = \arctan2(Y_{2i}, Y_{1i}), \quad i = 1, \ldots, n.$$

In the simulation study, the projected Gaussian spatial process is generated setting $\boldsymbol{\mu} = (1,1)^\top$ in (15) (to ensure unimodality of the errors and thus obtain homogeneous samples) and considering the cross-covariance function:

$$\mathbb{Cov}(\boldsymbol{w}_i, \boldsymbol{w}_j \mid \boldsymbol{X}_i, \boldsymbol{X}_j) = [\exp(-\|\boldsymbol{X}_i - \boldsymbol{X}_j\|/a_e)]\boldsymbol{T}, \ (16)$$

where a_e is the practical range. The variance and the parameter τ which controls the correlation between the linear variables are fixed to $\sigma = 1$ and $\tau = 0.9$, respectively, to better convey the dependence structure from the linear to the circular process (see Wang and

Gelfand, 2014, Fig. 4). As for the wrapping approach, the realization of the circular error process $\boldsymbol{\varepsilon}$ must be centered. Different degrees of spatial dependence are studied, considering values of $a_e = 0.1$ (weak correlation), $a_e = 0.3$ (medium correlation) and $a_e = 0.6$ (strong correlation). Fig. 3 shows a sample on a 15×15 grid of a simulated projected Gaussian spatial process for different values of a_e, with values of $\boldsymbol{\mu} = (1,1)^\top$, $\sigma = 1$ and $\tau = 0.9$. From left to right, the range increases, i.e., there is a stronger spatial dependence structure, and, consequently, the corresponding circular process realization shows a smoother pattern.

For the regression function r_1, numerical results are summarized in Table 3. As it was pointed out in Section 4.2 for the case of circular errors generated from wrapped Gaussian spatial processes, the CASE corresponding to the CV bandwidth matrix is the largest in all the scenarios. Regarding the MCV criterion, when the dependence structure is stronger, the value of l must be larger. For example, considering a weak dependence structure ($a_e = 0.1$) the use of $\boldsymbol{H}_{\mathrm{MCV}_2}$ seems to show a slightly better performance. If the dependence structure is stronger, $\boldsymbol{H}_{\mathrm{MCV}_3}$ provides better results. Table

4 shows the results for the regression function r_2, from where similar conclusions to those described when using r_1 can be deduced.

Numerical outputs are completed with some additional plots. Given that similar results were obtained in the previous simulations for $\hat{m}_{\boldsymbol{H}}(\boldsymbol{x};0)$ and $\hat{m}_{\boldsymbol{H}}(\boldsymbol{x};1)$, plots are only shown for $\hat{m}_{\boldsymbol{H}}(\boldsymbol{x};0)$. As an illustration of the appropriate performance of the estimator $\hat{m}_{\boldsymbol{H}}(\boldsymbol{x};0)$, Fig. 4 shows the theoretical regression functions r_1 and r_2 (left panels) and the corresponding average, over 500 replicates, of the fitted values using $\hat{m}_{\boldsymbol{H}}(\boldsymbol{x};0)$, considering samples of size $n = 400$ and circular errors generated from a wrapped Gaussian spatial process (center panels) and from a projected Gaussian spatial process (right panels). In this example, for both types of circular errors, an exponential covariance model is used with range parameter equal to 0.3. Estimates are computed employing the bandwidth matrix $\boldsymbol{H}_{\mathrm{MCV}_3}$. Notice that, for comparison purposes, the theoretical regression functions are plotted in a 30×30 regular grid on the covariate region (the same grid where the estimations are computed). Plots in the top row present the results for the data generated using the regression function r_1 and those in the bottom row using r_2. The estimation of the circular trend surfaces seems to be quite accurate, no matter the approach (wrapped or projected) used to generate the circular spatial errors.

5 Real data illustration

The performance of the proposed estimators is illustrated on the Adriatic Sea wave direction dataset presented in the Introduction. A brief description of this dataset, as well as the regression model considered, are provided in Section 5.1. As noted in the Introduction, this dataset (or part of it) has been mainly analyzed using parametric methods (Jona-Lasinio et al., 2012; Wang and Gelfand, 2014; Lagona et al., 2015; Mastrantonio et al., 2016), while we use now a nonparametric approach. Details on the kernel and bandwidth matrix employed in the estimation procedure are given in Section 5.2. Taking into account that the performance of the regression estimators $\hat{m}_{\boldsymbol{H}}(\boldsymbol{x};0)$ and $\hat{m}_{\boldsymbol{H}}(\boldsymbol{x};1)$ was similar in the simulation study, only results employing $\hat{m}_{\boldsymbol{H}}(\boldsymbol{x};1)$ are shown for this application. These results are included in Section 5.3. Some ideas on outliers diagnostics are provided in Section 5.4.

5.1 Wave direction dataset and regression model

Wave directions were recorded in 1494 grid points on the Adriatic Sea area from a calm period transitioning to a storm period at different times. These data outputs were derived from a wave model implemented by Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA) and they are available in the R package CircSpaceTime (Jona-Lasinio et al., 2019).

In this illustration, we only consider wave directions for a calm period, corresponding to measurements taken at 06:00 on April 2 at Adriatic Sea (Fig. 1 shows a random sample of 150 observations of this dataset). We assume the linear-circular regression model given in (1), where for $i = 1, \ldots, 1494$, $\boldsymbol{X}_i = (X_{i1}, X_{i2})$, represent the different locations, with X_{i1} the longitude and X_{i2} the latitude, and Θ_i the corresponding wave direction at that location.

5.2 Bandwidth matrix selection

To compute the nonparametric estimator of the circular regression function, given in (8), a multiplicative triweight kernel is considered. The bandwidth matrix is selected employing a cross-validation criterion. In order to decide if using CV or MCV (and, in that case, a suitable value for the radius l), the whole sample is split in two parts, a randomly selected training sample of size 1345 (90% of the data), denoted by $\{(\tilde{\boldsymbol{X}}_i, \tilde{\Theta}_i)\}_{i=1}^{1345}$, and a test sample, made up of the remaining observations, of size 149 (10% of the data), denoted by $\{(\check{\boldsymbol{X}}_j, \check{\Theta}_j)\}_{j=1}^{149}$. Then, estimations at each testing point $\check{\boldsymbol{X}}_j, j = 1, \ldots, 149$, with different bandwidths, are compared with the testing responses using the following prediction error:

$$\sum_{j=1}^{149} \left\{ 1 - \cos \left[\check{\Theta}_j - \hat{m}_{\hat{\boldsymbol{H}}}(\check{\boldsymbol{X}}_j;1) \right] \right\}, \tag{17}$$

where $\hat{m}_{\hat{\boldsymbol{H}}}(\check{\boldsymbol{X}}_j;1)$ is the LL-type circular regression estimator computed using the training sample and evaluated at the testing point $\check{\boldsymbol{X}}_j, j = 1, \ldots, 149$, and $\hat{\boldsymbol{H}}$ denotes the bandwidth matrix selected using CV or MCV, employing the training sample. In the case of the MCV criterion, different values of the radius l are considered. As in the simulation study, we set $l(b) = \sqrt{2}b/10$, now with $b = 1, \ldots, 10$. These bandwidth matrices are searched in the family of the symmetric and definite positive full bandwidth matrices, using an optimization algorithm based on the Nelder–Mead simplex method described in Lagarias et al. (1998). To apply this optimization procedure, we use the initial bandwidth matrix $\boldsymbol{H}_{\mathrm{init}} = 1.5 \cdot \mathrm{diag}\left\{\hat{\sigma}_{\tilde{X}_1}, \hat{\sigma}_{\tilde{X}_2}\right\}$, where $\hat{\sigma}_{\tilde{X}_1}$ and $\hat{\sigma}_{\tilde{X}_2}$, with $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \tilde{X}_2)$, are the training sample standard deviations of $\tilde{X}_1$ and $\tilde{X}_2$, respectively. Fig. 5 shows the prediction error given in (17) for each bandwidth matrix $\hat{\boldsymbol{H}}$. It can be seen that the minimum error is achieved
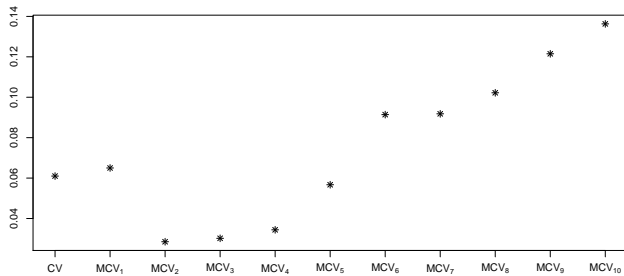
Fig. 5 Prediction errors given in (17) for each bandwidth matrix selected by CV and MCV, with different values of the radius $l(b) = \sqrt{2}b/10$, $b = 1, \ldots, 10$.

when MCV, with $b = 2$, is employed, converging the algorithm when using this criterion to

$$\boldsymbol{H}_{\mathrm{MCV}_2} = \begin{pmatrix} 0.4744 & 0.0081 \\ 0.0081 & 0.3529 \end{pmatrix}. \tag{18}$$

This solution is obtained from the application of a numerical optimization algorithm which, for our sample of relatively large size, took 4251.2983 seconds.

5.3 Circular trend surface estimates

The circular trend surface estimates using $\boldsymbol{H}_{\mathrm{MCV}_2}$, given in (18), are shown in Fig. 6. The estimation grid is constructed by overlying the survey values of longitude and latitude with a 100×100 grid and, then, dropping every grid point that did not satisfy at least one of the following two requirements: (a) it is within two "grid cell length" from an observation point, or (b) the calculation for the estimates of the sine and cosine components at that grid point uses a smoothing vector that is sufficiently stable. The sine and cosine of the detrended wave direction dataset were tested for isotropy and stationarity, following the proposals by Bowman and Crujeiras (2013). For both tests, p-values were larger than the usual significance levels (for isotropy: 0.3206 and 0.1271 for sine and cosine, respectively; for stationarity, p-values were larger than 0.99 for both processes).

From Fig. 6, it can be clearly seen the shoreline orientation of the waves (recall that our measurements correspond to a calm period), providing the different color pattern along the coastline. Something which is interesting to notice is the behavior in the Gulf, where waves rotate to different directions, and a main current can be also observed. According to this pattern, more variation can be observed in the North, something that was also pointed out by Jona-Lasinio et al. (2012), although for a storm period.
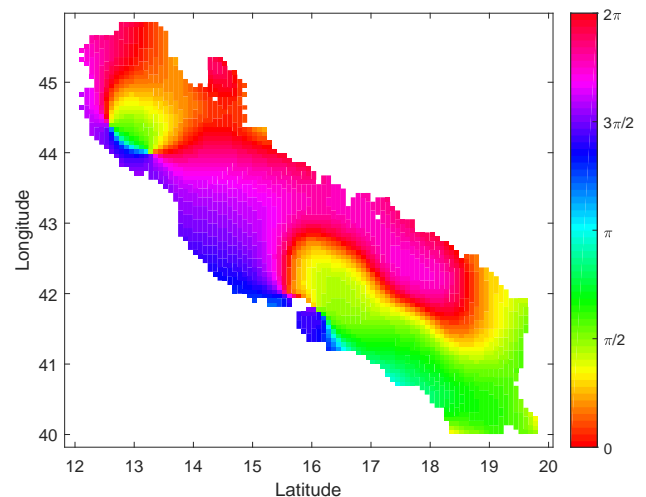

Fig. 6 Regression function estimation using the LL-type estimator $\hat{m}_{\hat{\boldsymbol{H}}}(\boldsymbol{x}; 1)$, using the bandwidth matrix $\boldsymbol{H}$ given in (18), selected with the MCV ($b = 2$) criterion.

5.4 Outlier diagnostics

Diagnostics tools for outlier detection are required in order to round off the data modeling. In this context, a residual analysis seeking for possible outliers must be carried out using circular data tools. Apart from the initial ideas in Jammalamadaka and Sengupta (2001), there have been some attempts to generalize Tukey's boxplot to the circular context, such as Anderson (1994), Abuzaid et al. (2012) and more recently, Buttarazzi et al. (2018) who devised a circular boxplot. Fig. 7 shows the circular boxplot of the residuals from the nonparametric fit in Fig. 6, where just three outliers (corresponding to anomalous values of the residuals) are detected in our data sample. When plotting the locations corresponding to these values (Fig. 8), it can be observed that they are three isolated points. The LL-type estimator has been fitted again deleting these three points and, as expected, not much differences have been found. In this case, the minimum of the error, given in (17), is achieved when MCV (with $b = 2$) is employed, converging the algorithm when using this criterion to

$$\boldsymbol{H}_{\mathrm{MCV}_2} = \begin{pmatrix} 0.5182 & 0.0065 \\ 0.0065 & 0.4001 \end{pmatrix},$$

yielding an almost identical estimation of the spatial trend to the one obtained with the whole dataset. This is actually expected given that this type of nonparametric estimators usually adapt satisfactorily in the presence of outliers. It should be noticed that if the circular boxplot is now recomputed when deleting the three points, no other outliers are identified.
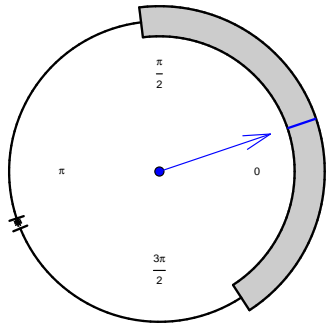
Fig. 7 Circular boxplot of the residuals obtained from the proposed LL-type estimator.
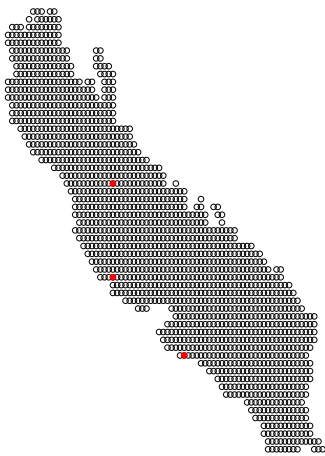

Fig. 8 Spatial locations in the Adriatic Sea area on April 2, 2010 at 6am during a calm period (black points) and spatial locations which are (possible) circular outliers (red points).

6 Conclusions, limitations, and further research

Nonparametric smoothing estimators of the regression function in a model with circular responses and real-valued covariates in the presence of spatial correlation are studied and applied to fit a circular trend surface for wave directions. The proposal considers two nonparametric regression models for the sine and cosine components of the circular responses, which are indeed regression models with real-valued responses. In particular, NW and LL estimators are used in these two real-valued regression models. The asymptotic conditional properties of the proposed kernel-type estimators are derived.

Although there is a substantial literature on modeling circular data with spatial dependence by introducing and formulating spatial processes for circular data, such as wrapped and projected Gaussian spatial processes, our proposal follows a different perspective. We consider an appropriate linear-circular regression model for spatially correlated data and estimate nonparamet-

rically the corresponding circular spatial trend. This is an alternative to model circular data at different spatial locations. It should be noted that no other direct competitor (up to our knowledge) has yet been proposed following these ideas, neither from a parametric nor from a nonparametric approach.

One of the advantages of the proposed procedure (if the bandwidth matrix is appropriately chosen, and a suitable bandwidth selector is also provided in this work) is that it relaxes parametric assumptions, and consequently enables one to explore and model the data more flexibly, avoiding misspecification problems. Moreover, this estimator can be employed as a first attempt to explore if a certain parametric family is appropriate or not to model the data. It is well known that goodness-of-fit evaluation of a parametric regression model is often performed using testing procedures, where the (parametric) fit of the model is compared to that obtained when considering a more general class of models. Nonparametric regression models are frequently employed in this setting (for instance, Härdle and Mammen, 1993; Meilán-Vila et al., 2020c). Following these ideas, goodness-of-fit tests for assessing a parametric model for a regression model with circular response and d-dimensional covariate, in both independent and spatially correlated frameworks have been proposed and analyzed (Meilán-Vila et al., 2020a).

Regarding the bandwidth matrix needed to compute the circular regression estimators given in (8), it can be selected by leave-one-out cross-validation, but this matrix is not necessarily a good one for spatially correlated data, given that $\mathbb{E}[\mathrm{CV}(\boldsymbol{H})]$ is severely affected by the correlation (Opsomer et al., 2001; Liu, 2001). In that context, it is advisable to employ other bandwidth selection criteria which take the spatial dependence structure into account. In our practical results, we also considered a modified cross-validation method suitably adjusted for the presence of spatial correlation, which considers to "leave $N(i)$ observations out". The idea of modifying the selection criterion in this manner is not new. An example of such adjustment is the "leave-$2l+1$-out" cross-validation approach (Hart and Vieu, 1990). In the case of the marine currents in the Adriatic Sea, a suitable parameter l in MCV, controlling the number of observations left out, has been selected by minimizing the prediction error given in (17). However, an interesting point would be to design a fully automatic procedure to compute the optimal radius in the MCV method. This approach should account for the spatial correlation of the covariates and, although it is out of the scope of the present paper, it would be an interesting topic of further research. On the other hand, cross-validation tech-

niques have the drawback of being unable to provide satisfactory results in a reasonable time for very large sample sizes due to its computational complexity. To overcome this problem, bagging cross-validation bandwidths studied for density and regression estimation with Euclidean data (Hall and Robinson, 2009) could be adapted to this context. Additionally, note that even though cross-validation bandwidths present appealing theoretical properties, in practice, their computation could present certain difficulties in a multidimensional framework.

Alternatively to the cross-validation methods previously described, the bandwidth could be selected as follows. First, as described before, the whole sample is split in two parts, a training and a testing samples. Then, using the overall dataset and a pilot bandwidth matrix, the nonparametric estimator is computed at each training data point. From this estimation, and using a wrapped (Jona-Lasinio et al., 2012) or a projected Gaussian spatial process (Wang and Gelfand, 2014), the residuals can be modeled, obtaining predictions at the testing locations. Finally, the bandwidth matrix can be selected by minimizing the corresponding prediction error computed with the testing sample.

It should be noted that just global bandwidths are considered in the numerical studies of this paper. A limitation derived from the use of global bandwidths is that the corresponding nonparametric estimators may provide spurious estimates in areas with sparse observations. In such regions, the number of observations within the neighborhood determined by the bandwidth may be too small, producing unstable estimates. A way to overcome this problem is employing local bandwidths that automatically adapt to the number of observations near the grid points where calculating the estimates. The problem of using local bandwidths is that a high computing time would be required, specially if the sample size is very large.

Even though it seems we do not need to worry about outliers in our real data analysis, in a general case, deriving such a conclusion may be not as clear as for our example. Since the data exhibit spatial dependence, it is difficult to assess whether a set of observations are really outliers or just correspond to an effect of the spatial dependence structure. Note that the definition of spatial outlier is not entirely precise and, besides, the available exploratory techniques for real-valued spatial processes, such as the Moran scatterplot (Anselin, 1995) or the variogram cloud (Cressie, 1993), do not allow a correct identification of such data. All this makes particularly difficult to generalize the definition and the existing methods to detect spatial outliers for circular

data. There is certainly an interesting issue that could be addressed as further research.

In the current setting, real-valued covariates for explaining the behavior of a circular response in presence of spatial correlation are considered. However, it might be the case that other types of covariates, such as other circular, or more generally, spherical covariates, may influence the circular response. For these more complex scenarios, there is a substantial research on modeling and on analyzing different inference approaches for random fields on spheres as well as on spheres across time. For instance, Porcu et al. (2016) developed cross-covariance functions of the great circle distances on the sphere. Alegría et al. (2019) proposed a flexible parametric family of matrix-valued covariance functions. To overcome the problem of generating samples from random fields, Emery and Porcu (2019) introduced an algorithm to generate isotropic vector-valued Gaussian random fields defined over the unit two-dimensional sphere embedded in the three-dimensional Euclidean space. Some of these approaches could be incorporated in our model, but these extensions are out of scope of the present paper and can be the focus of future researches.

Finally, it is worth noting that environmental processes usually present an asymmetric behavior, requiring sophisticated distribution models (such as the Birnbaum-Saunders (BS) distribution) for an appropriate fitting. In this context, Saulo et al. (2013) proposed a kernel method for estimating asymmetric densities based on a generalization of the BS model. From a regression perspective, Leiva et al. (2020) proposed a geostatistical model based on BS quantile regression and Martínez et al. (2019) formulated a regression model, considering a scalar response and functional covariates, supposing that its errors are spatially correlated and follow a BS distribution. In our context, no shape conditions are imposed regarding the (circular) error distribution, although the consideration of an extension of the BS distribution model to the circular context may enable the design of a parametric but relatively flexible regression model.

In practice, the numerical studies performed in this work were run in an Intel Core i7-9700K at 3.60Ghz. The simulations were implemented in the statistical environment `R` (R Development Core Team, 2020), using functions included in the `npsp` and `CircSpaceTime` packages (Fernández-Casal, 2019; Jona-Lasinio et al., 2019). The real data application was performed in MATLAB software (www.mathworks.com).

Acknowledgements

References

Abuzaid AH, Mohamed IB, Hussin AG (2012) Boxplot for circular variables. Computation Stat 27(3):381–392

Alegría A, Porcu E, Furrer R, Mateu J (2019) Covariance functions for multivariate Gaussian fields evolving temporally over planet earth. Stoch Env Res Risk A 33(8-9):1593–1608

Anderson CM (1994) Graphical methods for circular and cylindrical data. In: Tech. Report TR-94-05, Department of Statistical and Actuarial Sciences, University of Western Ontario

Anselin L (1995) Local indicators of spatial association-LISA. Geogr Anal 27(2):93–115

Bowman AW, Crujeiras RM (2013) Inference for variograms. Comput Stat Data Anal 66:19–31

Buttarazzi D, Pandolfo G, Porzio GC (2018) A boxplot for circular data. Biometrics 74(4):1492–1501

Carnicero JA, Ausín MC, Wiper MP (2013) Nonparametric copulas for circular–linear and circular–circular data: an application to wind directions. Stoch Env Res Risk A 27(8):1991–2002

Casson E, Coles S (1998) Extreme hurricane wind speeds: estimation, extrapolation and spatial smoothing. J Wind Eng Ind Aerod 74:131–140

Chacón JE, Duong T (2018) Multivariate kernel smoothing and its applications. CRC Press

Cressie NA (1993) Statistics for spatial data. Wiley, New York

Di Marzio M, Panzera A, Taylor CC (2012) Nonparametric smoothing and prediction for nonlinear circular time series. J Time Ser Anal 33(4):620–630

Di Marzio M, Panzera A, Taylor CC (2013) Nonparametric regression for circular responses. Scand J Stat 40(2):238–255

Duong T (2015) Spherically symmetric multivariate beta family kernels. Stat Probab Lett 104:141–145

Emery X, Porcu E (2019) Simulating isotropic vector-valued Gaussian random fields on the sphere through finite harmonics approximations. Stoch Env Res Risk A 33(8-9):1659–1667

Fernández-Casal R (2019) npsp: Nonparametric spatial (geo)statistics. URL http://cran.r-project.org/package=npsp, R package version 0.7-5

García-Portugués E, Barros AM, Crujeiras RM, González-Manteiga W, Pereira J (2014) A test for directional-linear independence, with applications to wildfire orientation and size. Stoch Env Res Risk A 28(5):1261–1275

Hall P, Robinson AP (2009) Reducing variability of crossvalidation for smoothing parameter choice. Biometrika 96:175–186

Härdle W, Mammen E (1993) Comparing nonparametric versus parametric regression fits. Ann Stat 21:1926–1947

Härdle W, Müller M (2012) Multivariate and semiparametric kernel regression, John Wiley & Sons, Ltd, chap 12, pp 357–391

Hart JD, Vieu P (1990) Data-driven bandwidth choice for density estimation based on dependent data. Ann Stat 18(2):873–890

Jammalamadaka SR, Sengupta A (2001) Topics in circular statistics, vol 5. World Scientific

Jona-Lasinio G, Gelfand A, Jona-Lasinio M (2012) Spatial analysis of wave direction data using wrapped Gaussian processes. Ann Appl Stat 6(4):1478–1498

Jona-Lasinio G, Mastrantonio G, Santoro M (2019) CircSpaceTime: spatial and spatio-temporal bayesian model for circular data. URL http://cran.r-project.org/package=CircSpaceTime, R package version 0.9.0

Kim S, SenGupta A (2017) Multivariate-multiple circular regression. J Stat Comput Sim 87(7):1277–1291

Lagarias JC, Reeds JA, Wright MH, Wright PE (1998) Convergence properties of the Nelder–Mead simplex method in low dimensions. SIAM J Optimz 9(1):112–147

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 Env Res Risk A 29(2):397–409

Leiva V, Sánchez L, Galea M, Saulo H (2020) Global and local diagnostic analytics for a geostatistical model based on a new approach to quantile regres-

sion. Stoch Env Res Risk A pp 1–15

Liu XH (2001) Kernel smoothing for spatially correlated data. PhD thesis, Department of Statistics, Iowa State University

Martínez S, Giraldo R, Leiva V (2019) Birnbaum–Saunders functional regression models for spatial data. Stoch Env Res Risk A 33(10):1765–1780

Mastrantonio G, Gelfand AE, Lasinio GJ (2016) The wrapped skew Gaussian process for analyzing spatio-temporal data. Stoch Env Res Risk A 30(8):2231–2242

Mastrantonio G, Pollice A, Fedele F (2018) Distributions-oriented wind forecast verification by a hidden Markov model for multivariate circular–linear data. Stoch Env Res Risk A 32(1):169–181

Meilán-Vila A, Francisco-Fernández M, Crujeiras R (2020a) Goodness-of-fit tests for parametric regression models with circular response. arXiv: 2008.13473

Meilán-Vila A, Francisco-Fernández M, Crujeiras R, Panzera A (2020b) Nonparametric multiple regression estimation for circular response. TEST DOI 10.1007/s11749-020-00736-w

Meilán-Vila A, Opsomer JD, Francisco-Fernández M, Crujeiras RM (2020c) A goodness-of-fit test for regression models with spatially correlated errors. TEST 29:728–749

Opsomer J, Wang Y, Yang Y (2001) Nonparametric regression with correlated errors. Stat Sci 16:134–153

Porcu E, Bevilacqua M, Genton MG (2016) Spatio-temporal covariance and cross-covariance functions of the great circle distance on a sphere. J Am Stat Assoc 111(514):888–898

R Development Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, URL http://www.R-project.org

Ruppert D, Wand MP (1994) Multivariate locally weighted least squares regression. Ann Stat 22:1346–1370

Saulo H, Leiva V, Ziegelmann FA, Marchant C (2013) A nonparametric method for estimating asymmetric densities based on skewed birnbaum–saunders distributions applied to environmental data. Stoch Env Res Risk A 27(6):1479–1491

Wang F, Gelfand AE (2014) Modeling space and space-time directional data using projected Gaussian processes. J Am Stat Assoc 109(508):1565–1580

Appendix. Proof of Theorem 1

Proof Before deriving the asymptotic variance of the estimator $\hat{m}_{\boldsymbol{H}}(\boldsymbol{x}, p)$, for $p = 0, 1$, stated in Theorem 1, some preliminary approximations are needed.

Firstly, defining $f_1(\boldsymbol{x}) = \sin[m(\boldsymbol{x})]$ and $f_2(\boldsymbol{x}) = \cos[m(\boldsymbol{x})]$, using sine and cosine addition formulas, the following relation between the covariance function $C_{n,3}$, defined from models (6) and (7), and the correlations $\rho_{k,n}$, $k = 1, 2, 3$, directly derived from model (1) and given in (2), (3) and (4), can be obtained:

$$\begin{aligned}
C_{n,3}(\boldsymbol{X}_i, \boldsymbol{X}_j) = {} & f_1(\boldsymbol{X}_i) f_2(\boldsymbol{X}_j) \sigma_2^2 \rho_{2,n}(\boldsymbol{X}_i - \boldsymbol{X}_j) \\
& - f_1(\boldsymbol{X}_i) f_1(\boldsymbol{X}_j) \sigma_{12} \rho_{3,n}(\boldsymbol{X}_i - \boldsymbol{X}_j) \\
& + f_2(\boldsymbol{X}_i) f_2(\boldsymbol{X}_j) \sigma_{12} \rho_{3,n}(\boldsymbol{X}_i - \boldsymbol{X}_j) \\
& - f_2(\boldsymbol{X}_i) f_1(\boldsymbol{X}_j) \sigma_1^2 \rho_{1,n}(\boldsymbol{X}_i - \boldsymbol{X}_j).
\end{aligned}$$

Moreover, denoting

$$k_{1,n}(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{X}_i - \boldsymbol{x}),$$

$$k_{2,n}(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{X}_i - \boldsymbol{x})(\boldsymbol{X}_i - \boldsymbol{x}),$$

$$k_{3,n}(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{X}_i - \boldsymbol{x})(\boldsymbol{X}_i - \boldsymbol{x})(\boldsymbol{X}_i - \boldsymbol{x})^{\top},$$

$$s_{1,n}(\boldsymbol{x}) = \frac{1}{n^2} \left[\sum_{i=1}^{n} K_{\boldsymbol{H}}^2(\boldsymbol{X}_i - \boldsymbol{x}) c(\boldsymbol{X}_i) + \sum_{i \neq j} K_{\boldsymbol{H}}(\boldsymbol{X}_i - \boldsymbol{x}) \right.$$
$$\left. \cdot K_{\boldsymbol{H}}(\boldsymbol{X}_j - \boldsymbol{x}) C_{n,3}(\boldsymbol{X}_i, \boldsymbol{X}_j) \right],$$

$$s_{2,n}(\boldsymbol{x}) = \frac{1}{n^2} \left[\sum_{i=1}^{n} K_{\boldsymbol{H}}^2(\boldsymbol{X}_i - \boldsymbol{x})(\boldsymbol{X}_i - \boldsymbol{x}) c(\boldsymbol{X}_i) \right.$$
$$+ \sum_{i \neq j} K_{\boldsymbol{H}}(\boldsymbol{X}_i - \boldsymbol{x}) K_{\boldsymbol{H}}(\boldsymbol{X}_j - \boldsymbol{x})(\boldsymbol{X}_i - \boldsymbol{x})$$
$$\left. \cdot C_{n,3}(\boldsymbol{X}_i, \boldsymbol{X}_j) \right],$$

$$s_{3,n}(\boldsymbol{x}) = \frac{1}{n^2} \left[\sum_{i=1}^{n} K_{\boldsymbol{H}}^2(\boldsymbol{X}_i - \boldsymbol{x})(\boldsymbol{X}_i - \boldsymbol{x})(\boldsymbol{X}_j - \boldsymbol{x})^{\top} c(\boldsymbol{X}_i) \right.$$
$$+ \sum_{i \neq j} K_{\boldsymbol{H}}(\boldsymbol{X}_i - \boldsymbol{x}) K_{\boldsymbol{H}}(\boldsymbol{X}_j - \boldsymbol{x})(\boldsymbol{X}_i - \boldsymbol{x})$$
$$\left. \cdot (\boldsymbol{X}_j - \boldsymbol{x})^{\top} C_{n,3}(\boldsymbol{X}_i, \boldsymbol{X}_j) \right],$$

and, after some calculations, it can be obtained that

$$k_{1,n}(\boldsymbol{x}) = f(\boldsymbol{x}) + o_{\mathbb{P}}(1), \tag{19}$$

$$k_{2,n}(\boldsymbol{x}) = \mu_2 \boldsymbol{\nabla} f(\boldsymbol{x}) \boldsymbol{H}^2 + o_{\mathbb{P}}(\boldsymbol{H}^2 \mathbf{1}_d), \tag{20}$$

$$k_{3,n}(\boldsymbol{x}) = \mu_2 f(\boldsymbol{x}) \boldsymbol{H}^2 + o_{\mathbb{P}}(\boldsymbol{H} \mathbf{1}_{d \times d} \boldsymbol{H}), \tag{21}$$

$$s_{1,n}(\boldsymbol{x}) = \frac{1}{n|\boldsymbol{H}|} \nu_0 f(\boldsymbol{x})[c(\boldsymbol{x}) + f(\boldsymbol{x}) C_3(\boldsymbol{x})]$$
$$+ o_{\mathbb{P}}\left(\frac{1}{n|\boldsymbol{H}|}\right), \tag{22}$$

$$s_{2,n}(\boldsymbol{x}) = \frac{1}{n|\boldsymbol{H}|} o_{\mathbb{P}}(\mathbf{1}_d) \tag{23}$$

$$s_{3,n}(\boldsymbol{x}) = \frac{1}{n|\boldsymbol{H}|} o_{\mathbb{P}}(\mathbf{1}_{d \times d}), \tag{24}$$

where $\mathbf{1}_d$ and $\mathbf{1}_{d \times d}$ denote the $d \times 1$ vector and the $d \times d$ matrix with every entry equal to 1, and

$$c(\boldsymbol{x}) = f_1(\boldsymbol{x}) f_2(\boldsymbol{x}) \sigma_2^2 - f_1^2(\boldsymbol{x}) \sigma_{12} + f_2^2(\boldsymbol{x}) \sigma_{12}$$
$$- f_1(\boldsymbol{x}) f_2(\boldsymbol{x}) \sigma_1^2, \tag{25}$$

$$C_3(\boldsymbol{x}) = f_1(\boldsymbol{x}) f_2(\boldsymbol{x}) \sigma_2^2 \rho_{c_2} - f_1^2(\boldsymbol{x}) \sigma_{12} \rho_{c_3} + f_2^2(\boldsymbol{x}) \sigma_{12} \rho_{c_3}$$
$$- f_1(\boldsymbol{x}) f_2(\boldsymbol{x}) \sigma_1^2 \rho_{c_1}. \tag{26}$$

To derive the variance of $\hat{m}_{\boldsymbol{H}}(\boldsymbol{x}; p)$, for $p = 0, 1$, denoting by $\boldsymbol{\mathcal{X}} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$, using Taylor expansions and following similar arguments to those used in Di Marzio et al. (2013) and Meilán-Vila et al. (2020b), it can be obtained that

$$\mathbb{V}\mathrm{ar}[\hat{m}_{\boldsymbol{H}}(\boldsymbol{x}; p) \mid \boldsymbol{\mathcal{X}}]$$
$$= \frac{m_1^2(\boldsymbol{x})}{[m_1^2(\boldsymbol{x}) + m_2^2(\boldsymbol{x})]^2} \mathbb{V}\mathrm{ar}[\hat{m}_{2,\boldsymbol{H}}(\boldsymbol{x}; p) \mid \boldsymbol{\mathcal{X}}]$$
$$+ \frac{m_2^2(\boldsymbol{x})}{[m_1^2(\boldsymbol{x}) + m_2^2(\boldsymbol{x})]^2} \mathbb{V}\mathrm{ar}[\hat{m}_{1,\boldsymbol{H}}(\boldsymbol{x}; p) \mid \boldsymbol{\mathcal{X}}]$$
$$- \frac{2 m_1(\boldsymbol{x}) m_2(\boldsymbol{x})}{[m_1^2(\boldsymbol{x}) + m_2^2(\boldsymbol{x})]^2} \mathbb{C}\mathrm{ov}[\hat{m}_{1,\boldsymbol{H}}(\boldsymbol{x}; p), \hat{m}_{2,\boldsymbol{H}}(\boldsymbol{x}; p) \mid \boldsymbol{\mathcal{X}}]$$
$$+ O\{[\hat{m}_{1,\boldsymbol{H}}(\boldsymbol{x}; p) - m_1(\boldsymbol{x})]^3\}$$
$$+ O\{[\hat{m}_{2,\boldsymbol{H}}(\boldsymbol{x}; p) - m_2(\boldsymbol{x})]^3\}. \tag{27}$$

The conditional variance of $\hat{m}_{j,\boldsymbol{H}}(\boldsymbol{x}; p)$, for $j = 1, 2$, and $p = 0, 1$, for spatially correlated data, can be derived using similar arguments to those given in Liu (2001), which yield

$$\mathbb{V}\mathrm{ar}[\hat{m}_{j,\boldsymbol{H}}(\boldsymbol{x}; p) \mid \boldsymbol{\mathcal{X}}] = \frac{\nu_0 [s_j^2(\boldsymbol{x}) + f(\boldsymbol{x}) C_j(\boldsymbol{x})]}{n|\boldsymbol{H}| f(\boldsymbol{x})}$$
$$+ o_{\mathbb{P}}\left(\frac{1}{n|\boldsymbol{H}|}\right), \tag{28}$$

where

$$s_1^2(\boldsymbol{x}) = f_1^2(\boldsymbol{x}) \sigma_2^2 + 2 f_1(\boldsymbol{x}) f_2(\boldsymbol{x}) \sigma_{12} + f_2^2(\boldsymbol{x}) \sigma_1^2, \tag{29}$$

$$s_2^2(\boldsymbol{x}) = f_2^2(\boldsymbol{x}) \sigma_2^2 - 2 f_2(\boldsymbol{x}) f_1(\boldsymbol{x}) \sigma_{12} + f_1^2(\boldsymbol{x}) \sigma_1^2, \tag{30}$$

$$C_1(\boldsymbol{x}) = f_1^2(\boldsymbol{x}) \sigma_2^2 \rho_{c_2} + 2 f_1(\boldsymbol{x}) f_2(\boldsymbol{x}) \sigma_{12} \rho_{c_3}$$
$$+ f_2^2(\boldsymbol{x}) \sigma_1^2 \rho_{c_1}, \tag{31}$$

$$C_2(\boldsymbol{x}) = f_2^2(\boldsymbol{x}) \sigma_2^2 \rho_{c_2} - 2 f_2(\boldsymbol{x}) f_1(\boldsymbol{x}) \sigma_{12} \rho_{c_3}$$
$$+ f_1^2(\boldsymbol{x}) \sigma_1^2 \rho_{c_1}. \tag{32}$$

Moreover, using (19) and (22), it is easy to obtain that the conditional covariance between $\hat{m}_{1,\boldsymbol{H}}(\boldsymbol{x}; 0)$ and $\hat{m}_{2,\boldsymbol{H}}(\boldsymbol{x}; 0)$ is

$$\mathbb{C}\mathrm{ov}[\hat{m}_{1,\boldsymbol{H}}(\boldsymbol{x}; 0), \hat{m}_{2,\boldsymbol{H}}(\boldsymbol{x}; 0) \mid \boldsymbol{\mathcal{X}}]$$
$$= \frac{\sum_{i=1}^n \sum_{j=1}^n K_{\boldsymbol{H}}(\boldsymbol{X}_i - \boldsymbol{x}) K_{\boldsymbol{H}}(\boldsymbol{X}_j - \boldsymbol{x})}{\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{X}_i - \boldsymbol{x}) \sum_{j=1}^n K_{\boldsymbol{H}}(\boldsymbol{X}_j - \boldsymbol{x})}$$
$$\cdot \mathbb{C}\mathrm{ov}[\sin(\Theta_i), \cos(\Theta_j) \mid \boldsymbol{\mathcal{X}}]$$
$$= \frac{\sum_{i=1}^n K_{\boldsymbol{H}}^2(\boldsymbol{X}_i - \boldsymbol{x}) c(\boldsymbol{X}_i)}{[\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{X}_i - \boldsymbol{x})]^2}$$
$$+ \frac{\sum_{i \neq j} K_{\boldsymbol{H}}(\boldsymbol{X}_i - \boldsymbol{x}) K_{\boldsymbol{H}}(\boldsymbol{X}_j - \boldsymbol{x}) C_{n,3}(\boldsymbol{X}_i, \boldsymbol{X}_j)}{[\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{X}_i - \boldsymbol{x})]^2}$$
$$= \frac{1}{n|\boldsymbol{H}| f(\boldsymbol{x})} \nu_0 [c(\boldsymbol{x}) + f(\boldsymbol{x}) C_3(\boldsymbol{x})]$$
$$+ o_{\mathbb{P}}\left(\frac{1}{n|\boldsymbol{H}|}\right). \tag{33}$$

On the other hand, the conditional covariance between $\hat{m}_{1,\boldsymbol{H}}(\boldsymbol{x}; 1)$ and $\hat{m}_{2,\boldsymbol{H}}(\boldsymbol{x}; 1)$ is

$$\mathbb{C}\mathrm{ov}[\hat{m}_{1,\boldsymbol{H}}(\boldsymbol{x}; 1), \hat{m}_{2,\boldsymbol{H}}(\boldsymbol{x}; 1) \mid \boldsymbol{\mathcal{X}}]$$
$$= \boldsymbol{e}_1^\top (\boldsymbol{X}_{\boldsymbol{x}}^\top \boldsymbol{W}_{\boldsymbol{x}} \boldsymbol{X}_{\boldsymbol{x}})^{-1} \boldsymbol{X}_{\boldsymbol{x}}^\top \boldsymbol{W}_{\boldsymbol{x}} \boldsymbol{\Sigma} \boldsymbol{W}_{\boldsymbol{x}} \boldsymbol{X}_{\boldsymbol{x}} (\boldsymbol{X}_{\boldsymbol{x}}^\top \boldsymbol{W}_{\boldsymbol{x}} \boldsymbol{X}_{\boldsymbol{x}})^{-1} \boldsymbol{e}_1,$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of $\sin(\Theta)$ and $\cos(\Theta)$, whose (i, j)-entry is $\boldsymbol{\Sigma}(i, j) = \mathbb{C}\mathrm{ov}[\sin(\Theta_i), \cos(\Theta_j)]$, $i, j = 1, \ldots, n$. Using (19), (20), (21), (22), (23) and (24), it follows that

$$\left(n^{-1} \boldsymbol{X}_{\boldsymbol{x}}^\top \boldsymbol{W}_{\boldsymbol{x}} \boldsymbol{X}_{\boldsymbol{x}}\right)^{-1}$$
$$= \begin{pmatrix} k_{1,n}(\boldsymbol{x}) & k_{2,n}^\top(\boldsymbol{x}) \\ k_{2,n}(\boldsymbol{x}) & k_{3,n}(\boldsymbol{x}) \end{pmatrix}^{-1}$$
$$= \begin{pmatrix} \frac{1}{f(\boldsymbol{x})} + o_{\mathbb{P}}(1) & \frac{-\boldsymbol{\nabla}^\top f(\boldsymbol{x})}{f^2(\boldsymbol{x})} + o_{\mathbb{P}}(\mathbf{1}_d^\top) \\ \frac{-\boldsymbol{\nabla} f(\boldsymbol{x})}{f^2(\boldsymbol{x})} + o_{\mathbb{P}}(\mathbf{1}_d) & \frac{1}{\mu_2 f(\boldsymbol{x}) \boldsymbol{H}^2} + o_{\mathbb{P}}(\boldsymbol{H} \mathbf{1}_{d \times d} \boldsymbol{H}) \end{pmatrix},$$

and that

$$\frac{1}{n^2} \boldsymbol{X}_{\boldsymbol{x}}^\top \boldsymbol{W}_{\boldsymbol{x}} \boldsymbol{\Sigma} \boldsymbol{W}_{\boldsymbol{x}} \boldsymbol{X}_{\boldsymbol{x}}$$
$$= \begin{pmatrix} s_{1,n}(\boldsymbol{x}) & s_{2,n}^\top(\boldsymbol{x}) \\ s_{2,n}(\boldsymbol{x}) & s_{3,n}(\boldsymbol{x}) \end{pmatrix}$$
$$= \frac{1}{n|\boldsymbol{H}|} \begin{pmatrix} \nu_0 f(\boldsymbol{x})[c(\boldsymbol{x}) + f(\boldsymbol{x}) C_3(\boldsymbol{x})] + o_{\mathbb{P}}(1) & o_{\mathbb{P}}(\mathbf{1}_d^\top) \\ o_{\mathbb{P}}(\mathbf{1}_d) & o_{\mathbb{P}}(\mathbf{1}_{d \times d}) \end{pmatrix}.$$

Consequently, by straightforward calculations, one gets that

$$\mathbb{C}\mathrm{ov}[\hat{m}_{1,\boldsymbol{H}}(\boldsymbol{x};1), \hat{m}_{2,\boldsymbol{H}}(\boldsymbol{x};1) \mid \boldsymbol{\mathcal{X}}]$$

$$= \frac{1}{n|\boldsymbol{H}|f(\boldsymbol{x})}\nu_0[c(\boldsymbol{x}) + f(\boldsymbol{x})C_3(\boldsymbol{x})]$$

$$+ o_{\mathbb{P}}\left(\frac{1}{n|\boldsymbol{H}|}\right). \tag{34}$$

Using (27), (28), (33) and (34), one gets that, for $p = 0, 1$,

$$\mathbb{V}\mathrm{ar}[\hat{m}_{\boldsymbol{H}}(\boldsymbol{x};p) \mid \boldsymbol{\mathcal{X}}] = \frac{1}{n|\boldsymbol{H}|}\frac{\nu_0}{f(\boldsymbol{x})}\frac{m_1^2(\boldsymbol{x})}{[m_1^2(\boldsymbol{x}) + m_2^2(\boldsymbol{x})]^2}$$

$$\cdot [s_2^2(\boldsymbol{x}) + f(\boldsymbol{x})C_2(\boldsymbol{x})]$$

$$+ \frac{1}{n|\boldsymbol{H}|}\frac{\nu_0}{f(\boldsymbol{x})}\frac{m_2^2(\boldsymbol{x})}{[m_1^2(\boldsymbol{x}) + m_2^2(\boldsymbol{x})]^2}$$

$$\cdot [s_1^2(\boldsymbol{x}) + f(\boldsymbol{x})C_1(\boldsymbol{x})]$$

$$- \frac{2}{n|\boldsymbol{H}|}\frac{\nu_0}{f(\boldsymbol{x})}\frac{m_1(\boldsymbol{x})m_2(\boldsymbol{x})}{[m_1^2(\boldsymbol{x}) + m_2^2(\boldsymbol{x})]^2}$$

$$\cdot [c(\boldsymbol{x}) + f(\boldsymbol{x})C_3(\boldsymbol{x})] + o_{\mathbb{P}}\left(\frac{1}{n|\boldsymbol{H}|}\right).$$

Notice that it holds that

$$m_1(\boldsymbol{x}) = f_1(\boldsymbol{x})\ell(\boldsymbol{x}) \quad \text{and} \quad m_2(\boldsymbol{x}) = f_2(\boldsymbol{x})\ell(\boldsymbol{x}). \tag{35}$$

Taking into account that $f_1^2(\boldsymbol{x}) + f_2^2(\boldsymbol{x}) = 1$, it can be easily deduced that $\ell(\boldsymbol{x}) = [m_1^2(\boldsymbol{x}) + m_2^2(\boldsymbol{x})]^{1/2}$. Therefore, using (25), (26), (29), (30), (31), (32) and (35), it follows that

$$m_1^2(\boldsymbol{x})[s_2^2(\boldsymbol{x}) + f(\boldsymbol{x})C_2(\boldsymbol{x})] + m_2^2(\boldsymbol{x})[s_1^2(\boldsymbol{x}) + f(\boldsymbol{x})C_1(\boldsymbol{x})]$$

$$- 2m_1(\boldsymbol{x})m_2(\boldsymbol{x})[c(\boldsymbol{x}) + f(\boldsymbol{x})C_3(\boldsymbol{x})]$$

$$= \ell^2(\boldsymbol{x})\sigma_1^2[1 + f(\boldsymbol{x})\rho_{c_1}].$$

Consequently, it can be directly obtained that

$$\mathbb{V}\mathrm{ar}[\hat{m}_{\boldsymbol{H}}(\boldsymbol{x};p) \mid \boldsymbol{\mathcal{X}}] = \frac{\nu_0\sigma_1^2[1 + f(\boldsymbol{x})\rho_{c_1}]}{n|\boldsymbol{H}|\ell^2(\boldsymbol{x})f(\boldsymbol{x})}$$

$$+ o_{\mathbb{P}}\left(\frac{1}{n|\boldsymbol{H}|}\right).$$