



Nonparametric multiple regression estimation for circular response

Andrea Meilán-Vila¹ · Mario Francisco-Fernández¹ · Rosa M. Crujeiras² · Agnese Panzera³

Received: 20 February 2020 / Accepted: 29 September 2020 / Published online: 12 October 2020
© Sociedad de Estadística e Investigación Operativa 2020

Abstract

Nonparametric estimators of a regression function with circular response and \mathbb{R}^d -valued predictor are considered in this work. Local polynomial estimators are proposed and studied. Expressions for the asymptotic conditional bias and variance of these estimators are derived, and some guidelines to select asymptotically optimal local bandwidth matrices are also provided. The finite sample behavior of the proposed estimators is assessed through simulations, and their performance is also illustrated with a real data set.

Keywords Linear–circular regression · Multiple regression · Local polynomial estimators

Mathematics Subject Classification 62G05 · 62G08 · 62G20 · 62H11

1 Introduction

New challenges on regression modeling appear when trying to describe relations between variables and some of them do not belong to an Euclidean space. This is

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11749-020-00736-w>) contains supplementary material, which is available to authorized users.

✉ Andrea Meilán-Vila
andrea.meilan@udc.es

¹ Research Group MODES, CITIC, Department of Mathematics, Faculty of Computer Science, Universidade da Coruña, Campus de Elviña s/n, 15071 A Coruña, Spain

² 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

³ Dipartimento di Statistica, Informatica, Applicazioni “G. Parenti”, Università degli Studi di Firenze, Viale Morgagni, 59, 50134 Firenze, Italy

the case for regression problems where some or all of the involved variables are circular ones. The special nature of circular data (points on the circumference of the unit circle; angles in $\mathbb{T} = [0, 2\pi)$) relies on their periodicity, which requires *ad hoc* statistical methods to analyze them. Circular statistics is an evolving discipline, and several statistical techniques for linear data now may claim their circular analogues. Comprehensive reviews on circular statistics (or more general, directional data) are provided in Fisher (1995), Jammalamadaka and SenGupta (2001) or Mardia and Jupp (2000). Some recent advances in directional statistics are collected in Ley and Verdebout (2017). Examples of circular data arise in many scientific fields such as biology, studying animal orientation (Batschelet 1981), environmental applications (SenGupta and Ugwuowo 2006) or oceanography (as in Wang et al. 2015, among others). When the circular variable is supposed to vary with respect to other covariates and the goal is to model such a relation, regression estimators for circular responses must be designed and analyzed.

Parametric regression approaches were originally considered in Fisher and Lee (1992) and Presnell et al. (1998), assuming a parametric (conditional) distribution model for the circular response variable. In this scenario, Euclidean covariates are supposed to influence the response via the parameters of the conditional distribution (e.g., through the location parameter, as the simplest case, or through location and concentration, if a von Mises distribution is chosen). Following the proposal in Presnell et al. (1998), Scapini et al. (2002) analyzed the orientation of two species of sand hoppers, considering parametric multiple regression methods for circular responses. A parametric multiple circular regression problem was also studied in Kim and SenGupta (2017), considering the functional relationship between a multivariate circular dependent variable and several circular covariates. Further, a multiple angular regression model for both angular and linear predictors was studied by Rivest et al. (2016). Maximum likelihood estimators for the parameters were derived for some von Mises error structures.

Beyond parametric restrictions, flexible approaches are also feasible in this context, just imposing some regularity conditions on the regression function, but avoiding the assumption of a specific parametric family for both the regression function and the conditional distribution. Nonparametric estimators of the regression function considering a model with a circular response and a single real-valued covariate were introduced in Di Marzio et al. (2013). The authors proposed *smooth* estimators for the regression function which are defined as the inverse tangent function of the ratio between two sample statistics, obtained as weighted sums of the sines and the cosines of the response observations, respectively. Specifically, they considered local constant and local linear weights.

The problem of nonparametrically estimating the conditional mean direction of a circular random variable, given a \mathbb{R}^d -valued covariate, is considered in this work. If the relation between both variables is viewed from a model-based approach, then our proposal aims to estimate the usual target regression function, given by the inverse tangent function of the ratio between the conditional expectations of the sine and cosine of the response variable. Our proposal considers two regression models for the sine and cosine components, which are indeed regression models with real-valued responses and d -dimensional covariates. Then, nonparametric estimators for the cir-

cular regression function are obtained as the inverse tangent function of the ratio of local polynomial estimators for the two regression functions of the sine and cosine models, respectively. The estimators obtained with this proposal generalize to both higher dimensions and higher polynomial degrees the proposals in Di Marzio et al. (2013). The approach of considering two flexible regression models for the sine and cosine components has been also explored in Jammalamadaka and Sarma (1993), where the objective is the estimation of the regression function in a model with circular response and circular covariate. In this case, the conditional expectations of the sine and the cosine of the response are approximated by trigonometric polynomials of a suitable degree. A similar approach has been also considered in Di Marzio et al. (2014), where the problem of nonparametrically estimating a regression function with spherical response and spherical covariate is addressed as a multi-output regression problem. In this case, each Cartesian coordinate of the spherical regression function is separately estimated.

This paper is organized as follows. In Sect. 2, the regression models for the sine and cosine components of the response are presented, jointly with a multiple regression model for the circular variable. Assuming that all these models simultaneously hold, certain relations between the first- and second-order moments of the involved errors are established. In Sect. 3, the nonparametric estimators of the regression function are proposed. Sections 3.1 and 3.2 contain the Nadaraya–Watson (NW) and local linear (LL) versions of these estimators, respectively, and include expressions for their asymptotic biases and variances. A local polynomial-type estimator with a general degree p , for the case of univariate predictor, is also analyzed in Sect. 3.3. The finite sample performance of the estimators is assessed through a simulation study provided in Sect. 4. Finally, Sect. 5 shows a real data application about sand hoppers orientation.

The proofs of all the theoretical results, along with some additional simulations experiments, are collected in the accompanying Supplementary Material.

2 Regression models for circular response

In this section, we will establish the rationale behind our estimation proposal. First, we will motivate the construction of our estimators, based on the expression of the conditional mean direction of a circular variable Θ given a d -dimensional covariate \mathbf{X} . Then, we will explain how our proposal can be related to a classical model-based approach, where the circular response variable admits a representation in terms of a regression function over the covariates plus a circular error term.

2.1 A general approach based on the conditional expectation

Let $\{(\mathbf{X}_i, \Theta_i)\}_{i=1}^n$ be a random sample from (\mathbf{X}, Θ) , where Θ is a circular random variable taking values on $\mathbb{T} = [0, 2\pi)$, and \mathbf{X} is a random variable with density f supported on $D \subseteq \mathbb{R}^d$. The dependence relation of Θ on \mathbf{X} can be modeled by the conditional mean direction of Θ given \mathbf{X} which, at a point $\mathbf{x} \in D$, is given by:

$$m(\mathbf{x}) = \text{atan2}[m_1(\mathbf{x}), m_2(\mathbf{x})], \tag{1}$$

where $m_1(\mathbf{x}) = E[\sin(\Theta) \mid \mathbf{X} = \mathbf{x}]$, $m_2(\mathbf{x}) = E[\cos(\Theta) \mid \mathbf{X} = \mathbf{x}]$ and the function $\text{atan2}(y, x)$ 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, respectively, having $\sin(\Theta)$ and $\cos(\Theta)$ as their responses. Specifically, we assume the models:

$$\sin(\Theta_i) = m_1(\mathbf{X}_i) + \xi_i, \quad i = 1, \dots, n, \tag{2}$$

and

$$\cos(\Theta_i) = m_2(\mathbf{X}_i) + \zeta_i, \quad i = 1, \dots, n, \tag{3}$$

where the ξ_i and the ζ_i are independent error terms, absolutely bounded by 1, satisfying $E(\xi \mid \mathbf{X} = \mathbf{x}) = E(\zeta \mid \mathbf{X} = \mathbf{x}) = 0$. Additionally, for every $\mathbf{x} \in D$, set $s_1^2(\mathbf{x}) = \text{Var}(\xi \mid \mathbf{X} = \mathbf{x})$, $s_2^2(\mathbf{x}) = \text{Var}(\zeta \mid \mathbf{X} = \mathbf{x})$ and $c(\mathbf{x}) = E(\xi\zeta \mid \mathbf{X} = \mathbf{x})$.

Considering models (2) and (3), a whole class of kernel-type estimators for $m(\mathbf{x})$ in (1), can be defined replacing in its expression the unknown functions $m_1(\mathbf{x})$ and $m_2(\mathbf{x})$ by suitable local polynomial estimators as follows:

$$\hat{m}_{\mathbf{H}}(\mathbf{x}; p) = \text{atan2}[\hat{m}_{1,\mathbf{H}}(\mathbf{x}; p), \hat{m}_{2,\mathbf{H}}(\mathbf{x}; p)], \tag{4}$$

where for any integer $p \geq 0$, $\hat{m}_{1,\mathbf{H}}(\mathbf{x}; p)$ and $\hat{m}_{2,\mathbf{H}}(\mathbf{x}; p)$ denote the p th-order local polynomial estimators (with bandwidth matrix \mathbf{H}) of $m_1(\mathbf{x})$ and $m_2(\mathbf{x})$, respectively. The special cases $p = 0$ and $p = 1$ yield a NW (or local constant)-type estimator and a LL-type estimator of $m(\mathbf{x})$, respectively.

It should be noted that models (2) and (3) can also be regarded as the components of a vector-valued regression model for the Cartesian coordinate representation of the circular response Θ . Hence, taking the representation of the circular response as the unit vector $[\cos(\Theta), \sin(\Theta)]$, these models amount to a regression model for vector-valued response whose error term is a random vector having zero conditional mean and conditional covariance matrix with diagonal entries $s_2^2(\mathbf{x})$ and $s_1^2(\mathbf{x})$, and off-diagonal entries both equal to $c(\mathbf{x})$. In this case, the dependence relation of $[\cos(\Theta), \sin(\Theta)]$ on \mathbf{X} can be modeled by the solution of the following minimization problem:

$$\arg \min_{\mathbf{u} \in \mathbb{R}^2: \|\mathbf{u}\|=1} E\{\|[\cos(\Theta), \sin(\Theta)] - \mathbf{u}\|^2 \mid \mathbf{X} = \mathbf{x}\},$$

where $\|\cdot\|$ stands for the Euclidean norm. The solution of this problem is given by the vector

$$\{ \|[m_2(\mathbf{x}), m_1(\mathbf{x})]\| \}^{-1} [m_2(\mathbf{x}), m_1(\mathbf{x})],$$

and its polar coordinate representation coincides with $m(\mathbf{x})$ as given in (1).

2.2 A model-based approach for polar representation

When trying to describe the relation between Θ and \mathbf{X} , apart from the approach described in the previous section, we can also focus directly on the polar coordinate representation of the response. With this perspective, using the random sample $\{(\mathbf{X}_i, \Theta_i)\}_{i=1}^n$, we assume the regression model:

$$\Theta_i = [m(\mathbf{X}_i) + \varepsilon_i] \pmod{2\pi}, \quad i = 1, \dots, n, \quad (5)$$

where \pmod stands for the modulo operation, and $\varepsilon_i, i = 1, \dots, n$, is an independent sample of a circular variable ε , satisfying $E[\sin(\varepsilon) \mid \mathbf{X} = \mathbf{x}] = 0$ and having finite concentration. In this setting, the circular regression function m in model (5) can be defined as the minimizer of the risk function $E\{1 - \cos[\Theta - m(\mathbf{X})] \mid \mathbf{X} = \mathbf{x}\}$, which is the analogue of the L2 risk. The minimizer of this cosine risk is given by (1). The assumption that model (5) simultaneously holds with the vector-valued regression model presented in the previous section leads to certain relations between the variances and covariances of the errors in models (2), (3) and (5), as will be described below.

Set $\ell(\mathbf{x}) = E[\cos(\varepsilon) \mid \mathbf{X} = \mathbf{x}]$, $\sigma_1^2(\mathbf{x}) = \text{Var}[\sin(\varepsilon) \mid \mathbf{X} = \mathbf{x}]$, $\sigma_2^2(\mathbf{x}) = \text{Var}[\cos(\varepsilon) \mid \mathbf{X} = \mathbf{x}]$ and $\sigma_{12}(\mathbf{x}) = E[\sin(\varepsilon)\cos(\varepsilon) \mid \mathbf{X} = \mathbf{x}]$. Then, using the sine and cosine addition formulas in model (5), it follows that, for $i = 1, \dots, n$:

$$\sin(\Theta_i) = \sin[m(\mathbf{X}_i)]\cos(\varepsilon_i) + \cos[m(\mathbf{X}_i)]\sin(\varepsilon_i) \quad (6)$$

and

$$\cos(\Theta_i) = \cos[m(\mathbf{X}_i)]\cos(\varepsilon_i) - \sin[m(\mathbf{X}_i)]\sin(\varepsilon_i). \quad (7)$$

Hence, defining $f_1(\mathbf{x}) = \sin[m(\mathbf{x})]$ and $f_2(\mathbf{x}) = \cos[m(\mathbf{x})]$ and applying conditional expectations in (6) and (7), it holds that:

$$m_1(\mathbf{x}) = f_1(\mathbf{x})\ell(\mathbf{x}) \quad \text{and} \quad m_2(\mathbf{x}) = f_2(\mathbf{x})\ell(\mathbf{x}). \quad (8)$$

Note that $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ correspond to the normalized versions of $m_1(\mathbf{x})$ and $m_2(\mathbf{x})$, respectively. Indeed, taking into account that $f_1^2(\mathbf{x}) + f_2^2(\mathbf{x}) = 1$, it can be easily deduced that $\ell(\mathbf{x}) = [m_1^2(\mathbf{x}) + m_2^2(\mathbf{x})]^{1/2}$. Hence, under model (5), $\ell(\mathbf{x})$ amounts to the mean resultant length of Θ given $\mathbf{X} = \mathbf{x}$, which, taking into account that $E[\sin(\varepsilon) \mid \mathbf{X} = \mathbf{x}] = 0$ is assumed, also corresponds to the mean resultant length of ε given $\mathbf{X} = \mathbf{x}$.

In addition, if models (2) and (3) simultaneously hold with model (5), equating expressions (2) and (6), and (3) and (7), and using (8), the errors in models (2) and (3) can be written as:

$$\xi_i = f_1(\mathbf{X}_i)[\cos(\varepsilon_i) - \ell(\mathbf{X}_i)] + f_2(\mathbf{X}_i)\sin(\varepsilon_i) \quad i = 1, \dots, n \quad (9)$$

and

$$\zeta_i = f_2(\mathbf{X}_i)[\cos(\varepsilon_i) - \ell(\mathbf{X}_i)] - f_1(\mathbf{X}_i)\sin(\varepsilon_i) \quad i = 1, \dots, n, \quad (10)$$

which satisfy that $E(\xi \mid \mathbf{X} = \mathbf{x}) = E(\zeta \mid \mathbf{X} = \mathbf{x}) = 0$. Then, the assumption that model (5) holds leads to a special case of error structure in models (2) and (3). As a consequence, the following explicit expressions for the conditional variances of the error terms involved in models (2) and (3), in terms of the conditional variances and covariance of the Cartesian coordinates of ε , can be obtained:

$$s_1^2(\mathbf{x}) = f_1^2(\mathbf{x})\sigma_2^2(\mathbf{x}) + 2f_1(\mathbf{x})f_2(\mathbf{x})\sigma_{12}(\mathbf{x}) + f_2^2(\mathbf{x})\sigma_1^2(\mathbf{x}), \tag{11}$$

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

as well as for the covariance between the error terms in (2) and (3):

$$c(\mathbf{x}) = f_1(\mathbf{x})f_2(\mathbf{x})\sigma_2^2(\mathbf{x}) - f_1^2(\mathbf{x})\sigma_{12}(\mathbf{x}) + f_2^2(\mathbf{x})\sigma_{12}(\mathbf{x}) - f_1(\mathbf{x})f_2(\mathbf{x})\sigma_1^2(\mathbf{x}). \tag{13}$$

3 Properties of kernel-type estimators

Asymptotic (conditional) bias and variance of the estimator given in (4) are derived in this section. We will focus on the cases in which $p = 0$ and $p = 1$. The asymptotic properties of the corresponding NW and LL estimators of $m_j(\mathbf{x})$, $j = 1, 2$, are firstly recalled just considering that models (2) and (3) hold. These results are then used to obtain the asymptotic properties of the estimator presented in (4) with polynomial degrees $p = 0$ and $p = 1$. When model (5) holds simultaneously with (2) and (3), some simplifications for the asymptotic bias and variance expressions can be obtained. Nevertheless, general results just assuming that (2) and (3) hold can be easily recovered from the stated theorems. Finally, asymptotic properties of local polynomial estimators with a higher-order p and $D \subseteq \mathbb{R}^d$ are also studied.

In what follows, $\nabla g(\mathbf{x})$ and $\mathcal{H}_g(\mathbf{x})$ will denote the vector of first-order partial derivatives and the Hessian matrix of a sufficiently smooth function g at \mathbf{x} , respectively. Moreover, for a vector $\mathbf{u} = (u_1, \dots, u_d)^T$ and an integrable function g , the multiple integral $\int \int \dots \int g(\mathbf{u}) du_1 du_2 \dots du_d$ will be simply denoted as $\int g(\mathbf{u}) d\mathbf{u}$. Finally, for any matrix \mathbf{A} , \mathbf{A}^T , $|\mathbf{A}|$ and $\text{tr}(\mathbf{A})$ denote its transpose, determinant and trace, respectively.

3.1 Nadaraya–Watson-type estimator

Considering models (2) and (3), local constant estimators for the regression functions m_j , $j = 1, 2$, at a given point $\mathbf{x} \in D \subseteq \mathbb{R}^d$, are, respectively, defined as:

$$\hat{m}_{j,\mathbf{H}}(\mathbf{x}; 0) = \begin{cases} \frac{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) \sin(\Theta_i)}{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})} & \text{if } j = 1, \\ \frac{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) \cos(\Theta_i)}{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})} & \text{if } j = 2, \end{cases} \tag{14}$$

where, for $\mathbf{u} \in \mathbb{R}^d$, $K_{\mathbf{H}}(\mathbf{u}) = |\mathbf{H}|^{-1}K(\mathbf{H}^{-1}\mathbf{u})$ is the rescaled version of a d -variate kernel function K , and \mathbf{H} is a $d \times d$ bandwidth matrix. The estimator $\hat{m}_{\mathbf{H}}(\mathbf{x}; 0)$ of $m(\mathbf{x})$, obtained by plugging (14) in (4), corresponds to the multivariate version of the local constant estimator proposed in Di Marzio et al. (2013).

Next, the asymptotic conditional bias and variance expressions for $\hat{m}_{\mathbf{H}}(\mathbf{x}; 0)$ are derived. First, using asymptotic theoretical results for the multivariate NW estimator (Härdle and Müller 2012), the asymptotic conditional bias and variance of $\hat{m}_{j,\mathbf{H}}(\mathbf{x}; 0)$, for $j = 1, 2$, are obtained. These preliminary results, along with the covariance between $\hat{m}_{1,\mathbf{H}}(\mathbf{x}; 0)$ and $\hat{m}_{2,\mathbf{H}}(\mathbf{x}; 0)$, are collected in Proposition 1. The following assumptions on the design density, the kernel function and the bandwidth matrix are required.

- (A1) The design density f is continuously differentiable at $\mathbf{x} \in D$, and satisfies $f(\mathbf{x}) > 0$. Moreover, s_j^2 and all second-order derivatives of the regression functions m_j , for $j = 1, 2$, are continuous at $\mathbf{x} \in D$, and $s_j^2(\mathbf{x}) > 0$.
- (A2) The kernel K is a spherically symmetric density function, twice continuously differentiable and with compact support (for simplicity with a nonzero value only if $\|\mathbf{u}\| \leq 1$). Moreover, $\int \mathbf{u}\mathbf{u}^T K(\mathbf{u})d\mathbf{u} = \mu_2(K)\mathbf{I}_d$, where $\mu_2(K) \neq 0$ and \mathbf{I}_d denotes the $d \times d$ identity matrix. It is also assumed that $R(K) = \int K^2(\mathbf{u})d\mathbf{u} < \infty$.
- (A3) The bandwidth matrix \mathbf{H} is symmetric and positive definite, with $\mathbf{H} \rightarrow 0$ and $n|\mathbf{H}| \rightarrow \infty$, as $n \rightarrow \infty$.

In assumption (A3), $\mathbf{H} \rightarrow 0$ means that every entry of \mathbf{H} goes to 0. Notice that, since \mathbf{H} is symmetric and positive definite, $\mathbf{H} \rightarrow 0$ is equivalent to $\lambda_{\max}(\mathbf{H}) \rightarrow 0$, where $\lambda_{\max}(\mathbf{H})$ denotes the maximum eigenvalue of \mathbf{H} . $|\mathbf{H}|$ is a quantity of order $O[\lambda_{\max}^d(\mathbf{H})]$ since $|\mathbf{H}|$ is equal to the product of all eigenvalues of \mathbf{H} .

Proposition 1 *Given the random sample $\{(\mathbf{X}_i, \Theta_i)\}_{i=1}^n$ from (\mathbf{X}, Θ) supported on $D \times \mathbb{T}$, assume models (2) and (3). Under assumptions (A1)–(A3), if \mathbf{x} is an interior point of the support of f , then, for $j = 1, 2$,*

$$\begin{aligned} E[\hat{m}_{j,\mathbf{H}}(\mathbf{x}; 0) - m_j(\mathbf{x}) \mid \mathbf{X}_1, \dots, \mathbf{X}_n] &= \frac{1}{2}\mu_2(K)\text{tr}[\mathbf{H}^2\mathcal{H}_{m_j}(\mathbf{x})] \\ &\quad + \frac{\mu_2(K)}{f(\mathbf{x})}\nabla^T m_j(\mathbf{x})\mathbf{H}^2\nabla f(\mathbf{x}) \\ &\quad + o_P[\text{tr}(\mathbf{H}^2)], \\ \text{Var}[\hat{m}_{j,\mathbf{H}}(\mathbf{x}; 0) \mid \mathbf{X}_1, \dots, \mathbf{X}_n] &= \frac{R(K)s_j^2(\mathbf{x})}{n|\mathbf{H}|f(\mathbf{x})} + o_P\left(\frac{1}{n|\mathbf{H}|}\right), \\ \text{Cov}[\hat{m}_{1,\mathbf{H}}(\mathbf{x}; 0), \hat{m}_{2,\mathbf{H}}(\mathbf{x}; 0) \mid \mathbf{X}_1, \dots, \mathbf{X}_n] &= \frac{R(K)c(\mathbf{x})}{n|\mathbf{H}|f(\mathbf{x})} + o_P\left(\frac{1}{n|\mathbf{H}|}\right). \end{aligned}$$

Now, using Proposition 1, the following theorem provides the asymptotic conditional bias and the asymptotic conditional variance of $\hat{m}_{\mathbf{H}}(\mathbf{x}; 0)$.

Theorem 1 *Given the random sample $\{(\mathbf{X}_i, \Theta_i)\}_{i=1}^n$ from (\mathbf{X}, Θ) supported on $D \times \mathbb{T}$, assume models (2), (3) and (5) hold. Under assumptions (A1)–(A3), the asymptotic*

conditional bias of estimator $\hat{m}_{\mathbf{H}}(\mathbf{x}; 0)$, at a fixed interior point \mathbf{x} in the support of f , is given by:

$$\begin{aligned} E[\hat{m}_{\mathbf{H}}(\mathbf{x}; 0) - m(\mathbf{x}) \mid \mathbf{X}_1, \dots, \mathbf{X}_n] &= \frac{1}{2} \mu_2(K) \text{tr}[\mathbf{H}^2 \mathcal{H}_m(\mathbf{x})] \\ &\quad + \frac{\mu_2(K)}{\ell(\mathbf{x}) f(\mathbf{x})} \nabla^T m(\mathbf{x}) \mathbf{H}^2 \nabla(\ell f)(\mathbf{x}) \\ &\quad + o_P[\text{tr}(\mathbf{H}^2)] \end{aligned}$$

and its asymptotic conditional variance is:

$$\text{Var}[\hat{m}_{\mathbf{H}}(\mathbf{x}; 0) \mid \mathbf{X}_1, \dots, \mathbf{X}_n] = \frac{R(K) \sigma_1^2(\mathbf{x})}{n |\mathbf{H}| \ell^2(\mathbf{x}) f(\mathbf{x})} + o_P\left(\frac{1}{n |\mathbf{H}|}\right).$$

Remark 1 Notice that the assumption that models (2) and (3) hold enable the definition of local estimators for $m_1(\mathbf{x})$ and $m_2(\mathbf{x})$, respectively, and then the definition of estimators of $m(\mathbf{x})$ having the form in (4). The further assumption that (5) holds lead to a special case where some simplifications in both the conditional bias and the conditional variance of the estimators are possible. In particular, as pointed out before, under model (5), in virtue of Eqs. (6) and (7), it holds that

$$[m_1^2(\mathbf{x}) + m_2^2(\mathbf{x})]^{1/2} = \ell(\mathbf{x}).$$

Further, if models (2), (3) and (5) simultaneously hold, due to the error structure in (9) and (10), and using (11), (12) and (13), it also holds that

$$\frac{m_1^2(\mathbf{x}) s_2^2(\mathbf{x}) + m_2^2(\mathbf{x}) s_1^2(\mathbf{x}) - 2m_1(\mathbf{x}) m_2(\mathbf{x}) c(\mathbf{x})}{m_1^2(\mathbf{x}) + m_2^2(\mathbf{x})} = \sigma_1^2(\mathbf{x}). \tag{15}$$

Results for the asymptotic bias and the asymptotic variance for the more general setting where just models (2) and (3) hold can be recovered by using the results of the above theorem with $[m_1^2(\mathbf{x}) + m_2^2(\mathbf{x})]^{1/2}$ in place of $\ell(\mathbf{x})$, in both the bias and variance expressions, and the left-hand side of (15) in place of $\sigma_1^2(\mathbf{x})$ in the variance expression.

Remark 2 Note that both the asymptotic conditional bias and the asymptotic conditional variance of $\hat{m}_{\mathbf{H}}(\mathbf{x}; 0)$ share the form of the corresponding quantities for the NW estimator of a regression function with real-valued response. In the asymptotic bias expression, both the gradient and the Hessian matrix of m refer to a circular regression function. In addition, under the assumption that models (2), (3) and (5) simultaneously hold, the asymptotic conditional variance depends on the ratio $\sigma_1^2(\mathbf{x})/\ell^2(\mathbf{x})$, accounting for the variability of the errors in model (5), which is related by (15) to the covariance and the variances of the error terms in models (2) and (3).

From Theorem 1, it is possible to derive the asymptotic (conditional) mean squared error (AMSE) of $\hat{m}_{\mathbf{H}}(\mathbf{x}; 0)$, defined as the sum of the square of the main term of the bias and the main term of the variance,

$$\begin{aligned}
\text{AMSE}[\hat{m}_{\mathbf{H}}(\mathbf{x}; 0)] &= \left\{ \frac{1}{2} \mu_2(K) \text{tr}[\mathbf{H}^2 \mathcal{H}_m(\mathbf{x})] + \frac{\mu_2(K)}{\ell(\mathbf{x})f(\mathbf{x})} \nabla^T m(\mathbf{x}) \mathbf{H}^2 \nabla(\ell f)(\mathbf{x}) \right\}^2 \\
&\quad + \frac{R(K) \sigma_1^2(\mathbf{x})}{n |\mathbf{H}| \ell^2(\mathbf{x}) f(\mathbf{x})} \\
&= \frac{1}{4} \mu_2^2(K) \text{tr}^2 \left(\mathbf{H}^2 \left\{ \frac{1}{\ell(\mathbf{x})f(\mathbf{x})} [\nabla(\ell f)(\mathbf{x}) \nabla^T m(\mathbf{x}) \right. \right. \\
&\quad \left. \left. + \nabla m(\mathbf{x}) \nabla^T(\ell f)(\mathbf{x})] + \mathcal{H}_m(\mathbf{x}) \right\} \right) + \frac{R(K) \sigma_1^2(\mathbf{x})}{n |\mathbf{H}| \ell^2(\mathbf{x}) f(\mathbf{x})}. \quad (16)
\end{aligned}$$

An asymptotically optimal local bandwidth matrix for $\hat{m}_{\mathbf{H}}(\mathbf{x}; 0)$ can be selected by minimizing (16) with respect to \mathbf{H} . Using Proposition 2.6 of Liu (2001), it can be obtained that this optimal local bandwidth is:

$$\begin{aligned}
\mathbf{H}_{\text{opt}}(\mathbf{x}) &= h^*(\mathbf{x}) \left[\tilde{\mathcal{B}}(\mathbf{x}) \right]^{-1/2} \\
&= \left[\frac{R(K) \sigma_1^2(\mathbf{x})}{nd \mu_2^2(K) f(\mathbf{x})} |\tilde{\mathcal{B}}(\mathbf{x})|^{1/2} \right]^{1/d+4} \cdot \left[\tilde{\mathcal{B}}(\mathbf{x}) \right]^{-1/2}, \quad (17)
\end{aligned}$$

where

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

with

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

Note that in the expression of $\mathbf{H}_{\text{opt}}(\mathbf{x})$, the matrix $\tilde{\mathcal{B}}(\mathbf{x})$ determines the shape and the orientation in the d -dimensional space of the covariate region which is used to locally compute the estimates. Such data regions are ellipsoids in \mathbb{R}^d , being the magnitude of the axes controlled by $\tilde{\mathcal{B}}(\mathbf{x})$. In the particular case of $\mathbf{H} = h \mathbf{I}_d$, the estimator $\hat{m}_{\mathbf{H}}(\mathbf{x}; 0)$, with \mathbf{x} being an interior point of the support, achieves an optimal convergence rate of $n^{-4/(d+4)}$, which is the same as the one for the multivariate NW estimator with real-valued response (Härdle and Müller 2012).

Despite deriving the previous explicit expression for the local optimal bandwidth (17), its use in practice is limited by the dependence on unknown functions, such as the design density f and the variance of the sine of the errors $\sigma_1^2(\mathbf{x})$. In addition, when the goal is to reconstruct the whole regression function and the focus is not only set on a specific point, it is more usual in practice to consider a global bandwidth for the estimation. An asymptotically optimal global bandwidth matrix \mathbf{H} could be obtained by minimizing a global error measurement (such as the integrated version of the AMSE). Again, this will depend on unknowns, leading to a non-trivial optimization problem, not being possible to obtain a closed-form solution. Alternatively, a suitable

adapted cross-validation criterion can be used to select the bandwidth matrix. This is indeed the bandwidth selection method employed in our numerical analysis and our real data application. More details will be provided in Sect. 4.

3.2 Local linear-type estimator

Similarly to the case when $p = 0$, the local linear case, corresponding to $p = 1$, is considered. Specifically, for models (2) and (3), the LL estimators of the regression functions $m_j, j = 1, 2$, at $\mathbf{x} \in D$, are defined by:

$$\hat{m}_{j,\mathbf{H}}(\mathbf{x}; 1) = \begin{cases} \mathbf{e}_1^T (\mathcal{X}_x^T \mathcal{W}_x \mathcal{X}_x)^{-1} \mathcal{X}_x^T \mathcal{W}_x \mathcal{S} & \text{if } j = 1, \\ \mathbf{e}_1^T (\mathcal{X}_x^T \mathcal{W}_x \mathcal{X}_x)^{-1} \mathcal{X}_x^T \mathcal{W}_x \mathcal{C} & \text{if } j = 2, \end{cases} \tag{18}$$

where \mathbf{e}_1 is a $(d + 1) \times 1$ vector having 1 in the first entry and 0 in all other entries, \mathcal{X}_x is a $n \times (d + 1)$ matrix having $(1, (\mathbf{X}_i - \mathbf{x})^T)$ as its i th row, $\mathcal{W}_x = \text{diag}\{K_{\mathbf{H}}(\mathbf{X}_1 - \mathbf{x}), \dots, K_{\mathbf{H}}(\mathbf{X}_n - \mathbf{x})\}$, $\mathcal{S} = (\sin(\Theta_1), \dots, \sin(\Theta_n))^T$ and $\mathcal{C} = (\cos(\Theta_1), \dots, \cos(\Theta_n))^T$.

Using asymptotic results for the multivariate local linear estimator (Ruppert and Wand 1994), the asymptotic conditional bias and variance of $\hat{m}_{j,\mathbf{H}}(\mathbf{x}; 1), j = 1, 2$, can be obtained. These expressions, along with the covariance between $\hat{m}_{1,\mathbf{H}}(\mathbf{x}; 1)$ and $\hat{m}_{2,\mathbf{H}}(\mathbf{x}; 1)$, are provided in the following result.

Proposition 2 *Given the random sample $\{(\mathbf{X}_i, \Theta_i)\}_{i=1}^n$ from (\mathbf{X}, Θ) supported on $D \times \mathbb{T}$, assume models (2) and (3). Under assumptions (A1)–(A3), if \mathbf{x} is an interior point of the support of f , then, for $j = 1, 2$,*

$$\begin{aligned} E[\hat{m}_{j,\mathbf{H}}(\mathbf{x}; 1) - m_j(\mathbf{x}) \mid \mathbf{X}_1, \dots, \mathbf{X}_n] &= \frac{1}{2} \mu_2(K) \text{tr}[\mathbf{H}^2 \mathcal{H}_{m_j}(\mathbf{x})] \\ &\quad + o_P[\text{tr}(\mathbf{H}^2)], \\ \text{Var}[\hat{m}_{j,\mathbf{H}}(\mathbf{x}; 1) \mid \mathbf{X}_1, \dots, \mathbf{X}_n] &= \frac{R(K) s_j^2(\mathbf{x})}{n |\mathbf{H}| f(\mathbf{x})} + o_P\left(\frac{1}{n |\mathbf{H}|}\right), \\ \text{Cov}[\hat{m}_{1,\mathbf{H}}(\mathbf{x}; 1), \hat{m}_{2,\mathbf{H}}(\mathbf{x}; 1) \mid \mathbf{X}_1, \dots, \mathbf{X}_n] &= \frac{R(K) c(\mathbf{x})}{n |\mathbf{H}| f(\mathbf{x})} + o_P\left(\frac{1}{n |\mathbf{H}|}\right). \end{aligned}$$

The estimator $\hat{m}_{\mathbf{H}}(\mathbf{x}; 1)$ of $m(\mathbf{x})$, obtained by plugging (18) in (4), corresponds to the multivariate version of the local linear estimator proposed in Di Marzio et al. (2013). The following theorem provides the asymptotic conditional bias and the asymptotic conditional variance of $\hat{m}_{\mathbf{H}}(\mathbf{x}; 1)$.

Theorem 2 *Given the random sample $\{(\mathbf{X}_i, \Theta_i)\}_{i=1}^n$ from (\mathbf{X}, Θ) supported on $D \times \mathbb{T}$, assume models (2), (3) and (5) hold. Under assumptions (A1)–(A3), the asymptotic conditional bias of estimator $\hat{m}_{\mathbf{H}}(\mathbf{x}; 1)$, with \mathbf{x} being a fixed interior point in the support of f , is given by:*

$$\begin{aligned} E[\hat{m}_{\mathbf{H}}(\mathbf{x}; 1) - m(\mathbf{x}) \mid \mathbf{X}_1, \dots, \mathbf{X}_n] &= \frac{1}{2} \mu_2(K) \text{tr}[\mathbf{H}^2 \mathcal{H}_m(\mathbf{x})] \\ &\quad + \frac{\mu_2(K)}{\ell(\mathbf{x})} \nabla^T m(\mathbf{x}) \mathbf{H}^2 \nabla \ell(\mathbf{x}) + o_P[\text{tr}(\mathbf{H}^2)] \end{aligned}$$

and its asymptotic conditional variance is:

$$\text{Var}[\hat{m}_{\mathbf{H}}(\mathbf{x}; 1) \mid \mathbf{X}_1, \dots, \mathbf{X}_n] = \frac{R(K) \sigma_1^2(\mathbf{x})}{n |\mathbf{H}| \ell^2(\mathbf{x}) f(\mathbf{x})} + o_P\left(\frac{1}{n |\mathbf{H}|}\right).$$

Remark 3 Notice that the same comments included in Remark 1 also apply for Theorem 2.

Remark 4 Estimators $\hat{m}_{\mathbf{H}}(\mathbf{x}; 0)$ and $\hat{m}_{\mathbf{H}}(\mathbf{x}; 1)$ have the same leading terms in their asymptotic conditional variances, while their asymptotic conditional biases, also being of the same order, have different leading terms. In particular, the main term of the asymptotic conditional bias of $\hat{m}_{\mathbf{H}}(\mathbf{x}; 1)$ does not depend on the design density, f . Moreover, as a consequence of its definition, the LL-type estimator, differently from the NW-type one, automatically adapts to boundary regions, in the sense that for compactly supported f , the asymptotic conditional bias has the same order both for the interior and for the boundary of the support of f (Ruppert and Wand 1994).

Remark 5 For $d = 1$, asymptotic results for estimators having the same form as the univariate version of estimator (4) with $p = 0$ and $p = 1$, are provided in Di Marzio et al. (2013). Despite they used slightly different formulations for their nonparametric estimators, their results, at interior points, can be directly compared with those obtained in Theorems 1 and 2. This correspondence is immediately clear for the asymptotic bias terms. For the asymptotic variance, the equivalence between the expressions can be obtained considering the relations between the variance of the error term in model (5) with the variance of the error terms in models (2) and (3), as stated in (15).

As a consequence of Theorem 2, and similarly to the NW case, an asymptotically optimal local bandwidth can also be obtained for $\hat{m}_{\mathbf{H}}(\mathbf{x}; 1)$, which coincides with (17), but taking $\mathcal{B}(\mathbf{x}) = \ell^{-1}(\mathbf{x})[\nabla \ell(\mathbf{x}) \nabla^T m(\mathbf{x}) + \nabla m(\mathbf{x}) \nabla^T \ell(\mathbf{x})] + \mathcal{H}_m(\mathbf{x})$.

3.3 Higher-order polynomials

Asymptotic theory on local polynomial estimators (Fan and Gijbels, 1996) can be used to generalize the previous results to the case of an arbitrary polynomial degree p . Similar arguments to those used to prove Theorems 1 and 2, can be applied to derive that the conditional bias of the p th-order polynomial-type estimator given in (4) will be of order $O_P\{[\text{tr}(\mathbf{H}^2)]^{(p+1)/2}\}$. Moreover, if p is even, f has a continuous derivative in a neighborhood of \mathbf{x} , and \mathbf{x} is an interior point of the support of the design density f , then the bias will be of order $O_P\{[\text{tr}(\mathbf{H}^2)]^{(p/2+1)}\}$. Here, as in Ruppert and Wand (1994), we will only focus on the case $d = 1$ to analyze asymptotically the nonparametric regression estimator given in (4) for $p > 1$. In particular, the p th

degree local polynomial estimators for $m_j, j = 1, 2$, at $x \in D \subseteq \mathbb{R}$, are:

$$\hat{m}_{j,h}(x; p) = \begin{cases} \mathbf{e}_1^T (\mathcal{X}_{x,p}^T \mathcal{W}_x \mathcal{X}_{x,p})^{-1} \mathcal{X}_{x,p}^T \mathcal{W}_x \mathcal{S} & \text{if } j = 1, \\ \mathbf{e}_1^T (\mathcal{X}_{x,p}^T \mathcal{W}_x \mathcal{X}_{x,p})^{-1} \mathcal{X}_{x,p}^T \mathcal{W}_x \mathcal{C} & \text{if } j = 2, \end{cases} \tag{19}$$

where, in this case, \mathbf{e}_1 is a $(p + 1) \times 1$ vector having 1 in the first entry and zero elsewhere, $\mathcal{X}_{x,p}$ is for $n \times p$ matrix with the (i, k) th entry equal to $(X_i - x)^{k-1}$, and \mathcal{W}_x is a diagonal matrix of order n with (i, i) th entry equal to $K_h(X_i - x)$, where $K_h(u) = 1/h K(u/h)$, being K a univariate kernel function, and h the bandwidth or smoothing parameter. In this univariate framework, the p th degree local polynomial-type estimator of m at x , denoted by $\hat{m}_h(x; p)$, has the same expression as the one given in (4), but using estimators $\hat{m}_{j,h}(x; p), j = 1, 2$, defined in (19), as the arguments of the atan2 function.

Let $K_{(p)}$ be the equivalent kernel function defined in Lejeune and Sarda (1992), which is a kernel of order $p + 2$ when p is even and of order $p + 1$ otherwise. Let $\mu_j(K_{(p)})$ and $R(K_{(p)})$ denote the moment of order j and the roughness of $K_{(p)}$, respectively. Under suitable adaptations of assumptions (A1)–(A3) to the univariate case and using asymptotic results for local polynomial estimators of an arbitrary order p , the asymptotic conditional bias and variance of $\hat{m}_{j,h}(x; p), j = 1, 2$, can be obtained. In the following theorems, we derive the asymptotic bias and variance expressions of $\hat{m}_h(x; p), x \in D$, only for $p = 2$ and $p = 3$. However, following similar arguments, these results could be extended with tedious calculations for higher-order polynomial degrees. It should be noted that for local polynomial regression in an Euclidean context, Fan and Gijbels (1996) recommended the use of polynomial orders $p = 1$ or $p = 3$.

Theorem 3 *Let $\{(X_i, \Theta_i)\}_{i=1}^n$ be a random sample from (X, Θ) supported on $D \times \mathbb{T}$, with $D \subseteq \mathbb{R}$, assume models (2), (3) and (5) hold (with $d = 1$), and let x be an interior point of the support of the design density f . Under assumptions (A1)–(A3) (adapted for $d = 1$) and assuming that $m_j, j = 1, 2$, admits continuous derivatives up to order four in a neighborhood of x , then,*

$$\begin{aligned} E[\hat{m}_h(x; 2) - m(x) \mid X_1, \dots, X_n] &= \frac{h^4 \mu_4(K_{(2)}) f^{(1)}(x)}{3! f(x)} [m^{(3)}(x) + a(x)] \\ &\quad + \frac{h^4 \mu_4(K_{(2)})}{4!} [m^{(4)}(x) + b(x)] + o_P(h^4) \end{aligned}$$

and

$$\text{Var}[\hat{m}_h(x; 2) \mid X_1, \dots, X_n] = \frac{R(K_{(2)})}{nh \ell^2(x) f(x)} \sigma_1^2(x) + o_P\left(\frac{1}{nh}\right),$$

where

$$a(x) = \frac{2\ell^{(2)}(x)m^{(1)}(x) + 4\ell^{(1)}(x)m^{(2)}(x)}{\ell(x)} + \frac{m_2^{(2)}(x)m_1^{(1)}(x) - m_1^{(2)}(x)m_2^{(1)}(x) + 2\ell^{(1)2}(x)m^{(1)}(x)}{\ell^2(x)}$$

and

$$b(x) = \frac{2\ell^{(3)}(x)m^{(1)}(x) + 6\ell^{(1)}(x)m^{(3)}(x) + 6\ell^{(2)}(x)m^{(2)}(x)}{\ell(x)} + \frac{2m_2^{(3)}(x)m_1^{(1)}(x) - 2m_1^{(3)}(x)m_2^{(1)}(x)}{\ell^2(x)} + \frac{6\ell^{(1)2}(x)m^{(2)}(x) + 6\ell^{(1)}(x)\ell^{(2)}(x)m^{(1)}(x)}{\ell^2(x)}.$$

Theorem 4 Let $\{(X_i, \Theta_i)\}_{i=1}^n$ be a random sample from (X, Θ) supported on $D \times \mathbb{T}$, with $D \subseteq \mathbb{R}$, assume models (2), (3) and (5) hold (with $d = 1$), and let x be an interior point of the support of the design density f . Under assumptions (A1)–(A3) (adapted for $d = 1$) and assuming that m_j , $j = 1, 2$, admits continuous derivatives up to order five in a neighborhood of x , then,

$$E[\hat{m}_h(x; 3) - m(x) \mid X_1, \dots, X_n] = \frac{h^4 \mu_4(K_{(3)})}{4!} [m^{(4)}(x) + b(x)] + o_P(h^4)$$

and

$$\text{Var}[\hat{m}_h(x; 3) \mid X_1, \dots, X_n] = \frac{R(K_{(3)})}{nh\ell^2(x)f(x)} \sigma_1^2(x) + o_P\left(\frac{1}{nh}\right).$$

Remark 6 Similar comments to those included in Remark 1 can be considered for Theorems 3 and 4.

4 Simulation study

In order to illustrate the performance of the estimators proposed in Sect. 3, a simulation study considering different scenarios and model (5) is carried out for $d = 2$ (that is, considering a circular response and a bidimensional covariate). For each scenario, 500 samples of size n ($n = 64, 100, 225$ and 400) are generated on a bidimensional regular grid in the unit square considering the following regression models, for $i = 1, \dots, n$:

$$\text{M1. } \Theta_i = [\text{atan2}(6X_{i1}^5 - 2X_{i1}^3 - 1, -2X_{i2}^5 - 3X_{i2} - 1) + \varepsilon_i] \pmod{2\pi},$$

$$\text{M2. } \Theta_i = [\text{acos}(X_{i1}^5 - 1) + \frac{3}{2} \text{asin}(X_{i2}^3 - X_{i2} + 1) + \varepsilon_i] \pmod{2\pi},$$

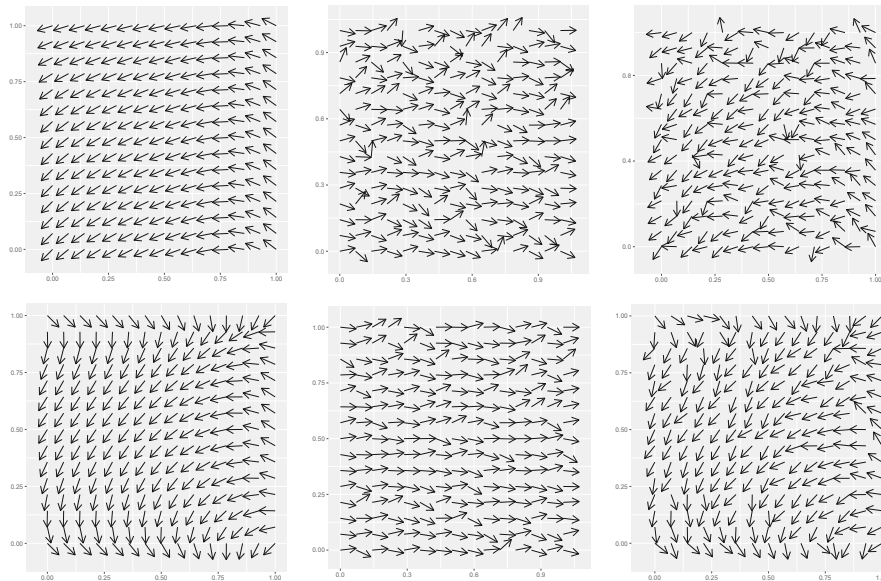


Fig. 1 Illustration of model generation (model M1: top row; model M2: bottom row) on a 15×15 grid. In left panels, regression functions evaluated at the grid points. In center panels, independent errors from a von Mises distribution with zero mean and concentration $\kappa = 5$, for model M1, and $\kappa = 15$, for model M2. In right panels, random response variables obtained by adding the two previous plots

where $\{(X_{i1}, X_{i2})\}_{i=1}^n$ denotes a sample of the bidimensional covariate $\mathbf{X} = (X_1, X_2)$, and the circular errors ε_i are drawn from a von Mises distribution $vM(0, \kappa)$ with different values of κ (5, 10 and 15).

Figure 1 shows two realizations of simulated data (model M1 in top row and model M2 in bottom row). In both cases, the sample size is $n = 225$. Left plots show the regression functions evaluated in the regularly spaced sample $\{(X_{i1}, X_{i2})\}_{i=1}^n$. Central panels present the random errors generated from a von Mises distribution with zero mean direction and concentration $\kappa = 5$, for model M1, and $\kappa = 15$, for model M2. Right panels show the values of the response variables, obtained adding regression functions and circular errors. It can be seen that the errors in the top row, corresponding to $\kappa = 5$, present more variability than the ones generated with $\kappa = 15$.

Numerical and graphical outputs summarize the finite sample performance of NW- and LL-type estimators in the different scenarios. The bandwidth matrix is chosen by cross-validation, selecting \mathbf{H} that minimizes the function:

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

where $\hat{m}_{\mathbf{H}}^{(i)}(\cdot; p)$ stands for the NW-type estimator ($p = 0$) or the LL-type estimator ($p = 1$), computed using all observations except $(\mathbf{X}_i, \vartheta_i)$. Taking into account the type of regression functions considered in models M1 and M2 and to speed up the computing

Table 1 Average error (over 500 replicates) of the CASE given in (20), for regression model M1, using NW- and LL-type estimators

| κ | n | NW | | LL | |
|----------|-----|-------------------|---------------------|-------------------|---------------------|
| | | \mathbf{H}_{CV} | \mathbf{H}_{CASE} | \mathbf{H}_{CV} | \mathbf{H}_{CASE} |
| 5 | 64 | 0.0610 | 0.0152 | 0.0672 | 0.0147 |
| | 100 | 0.0280 | 0.0111 | 0.0358 | 0.0100 |
| | 225 | 0.0124 | 0.0066 | 0.0158 | 0.0051 |
| | 400 | 0.0075 | 0.0047 | 0.0091 | 0.0033 |
| 10 | 64 | 0.0094 | 0.0092 | 0.0071 | 0.0066 |
| | 100 | 0.0102 | 0.0072 | 0.0055 | 0.0043 |
| | 225 | 0.0065 | 0.0042 | 0.0028 | 0.0026 |
| | 400 | 0.0042 | 0.0029 | 0.0019 | 0.0016 |
| 15 | 64 | 0.0182 | 0.0072 | 0.0201 | 0.0056 |
| | 100 | 0.0091 | 0.0054 | 0.0110 | 0.0041 |
| | 225 | 0.0046 | 0.0032 | 0.0050 | 0.0021 |
| | 400 | 0.0032 | 0.0023 | 0.0029 | 0.0014 |

Errors are generated from a von Mises distribution with different concentration parameters ($\kappa = 5, 10, 15$). Bandwidth matrix is selected by cross-validation, \mathbf{H}_{CV} . Additionally, results when using the optimal bandwidth \mathbf{H}_{CASE} are also included

times, in this simulation study, the bandwidth matrix is restricted to be diagonal with possibly different elements. A multivariate Epanechnikov kernel is considered for simulations.

Table 1 shows, for model M1 and in the different scenarios, the average (over the 500 replicates) of the circular average squared error (CASE), defined as (Kim and SenGupta 2017):

$$\text{CASE}[\hat{m}_{\mathbf{H}}(\cdot; p)] = \frac{1}{n} \sum_{i=1}^n \{1 - \cos [m(\mathbf{X}_i) - \hat{m}_{\mathbf{H}}(\mathbf{X}_i; p)]\}, \quad (20)$$

with $p = 0$ (NW) and $p = 1$ (LL), when \mathbf{H} is selected by cross-validation. For comparative purposes, the diagonal optimal bandwidth matrix \mathbf{H}_{CASE} minimizing (20) (obtained by intensive search) is also computed. The corresponding averages of the minimum values of the CASE are also included in Table 1. It can be seen that the average errors decrease when the sample size increase, and it is smaller for the LL-type estimator for most of the scenarios. Additionally, as expected, results are generally better when the error concentration gets larger. Average errors of the CASE for model M2 are shown in Table 2.

Numerical outputs are completed with some additional plots. As an illustration of the correct performance of NW- and LL-type estimators, Fig. 2 shows the theoretical regression functions for models M1 and M2 (left panels) and the corresponding average, over 500 replicates, of the estimates, using the specific scenarios considered in Fig. 1 (NW and LL estimates in the center and right panels, respectively). Notice that, for comparison purposes, the theoretical regression functions are plotted in a 100×100

Table 2 Average error (over 500 replicates) of the CASE given in (20), for regression model M2, using NW- and LL-type estimators

| κ | n | NW | | LL | |
|----------|-----|-------------------|---------------------|-------------------|---------------------|
| | | \mathbf{H}_{CV} | \mathbf{H}_{CASE} | \mathbf{H}_{CV} | \mathbf{H}_{CASE} |
| 5 | 64 | 0.0638 | 0.0303 | 0.0684 | 0.0209 |
| | 100 | 0.0330 | 0.0239 | 0.0369 | 0.0154 |
| | 225 | 0.0190 | 0.0158 | 0.0170 | 0.0089 |
| | 400 | 0.0141 | 0.0120 | 0.0102 | 0.0061 |
| 10 | 64 | 0.0297 | 0.0184 | 0.0315 | 0.0118 |
| | 100 | 0.0181 | 0.0151 | 0.0172 | 0.0091 |
| | 225 | 0.0131 | 0.0106 | 0.0085 | 0.0054 |
| | 400 | 0.0109 | 0.0086 | 0.0054 | 0.0038 |
| 15 | 64 | 0.0198 | 0.0139 | 0.0206 | 0.0088 |
| | 100 | 0.0138 | 0.0116 | 0.0118 | 0.0068 |
| | 225 | 0.0114 | 0.0087 | 0.0061 | 0.0041 |
| | 400 | 0.0100 | 0.0075 | 0.0041 | 0.0029 |

Errors are generated from a von Mises distribution with different concentration parameters ($\kappa = 5, 10, 15$). Bandwidth matrix is selected by cross-validation, \mathbf{H}_{CV} . Additionally, results when using the optimal bandwidth \mathbf{H}_{CASE} are also included

regular grid of the explanatory variables (the same grid where the estimations were computed). Plots in the top row present the results for the data generated from model M1 and those in the bottom row for model M2. Although both estimators have a similar and correct behavior, the LL-type estimator seems to show a slightly better performance, at least, for these samples. More reliable comparisons between NW- and LL-type estimators can be performed computing the circular bias (CB), the circular variance (CVAR) and the circular mean squared error (CMSE) for both estimators, in a grid of values of the explanatory variables. These quantities, at a point \mathbf{x} , are defined as:

$$CB[\hat{m}_{\mathbf{H}}(\mathbf{x}; p)] = E\{\sin[\hat{m}_{\mathbf{H}}(\mathbf{x}; p) - m(\mathbf{x})]\}, \tag{21}$$

$$CVAR[\hat{m}_{\mathbf{H}}(\mathbf{x}; p)] = E\{1 - \cos[\hat{m}_{\mathbf{H}}(\mathbf{x}; p) - \mu(\mathbf{x}; \mathbf{p})]\}, \tag{22}$$

$$CMSE[\hat{m}_{\mathbf{H}}(\mathbf{x}; p)] = E\{1 - \cos[m(\mathbf{x}) - \hat{m}_{\mathbf{H}}(\mathbf{x}; p)]\}, \tag{23}$$

where $\mu(\mathbf{x}; \mathbf{p})$ in CVAR denotes the circular mean of $\hat{m}_{\mathbf{H}}(\mathbf{x}; p)$. Notice that, using Taylor expansions, Eqs. (21), (22) and (23) are equivalent to the Euclidean versions of these expressions (Kim and SenGupta 2017).

Figures 3 and 4 show, in the scenarios considered in Fig. 1, the CB, CVAR and CMSE computed in a 100×100 regular grid of the explanatory variables, when using NW (top row) and LL (bottom row)-type estimators, for models M1 and M2, respectively. The expectations in (21), (22) and (23) are approximated by the averages over the 500 replicates generated. It can be seen that the NW-type estimator ($p = 0$) provides larger biases and smaller variances than the LL-type estimator ($p = 1$) in both settings. However, the CMSE is smaller for the LL fit in most of the grid points.

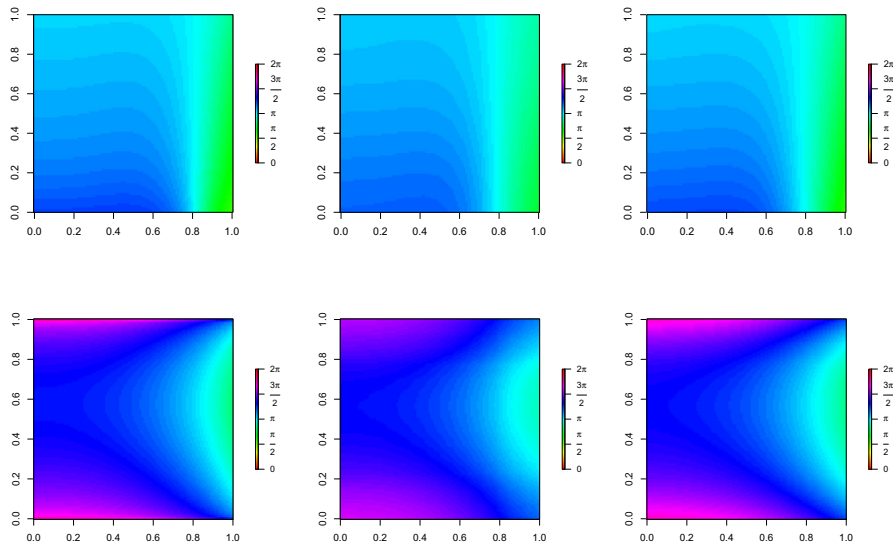


Fig. 2 Theoretical regression function (left), jointly with the average, over 500 replicates, of NW (center) and LL (right)-type estimates, using the specific scenarios considered in Fig. 1, for model M1 (top row) and model M2 (bottom row)

Similar results for the CB, CVAR and CMSE for both estimators were obtained in other scenarios.

5 Real data example

A real data example is presented in order to illustrate the application of the proposed estimators. Based on the simulation study, where the LL-type estimator presented a slightly better performance than the NW one, just results corresponding to $\hat{m}_{\mathbf{H}}(\mathbf{x}; 1)$ are provided for real data. Orientations of two species of sand hoppers (*Talorchestia brito* and *Talitrus saltator*) on the Zouara beach in northwestern Tunisia are considered. Following the proposal in Presnell et al. (1998), these observations were analyzed in Scapini et al. (2002). They used a parametric approach that assumes a projected normal distribution for the scape directions and the corresponding parameters (circular mean and mean resultant vector) depend on the explanatory variables through a linear model. We refer to Scapini et al. (2002) and Marchetti and Scapini (2003) for details on the experiment, a thorough data analysis and sound biological conclusions. Dealing with the same data set, in Marchetti and Scapini (2003), the authors conclude that the orientation is different for the two sexes (males and females) and they explicitly mention that nonparametric smoothers are flexible tools that may suggest unexpected features of the data. So, the illustration with our proposal is a first attempt to analyze this data set with nonparametric tools in order to check how orientation (in degrees) behaves when temperature (in Celsius degrees) and (relative) humidity (in percentage) are included as covariates. For illustration purposes, only observations corresponding

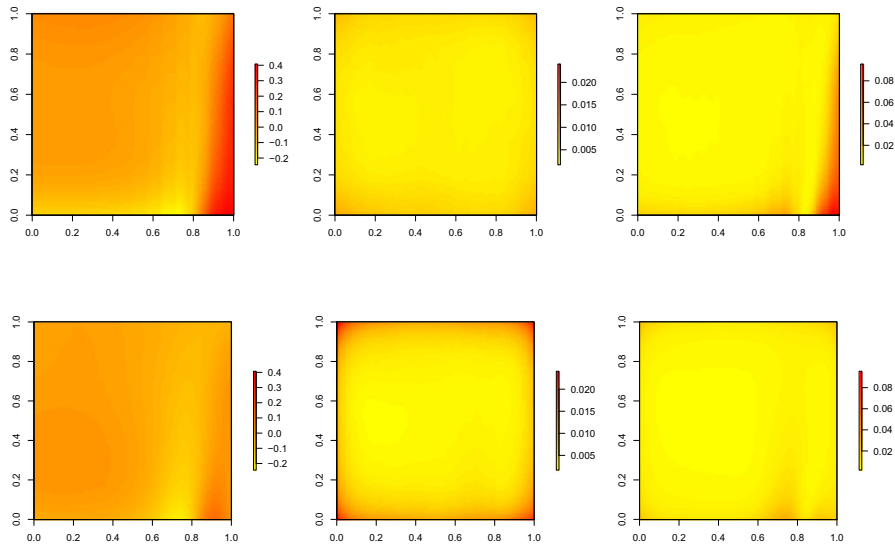


Fig. 3 Circular bias (left), circular variance (center) and CMSE (right) surfaces for model M1 for a 100×100 regular grid, using NW (top row) and LL (bottom row) fits. $n = 225$ and von Mises errors with zero mean and $\kappa = 5$

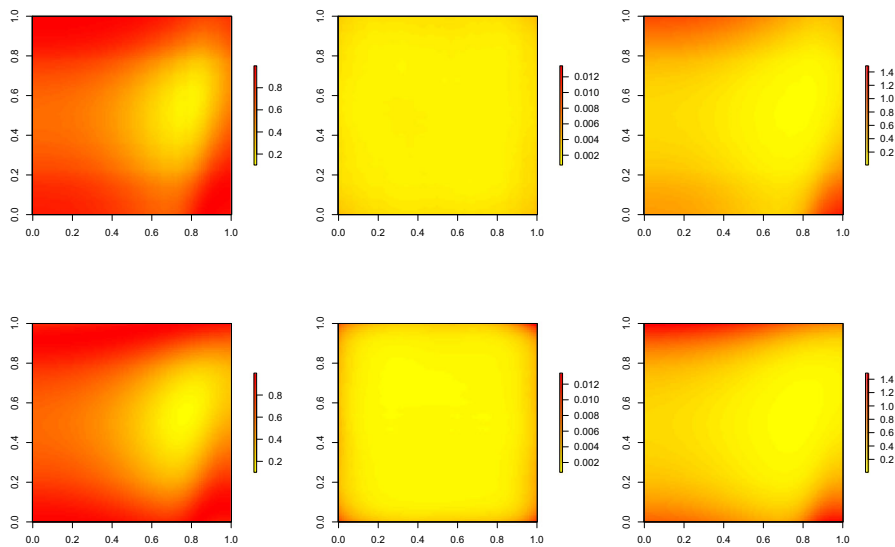


Fig. 4 Circular bias (left), circular variance (center) and CMSE (right) surfaces for model M2 for a 100×100 regular grid, using NW (top row) and LL (bottom row) fits. $n = 225$ and von Mises errors with zero mean and $\kappa = 15$

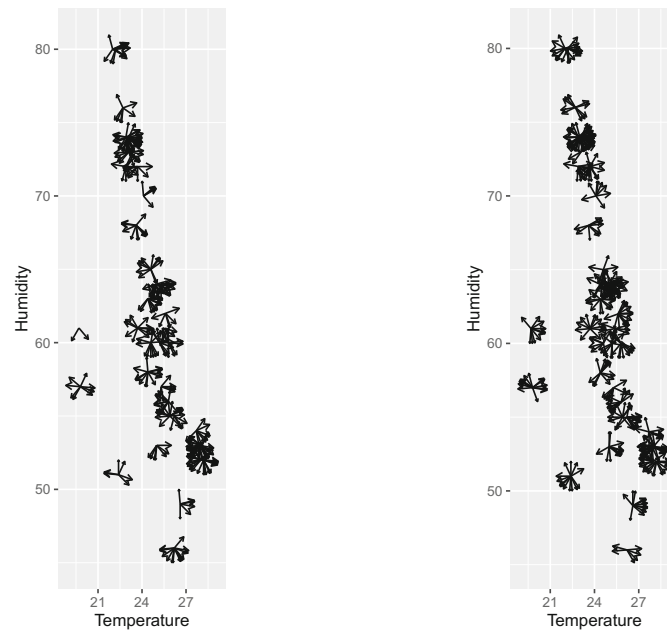


Fig. 5 Observed orientation of male (left) and female (right) sand hoppers as a function of temperature and relative humidity

to (relative) humidity values larger than 45% are considered in this analysis. The corresponding data sets are plotted in Fig. 5 (males in the left panel and females in the right panel), being the sample sizes $n = 330$ and $n = 404$, for male and female sand hoppers, respectively.

Figure 6 shows the LL-type estimates for male (left) and female (right) mean orientations, considering temperature (horizontal axis) and relative humidity (vertical axis) as covariates. Note that measurements of temperature and humidity are the same for males and females, given that these values correspond to experimental conditions. In this example, unlike in the simulation experiments, the CV bandwidth matrix has been 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). Using the initial bandwidth matrix $\mathbf{H}_{\text{init}} = 1.5 \cdot \text{diag} \{ \hat{\sigma}_{X_1}, \hat{\sigma}_{X_2} \}$, the algorithm converged to

$$\mathbf{H}_{\text{CV}}^m = \begin{bmatrix} 2.7781 & 0.0001 \\ 0.0001 & 15.2529 \end{bmatrix},$$

for males, and to

$$\mathbf{H}_{\text{CV}}^f = \begin{bmatrix} 4.0930 & -0.0009 \\ -0.0009 & 13.1937 \end{bmatrix},$$

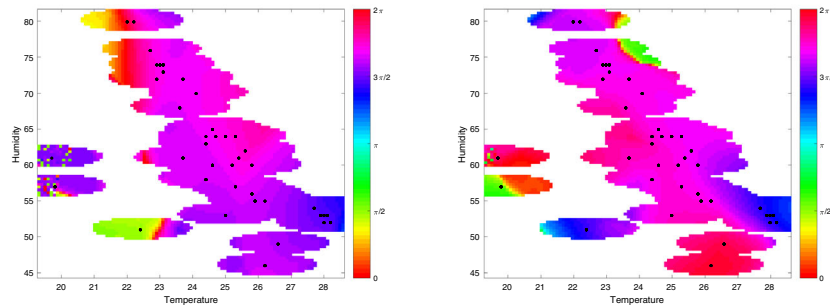


Fig. 6 Estimates of the mean orientation of males (left) and females (right) sand hoppers, considering a LL-type estimator with a cross-validation bandwidth matrix. Horizontal axis: temperature, in °C. Vertical axis: relative humidity, in percentage

for females, where $\hat{\sigma}_{X_1}$ and $\hat{\sigma}_{X_2}$ denote the sample standard deviations of the covariates $X_1 = \text{“temperature”}$ and $X_2 = \text{“humidity,”}$ respectively. As in the previous section, a multivariate Epanechnikov kernel is considered. Note that the estimation grid of explanatory variables on which the estimates of the mean were computed was constructed by overlying the survey values of temperature and humidity with a 100×100 grid and, then, dropping every grid point that did not satisfy one of the following two requirements: (a) it is within 15 “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. Both requirements are admittedly somewhat arbitrary, but they represent a compromise between coverage over the region of interest and ability to avoid singular design matrices. Even with these restrictions, some of the estimates for low temperature values (around 20 °C) seem to be spurious, specially in the case of male individuals. This can be due to data sparseness or a boundary effect, two well-known situations where kernel-based smoothing methods may present certain drawbacks. Trying to avoid some of these problems and taking into account that there are repeated values of the covariates, additional estimates have been obtained after jittering the original data (the corresponding plots are not shown), obtaining estimates that follow similar patterns to those shown in Fig. 6. The mean direction followed by male and female sand hoppers is different for some temperature and humidity conditions. Seaward orientation was roughly $7\pi/4$, so it can be seen that females are more seaward oriented than males, specially for mid to low values of temperature.

6 Discussion

Nonparametric estimation of the conditional mean direction (or the regression function, from a model-based approach) of a circular random variable, given a \mathbb{R}^d -valued covariate, is studied in this paper. Our proposal considers kernel-based approaches, with special attention on NW- and LL-type estimators in general dimension, and for higher-order polynomials in the one-dimensional case. Asymptotic conditional bias

and variance are derived and the performance of the estimators is assessed in a simulation study.

For practical implementation, the selection of a d -dimensional bandwidth matrix is required. In the regression Euclidean context, the bandwidth selection problem has been widely addressed in the last decades (see, for example, Köhler et al. 2014, where a review on bandwidth selection methods for kernel regression is provided). More related to the topic of the present paper, a rule-of-thumb and a bandwidth rule for selecting scalar or diagonal bandwidth matrices for the multivariate local linear regression estimator with real-valued response and \mathbb{R}^d -valued covariate is derived in Yang and Tschernig (1999). Also in this setting, in González-Manteiga et al. (2004), a bootstrap method to estimate the mean squared error and the smoothing parameter for the multidimensional regression local linear estimator is proposed. However, in the framework of nonparametric regression methods for circular variables, the research on bandwidth selection is very scarce or nonexistent. Our practical results are derived with a cross-validation bandwidth given that, up to our knowledge, there are no other bandwidth selectors available in this context. The design of alternative procedures to select the bandwidth matrix for the estimators studied in this paper based, for example, on bootstrap methods are indeed of great interest. This problem is out of the scope of the present paper, but it is an interesting topic of research for a future study.

Once the problem of including a \mathbb{R}^d -valued covariate for explaining the behavior of a circular response is solved, it seems natural to think about the consideration of covariates of different nature. Since the proposed estimator is constructed by considering the atan2 of the smooth estimators of the regression functions for the sine and cosine components of the response, an adaptation of our proposal for different types of covariates implies the use of suitable weights. For instance, if a spherical (circular, as a particular case) or a mixture of spherical and real-valued covariates are considered to influence a circular response, weights for estimating the sine and cosine components could be constructed following the ideas in García-Portugués et al. (2013) for cylindrical density estimation. If a categorical covariate is included in the model, a similar approach to the one in Racine and Li (2004) or in Li and Racine (2004) could be also followed. In all these cases, bandwidth matrices should be selected, and cross-validation techniques could be applied.

The results obtained in Theorems 3 and 4 can be extended to an arbitrary dimension d of the space of the covariates by using the asymptotic properties for $\hat{m}_{j,\mathbf{H}}(\mathbf{x}; p)$, provided in Gu et al. (2015), who considered the leading term of the bias and the variance of the multivariate local polynomial estimator of general order p . Results on the asymptotic distribution of the multivariate local polynomial estimator (for a general p) is also provided in Gu et al. (2015). The joint asymptotic normality of $\hat{m}_{1,\mathbf{H}}(\mathbf{x}; p)$ and $\hat{m}_{2,\mathbf{H}}(\mathbf{x}; p)$ can be used to derive, via the delta-method, the asymptotic distribution of statistics which can be expressed in terms of $\hat{m}_{1,\mathbf{H}}(\mathbf{x}; p)$ and $\hat{m}_{2,\mathbf{H}}(\mathbf{x}; p)$. For example, a suitable adaptation of Proposition 3.1 of Jammalamadaka and SenGupta (2001) can be used to derive the limiting distribution of the tangent of $\hat{m}_{\mathbf{H}}(\mathbf{x}; p)$.

In our scenario, data generated from the regression model are assumed to be independent. However, in many practical situations, this assumption does not seem reasonable (e.g., data area collected over time or space). The simple construction scheme behind the proposed class of estimators makes possible to easily obtain asymptotic

properties in more general frameworks. As an example, when data are not independent but are realizations of stationary processes satisfying some mixing conditions, the results provided in Masry (1996) can be used. It should be also noted that, when the data exhibit some kind of dependence, although the expression for the estimator will be the same, this structure will affect the estimator variance and should be taking into account to select properly the bandwidth parameter, as in Francisco-Fernandez and Opsomer (2005).

Acknowledgements The authors acknowledge the support from the Xunta de Galicia Grant ED481A-2017/361 and the European Union (European Social Fund—ESF). This research has been partially supported by MINECO Grants MTM2016-76969-P and MTM2017-82724-R, and by the Xunta de Galicia (Grupo de Referencia Competitiva ED431C-2017-38, and Centro de Investigación de Galicia “CITIC” ED431G 2019/01), all of them through the ERDF. The authors thank Prof. Felicita Scapini and his research team who kindly provided the sand hoppers data that are used in this work. Data were collected within the Project ERB ICI8-CT98-0270 from the European Commission, Directorate General XII Science. The authors also thank two anonymous referees for numerous useful comments that significantly improved this article.

References

- Batschelet E (1981) Circular statistics in biology. Mathematics in biology. Academic Press, London
- Di Marzio M, Panzera A, Taylor CC (2013) Non-parametric regression for circular responses. *Scand J Stat* 40(2):238–255
- Di Marzio M, Panzera A, Taylor CC (2014) Nonparametric regression for spherical data. *J Am Stat Assoc* 109(506):748–763
- Fan J, Gijbels I (1996) Local polynomial regression. Chapman and Hall, London
- Fisher NI (1995) Statistical analysis of circular data. Cambridge University Press, Cambridge
- Fisher NI, Lee AJ (1992) Regression models for an angular response. *Biometrics* 48(3):665–677
- Francisco-Fernandez M, Opsomer JD (2005) Smoothing parameter selection methods for nonparametric regression with spatially correlated errors. *Can J Stat* 33(2):279–295
- García-Portugués E, Crujeiras RM, González-Manteiga W (2013) Kernel density estimation for directional-linear data. *J Multivariate Anal* 121:152–175
- González-Manteiga W, Martínez-Miranda M, Pérez-González A (2004) The choice of smoothing parameter in nonparametric regression through Wild Bootstrap. *Comput Stat Data An* 47(3):487–515
- Gu J, Li Q, Yang JC (2015) Multivariate local polynomial kernel estimators: leading bias and asymptotic distribution. *Economet Rev* 34(6–10):979–1010
- Härdle W, Müller M (2012) Multivariate and semiparametric kernel regression. Wiley, London, pp 357–391
- Jammalamadaka SR, Sarma YR (1993) Circular regression. In: Matsusita K (ed) Statistical science and data analysis. VSP, Utrecht, pp 109–128
- Jammalamadaka SR, SenGupta A (2001) Topics in circular statistics, vol 5. World Scientific, Singapore
- Kim S, SenGupta A (2017) Multivariate-multiple circular regression. *J Stat Comput Sim* 87(7):1277–1291
- Köhler M, Schindler A, Sperlich S (2014) A review and comparison of bandwidth selection methods for kernel regression. *Int Stat Rev* 82(2):243–274
- Lagarias JC, Reeds JA, Wright MH, Wright PE (1998) Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM J Optim* 9(1):112–147
- Lejeune M, Sarda P (1992) Smooth estimators of distribution and density functions. *Comput Stat Data Anal* 14(4):457–471
- Ley C, Verdebout T (2017) Modern directional statistics. Chapman and Hall/CRC, London
- Li Q, Racine J (2004) Cross-validated local linear nonparametric regression. *Stat Sin* 14(2):485–512
- Liu XH (2001) Kernel smoothing for spatially correlated data. PhD thesis, Department of Statistics, Iowa State University
- Marchetti GM, Scapini F (2003) Use of multiple regression models in the study of sandhopper orientation under natural conditions. *Estuar Coast Shelf Sci* 58:207–215
- Mardia KV, Jupp PE (2000) Directional statistics, vol 494. Wiley, London

- Masry E (1996) Multivariate regression estimation local polynomial fitting for time series. *Stoch Proc Appl* 65(1):81–101
- Presnell B, Morrison SP, Littell RC (1998) Projected multivariate linear models for directional data. *J Am Stat Assoc* 93(443):1068–1077
- Racine J, Li Q (2004) Nonparametric estimation of regression functions with both categorical and continuous data. *J Econom* 119(1):99–130
- Rivest LP, Duchesne T, Nicosia A, Fortin D (2016) A general angular regression model for the analysis of data on animal movement in ecology. *J R Stat Soc C-Appl* 65(3):445–463
- Ruppert D, Wand MP (1994) Multivariate locally weighted least squares regression. *Ann Stat* 22(3):1346–1370
- Scapini F, Aloia A, Bouslama MF, Chelazzi L, Colombini I, ElGtari M, Fallaci M, Marchetti GM (2002) Multiple regression analysis of the sources of variation in orientation of two sympatric sandhoppers, *Talitrus saltator* and *Talorchestia brito*, from an exposed Mediterranean beach. *Behav Ecol Sociobiol* 51(5):403–414
- SenGupta A, Ugwuowo FI (2006) Asymmetric circular-linear multivariate regression models with applications to environmental data. *Environ Ecol Stat* 13(3):299–309
- Wang F, Gelfand AE, Jona-Lasinio G (2015) Joint spatio-temporal analysis of a linear and a directional variable: space–time modeling of wave heights and wave directions in the Adriatic Sea. *Stat Sin* 25(1):25–39
- Yang L, Tschernig R (1999) Multivariate bandwidth selection for local linear regression. *J R Stat Soc B* 61(4):793–815

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

AUTOCERTIFICAZIONE
per l'attribuzione delle parti nelle pubblicazioni con coautori

La sottoscritta AGNESE PANZERA nata a Isernia il 20.12.1977, per consentire alla Commissione la valutazione del contributo individuale nelle pubblicazioni con coautori ai fini del giudizio previsto dal D.M. n. 120/2016, consapevole che ai sensi degli artt. 46 e 47 D.P.R. n. 445/2000 le dichiarazioni mendaci, la falsità negli atti e l'uso di fatti falsi sono puniti ai sensi del codice penale e delle leggi speciali in materia, secondo le disposizioni richiamate dall'art. 76 del D.P.R. n. 445 indicato,

DICHIARA

che per la pubblicazione

Meilán-Vila, Andrea; Francisco-Fernández, Mario; Crujeiras, Rosa M.; Panzera, Agnese (2021).
Nonparametric multiple regression estimation for circular response. TEST, vol. 30, pp. 650-672,
ISSN:1133-0686

sebbene abbia partecipato ad ogni fase dell'attività di ricerca per la preparazione della stessa, il proprio contributo in merito alla creazione materiale ha riguardato in particolare le parti di seguito indicate:

Sezioni 1, 2.1, 2.2. e 6.

Firenze, 7 Giugno 2023

Agnese Panzera

