Received: 8 January 2025 | Revised: 1 August 2025 | Accepted: 1 October 2025

DOI: 10.1111/sjos.70027

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Semiparametric regression for circular response with application in ecology

Jose Ameijeiras-Alonso[1] | **Irène Gijbels**[2]

[1]CITMAga, Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, Santiago de Compostela, A Coruña, Spain
[2]Department of Mathematics, KU Leuven, Leuven, Belgium

Correspondence
Jose Ameijeiras-Alonso, CITMAga, Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, Santiago de Compostela, A Coruña, Spain.
Email: jose.ameijeiras@usc.es

Funding information
Ministerio de Ciencia, Innovación y Universidades, Grant/Award Number: PID2020-116587GB-I00; Xunta de Galicia, Grant/Award Number: ED431C 2021/24; KU Leuven, Grant/Award Number: C16/20/002; Research Foundation Flanders, Sabbatical Bench Fee, Grant/Award Number: K803924N

[Correction added on 10 December 2025, after the first online publication: In the Funding information section, the funder and grant number "Research Foundation Flanders, Sabbatical Bench Fee, Grant/Award Number: K803924N" has been added in this version.]

ABSTRACT

A regression model for a circular response variable depending on a linear or a circular predictor is presented in this paper. The conditional density belongs to a parametric flexible family that allows for asymmetry and varying peakedness around the modal direction. The modal direction and concentration depend on the covariate and are nonparametrically modeled via local polynomial fitting with a kernel weight. The asymptotic normality of the estimators for the conditional modal direction and concentration is established. Furthermore, from these theoretical results, the expression of the optimal smoothing parameter and a proposed data-driven estimator are derived. An application concerns the orientation of migratory birds according to the flight altitude and the wind direction.

KEYWORDS

directional statistics, flexible modeling, local likelihood, modal direction regression

1 | INTRODUCTION

Regression methods provide a way to estimate the relationship between two random variables. Several regression models were proposed in the statistical literature, but new challenges appear

Jose Ameijeiras-Alonso and Irène Gijbels equally contributed to this manuscript.

when at least one of the variables does not belong to the Euclidean space. This is the case when periodicity needs to be taken into account and data can be represented in the unit circumference, i.e., when having circular data.

In this paper, we focus on the case where the dependent variable is circular and the predictor is either linear (real-valued) or circular. Several parametric models were proposed since the seminal work of Gould (1969). To avoid the need for specifying a concrete parametric family, more flexible nonparametric models were studied in this context (see, e.g., Di Marzio et al., 2013; Alonso-Pena & Crujeiras, 2023). Most of the proposed models for circular responses assume that only the location parameter depends on the covariate, with the notable exception of the parametric model proposed by Fisher and Lee (1992), where also the concentration ("dispersion") parameter depends on the linear predictor. However, one can find many examples where not only the location parameter but also the concentration depends on the predictor, and a flexible model is needed to estimate the relation between the concentration and the covariate.

This is the case for the data analyzed in Sjöberg and Nilsson (2015b), which contains the flight orientation of nocturnal migratory songbirds. The authors examine how flight altitude (a linear predictor) and wind direction (a circular predictor) influence the birds' orientation (a circular response). By categorizing the explanatory variables, Sjöberg and Nilsson (2015b) found that the mean direction of the birds is not influenced by altitude but is influenced by wind direction. The concentration varies with the predictor in both cases, and Sjöberg and Nilsson (2015b) also tested and rejected the hypothesis of equal concentrations across the different groups. In Figure 1, we replicated their plots to illustrate these effects. However, as seen in Figure 2 (left), when plotting the conditional mean direction on its support, it seems unreasonable for the mean direction to change so drastically from SE to SW winds. Since the data is available on a continuous scale, it would be more insightful to study the behavior of the location and concentration parameters as a function of the continuous predictor. The results of applying a parametric regression model (see Section 2 for details) are shown in Figure 2 (center), where we unwrapped the support onto a planar plot for better visualization. There, we can observe that these types of models allow for continuous plots on the torus; specifically, the conditional mean direction (green triangle-dashed curve) maintains the same value at the South wind direction ($-\pi$ and π).

Also, since the focus is on the conditional main direction of the birds, one could study the conditional modal direction instead of the conditional mean direction. The modal direction is defined as the point (within a continuous interval of length 2π) where the density function reaches its global maximum. As observed from the rose diagrams and the mean resultant vector, the mean and modal directions do not generally coincide. In Figure 2 (center), we employed the asymmetric generalized Batschelet distribution (see Ameijeiras-Alonso et al., 2022, or Section 2) to estimate the modal direction (dark blue thin lines) in each category, which differs from the estimated mean direction.

The first objective of this paper is to model the modal direction and concentration as functions of a linear or circular variable. The second objective is to do this in a flexible way. As shown in Figure 2 (bottom-right), the parametric model fails, e.g., to capture the oscillations of the conditional concentration with respect to wind direction. Therefore, this paper presents a method that allows for the flexible estimation of the conditional modal direction and concentration using a semiparametric regression model, the results of which are plotted in Figure 2 (center and right, blue dashed curves).

The model is semiparametric in the sense that the conditional density belongs to the flexible unimodal family introduced by Ameijeiras-Alonso et al. (2022). This parametric family allows

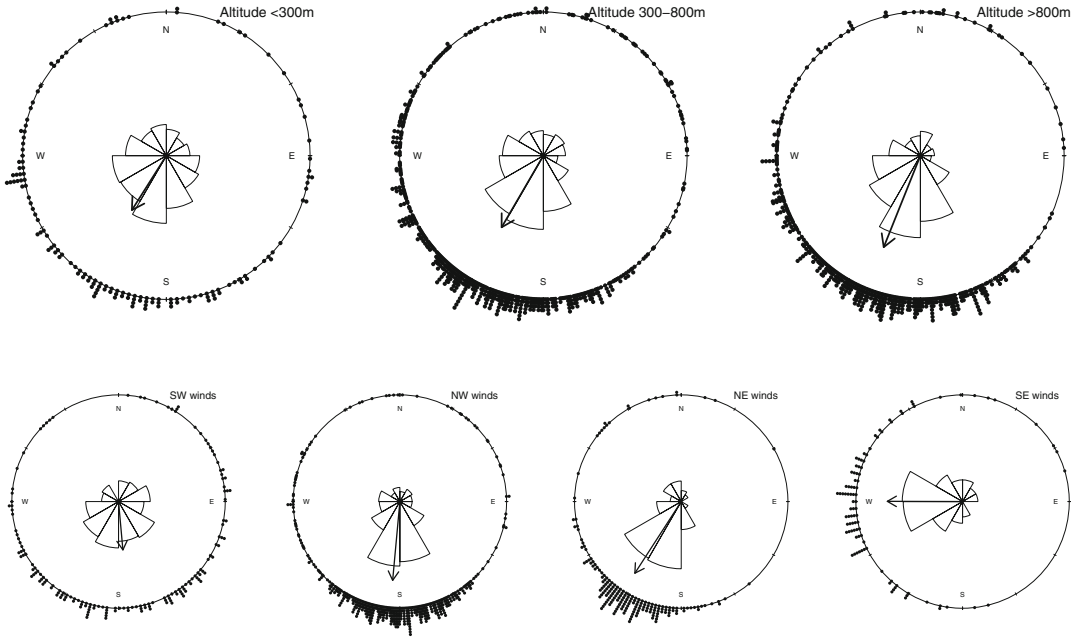

FIGURE 1 Rose diagram and sample mean resultant vector (the orientation represents the circular mean, and the length is proportional to the concentration). Orientation as a function of altitude (top) or as a function of wind direction (bottom).

for asymmetry and varying peakedness around the modal direction. However, the conditional modal direction and concentration are estimated nonparametrically using a local polynomial approach. We consider cases where the modal direction and concentration depend on a linear or circular predictor. More details about the local polynomial approach for linear data can be found in Fan and Gijbels (1996).

Other examples that highlight the need for flexible regression models with location and concentration parameters depending on the covariates can be found when studying the timing (in annual terms) of growth of tropical trees, depending on their size (Hogan et al., 2019); when analyzing the orientation of small blue periwinkles, depending on the moved distance (Fisher & Lee, 1992); when considering the wind direction at noon at a weather station in Texas, depending on the wind direction at 6 a.m. (Kato & Jones, 2010); or when studying the date (of the year) of onset of acute primary angle-closure glaucoma according to the age of the patient (Gao et al., 2006).

The paper is organized as follows. In Section 2, we review some of the main parametric and nonparametric regression models for the circular response case. The proposed semiparametric model is introduced in Section 3. The main theoretical results are established in Section 4. In Section 5, we provide some practical advice regarding the kernel and the smoothing parameter choices. In Section 6, we show the applicability of the proposed method in an example in ecology. Some final comments and discussion are given in Section 7. Additional details regarding the notation and assumptions, a simulation study comparing the performance of the semiparametric estimator with a previously proposed nonparametric approach, further information on the data application, an extension to the multivariate regression case, and the proofs of the theoretical results are provided in the Appendix. The Supplementary Material includes the R codes for replicating the presented results.

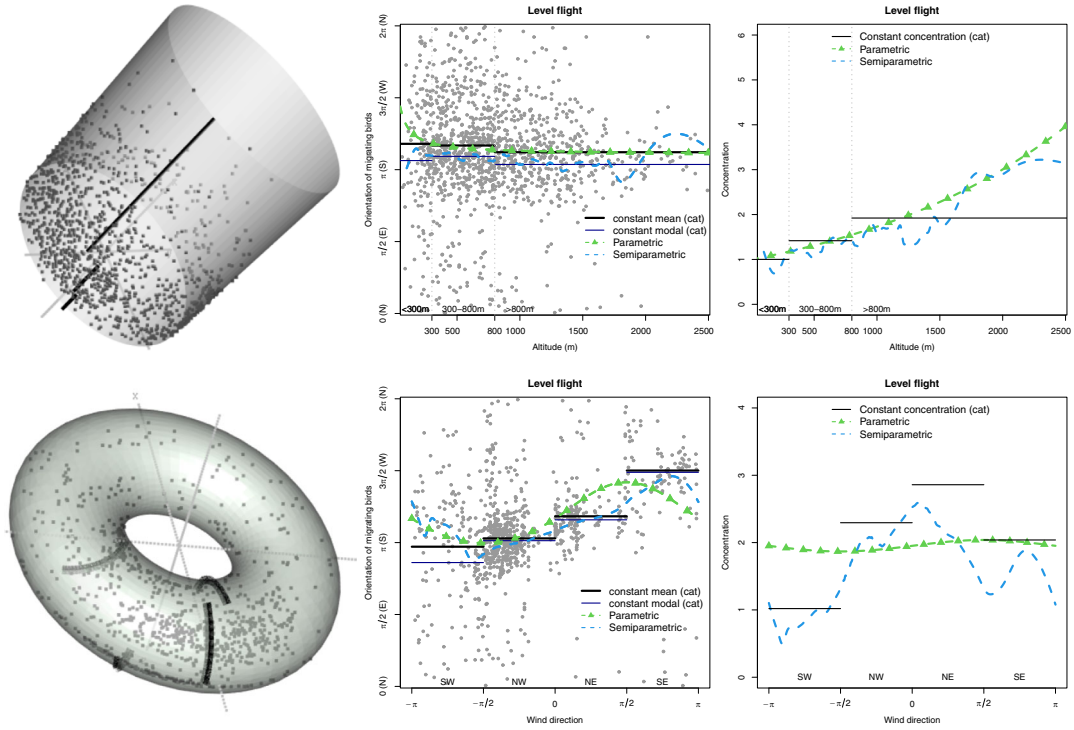

FIGURE 2 Orientation by altitude (top) or wind direction (bottom). Left: Sample on its support (cylinder/torus) with constant mean direction (line) per category. Center: Conditional location estimator plots. Right: Conditional concentration plots.

2 | REGRESSION FOR CIRCULAR RESPONSE

Before introducing the regression models, we focus on the flexible two-piece family of circular distributions introduced by Ameijeiras-Alonso et al. (2022). A review of other circular models can be found there or in Pewsey et al. (2013, Ch. 4).

The model of Ameijeiras-Alonso et al. (2022) is flexible in the sense that using any (reflectively) symmetric and unimodal density as a basis, with two parameters, modal direction m ($-\pi \leq m < \pi$) and concentration c ($c \geq 0$), it allows to model wider ranges of peakedness (curvature around the modal direction) and asymmetry. The proposed family also depends on two extra parameters: Peakedness at left ($p_L \in \mathbb{R}$) and at right ($p_R \in \mathbb{R}$) of the modal direction m. This model was constructed with the objective of keeping unimodality, with the modal direction at m. Also, symmetry is only obtained when the peakedness parameters are equal. Other properties are provided in Ameijeiras-Alonso et al. (2022).

Given a circular symmetric and unimodal density (with modal direction 0), denoted $f_{0,c}$, and a weight function w, the two-piece density function is defined as

$$g_{\Phi;p_L,p_R}(\phi; m, c) = \frac{1}{C_{c,p_L,p_R}} \begin{cases} f_{0,c}[(\phi - m) + p_L w(\phi - m)] & \text{if } \phi \in I_{m,1}, \\ f_{0,c}[(\phi - m) + p_R w(\phi - m)] & \text{if } \phi \in I_{m,2}, \end{cases} \tag{1}$$

where $C_{c,p_L,p_R} = C_{c,p_L} + C_{c,p_R}$ is the normalizing constant, with $C_{c,p_L} = \int_{I_{0,1}} f_{0,c}[\phi + p_L w(\phi)]d\phi$, $C_{c,p_R} = \int_{I_{0,2}} f_{0,c}[\phi + p_R w(\phi)]d\phi$, $I_{m,2} = [-\pi, \pi) \setminus I_{m,1}$, and $I_{m,1}$ defined as

$$I_{m,1} = [-\pi + m, m), \text{ if } m \geq 0, \qquad I_{m,1} = [-\pi, m) \cup [\pi + m, \pi), \text{ if } m < 0.$$

One of the main examples of a distribution belonging to family (1) is the generalized Batschelet, which is obtained when $w(\theta) = \sin(\theta)$ and $f_{0,c}$ is the von Mises density. With these choices of $(f_{0,c}, w)$, the density in Equation (1) can be expressed as:

$$g_{\Phi;p_L,p_R}(\phi; m, c) = \frac{1}{2\pi \mathcal{I}_0(c) C_{c,p_L,p_R}} \begin{cases} \exp[c\cos((\phi - m) + p_L \sin(\phi - m))] & \text{if } \phi \in I_{m,1}, \\ \exp[c\cos((\phi - m) + p_R \sin(\phi - m))] & \text{if } \phi \in I_{m,2}, \end{cases} \tag{2}$$

with $\mathcal{I}_k(c)$ denoting the modified Bessel function of the first kind of order k. The generalized Batschelet model is depicted in Figure 3 for various parameter configurations, with the modal direction set at $m = 0$. It has two important submodels: The von Mises distribution, which occurs when $p_L = p_R = 0$ (in this instance, $C_{c,p_L,p_R} = 1$), and the Batschelet distribution (see, e.g., Pewsey et al., 2011), which arises when $p_L = p_R$. In Figure 2 (center), the constant modal direction m is estimated using this model for the orientation of the migrating birds in each category.

Let $(\Delta_1, \Phi_1), \ldots, (\Delta_n, \Phi_n)$ be an i.i.d. sample from (Δ, Φ), where Φ is a circular random variable and Δ is either a linear or a circular random variable. Consider now the regression setting where Φ depends on Δ. Then, in this paper, we consider the conditional density of Φ given $\Delta = \delta$, the density $g_{\Phi|\Delta;p_L,p_R}$, where the modal direction parameter m and the concentration parameter c depend on δ. Once this model is imposed, the key issue lies in the estimation of the functions $m(\cdot)$ and $c(\cdot)$.

With the objective of introducing the main ideas of our semiparametric approach, we briefly review, in the following section, some parametric and nonparametric regression models for circular response. Additional parametric models are discussed in Pewsey et al. (2013, Ch. 8) or Kim and SenGupta (2016). For a comprehensive review of nonparametric regression models, we refer to Alonso-Pena and Crujeiras (2024).

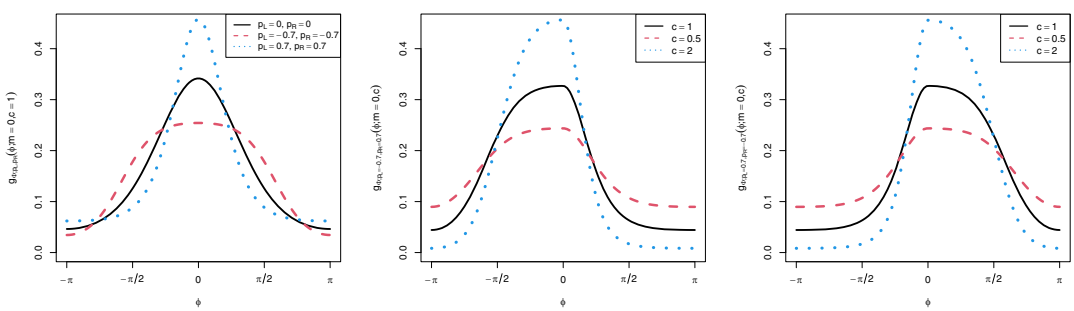

FIGURE 3 Generalized Batschelet density function (2) with $m = 0$. Left: Symmetric submodels (with $c = 1$) von Mises density ($p_L = p_R = 0$) and Batschelet densities ($p_L = p_R = -0.7$ and $p_L = p_R = 0.7$). Center and right: Asymmetric submodels (with varying concentration c); center: $p_L = -0.7$, and $p_R = 0.7$; and right: $p_L = 0.7$, and $p_R = -0.7$.

2.1 | Parametric regression

Employing the conditional von Mises density, i.e., density (2), with $p_L = p_R = 0$, classical parametric regression models focus on how to estimate $m(\delta)$ and $c(\delta)$ using a simple parametric model. For simplicity, let us first focus on the case where the explanatory variable is linear ($\Delta = X$). Then, following the classical regression literature, one could model $m(x)$ and $c(x)$ with a polynomial, of degree p_1 and p_2, respectively, i.e., for $r \in \{1, 2\}$, consider the polynomials $\check{\eta}_{r,p_r}(x) = \check{\eta}_{r;0} + \check{\eta}_{r;1}x + \ldots + \check{\eta}_{r;p_r}x^{p_r}$. This regression model, with $m(x) = \check{\eta}_{1,p_1}(x)$ and constant $c(x) = \check{\eta}_{2,0}$, was originally proposed by Gould (1969).

The main issue with the approach of Gould (1969) is that the parameter space of the modal direction and the concentration should be restricted to their support. In contrast, polynomial regression assumes that the response can take any value on the real line. Following Fisher and Lee (1992), this issue can be solved by considering strictly monotonic transformations $h_1(m(x)) = \check{\eta}_{1,p_1}(x)$ and $h_2(c(x)) = \check{\eta}_{2,p_2}(x)$, such that $m(x) = h_1^{-1}(\check{\eta}_{1,p_1}(x)) \in [-\pi, \pi)$ or another interval of range 2π and $c(x) = h_2^{-1}(\check{\eta}_{2,p_2}(x)) \in S_c$, where S_c denotes the support of c for the base density $f_{0,c}$. In particular, for the conditional von Mises density, Fisher and Lee (1992) give as an example, the functions $\check{h}_1^{-1}(u; l, k) = l + 2\tan^{-1}(sgn(u)|u|^k)$ and $\check{h}_2^{-1}(u) = \exp(u)$. In Figure 2 (top, center and right), we used this model with $p_1 = p_2 = 1$ and the link functions $\check{h}_1^{-1}(u; l, 1)$ and $\check{h}_2^{-1}(u)$ to estimate the conditional location $m(x)$ and concentration $c(x)$ of the birds' orientation depending on the altitude.

If the explanatory variable is circular ($\Delta = \Theta$), a second complication is that the p_rth degree linear polynomial is not a valid model. In that case, the assumption of periodicity (in 2π) needs to be satisfied, i.e., $m(\theta) = m(\theta + 2\pi)$ and $c(\theta) = c(\theta + 2\pi)$. Thus, a circular-linear regression model could be employed before applying the transformations. In particular, one possible approach is, first, to consider the p_rth degree circular sine model, $\check{\eta}^{\circ}_{r,p_r}(\theta) = \check{\eta}_{r;0} + \check{\eta}_{r;1}\sin(\theta) + \ldots + \check{\eta}_{r;p_r}\sin^{p_r}(\theta)$. For simplicity of exposition, in this manuscript, we name this function the circular polynomial model. Second, as in the previous case, the transformations $m(\theta) = h_1^{-1}(\check{\eta}^{\circ}_{1,p_1}(\theta))$ and $c(\theta) = h_2^{-1}(\check{\eta}^{\circ}_{2,p_2}(\theta))$ are employed. A similar approach was proposed by Kim and SenGupta (2017) in their arc-tangent-sine link model, by considering $p_1 = 1$, $\check{h}_1^{-1}(u; l, 1) = l + 2\tan^{-1}(u)$ and c being constant. In Figure 2 (bottom, center and right), we applied this model with $p_1 = p_2 = 1$ and the link functions $\check{h}_1^{-1}(u; l, 1)$ and $\check{h}_2^{-1}(u) = \exp(u)$ to estimate the conditional location $m(\theta)$ and concentration $c(\theta)$ of the birds' orientation as a function of wind direction.

2.2 | Nonparametric regression

The parametric models discussed in Section 2.1 assume a von Mises distribution for the conditional distribution, making them suitable for modeling both the conditional mean and the modal direction. In the nonparametric (kernel-based) literature, when the objective is to model the conditional location parameter, one must distinguish whether the focus is on the conditional mean or modal direction. Di Marzio et al. (2013) proposed a local polynomial model for estimating the conditional mean direction, while the study of the conditional modal direction was conducted by Alonso-Pena and Crujeiras (2023). However, both approaches do not address the study of conditional concentration.

As our emphasis is on the conditional modal direction, we briefly introduce the nonparametric conditional (multi)modal estimator for circular responses from Alonso-Pena and Crujeiras (2023).

They initiated this with the development of a conditional density estimator:

$$\hat{f}(\phi|\delta) = \frac{\sum_{i=1}^{n} \mathcal{W}(\delta - \Delta_i) K_\rho(\phi - \Phi_i)}{\sum_{i=1}^{n} \mathcal{W}(\delta - \Delta_i)}, \tag{3}$$

where the weight $\mathcal{W}$ depends on the nature of the predictor, $\mathcal{W}(\delta)$ is a scalar kernel $L_h(x) = h^{-1}L(x/h)$, if the predictor is linear, or a circular kernel $K_\nu(\theta)$, if the predictor is circular. Using the conditional density estimator in Equation (3), the modal regression (multi)function is given by

$$\widehat{M}(\delta) = \left\{ \phi \,:\, \frac{\partial}{\partial \phi} \hat{f}(\phi|\delta) = 0, \frac{\partial^2}{\partial \phi^2} \hat{f}(\phi|\delta) < 0 \right\}.$$

Given a pair of smoothing parameters, (h, ρ), in the linear predictor case, or (ν, ρ), in the circular case, a circular version of the mean shift algorithm is employed to calculate $\widehat{M}(\delta)$. In Alonso-Pena and Crujeiras (2023), a modal cross-validation method is employed to obtain a data-driven pair of smoothing parameters. Note that this method does not guarantee that the conditional estimation is unimodal. According to the authors' knowledge, there are no nonparametric conditional unimodal estimators for circular responses discussed in the literature. But, if for a given pair of smoothing parameters, $\widehat{M}(\delta)$ is not unimodal for all the values of δ, one can always try to oversmooth the estimator to obtain a conditional unimodal density. This is what we did in Appendix B, to obtain a competitor for our proposal. We use the cross-validation smoothing parameters of Alonso-Pena and Crujeiras (2023) if the conditional density is unimodal. Otherwise, we decrease the value of ρ (the concentration parameter associated with the response Φ) until we get a conditional unimodal density estimator, for all the analyzed values of δ.

3 | SEMIPARAMETRIC REGRESSION FOR CIRCULAR RESPONSE

For semiparametric modeling, we use the local polynomial fitting approach. For doing so, we employ the parametric approach described in Section 2.1, where the functions η_1 and η_2 are unknown. The employed conditional density of Φ given Δ, either linear or circular, is

$$\begin{aligned} &g_{\Phi|\Delta;p_L,p_R}(\phi; m(\Delta), c(\Delta)|\Delta = \delta) \\ &\quad = \frac{1}{C_{c(\delta),p_L,p_R}} \begin{cases} f_{0,c(\delta)}[(\phi - m(\delta)) + p_L w(\phi - m(\delta))] & \text{if } \phi \in I_{m(\delta),1}, \\ f_{0,c(\delta)}[(\phi - m(\delta)) + p_R w(\phi - m(\delta))] & \text{if } \phi \in I_{m(\delta),2}, \end{cases} \end{aligned} \tag{4}$$

where $m(\delta) = h_1^{-1}(\eta_1(\delta))$ and $c(\delta) = h_2^{-1}(\eta_2(\delta))$. As in Section 2, the strictly monotonic transformations $h_1 : [-\pi, \pi) \to \mathbb{R}$ and $h_2 : S_c \to \mathbb{R}$ are assumed to be known. For convenience, in practice, we will use the conditional generalized Batschelet distribution (see (2)), where the conditional modal direction and concentration are equal to

$$\begin{aligned} m(\delta) &= h_1^{-1}(\eta_1(\delta)) = \check{h}_1^{-1}(\eta_1(\delta); 0, 1) = 2\tan^{-1}(\eta_1(\delta)), \\ c(\delta) &= h_2^{-1}(\eta_2(\delta)) = \check{h}_2^{-1}(\eta_2(\delta)) = \exp(\eta_2(\delta)). \end{aligned}$$

For a given value δ_0, assume that the $(\mathfrak{p}_r + 1)$th derivative of the unknown function η_r exists. Then, $\eta_r(\delta)$, with $r \in \{1, 2\}$ and δ in a neighborhood of δ_0, can be approximated, using a Taylor expansion of order $\mathfrak{p}_r$, as follows:

$$\begin{aligned}
\eta_r(\delta) &\approx \eta_r(\delta_0) + \eta_r'(\delta_0)(\delta - \delta_0) + \ldots + \frac{\eta_r^{(\mathfrak{p}_r)}(\delta_0)}{\mathfrak{p}_r!}(\delta - \delta_0)^{\mathfrak{p}_r},\\
\eta_r(x) &\overset{\text{linear}}{\approx} \beta_{r;0} + \beta_{r;1}(x - x_0) + \ldots + \beta_{r;\mathfrak{p}_r}(x - x_0)^{\mathfrak{p}_r},\\
\eta_r(\theta) &\overset{\text{circular}}{\approx} \beta_{r;0} + \beta_{r;1}\sin(\theta - \theta_0) + \ldots + \beta_{r;\mathfrak{p}_r}\sin^{\mathfrak{p}_r}(\theta - \theta_0).
\end{aligned}$$

The superscripts "linear" and "circular" in the mathematical symbols are used throughout this paper to clarify whether its value corresponds to the linear or the circular predictor case. In the first equation, however, no superscript is included, as the expression holds for δ being either linear or circular.

Remark 1. In this paper, we provide the general theory for any positive integer values $\mathfrak{p}_1$ and $\mathfrak{p}_2$. But, in practice, we will consider the simpler case $\mathfrak{p}_1 = \mathfrak{p}_2 = 1$. Thus, for the circular case, we use the approximation $\sin(\theta) \approx \theta$, which is also employed in the local polynomial circular-linear regression literature (see, e.g. Di Marzio et al., 2009). Note that considering larger values of $\mathfrak{p}_r$ may have practical consequences for the circular case. One of them is that the equality $\beta_{r;v} = \eta_r^{(v)}(\delta_0)/v!$ is only valid in the linear case and in the circular case if $v \leq 2$. For the circular case, if $v > 2$, $\beta_{r;v}$ is a linear combination of $\eta_s^{(j)}(\theta_0), j \in \{1, \ldots, v\}$. For example, from the Taylor expansion of order $\mathfrak{p}_r = 3$, we obtain that, for θ in a neighborhood of θ_0,

$$\begin{aligned}
\eta_r(\theta) &\overset{\text{circular}}{=} \eta_r(\theta_0) + \eta_r'(\theta_0)(\theta - \theta_0) + \frac{\eta_r''}{2}(\theta_0)(\theta - \theta_0)\\
&\quad + \frac{\eta_r^{(3)}(\theta_0)}{6}(\theta - \theta_0)^3 + O\big((\theta - \theta_0)^4\big)\\
&= \eta_r(\theta_0) + \eta_r'(\theta_0)\sin(\theta - \theta_0) + \frac{\eta_r''(\theta_0)}{2}\sin^2(\theta - \theta_0)\\
&\quad + \frac{\eta_r^{(3)}(\theta_0) + \eta_r'(\theta_0)}{6}\sin^3(\theta - \theta_0) + O\big((\theta - \theta_0)^4\big).
\end{aligned}$$

We will employ the star notation, $\eta_{r,v}^*(\theta_0)$, to refer to the linear combinations of $\eta_s^{(j)}(\theta_0)$ that are needed to obtain $\beta_{r;v}$. From the example above, we would obtain that,

$$\begin{aligned}
\eta_r(\theta) &\overset{\text{circular}}{=} \beta_{r;0} + \beta_{r;1}\sin(\theta - \theta_0) + \beta_{r;2}\sin^2(\theta - \theta_0) + \beta_{r;3}\sin^3(\theta - \theta_0) + O\big((\theta - \theta_0)^4\big),\\
\beta_{r;v} &\overset{\text{circular}}{=} \eta_{r,v}^*(\theta_0) = \eta_r^{(v)}(\delta_0)/v! \quad \text{, if } v \leq 2,\\
\beta_{r;3} &\overset{\text{circular}}{=} \eta_{r,3}^*(\theta_0) = (\eta_r^{(3)}(\theta_0) + \eta_r'(\theta_0))/3!.
\end{aligned}$$

Consider now the log-likelihood of the density function (4):

$$\begin{aligned}
&\sum_{i=1}^{n} \ln g_{\Phi|\Delta;p_L,p_R}\big(\Phi_i; h_1^{-1}(\eta_1(\Delta)), h_2^{-1}(\eta_2(\Delta))|\Delta = \Delta_i\big)\\
&\quad = \sum_{i=1}^{n} \ell(h_1^{-1}(\eta_1(\Delta_i)), h_2^{-1}(\eta_2(\Delta_i)); \Phi_i).
\end{aligned}$$

If Δ_i is "close" to δ_0, the contribution of (Δ_i, Φ_i) to the log-likelihood can be approximated by

$$\begin{aligned}
&\ell(h_1^{-1}(\eta_1(\Delta_i)), h_2^{-1}(\eta_2(\Delta_i)); \Phi_i) \\
&\quad \approx \ell\left(h_1^{-1}\left(\sum_{j_1=0}^{\mathfrak{p}_1} \frac{\eta_1^{(j_1)}(\delta_0)}{j_1!}(\Delta_i - \delta_0)^{j_1}\right), h_2^{-1}\left(\sum_{j_2=0}^{\mathfrak{p}_2} \frac{\eta_2^{(j_2)}(\delta_0)}{j_2!}(\Delta_i - \delta_0)^{j_2}\right); \Phi_i\right) \\
&\ell(h_1^{-1}(\eta_1(X_i)), h_2^{-1}(\eta_2(X_i)); \Phi_i) \\
&\quad \overset{\text{linear}}{\approx} \ln g_{\Phi|X;p_L,p_R}\left(\Phi_i; h_1^{-1}\left(\sum_{j_1=0}^{\mathfrak{p}_1} \beta_{1;j_1}(X - x_0)^{j_1}\right), h_2^{-1}\left(\sum_{j_2=0}^{\mathfrak{p}_2} \beta_{2;j_2}(X - x_0)^{j_2}\right)|X = X_i\right), \\
&\ell(h_1^{-1}(\eta_1(\Theta_i)), h_2^{-1}(\eta_2(\Theta_i)); \Phi_i) \\
&\quad \overset{\text{circular}}{\approx} \ln g_{\Phi|\Theta;p_L,p_R}\left(\Phi_i; h_1^{-1}\left(\sum_{j_1=0}^{\mathfrak{p}_1} \beta_{1;j_1}\sin^{j_1}(\Theta - \theta_0)\right), h_2^{-1}\left(\sum_{j_2=0}^{\mathfrak{p}_2} \beta_{2;j_2}\sin^{j_2}(\Theta - \theta_0)\right)|\Theta = \Theta_i\right).
\end{aligned}$$

As mentioned earlier, the aforementioned approximation holds true only when Δ_i lies in a neighborhood of δ_0. This concern can be addressed by incorporating a kernel weight function $\mathcal{W}$ into the multiplication of the log-likelihood contribution. The kernel puts more weight on the log-likelihood contribution of (Δ_i, Φ_i) when Δ_i is close to δ_0. Depending on the nature of the predictor, $\mathcal{W}(\delta)$ is either a scalar kernel $L_h(x) = h^{-1}L(x/h)$ or a circular kernel $K_\nu(\theta)$. Thus, the corresponding local log-likelihood function is

$$\begin{aligned}
&\mathcal{L}_n(x_0; \boldsymbol{\beta_1}, \boldsymbol{\beta_2}) \\
&\quad \overset{\text{linear}}{=} \sum_{i=1}^{n} \ell\left(h_1^{-1}\left(\sum_{j_1=0}^{\mathfrak{p}_1} \beta_{1;j_1}(X_i - x_0)^{j_1}\right), h_2^{-1}\left(\sum_{j_2=0}^{\mathfrak{p}_2} \beta_{2;j_2}(X_i - x_0)^{j_2}\right); \Phi_i\right) L_h(X_i - x_0), \\
&\mathcal{L}_n(\theta_0; \boldsymbol{\beta_1}, \boldsymbol{\beta_2}) \\
&\quad \overset{\text{circular}}{=} \sum_{i=1}^{n} \ell\left(h_1^{-1}\left(\sum_{j_1=0}^{\mathfrak{p}_1} \beta_{1;j_1}\sin^{j_1}(\Theta_i - \theta_0)\right), h_2^{-1}\left(\sum_{j_2=0}^{\mathfrak{p}_2} \beta_{2;j_2}\sin^{j_2}(\Theta_i - \theta_0)\right); \Phi_i\right) K_\nu(\Theta_i - \theta_0),
\end{aligned}$$

where, for $r \in \{1, 2\}$, we denoted $\boldsymbol{\beta_r} = (\beta_{r;0}, \ldots, \beta_{r;p_r})$. Maximization of $\mathcal{L}_n(\delta_0; \boldsymbol{\beta_1}, \boldsymbol{\beta_2})$ with respect to $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ provides the local (in δ_0) polynomial maximum log-likelihood estimators

$$\left(\hat{\boldsymbol{\beta}}_1(\delta_0), \hat{\boldsymbol{\beta}}_2(\delta_0)\right) = \underset{\boldsymbol{\beta_1}, \boldsymbol{\beta_2}}{\operatorname{argmax}}\, \mathcal{L}_n(\delta_0; \boldsymbol{\beta_1}, \boldsymbol{\beta_2}). \tag{5}$$

For $r \in \{1, 2\}$, we can derive the estimators of $\eta_r(\delta_0)$ and its derivatives from $\hat{\boldsymbol{\beta}}_r(\delta_0)$ using the relations stated in Remark 1. The estimators for the modal direction and concentration at the value δ_0 are obtained as $\hat{m}(\delta_0) = h_1^{-1}(\hat{\beta}_{1;0}(\delta_0))$ and $\hat{c}(\delta_0) = h_2^{-1}(\hat{\beta}_{2;0}(\delta_0))$, respectively. The entire function estimates $\hat{m}(\delta)$ and $\hat{c}(\delta)$ can be derived by considering a grid of δ values within the definition domain of the regressor Δ and solving the maximization problem (5) for each point on the grid.

Note that in the case where $p_L = p_R$, $\hat{m}(\delta_0)$ will estimate both the conditional mean and modal direction as the conditional density is symmetric. Consequently, to model the conditional mean direction as in Di Marzio et al. (2013), one can use $\hat{m}(\delta)$ in the semiparametric model if $p_L = p_R$. For any values of p_L and p_R, the value of the conditional density at δ_0 is given by

$g_{\Phi|\Delta;p_L,p_R}(\cdot;\hat{m}(\Delta),\hat{c}(\Delta)|\Delta=\delta_0)$. Therefore, when $p_L \neq p_R$, the conditional mean direction can be computed from the (estimated) conditional density.

In this paper, we have restricted ourselves to the case where the peakedness parameters p_L and p_R are constant. Of course, similar steps to the ones in this section could be followed to obtain conditional peakedness parameters $p_L(\delta) = h_3^{-1}(\eta_3(\delta))$ and $p_R(\delta) = h_3^{-1}(\eta_4(\delta))$, using a strictly monotonic transformation $h_3 : [-1,1] \to \mathbb{R}$ and estimating η_3 and η_4. However, by considering this more general case, the theory developed in Section 4 would become more tedious, from a theoretical and implementation viewpoint. See also the discussion in Section 7. In this paper, we restricted our focus to the case where only the modal direction and the concentration depend on the regressor, while p_L and p_R are known constants. In Section 6, we also show results when the peakedness parameters are estimated from the whole sample in a practical setting.

4 | THEORETICAL RESULTS

We next investigate the asymptotic behavior of the local polynomial maximum log-likelihood estimators $(\hat{\boldsymbol{\beta}}_1(\delta_0), \hat{\boldsymbol{\beta}}_2(\delta_0))$ in Equation (5). Relying on these asymptotic convergence results, we also provide the asymptotic normality of the conditional modal direction and concentration estimators, and discuss optimal smoothing parameters.

All the asymptotic results will depend on the quantities below.

$$\begin{aligned}
\mu_k &\overset{\text{circular}}{=} \int \sin^k(\theta)K_\nu(\theta)d\theta, & \mu_k &\overset{\text{linear}}{=} \int x^k L(x)dx,\\
\tau_k &\overset{\text{circular}}{=} \int \sin^k(\theta)K_\nu^2(\theta)d\theta, & \tau_k &\overset{\text{linear}}{=} \int x^k L^2(x)dx,\\
\omega_{k,l} &\overset{\text{circular}}{=} \int \theta^l \sin^k(\theta)K_\nu(\theta)d\theta.
\end{aligned}$$

Some clarifications and remarks about these quantities are provided in Appendix A. In that section we also introduce the following matrices: $\mathfrak{I}(\delta_0)$, the Fisher information matrix of $m(\delta_0)$ and $c(\delta_0)$; $\boldsymbol{H}_{\mathfrak{p}_r}$, a matrix that depends on h; $\boldsymbol{T}$ depending on τ_k; $\boldsymbol{\Sigma}_{\delta_0}$, depending on $g_\Delta(\delta_0)$ (the marginal density of Δ), $\mathfrak{I}(\delta_0)$, and $\boldsymbol{T}$; and $\boldsymbol{A}_{\delta_0}$, depending on $g_\Delta(\delta_0)$, $\mathfrak{I}(\delta_0)$, and $\boldsymbol{U}$, where the latter matrix $\boldsymbol{U}$ is equal to

$$\boldsymbol{U} = \begin{pmatrix} \boldsymbol{U}_{\mathfrak{p}_1\mathfrak{p}_1} & \boldsymbol{U}_{\mathfrak{p}_1\mathfrak{p}_2} \\ \boldsymbol{U}_{\mathfrak{p}_2\mathfrak{p}_1} & \boldsymbol{U}_{\mathfrak{p}_2\mathfrak{p}_2} \end{pmatrix} = \begin{pmatrix} (\mu_{k+l-2})_{1\le k,l\le \mathfrak{p}_1+1} & (\mu_{k+l-2})_{1\le k\le \mathfrak{p}_1+1,1\le l\le \mathfrak{p}_2+1} \\ (\mu_{k+l-2})_{1\le k\le \mathfrak{p}_2+1,1\le l\le \mathfrak{p}_1+1} & (\mu_{k+l-2})_{1\le k,l\le \mathfrak{p}_2+1} \end{pmatrix}.$$

Below, $\boldsymbol{V}_{k,\mathfrak{p}_r}(\delta_0)$ denotes the matrix $\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}$, but with the $(k+1)$st column replaced by the vector $(1,\delta_0,\ldots,\delta_0^{\mathfrak{p}_r})^\top$, where the superscript $\top$ denotes the transpose of a vector or a matrix. For the circular case, the matrix $\boldsymbol{V}^\circ_{k,\mathfrak{p}_r}(\delta_0)$ coincides with $\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}$, but the $(k+1)$st column is replaced by the vector $(1,\sin(\theta_0),\ldots,\sin^{\mathfrak{p}_r}(\theta_0))^\top$. When the determinant $|\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}| \neq 0$, we also need to define

$$\begin{aligned}
W_{k,\mathfrak{p}_r}(x_0) &\overset{\text{linear}}{=} k!(|\boldsymbol{V}_{k,\mathfrak{p}_r}(x_0)|/|\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}|)L(x_0), & W_{k,\mathfrak{p}_r}(\theta_0) &\overset{\text{circular}}{=} k!(|\boldsymbol{V}_{k,\mathfrak{p}_r}(\theta_0)|/|\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}|)K_\nu(\theta_0),\\
& & W^\circ_{k,\mathfrak{p}_r}(\theta_0) &\overset{\text{circular}}{=} k!(|\boldsymbol{V}^\circ_{k,\mathfrak{p}_r}(\theta_0)|/|\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}|)K_\nu(\theta_0),\\
\zeta_{j,k,\mathfrak{p}_r} &\overset{\text{linear}}{=} \int x^k W_{j,\mathfrak{p}_r}(x)dx, & \zeta_{j,k,\mathfrak{p}_r} &\overset{\text{circular}}{=} \int \sin^k(\theta)W_{j,\mathfrak{p}_r}(\theta)d\theta,\\
\xi_{k,\mathfrak{p}_r} &\overset{\text{linear}}{=} \int \left(W_{k,\mathfrak{p}_r}(x)\right)^2 dx, & \xi_{k,\mathfrak{p}_r} &\overset{\text{circular}}{=} \int \left(W^\circ_{k,\mathfrak{p}_r}(\theta)\right)^2 d\theta.
\end{aligned}$$

TABLE 1 First block: Linear kernels.

Kernel	Expression of $L(x)$	Smoothing parameter	$\zeta_{0,2,1}$	$\xi_{0,1}$
Epanechnikov	$3(1-x^2)/4$, $x \in [-1,1]$	$h \geq 0$	1/5	3/5
uniform	$1/2$, $x \in [-1,1]$	$h_n \to 0$	1/3	1/2
Kernel	**Expression of $K_\nu(\theta)$**	**Smoothing parameter**	ν_2	τ_0
von Mises	$\frac{1}{2\pi}\left(1+2\sum_{j=1}^{\infty}\frac{\mathcal{I}_k(\nu)}{\mathcal{I}_0(\nu)}\cos(j\theta)\right)$	$\nu \geq 0$	$\frac{1}{2}-\frac{\mathcal{I}_2(\nu)}{2\mathcal{I}_0(\nu)}$	$\frac{1}{2\sqrt{\pi\mu_{2;n}}}$
	$=\exp(\nu\ \cos(\theta))/(2\pi\ \mathcal{I}_0(\nu))$	$\nu_n \to \infty$	$\approx 1/\nu_n$	
wrapped normal	$\frac{1}{2\pi}\left(1+2\sum_{j=1}^{\infty}\nu^{j^2}\cos(j\theta)\right)$	$\nu \in [0,1]$	$(1-\nu^4)/2$	
		$\nu_n \to 1$		

Note: Second block: Circular kernels. From left to right: Name of the kernel, density expression, smoothing parameter, and quantities related to the optimal smoothing parameter. A subindex n indicates the behavior of the parameter as $n \to \infty$.

Values of some of the previous quantities for some relevant kernels are provided in Table 1 in Section 5.1. An important point of difference between the linear and circular cases, related to some of these quantities, is discussed in Remark 2.

Remark 2. The quantities μ_k, τ_k, $\omega_{k,l}$, $\zeta_{j,k,\mathfrak{p}_r}$, and $\xi_{k,\mathfrak{p}_r}$ depend on the smoothing parameter ν in the circular case, whereas they do not depend on h in the linear case. So, when considering a smoothing parameter depending on the sample, denoted by ν_n, an abuse of notation will be made, denoting $\mu_k = \mu_{k;n}$ in the same way that h denotes h_n. This dependence has important implications, as, just for the circular case, the above quantities need to tend to zero, as $n \to \infty$.

Using the above notations, we establish the theoretical results by assuming some of the conditions listed below. The blocks (A1)–(A3), (B1)–(B3), and (C1)–(C4) correspond to conditions needed for the base density and the weight function, also in the unconditional case. Thus, we refer to Ameijeiras-Alonso et al. (2022) where several examples of base densities and weight functions satisfying these conditions are given. A representative example of a density satisfying these conditions is the conditional generalized Batschelet distribution (see 2).

Conditions (D1)–(D6) are required for the local likelihood approach, while Conditions (D7)–(D9) are used to derive additional results. The functions ψ_r, ψ_{rs}, and ψ_{rst} denote partial derivatives of the log-likelihood function, and their explicit forms are provided in Appendix A. Conditions (D1)–(D5) and (D8) are similar to the standard assumptions in the local quasi-likelihood estimation literature (Fan & Gijbels, 1996, Ch. 5). In the linear case, Conditions (D6) and (D7) are also standard in this context. For the circular case, alternative formulations are required since scaled kernels are not available, but the assumptions resemble their linear counterparts. Condition (D9) is introduced to obtain a simple, explicit expression for the optimal smoothing parameter in the circular case. It is satisfied by two of the most commonly used circular kernels: The von Mises and the wrapped normal densities (see Section 5.1 for further details).

4.1 | Regularity conditions on the base density

A1. $f_{0,c}$ belongs to a location family, i.e., $f_{m,c}(\phi) = f_{0,c}(\phi + m)$, for any ϕ and m.

A2. $f_{0,c}$ is a periodic density function with period 2π, i.e., for all integers k, $f_{0,c}(\phi) = f_{0,c}(\phi + 2k\pi)$; $f_{0,c}(\phi) > 0$, and $\int_{-\pi}^{\pi} f_{0,c}(\phi)d\phi = 1$.
A3. $f_{0,c}$ has a continuous derivative satisfying $f'_{0,c}(\phi) > 0$ if $\phi \in (-\pi, 0)$ and $f'_{0,c}(\phi) < 0$ if $\phi \in (0, \pi)$. $f_{0,c}$ is an even function, i.e., $f_{0,c}(-\phi) = f_{0,c}(\phi)$, for all ϕ.

4.2 | Regularity conditions on the weight function

B1. w is periodic, with period 2π.
B2. w is a non-constant odd function, i.e., for all ϕ, $w(\phi) = -w(-\phi)$.
B3. w is a bounded function with continuous derivative, satisfying, for all $\phi \in [-\pi, \pi)$ and for some $l > 0$: (i) $l|w(\phi)| \leq |\phi|$; (ii) $l|w(\phi)| \leq \pi - |\phi|$; (iii) $l|w'(\phi)| < 1$, if $\phi \neq k\pi$, for any integer k. For the largest value of l verifying the previous conditions, we will assume that the peakedness parameters p_L and p_R satisfy $-l \leq p_L, p_R \leq l$.

4.3 | Assumptions for the Fisher information matrix

C1. The (conditional) parameters of (4) have a common support $\mathbf{\Lambda}$, for all δ. The true parameters $(m(\delta), c(\delta), p_L, p_R)$ are in the interior of $\mathbf{\Lambda}_R$, a compact subset of $\mathbf{\Lambda}$, for all δ.
C2. If $c_1 > c_2$, the base density satisfies that $f_{0,c_1}(0) > f_{0,c_2}(0)$ and $f_{0,c_1}(-\pi) < f_{0,c_2}(-\pi)$.
C3. $f_{0,c}$ and w have a bounded continuous second-order derivative with respect to ϕ and also with respect to c (in the case of $f_{0,c}$).
C4. $(\partial/\partial c)f_{0,c}$ is an even function.

4.4 | Assumptions for the local likelihood approach

D1. For $r \in \{1, 2\}$ and for all δ, $\mathfrak{I}_{rr}(\delta)$ has a non-zero value.
D2. For $r, s, t \in \{1, 2\}$, and for all δ and ϕ, the following functions are continuous in δ: $g'_{\Delta}(\delta)$, $\eta_r^{(\mathfrak{p}_r+2)}(\delta)$, $\psi_{rst}(\eta_1(\delta), \eta_2(\delta); \phi)$.
D3. For all ϕ, for all δ, and for $r, s, t \in \{1, 2\}$; there exist two functions $M_1(\phi) < \infty$ and $M_2(\phi) < \infty$, such that $|\psi_{rs}(\eta_1(\delta), \eta_2(\delta); \phi)| < M_1(\phi)$ and $|\psi_{rst}(\eta_1(\delta), \eta_2(\delta); \phi)| < M_2(\phi)$.
D4. For each point in the boundary of the support of Δ, there exists an interval (arc in the circular case) C containing that point and with a non-null interior such that $\inf_{\delta \in C} g_{\Delta}(\delta) > 0$.
D5. The kernels L and K_ν are symmetric density functions. The kernel L has a bounded support, which, for simplicity, is assumed to be $[-1, 1]$. The kernel K_ν is a circular density (see Condition A2). L and K_ν are bounded functions.
D6. As $n \to \infty$, the following results hold. In the linear case, the bandwidth $h = h_n$ satisfies $h_n \to 0$ and $nh_n \to \infty$. In the circular case, for $r \in \{1, 2\}$ and integer values of k, the values of $\mu_k = \mu_{k;n}$ and $\xi_{k,\mathfrak{p}_r} = \xi_{k,\mathfrak{p}_r;n}$ depending on the concentration $\nu = \nu_n$ satisfy $\mu_{k;n} \to 0$, for all $k > 0$, and $n/\xi_{k,\mathfrak{p}_r;n} \to \infty$, where k will be the analyzed order of the derivative $0 \leq k \leq p_r$.
D7. As $n \to \infty$, the following results hold. In the linear case, $nh_n^3 \to \infty$. In the circular case, for any integer k, with $0 \leq k \leq p_r$, the quantity $\omega_{k,1} = \omega_{k,1;n}$ depending on $\nu = \nu_n$ satisfies $\omega_{k,1;n} = O(b_n)$, with b_n verifying $b_n \to 0$ and $nb_n^2 \to \infty$.
D8. For all δ, both $h'_1(m(\delta))$ and $h'_2(c(\delta))$ exist and are non-null.
D9. As $n \to \infty$, the quantity $\tau_0 = \tau_{0;n}$ depending on $\nu = \nu_n$ satisfies $\tau_{0;n} = (\pi\mu_{2;n})^{-1/2}/2$.

The main theoretical result of the paper is stated in Theorem 1, which establishes the weak convergence of the local polynomial maximum log-likelihood estimators, $\hat{\boldsymbol{\beta}}_1(\delta_0)$ and $\hat{\boldsymbol{\beta}}_2(\delta_0)$. As usual, we use the notation "$\xrightarrow{D}$" to denote weak convergence. The concept of weak convergence is the same in both the linear and circular cases; however, we include the subscripts "linear" or "circular" to distinguish between the asymptotic results, in terms of bias and variance, corresponding to each setting. The proofs of all theoretical results in this section are provided in Appendix D.

Theorem 1. *Assume that Conditions* $(A1)$–$(A3)$, $(B1)$–$(B3)$, $(C1)$–$(C4)$, *and* $(D1)$–$(D6)$ *are satisfied. Given a value* $\delta_0 \in S_\Delta$ *and for fixed values of* p_L *and* p_R*; then, as* $n \to \infty$,

$$\left(\sqrt{nh}\left(\boldsymbol{H}_{\mathfrak{p}_1}(\hat{\boldsymbol{\beta}}_1(x_0)-\boldsymbol{\beta}_1(x_0)), \boldsymbol{H}_{\mathfrak{p}_2}(\hat{\boldsymbol{\beta}}_2(x_0)-\boldsymbol{\beta}_2(x_0))\right)^\top - \boldsymbol{\mathcal{B}}_{x_0}\right) \xrightarrow[\text{linear}]{D} N\left(\mathbf{0}, \boldsymbol{A}_{x_0}^{-1}\boldsymbol{\Sigma}_{x_0}\boldsymbol{A}_{x_0}^{-1}\right),$$

$$\left(\sqrt{n}\left(\hat{\boldsymbol{\beta}}_1(\theta_0)-\boldsymbol{\beta}_1(\theta_0), \hat{\boldsymbol{\beta}}_2(\theta_0)-\boldsymbol{\beta}_2(\theta_0)\right)^\top - \boldsymbol{\mathcal{B}}_{\theta_0}\right) \xrightarrow[\text{circular}]{D} N\left(\mathbf{0}, \boldsymbol{A}_{\theta_0}^{-1}\boldsymbol{\Sigma}_{\theta_0}\boldsymbol{A}_{\theta_0}^{-1}\right).$$

Given a value of $k \in \{0,\ldots,\mathfrak{p}_r\}$, *with* $r \in \{1,2\}$, *the asymptotic value of the* $(k+1)$*th component of the bias* $\boldsymbol{\mathcal{B}}_{\delta_0}$ *depends on the parity of* $(\mathfrak{p}_r - k)$. *If* $(\mathfrak{p}_r - k)$ *is even,* x_0 *is a point in the interior of* S_X *(linear case) and Assumption* $(D7)$ *is satisfied, then*

$$\begin{aligned}
&\left(\boldsymbol{\mathcal{B}}_{x_0}\right)_{r,k+1} \\
&\overset{\text{linear}}{=} \frac{\sqrt{nh}h^{\mathfrak{p}_r+2}}{k!}\left(\frac{\eta_r^{(\mathfrak{p}_r+2)}(x_0)}{(\mathfrak{p}_r+2)!}\zeta_{k,\mathfrak{p}_r+2,\mathfrak{p}_r} + \frac{\frac{\partial}{\partial x_0}\left[g_X(x_0)\mathfrak{T}_{rr}(x_0)\right]}{g_X(x_0)\mathfrak{T}_{rr}(x_0)}\frac{\eta_r^{(\mathfrak{p}_r+1)}(x_0)}{(\mathfrak{p}_r+1)!}\left(\zeta_{k,\mathfrak{p}_r+2,\mathfrak{p}_r} - k\zeta_{k-1,\mathfrak{p}_r+1,\mathfrak{p}_r}\right)\right) \\
&\quad + o\left(\sqrt{nh}h^{\mathfrak{p}_r+2}\right), \\
&\left(\boldsymbol{\mathcal{B}}_{\theta_0}\right)_{r,k+1} \\
&\overset{\text{circular}}{=} \frac{\sqrt{n}}{k!}\left(\eta^*_{r,\mathfrak{p}_r+2}(\theta_0)\zeta_{k,\mathfrak{p}_r+2,\mathfrak{p}_r} + \frac{\frac{\partial}{\partial\theta_0}\left[g_\Theta(\theta_0)\mathfrak{T}_{rr}(\theta_0)\right]}{g_\Theta(\theta_0)\mathfrak{T}_{rr}(\theta_0)}\eta^*_{r,\mathfrak{p}_r+1}(\theta_0)\right. \\
&\quad \left.\times\left(\zeta_{k,\mathfrak{p}_r+2,\mathfrak{p}_r} - k\zeta_{k-1,\mathfrak{p}_r+1,\mathfrak{p}_r}\right)\right) + o\left(\sqrt{n}\omega_{\mathfrak{p}_r+3,k-1}\right).
\end{aligned}$$

Otherwise,

$$\left(\boldsymbol{\mathcal{B}}_{x_0}\right)_{r,k+1} \overset{\text{linear}}{=} \frac{\sqrt{nh}h^{\mathfrak{p}_r+1}}{k!}\frac{\eta_r^{(\mathfrak{p}_r+1)}(x_0)}{(\mathfrak{p}_r+1)!}\zeta_{k,\mathfrak{p}_r+1,\mathfrak{p}_r} + o\left(\sqrt{nh}h^{\mathfrak{p}_r+1}\right),$$

$$\left(\boldsymbol{\mathcal{B}}_{\theta_0}\right)_{r,k+1} \overset{\text{circular}}{=} \frac{\sqrt{n}}{k!}\eta^*_{r,\mathfrak{p}_r+1}(\theta_0)\zeta_{k,\mathfrak{p}_r+1,\mathfrak{p}_r} + o\left(\sqrt{n}\omega_{\mathfrak{p}_r+2,k-1}\right).$$

Remark 3. As shown in Theorem 1, and as is standard in local polynomial kernel estimation, using odd values of $(\mathfrak{p}_r - k)$ in the linear covariate case offers advantages: It avoids boundary modifications and yields simpler expressions. The minor disadvantage is that a higher-order polynomial is needed to achieve the same asymptotic bias order. For instance, both local quadratic and cubic estimators give the same asymptotic bias order. In the circular case, despite the absence of boundaries, odd values still

lead to simpler asymptotic expressions, which is especially useful when deriving the optimal smoothing parameter due to the lack of a kernel scaling factor (see Appendix D3). It can be shown that the asymptotic bias when estimating $m(\theta_0)$ or $c(\theta_0)$ is of order $O(\mu_{\mathfrak{p}_r+2})$ in the even case and $O(\mu_{\mathfrak{p}_r+1})$ in the odd case. If $\mu_{k+2} = o(\mu_k)$, for even positive integers k (see Section 5.1), then lower-order even polynomials can achieve the same asymptotic bias order as higher-order odd ones.

The explicit formula of the asymptotic normality of the maximum log-likelihood estimator of the kth derivative of $\eta_r(\delta_0)$, $\hat{\eta}_r^{(k)}(\delta_0)$, can be obtained as a consequence of Theorem 1, see Appendix D2. Also, from Theorem 1, the asymptotic normality of the conditional modal direction and concentration can be derived, as stated in Corollary 1.

Corollary 1. *Suppose that the conditions of Theorem 1 hold, and assume also Condition (D8) holds. Then, conditionally on a value $\delta_0 \in S_\Delta$, the following convergences, as $n \to \infty$, are obtained,*

$$(nh_n g_X(x_0)\mathfrak{I}_{11}(x_0))^{1/2}\xi_{0,\mathfrak{p}_1}^{-1/2}h_1'(m(x_0))\left(\hat{m}(x_0) - m(x_0) - \frac{\left(\boldsymbol{\mathcal{B}}_{x_0}\right)_{1,1}}{\sqrt{nh_n}h_1'(m(x_0))}\right) \xrightarrow[\text{linear}]{D} N(0,1),$$

$$(nh_n g_X(x_0)\mathfrak{I}_{22}(x_0))^{1/2}\xi_{0,\mathfrak{p}_2}^{-1/2}h_2'(c(x_0))\left(\hat{c}(x_0) - c(x_0) - \frac{\left(\boldsymbol{\mathcal{B}}_{x_0}\right)_{2,1}}{\sqrt{nh_n}h_2'(c(x_0))}\right) \xrightarrow[\text{linear}]{D} N(0,1),$$

$$(ng_\Theta(\theta_0)\mathfrak{I}_{11}(\theta_0))^{1/2}\xi_{0,\mathfrak{p}_1}^{-1/2}h_1'(m(\theta_0))\left(\hat{m}(\theta_0) - m(\theta_0) - \frac{\left(\boldsymbol{\mathcal{B}}_{\theta_0}\right)_{1,1}}{\sqrt{n}h_1'(m(\theta_0))}\right) \xrightarrow[\text{circular}]{D} N(0,1),$$

$$(ng_\Theta(\theta_0)\mathfrak{I}_{22}(\theta_0))^{1/2}\xi_{0,\mathfrak{p}_2}^{-1/2}h_2'(c(\theta_0))\left(\hat{c}(\theta_0) - c(\theta_0) - \frac{\left(\boldsymbol{\mathcal{B}}_{\theta_0}\right)_{2,1}}{\sqrt{n}h_2'(c(\theta_0))}\right) \xrightarrow[\text{circular}]{D} N(0,1),$$

where the expressions of $\boldsymbol{\mathcal{B}}_{\delta_0}$ coincide with those of Theorem 1 depending on the parity of $\mathfrak{p}_r$, under the respective assumptions.

An optimal smoothing parameter for the estimation of m and c can be obtained by minimizing the asymptotic mean integrated squared error (AMISE) of the corresponding estimators:

$$\text{AMISE}(\hat{m}) = \mathbb{E}\left[\int (\hat{m}(\delta) - m(\delta))^2 w_1(\delta)d\delta\right] \text{ and } \text{AMISE}(\hat{c}) = \mathbb{E}\left[\int (\hat{c}(\delta) - c(\delta))^2 w_2(\delta)d\delta\right].$$

In the AMISE expression, we introduce two weight functions: w_1 for the conditional modal direction and w_2 for the conditional concentration. These weights are included to improve numerical stability when needed and to allow differential emphasis if a shared error criterion is used. Since both conditional parameter functions are unknown, we propose using a (weighted) sum of the AMISE functions, defined as $\text{AMISE}(\hat{m}, \hat{c}) = \text{AMISE}(\hat{m}) + \text{AMISE}(\hat{c})$, which serves as the basis for deriving a common optimal smoothing parameter. The expressions for these quantities follow from Corollary 1, and the optimal smoothing parameter that minimizes $\text{AMISE}(\hat{m}, \hat{c})$ is derived in Corollary 2.

Corollary 2. *Under the main assumptions of Theorem 1 and Condition (D8), and given the weight functions w_1 and w_2, the optimal bandwidth parameter, for any positive odd*

value of $\mathfrak{p} = \mathfrak{p}_1 = \mathfrak{p}_2$ *in the linear case is*

$$h_{p_L,p_R;\text{opt}} \overset{\text{linear}}{=} \left(\frac{((\mathfrak{p}+1)!)^2 \xi_{0,\mathfrak{p}} \int \frac{w_1(x)\left(\mathfrak{I}_{11}(x)\right)^{-1}\left(h_1'(m(x))\right)^{-2} + w_2(x)\left(\mathfrak{I}_{22}(x)\right)^{-1}\left(h_2'(c(x))\right)^{-2}}{g_X(x)} dx}{(2\mathfrak{p}+1)\zeta_{0,\mathfrak{p}+1,\mathfrak{p}} \int \left(w_1(x)\left(\frac{\eta_1^{(\mathfrak{p}+1)}(x)}{h_1'(m(x))}\right)^2 + w_2(x)\left(\frac{\eta_2^{(\mathfrak{p}+1)}(x)}{h_2'(c(x))}\right)^2 \right) dx} \right)^{1/(2\mathfrak{p}+3)} n^{-1/(2\mathfrak{p}+3)}.$$

Under the previous assumptions, the optimal AMISE is of order $O(n^{-(2\mathfrak{p}+2)/(2\mathfrak{p}+3)})$.

Assuming also Condition (*D*9), *in the circular case, for the local-linear case* ($\mathfrak{p}_1 = \mathfrak{p}_2 = 1$), *we obtain that, as* $n \to \infty$, *the optimal concentration parameter can be derived from*

$$\mu_{2;p_L,p_R;\text{opt}} \overset{\text{circular}}{=} \left(\frac{\int \frac{w_1(\theta)\left(\mathfrak{I}_{11}(\theta)\right)^{-1}\left(h_1'(m(\theta))\right)^{-2} + w_2(\theta)\left(\mathfrak{I}_{22}(\theta)\right)^{-1}\left(h_2'(c(\theta))\right)^{-2}}{g_\Theta(\theta)} d\theta}{2\sqrt{\pi} \int \left(w_1(\theta)\left(\frac{\eta_1''(\theta)}{h_1'(m(\theta))}\right)^2 + w_2(\theta)\left(\frac{\eta_2''(\theta)}{h_2'(c(\theta))}\right)^2 \right) d\theta} \right)^{2/5} n^{-2/5}.$$

Under the previous assumptions, for $\mathfrak{p}_1 = \mathfrak{p}_2 = 1$, *both in the linear and the circular case, the optimal AMISE is of order* $O(n^{-4/5})$.

Note that while for the linear case Corollary 2 already provides the expression for the optimal bandwidth, it does not provide the optimal concentration parameter $\nu_{p_L,p_R;\text{opt}}$ for the circular case. In Section 5.1, we further comment on how to obtain the value of the concentration parameter $\nu_{p_L,p_R;\text{opt}}$ from $\mu_{2;p_L,p_R;\text{opt}}$.

Appendix D3 provides the value of the local smoothing parameter as a function of δ, derived from the asymptotic mean squared error (AMSE) expression. A similar approach to that in Appendix D3 could be used to obtain two separate smoothing parameters, one for the conditional modal regression and another for the concentration.

The extension of the optimal concentration parameter in the circular case, for a general value of $\mathfrak{p}$, is generally not straightforward to obtain. In Appendix D3, we comment on this topic and how Condition (D9) could be relaxed.

5 | PRACTICAL CONSIDERATIONS

In the following, we provide some practical considerations when employing the results of Section 4. In the first part, some advice regarding the kernel choice is given, where we also study whether assumptions are fulfilled for some standard kernel choices. In the second part, we study how to obtain data-driven smoothing parameters.

5.1 | Regarding the kernel

In this part, we focus on some special considerations related to the kernel choice in the circular case. For the linear case, we refer to Fan et al. (1997), where the linear kernels are analyzed in detail. In their Theorem 2.1, it is found that the optimal kernel in terms of the AMISE is the

Epanechnikov kernel, and the minimum variance kernel is the uniform kernel. Among the set of assumptions in Appendix A, only Condition (D5) refers to linear kernels, and this assumption is satisfied by both kernels. The optimal bandwidth parameter, for the local-linear case, only depends on two values related to the chosen kernel (see Corollary 2), namely $\zeta_{0,2,1}$ and $\xi_{0,1}$. Their values are given in Table 1.

The conditions on the circular kernel and the quantities μ_k, τ_k, and $\omega_{k,l}$ can be analyzed in a simple way when the symmetric kernel admits the following convergent Fourier representation,

$$K_\nu(\theta) \overset{\text{circular}}{=} \frac{1}{2\pi}\left(1 + 2\sum_{j=1}^{\infty}\alpha_j(\nu)\cos(j\theta)\right). \tag{6}$$

In that case, using trigonometric identities and integrals involving sine and cosine, we get that, for k an even strictly positive integer,

$$\mu_k \overset{\text{circular}}{=} \frac{1}{2^k}\binom{k}{k/2} + \frac{1}{2^{k-1}}\sum_{j=0}^{k/2-1}(-1)^{j+k/2}\binom{k}{j}\alpha_{k-2j}(\nu), \qquad \mu_2 \overset{\text{circular}}{=} (1-\alpha_2(\nu))/2, \tag{7}$$

$$\tau_0 \overset{\text{circular}}{=} \frac{1}{2\pi} + \frac{1}{\pi}\sum_{j=1}^{\infty}\alpha_j^2(\nu). \tag{8}$$

Among the choices for circular kernels that admit the convergent Fourier representation (6), we analyzed in Table 1 two classical choices: The von Mises and the wrapped normal kernels. For these kernels, the expression of α_k is given by $\alpha_k(\nu_n) = \mathcal{I}_k(\nu_n)/\mathcal{I}_0(\nu_n)$, for the von Mises kernel, and by $\alpha_k(\nu_n) = \nu_n^{k^2}$, for the wrapped normal kernel.

Using the equality in Equation (7) for these two kernels (see Table 1), we obtain that the value of the optimal concentration can be obtained from the quantity $\mu_{2;p_L,p_R;\text{opt}}$ in Corollary 2, as follows,

$$\nu_{p_L,p_R;\text{opt}} = \left(\mu_{2;p_L,p_R;\text{opt}}\right)^{-1}, \text{ for the von Mises kernel,}$$
$$\nu_{p_L,p_R;\text{opt}} = \left(1 - 2\mu_{2;p_L,p_R;\text{opt}}\right)^{1/4}, \text{ for the wrapped normal kernel.}$$

These kernels are bounded and, thus, satisfy Condition (D5). Regarding the smoothing parameter, in the linear case, we only need to guarantee that the smoothing parameter satisfies $h_n \to 0$ and $nh_n \to \infty$, to derive Corollary 2. On the contrary, for the circular case, Conditions (D6) and (D9) also depend on the employed kernel.

For all the circular kernels that admit the expression (6), $\mu_{k;n} \to 0$ (first part of Condition (D6)) if $\alpha_k(\nu_n) \to 1$. Thus, as $n \to \infty$, we will impose that $\nu_n \to \infty$, for the von Mises kernel; and $\nu_n \to 1$, for the wrapped normal kernel. From Equality (8), we can see that τ_0 converges to the value in Table 1, which implies Condition (D9). Since $\xi_{0,1} = \tau_0$, the second part of Condition (D6) ($k = 0$, $\mathfrak{p}_r = 1$) is satisfied if, as $n \to \infty$, $n^2/\nu_n \to \infty$, for the von Mises kernel; and $n^2(1-\nu_n^4) \to \infty$, for the wrapped normal kernel. Note also that, in the theoretical results, for these two kernels $\omega_{j,l}$ can be replaced by μ_{j+l} as $\mu_{k+2} = o(\mu_k)$, when k is an even positive integer (see Remark A2 in Appendix A).

We observe the following regarding two other classical choices of symmetric circular densities: Cardioid kernel, $K_{C;\nu}(\theta) = (1 + 2\nu\cos(\theta))/(2\pi)$, with $|\nu| < 1/2$; and wrapped Cauchy kernel

$K_{\mathrm{WC};\nu}(\theta) = (1-\nu^2)/(2\pi(1+\nu^2-2\nu\cos(\theta)))$, with $\nu \in [0,1]$. For the cardioid kernel, we obtain that, independently of ν, $\mu_2 = 0.5$. Thus, the cardioid density should not be employed as a kernel as it does not satisfy Condition (D6).

For the wrapped Cauchy density, we obtain that $\alpha_k(\nu) = \nu^k$, and thus, $\alpha_k(\nu_n) \to 1$, if $\nu_n \to 1$ as $n \to \infty$. In that case $\lim_{\nu_n \to 1} \mu_{2;n}/\mu_{4;n} = 2$. Thus, the extended asymptotic theory in Appendix D needs to be employed as μ_{j+l} cannot be replaced by $\omega_{j,l}$ (see Remark AP2 in Appendix A). Also, it can be seen that Condition (D9) is not satisfied for this kernel. Thus, the expression of the optimal smoothing parameter cannot be derived from Corollary 2. The optimal concentration is obtained in this case by using that, as $n \to \infty$, $\tau_{0;n} = 1/(2\pi\mu_{2;n})$. Consequently, applying the same reasoning as the one employed in Appendix D3, we obtain that the optimal AMISE for this kernel is $O(n^{-2/3})$.

By employing Corollary 2, we can see that the optimal AMISE for the von Mises and wrapped normal kernels is of order $O(n^{-4/3})$. Thus, better results in terms of AMISE are expected when employing one of these two kernels when compared when using the wrapped Cauchy kernel. In practice, we recommend to employ the von Mises kernel over the wrapped normal kernel for two reasons: (i) it is the standard kernel employed in the nonparametric circular literature (see, e.g., Di Marzio et al., 2009), and (ii) it is computationally more efficient as it does not involve an infinite sum (see Table 1).

5.2 | Regarding the smoothing parameter

When considering the asymptotic optimal smoothing parameters of Corollary 2, we can see that the values depend on some quantities that are unknown in practice: $m(\cdot)$, $c(\cdot)$, $\eta_1''(\cdot)$, $\eta_2''(\cdot)$ and $g_\Delta(\cdot)$. In this section, we propose a practical rule of thumb smoothing parameter that follows the ideas of Fan and Gijbels (1996, Ch. 4). First, for $r \in \{1,2\}$, η_r is estimated by globally fitting a polynomial of order $\mathfrak{p}_r + 3$. In the local-linear case ($\mathfrak{p}_1 = \mathfrak{p}_2 = 1$), we suggest to estimate η_r with the parametric fit

$$\check{\eta}_r(x) \overset{\text{linear}}{=} \check{\eta}_{r,4}(x) = \check{\eta}_{r;0} + \check{\eta}_{r;1}x + \check{\eta}_{r;2}x^2 + \check{\eta}_{r;3}x^3 + \check{\eta}_{r;4}x^4,$$
$$\check{\eta}_r(\theta) \overset{\text{circular}}{=} \check{\eta}^{\circ}_{r,4}(x) = \check{\eta}_{r;0} + \check{\eta}_{r;1}\sin(\theta) + \check{\eta}_{r;2}\sin^2(\theta) + \check{\eta}_{r;3}\sin^3(\theta) + \check{\eta}_{r;4}\sin^4(\theta).$$

In the expressions of Corollary 2, consider the weight $w_1(\delta) = w_2(\delta) = w_0(\delta)g_\Delta(\delta)$, for some specific function w_0. Then, in the numerator, the quantity $\mathfrak{F}_{rr}(\delta)$ can be roughly approximated by the constant value

$$\check{\mathfrak{F}}_{rr} = \frac{1}{n}\sum_{i=1}^{n}\check{\mathfrak{F}}_{rr}(\Delta_i), \text{ where } \check{\mathfrak{F}}_{rr}(\delta) = \mathfrak{F}_{rr}(h_2^{-1}(\check{\eta}_2(\delta))).$$

The other values of the numerator can be estimated replacing $h_1'(m(\delta))$ and $h_2'(c(\delta))$ by $h_1'(h_1^{-1}(\check{\eta}_1(\delta)))$ and $h_2'(h_2^{-1}(\check{\eta}_2(\delta)))$. In the denominator, we can estimate

$$\int \left(\frac{\eta_1''(\delta)}{h_1'(m(\delta))}\right)^2 w_1(\delta)d\delta \text{ and } \int \left(\frac{\eta_2''(\delta)}{h_2'(c(\delta))}\right)^2 w_2(\delta)d\delta \text{ with } \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\check{\eta}_r''(\Delta_i)}{h_r'(h_r^{-1}(\check{\eta}_r(\Delta_i)))}\right)^2 w_0(\Delta_i).$$

Given a weight function w_0, replacing the previous estimators in the optimal smoothing parameters, under the hypothesis of Corollary 2, we obtain the following rule of thumb (RoT) selectors for the local-linear case,

$$h^{\text{RoT}}_{p_L,p_R;n} \overset{\text{linear}}{=} \left(\frac{4\xi_{0,1}\sum_{r=1}^{2}\breve{\mathfrak{F}}_{rr}^{-1}\int \left(h_r'(h_r^{-1}(\breve{\eta}_r(x)))\right)^{-2} w_0(x)dx}{3\zeta_{0,2,1}\sum_{r=1}^{2}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{2\breve{\eta}_{r;2}+6\breve{\eta}_{r;3}X_i+12\breve{\eta}_{r;4}X_i^2}{h_r'(h_r^{-1}(\breve{\eta}_r(X_i)))}\right)^2 w_0(X_i)} \right)^{1/5} n^{-1/5}, \tag{9}$$

$$\mu^{\text{RoT}}_{2;p_L,p_R;n} \overset{\text{circular}}{=} \left(\frac{\sum_{r=1}^{2}\breve{\mathfrak{F}}_{rr}^{-1}\int \left(h_r'(h_r^{-1}(\breve{\eta}_r(\theta)))\right)^{-2} w_0(\theta)d\theta}{2\sqrt{\pi}\sum_{r=1}^{2}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\breve{\eta}_r''(\Theta_i)}{h_r'(h_r^{-1}(\breve{\eta}_r(\Theta_i)))}\right)^2 w_0(\Theta_i)} \right)^{2/5} n^{-2/5}, \tag{10}$$

$$\begin{aligned}\breve{\eta}_r''(\Theta_i) \overset{\text{circular}}{=}\ & 2\cos^2(\Theta_i)\left(\breve{\eta}_{r;2}+3\breve{\eta}_{r;3}\sin(\Theta_i)+6\breve{\eta}_{r;4}\sin^2(\Theta_i)\right)-\sin(\Theta_i)\\ & \times\left(\breve{\eta}_{r;1}+2\breve{\eta}_{r;2}\sin(\Theta_i)+3\breve{\eta}_{r;3}\sin^2(\Theta_i)+4\breve{\eta}_{r;4}\sin^3(\Theta_i)\right).\end{aligned}$$

Thus, a data-driven bandwidth for the linear case with an Epanechnikov kernel can be directly obtained from a sample using $h^{\text{RoT}}_{p_L,p_R;n}$, as given in Equation (9), by substituting $\xi_{0,1}$ with $3/5$ and $\zeta_{0,2,1}$ with $1/5$ (see Table 1). Meanwhile, for the circular case with a von Mises kernel, one can directly employ (10), and the data-driven concentration parameter would be equal to $\nu^{\text{RoT}}_{p_L,p_R;n} = 1/\mu^{\text{RoT}}_{2;p_L,p_R;n}$.

The weight function w_0 is assumed to be known. For example, a standard choice, and the one employed in Section 6 and Appendix B, is to use w_0 as the indicator function in the interval of interest $[a, b]$ in the linear predictor case, or the arc $[-\pi, \pi)$ in the circular case.

A simulation study is presented in Appendix B to assess the performance of the estimators of the conditional modal direction in the proposed semiparametric regression model. In this study, we use either the Epanechnikov or the von Mises kernel along with the proposed data-driven smoothing parameter. We also compare the performance of our estimator with that of the nonparametric modal estimator introduced by Alonso-Pena and Crujeiras (2023) (see Section 2.2). Specifically, we consider two scenarios: One in which the data-generating process satisfies the assumptions underlying the proposed semiparametric model, and another in which these assumptions are violated (i.e., the conditional density cannot be well approximated by (4)). This allows us to demonstrate the good practical performance and robustness of our method under both correctly specified and misspecified conditional densities.

6 | REAL DATA APPLICATION

The objective of this section is to analyze the flight orientation of nocturnal migratory songbirds. Specifically, building upon the studies conducted by Sjöberg and Nilsson (2015b), our interest lies in examining how flight altitude and wind direction affect the birds' orientation during level flight, as well as climbing flights (see later), and this with regard to modal direction and concentration.

In general, it is assumed that migratory birds choose their direction before leaving their stopovers. This would mean that the orientation of the birds should be the same, regardless of the altitude of the flight. In Sjöberg and Nilsson (2015b), discretizing the birds' flight altitude (see also Section 1), they found that the track directions of birds at high altitudes were more concentrated, which would be indicative that the birds adjust their traveling direction once aloft. Using our semiparametric model, we study how the concentration varies depending on the flight altitude, both when they are climbing or in level flight.

Differently to the birds' altitude, wind conditions have been shown to have an important effect on the birds' orientation. The second objective of this study is to see how the wind orientation alters the modal direction in the distribution of the passerines' orientation and to analyze if it also produces changes in the concentration.

The data containing the orientation of songbird migrants can be found in Sjöberg and Nilsson (2015a). This dataset was obtained using an X-band tracking radar located at Falsterbo peninsula, Sweden (55°38′ N, 12°82′ E). The radar was operated from sunset to a few hours after midnight on selected nights during the autumns of 2009–2011. In the collected data, flights were subdivided into climbing flights (vertical speed greater than 0.75ms^{-1}, with a sample size of $n_c = 356$) and level flights (vertical speed between -0.75ms^{-1} and 0.75ms^{-1}, with a sample size of $n_l = 1266$). Descending flights (vertical speed lower than -0.75ms^{-1}) were excluded. Wind orientation was recorded by tracking weather balloons. More details on the radar operation and data processing can be found in Nilsson et al. (2014). Note that, as in Sjöberg and Nilsson (2015b), we assumed that the collected flight orientations are independent. In our analysis, we also distinguish between climbing flights and level flights.

We first assume that the flight altitude and the wind direction do not influence the modal direction or the concentration. Then, by observing the histograms and the circular kernel density estimates in Figures 4 and 5 (right panels), we see that, in all the studied cases, the distribution is unimodal and well-approximated when using the generalized Batschelet density (2), with parameters estimated from the whole sample via maximum likelihood (dashed lines). Note that we have employed linear/planar plots to improve visualization, but it is important to remember that $-\pi$ is connected to π. The circular equivalent of the histograms, the rose diagrams, are shown in Figure 1. As mentioned before, the objective is to study the effect that the altitude or the wind direction has on the modal direction and the concentration of the birds' orientation. Thus, with the objective of comparing with the non-effect model, in our semiparametric approach, we employ the conditional generalized Batschelet density with peakedness parameters equal to those estimated from the whole sample (see dashed lines in Figures 4 and 5). In Figures 4 and 5 (left and central panels, triangle-dashed lines), we have also included the parametric estimators discussed in Section 2.1. For the altitude predictor (Figure 4), we used the Fisher and Lee (1992) regression model with $p_1 = p_2 = 1$ and the link functions $\check{h}_1^{-1}(u; l, 1)$ and $\check{h}_2^{-1}(u)$. For the wind orientation predictor (Figure 5), we employed the arc-tangent-sine link model from Kim and SenGupta (2017), allowing for a conditional concentration with $c(\theta) = \exp(\check{\eta}_{r;0} + \check{\eta}_{r;1}\sin(\theta))$. More complex polynomial parametric models, with larger values of p_1 and p_2, are presented in Appendix C. Figures 4 and 5 (left panels, dot-dashed curves) also include the nonparametric conditional (uni)modal direction estimator. On these left panels, we have modified the display interval for flight orientations to the range $[0, 2\pi)$ to enhance the visibility of the conditional modal direction.

From these estimators, we see that the obtained conditional location estimators are almost constant for the altitude predictor (Figure 4, left panels), while they vary with the wind orientation

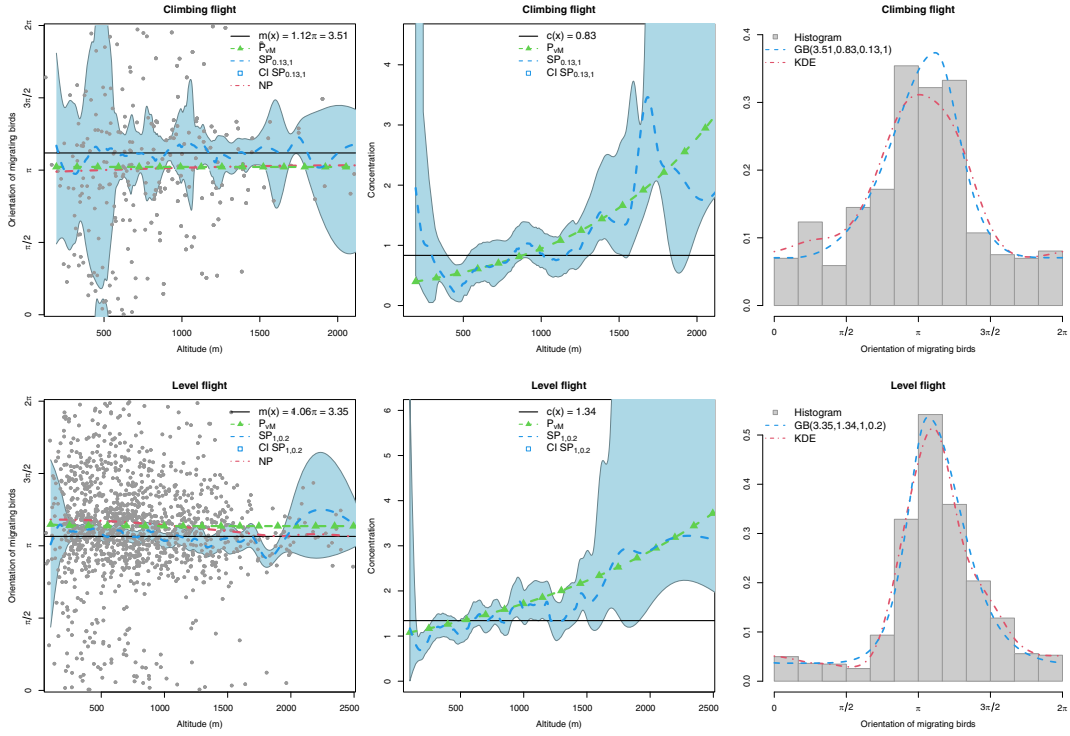

FIGURE 4 Gray dots: Flight orientation and altitude for the migration birds in climbing (top) and level (bottom) flight. Modal direction (left) and concentration (center), assuming that they do (discontinuous lines) or not (continuous lines), depend on the regressor. Parametric (triangle-dashed) and semiparametric (dashed) estimators. Colored (light blue) area: 95% confidence bands based on the semiparametric approach. Dot-dashed lines (NP, left): Nonparametric modal direction estimation. Right: Non-regression flight orientation.

(Figure 5, left panels). This could be considered as evidence against a constant modal orientation of birds' orientation for the wind direction regressor. For wind orientation, the parametric model captures the change in pattern but does not account for the fact that the most significant change (compared to the unconditional case) occurs near the SE wind orientation. For the altitude predictor (Figure 4, left panels), when comparing the parametric and the semiparametric location estimators, we also observe that the two curves are nearly parallel but do not coincide. This discrepancy arises because the conditional generalized Batschelet distribution is asymmetric, causing the conditional mean and modal direction to differ. Since the peakedness parameters are selected from the unconditional sample, below, we conduct a goodness-of-fit test to show that the employed semiparametric asymmetric distribution is a plausible model.

For the conditional concentration (Figure 4, center panels), both estimators indicate that concentration increases with altitude (at least for altitudes between 500 and 1,700m). However, when examining the relationship between concentration and wind orientation (Figure 5, center panels), the parametric model returns an almost constant concentration, whereas the semiparametric model shows higher values for wind orientations closer to 0 (N) and lower values closer to π (S). This latter result seems to support the hypothesis of a change in concentration with the predictor as suggested by Sjöberg and Nilsson (2015a).

To investigate whether our semiparametric approach provides a reasonable fit, we performed the following goodness-of-fit test over the estimation obtained when the peakedness parameters

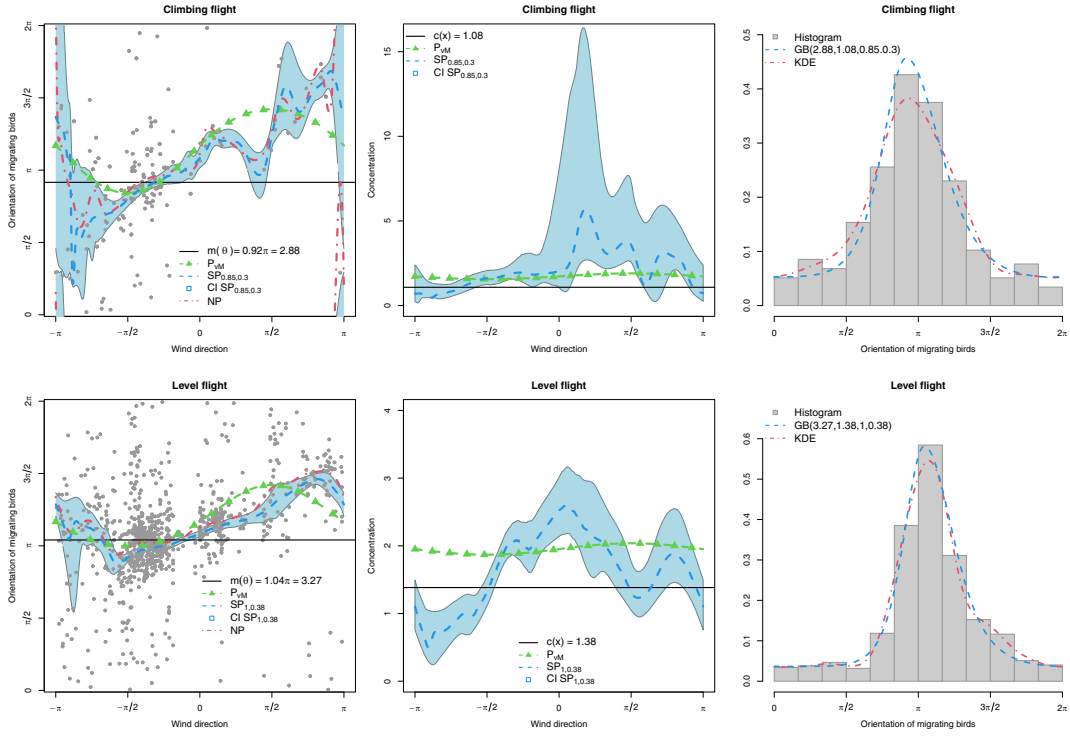

FIGURE 5 Gray dots: Flight orientation and wind direction for the migration birds in climbing (top) and level (bottom) flight. Modal direction (left) and concentration (center), assuming that they do (discontinuous lines) or not (continuous lines), depend on the regressor. Parametric (triangle-dashed) and semiparametric (dashed) punctual estimations. Colored (light blue) area: 95% confidence bands based on the semiparametric approach. Dot-dashed lines (NP, left): Nonparametric modal direction estimation. Right: Non-regression flight orientation.

are estimated from the whole sample. First, since we are working here with a conditional generalized Batschelet distribution, we considered its cumulative distribution function:

$$G_{\Phi|\Delta;p_L,p_R}(\phi;\hat{m}(\Delta),\hat{c}(\Delta)|\Delta=\delta)=\int_{-\pi}^{\phi} g_{\Phi|\Delta;p_L,p_R}(\varphi;\hat{m}(\delta),\hat{c}(\delta)|\Delta=\delta)d\varphi.$$

Then, from the original sample, $(\Phi_1,\Delta_1),\ldots,(\Phi_n,\Delta_n)$, considering the circular probability integral transformation, the following values would be obtained,

$$\Upsilon=\left(2\pi G_{\Phi|\Delta,p_L,p_R}(\Phi_1;\hat{m}(\Delta),\hat{c}(\Delta)|\Delta=\Delta_1),\ldots,2\pi G_{\Phi|\Delta,p_L,p_R}(\Phi_n;\hat{m}(\Delta),\hat{c}(\Delta)|\Delta=\Delta_n)\right). \quad (11)$$

Using the sample Υ, we calculated the test statistic U^2 of Watson (1961) testing for uniformity. As mentioned, e.g., in Pewsey et al. (2013, Section 6.2.3), the usual critical values of the test for circular uniformity should only be employed for "large" sample sizes. For that reason, we have also performed a bootstrap procedure. Conditionally on each Δ_i, $i\in\{1,\ldots,n\}$, we have generated 1000 values Φ_i^{j*}, $j\in\{1,\ldots,1000\}$, from the distribution associated to $g_{\Phi|\Delta;p_L,p_R}(\cdot;\hat{m}(\Delta),\hat{c}(\Delta)|\Delta=\Delta_i)$. For each j, the bootstrap resample Υ^{j*}, is calculated as in Equation (11), replacing the values of Φ_i by Φ_i^{j*}. Finally, the value of the test statistic of Watson (1961), $U^{2\,j*}$, is computed for each

TABLE 2 Obtained p-values when testing if the flight orientation data is generated from the conditional generalized Batschelet distribution.

Regressor	**Altitude**				**Wind direction**			
Flight	**Climbing**		**Level**		**Climbing**		**Level**	
Calibration	Asym	Boot	Asym	Boot	Asym	Boot	Asym	Boot
p-value	0.257	0.252	0.653	0.669	0.619	0.624	0.644	0.633

Note: Results are obtained using the usual critical values of the test for circular uniformity (Asym) and employing the proposed bootstrap procedure (Boot).

bootstrap resample Υ^{j*}. Since the null hypothesis is rejected for large values of the test statistic, the p-values are equal to $\sum_{j=1}^{1000}\mathbb{I}(U^{2\,j*} > U^2)/1000$, where $\mathbb{I}$ is the indicator function. The results employing the usual (asymptotic) critical values and the bootstrap approach are reported in Table 2. Similar p-values are obtained using both approaches, which could be an indication that standard critical values can be employed in this case. For a significance level of $\alpha = 0.05$, the results shown in Table 2 indicate that it cannot be rejected that the semiparametric regression model provides a reasonable fit for this data in the four considered scenarios (two predictors, and two groups of flight types). Note that this testing procedure may be quite conservative. In Appendix B, we discuss this in more detail and assess its calibration and power through a brief simulation study.

In the following, we explore in more detail whether we can assume a non-effect model for the modal direction and concentration. It is noteworthy that if the focus is on the mean direction rather than the modal direction, the non-effect test proposed by Alonso-Pena et al. (2021) could be employed. However, to the best of the authors' knowledge, there is no non-effect test available specifically for the modal direction. Since the objective is to compare with the non-dependent model, we have constructed 95% confidence bands (pointwise confidence intervals) for the modal direction and concentration (light blue area in Figures 4 and 5) for the semiparametric model (with peakedness parameters estimated from the entire sample). To compute the confidence intervals, one strategy would be to use the asymptotic normality results established in Corollary 1. Given a smoothing parameter, the asymptotic confidence intervals could be obtained by replacing the unknowns in Corollary 1 with a parametric global fitting, as done to obtain the rule of thumb in Section 5.2. However, these confidence intervals would depend on the initial choice of the smoothing parameter, and they may not be accurate, as, in some of the analyzed scenarios, the sample size is rather small. For that reason, we decided to employ a bootstrap procedure. Different resampling strategies could be employed, such as generating resamples from the estimated conditional distribution. Since the objective is only illustrative, we simply generated 1000 replicates by sampling with replacement from the original bivariate samples containing the birds' orientation and the employed regressor. The confidence bands were then constructed by estimating the modal direction and concentration curves for each bootstrap resample, employing the peakedness and smoothing parameters of the original sample. For each value of δ, confidence limits of $c(\delta)$ are obtained by considering the 0.025 and 0.975 sample quantiles of the bootstrap estimators $\hat{c}^{*1}(\delta),\ldots,\hat{c}^{*1000}(\delta)$. Constructing the confidence intervals of $m(\delta)$ is not straightforward due to the lack of natural order on the circle. For that reason, we decided to use as a criteria the geodesic distance to $\hat{m}(\delta)$. Then, the confidence interval is obtained from the smallest arc containing 95% of the bootstrap estimators $\hat{m}^{*1}(\delta),\ldots,\hat{m}^{*1000}(\delta)$. The performance of this bootstrap confidence interval procedure, in terms of coverage, is analyzed in more detail in Appendix B through a simulation study.

According to these 95% confidence bands, we can assume that the modal direction is constant in the climbing flights, with a value around 1.1π (Figure 4, top-left panel). For level flights (Figure 4, bottom-left panel), the modal direction of migratory birds changes slightly: It is close to 1.1π when the altitude is "low" and close to π when the altitude is "high." Regarding the other regressor, we observe that the modal direction clearly changes with the wind direction (Figure 5, left panels). For the climbing flights (Figure 5, top-left panel), it can be assumed that the modal direction changes "linearly" in the arcs $(-0.9\pi, 0.1\pi)$ and $(\pi/2, 0.9\pi)$ (with modal orientation going from $\pi/2$ to $5\pi/3$). In the level flights (Figure 5, bottom-left panel), it appears that the modal direction changes linearly within the arc $(-\pi/2, 3\pi/4)$, with the modal orientation shifting from 0.9π to $3\pi/2$. For other wind directions, we can assume that the modal direction changes linearly in the opposite orientation. The range of modal orientations depending on the wind direction is bigger for climbing than for level flights. In Appendix C, we explore whether a parametric fit, such as those discussed in Section 2.1, can estimate the conditional modal direction.

Regarding the concentration, by observing the 95% confidence bands for the concentration (light blue area in the center panels of Figures 4 and 5), we obtain the same conclusion as in Sjöberg and Nilsson (2015b), which is that the concentration depends on the altitude and on the wind. In climbing flights, when considering the altitude as a regressor (Figure 4, top-center panel), we can assume that the parametric model provides a reasonable fit as it is within the 95% confidence bands. In the level flights (Figure 4, bottom-center panel), we can assume a linear increasing relationship between both variables. When considering the wind direction regressor (Figure 5, center panels), we observe that the conditional concentration increases between $-7\pi/8$, with a punctual value of 0.5, and $\pi/8$, with a punctual value of 5 for the climbing flights, and 2.7 for the level flights. For other wind directions, a decreasing pattern could be assumed, but it is noteworthy that the point estimation reveals a second peak around $3\pi/4$. In Appendix C, we also examine the behavior of the parametric conditional concentration estimator.

7 | DISCUSSION

In this paper, we proposed a new local polynomial regression methodology for circular responses. Several advantages can be found in the proposed semiparametric model. It allows for flexible modeling of both the modal direction and the concentration depending on a linear or circular regressor. It also allows for an asymmetric conditional density, controlling the peakedness at the left and at the right of the conditional modal direction. Arguable disadvantages of the proposed semiparametric methodology are the lower computational efficiency and that a conditional distribution is assumed. But the latter may also be useful when alternative density features, such as the dependent mean direction, need to be explored.

As is standard in kernel-based methods, the performance of the estimators depends on the smoothing parameter (bandwidth or concentration). However, the asymptotic theory developed in Section 4 allowed us to determine an optimal smoothing parameter, which led to a practical rule-of-thumb smoothing parameter. A common optimal smoothing parameter was proposed, but as noted in Section 4, similar arguments can be used to estimate separate smoothing parameters for the conditional modal direction and the concentration. The proposed criterion was employed because the primary aim of this paper is to estimate both these functions. However, if the objective were to estimate the full conditional density, one could instead use the AMSE/AMISE of the conditional density function, $g_{\Phi|\Delta;p_L,p_R}(\phi; \hat{m}(\Delta), \hat{c}(\Delta)|\Delta = \delta)$, to evaluate the performance of the estimator.

The developed theory already covers important subcases, such as the semiparametric model when assuming that the conditional response follows a von Mises distribution (or any other standard two-parameter circular distribution). When the conditional distribution is asymmetric, as observed in the simulation study of Appendix B and the real data application of Section 6, the selection of the peakedness parameters plays a significant role. Further investigation needs to be conducted to determine the most suitable peakedness parameter estimates from the sample. One idea, employed in Section 6, is to estimate them from the whole sample. This could be done in the real data application, as, by default, it is assumed that the modal and concentration parameters do not vary with the regressor. However, note that if the true model is similar to Model C in the simulation study of Appendix B, the unconditional data would be close to the uniform distribution. Thus, by using this strategy, the peakedness parameters would not be correctly estimated. A more promising direction would be to extend the ideas presented in this paper in the spirit discussed at the end of Section 3, considering the case where the peakedness parameters depend on the regressor. Specifically, it would be of interest to explore the case where $p_L(\delta) = h_3^{-1}(\eta_3(\delta))$ and $p_R(\delta) = h_3^{-1}(\eta_4(\delta))$. In this scenario, it is important to note that not only the theoretical aspects but also the numerical implementation of this extension would be more complex due to the increased number of parameter functions that would need to be estimated.

In this paper, we focused on the case where the circular response depends on a single linear or circular covariate Δ. For readers interested in local likelihood regression, where the response is linear (belonging to the double exponential family) and the predictor is circular, we refer to Alonso-Pena et al. (2023, 2024).

An interesting extension of the current paper would be the multivariate regression problem, where the response variable Φ depends on a vector of regressors $\boldsymbol{\Delta} = (\Delta_1, \ldots, \Delta_d)^\top$. For example, related to our case study in Section 6, it would be of interest to study the flight orientation for the migrating birds, depending at the same time on both the altitude and the wind direction. In that case, similar ideas to those in this paper could be employed by considering a product kernel and the following Taylor expansion around $\boldsymbol{\delta}_0$,

$$\eta_r(\boldsymbol{\delta}) \approx \beta_{r;0} + \sum_{j_1 : \Delta_{j_1} \text{ is a linear r.v.}} \beta_{r;j_1}(x_{j_1} - x_{0;j_1}) + \sum_{j_2 : \Delta_{j_2} \text{ is a circular r.v.}} \beta_{r;j_2} \sin(\theta_{j_2} - \theta_{0;j_2}),$$

with $r \in \{1, 2\}$. We further elaborate on this idea in Appendix C. The results obtained using the multivariate semiparametric estimator with the real data from Section 6 are depicted in Figure C2 in Appendix C.

Another area for future investigation could involve obtaining an alternative data-driven smoothing parameter. In this paper, a rule-of-thumb selection rule was proposed. An alternative would be to employ more flexible parametric models or a pilot estimation for the derivatives to replace the unknown quantities $\eta_r''(\delta)$, with $r \in \{1, 2\}$, in Corollary 2. Other alternatives include cross-validation or a method based on the Akaike information criterion. For further details, in the linear setting, see Loader (2006, Ch. 10 and 11).

Finally, a challenging future research topic is the construction of data-driven asymptotic confidence bands for the dependent modal direction and concentration. For each analyzed value of δ, the confidence intervals could be derived from Corollary 1. In that case, further investigation is needed to determine how to estimate the unknowns and which smoothing parameter should be employed to estimate the bias and the variance. We refer to Fan and Gijbels (1996, Section 4.4) for a discussion on this topic in the linear setting.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Editor, an Associate Editor, and two reviewers for valuable comments that led to an improvement of the manuscript. This work is supported by Grant PID2020-116587GB-I00 funded by MICIU/AEI/10.13039/501100011033, the Competitive Reference Groups 2021–2024 (ED431C 2021/24) from the Xunta de Galicia for J. Ameijeiras-Alonso, and the C16/20/002 project (Research Fund KU Leuven) for I.Gijbels. This research was initiated when the first author was a postdoctoral researcher at KU Leuven.

CONFLICT OF INTEREST STATEMENT

The authors declare no potential conflict of interest.

FUNDING INFORMATION

None reported.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the supplementary material of this article.

ORCID

Jose Ameijeiras-Alonso https://orcid.org/0000-0002-4122-6992

Irène Gijbels https://orcid.org/0000-0002-4443-9803

REFERENCES

Alonso-Pena, M., Ameijeiras-Alonso, J., & Crujeiras, R. M. (2021). Nonparametric tests for circular regression. *Journal of Statistical Computation and Simulation*, *91*(3), 477–500.

Alonso-Pena, M., & Crujeiras, R. M. (2023). Analyzing animal escape data with circular nonparametric multimodal regression. *The Annals of Applied Statistics*, *17*(1), 130–152.

Alonso-Pena, M., & Crujeiras, R. M. (2025). *A review on nonparametric circular regression*. In S. Kumar, B. C. Arnold, K. Shimizu, & A. K. Laha (Eds.), *Directional and Multivariate Statistics*. Springer. https://doi.org/10.1007/978-981-96-2004-3_9.

Alonso-Pena, M., Gijbels, I., & Crujeiras, R. M. (2023). Flexible joint modeling of mean and dispersion for the directional tuning of neuronal spike counts. *Biometrics*, *79*(4), 3431–3444.

Alonso-Pena, M., Gijbels, I., & Crujeiras, R. M. (2024). A general framework for circular local likelihood regression. *Journal of the American Statistical Association*, *119*(548), 2709–2721.

Ameijeiras-Alonso, J., Gijbels, I., & Verhasselt, A. (2022). On a family of two–piece circular distributions. *Computational Statistics & Data Analysis*, *168*, 107403.

Di Marzio, M., Panzera, A., & Taylor, C. C. (2009). Local polynomial regression for circular predictors. *Statistics & Probability Letters*, *79*(19), 2066–2075.

Di Marzio, M., Panzera, A., & Taylor, C. C. (2013). Non-parametric regression for circular responses. *Scandinavian Journal of Statistics*, *40*(2), 238–255.

Downs, T. D., & Mardia, K. (2002). Circular regression. *Biometrika*, *89*(3), 683–698.

Fan, J., Gasser, T., Gijbels, I., Brockmann, M., & Engel, J. (1997). Local polynomial regression: Optimal kernels and asymptotic minimax efficiency. *Annals of the Institute of Statistical Mathematics*, *49*(1), 79–99.

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications*. Chapman and Hall.

Fan, J., Heckman, N. E., & Wand, M. P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *Journal of the American Statistical Association*, *90*(429), 141–150.

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

Gao, F., Chia, K. S., Krantz, I., Nordin, P., & Machin, D. (2006). On the application of the von Mises distribution and angular regression methods to investigate the seasonality of disease onset. *Statistics in Medicine*, *25*(9), 1593–1618.

Gould, A. L. (1969). A regression technique for angular variates. *Biometrics*, *25*, 683–700.

Hogan, J. A., McMahon, S. M., Buzzard, V., Michaletz, S. T., Enquist, B. J., Thompson, J., Swenson, N. G., & Zimmerman, J. K. (2019). Drought and the interannual variability of stem growth in an aseasonal, everwet forest. *Biotropica*, *51*(2), 139–154.

Johnson, R. A., & Wehrly, T. E. (1978). Some angular-linear distributions and related regression models. *Journal of the American Statistical Association*, *73*(363), 602–606.

Kato, S., & Jones, M. (2010). A family of distributions on the circle with links to, and applications arising from, Möbius transformation. *Journal of the American Statistical Association*, *105*(489), 249–262.

Kato, S., & Jones, M. (2015). A tractable and interpretable four–parameter family of unimodal distributions on the circle. *Biometrika*, *102*(1), 181–190.

Kim, S., & SenGupta, A. (2016). *Regressions involving circular variables: An overview*. In *Platinum Jubilee International Conference on Applications of Statistics* (pp. 25–33). Springer.

Kim, S., & SenGupta, A. (2017). Multivariate-multiple circular regression. *Journal of Statistical Computation and Simulation*, *87*(7), 1277–1291.

Loader, C. (2006). *Local regression and likelihood*. Springer Science & Business Media.

Meilán-Vila, A., Francisco-Fernández, M., Crujeiras, R. M., & Panzera, A. (2021). Nonparametric multiple regression estimation for circular response. *TEST*, *30*(3), 650–672.

Nilsson, C., Bäckman, J., & Alerstam, T. (2014). Seasonal modulation of flight speed among nocturnal passerine migrants: Differences between short-and long-distance migrants. *Behavioral Ecology and Sociobiology*, *68*(11), 1799–1807.

Oliveira, M., Crujeiras, R. M., & Rodríguez-Casal, A. (2014). NPCirc: An R package for nonparametric circular methods. *Journal of Statistical Software*, *61*(9), 1–26.

Pewsey, A., Neuhäuser, M., & Ruxton, G. D. (2013). *Circular statistics in R*. Oxford University Press.

Pewsey, A., Shimizu, K., & de la Cruz, R. (2011). On an extension of the von Mises distribution due to Batschelet. *Journal of Applied Statistics*, *38*(5), 1073–1085.

Rivest, L. P., Duchesne, T., Nicosia, A., & Fortin, D. (2016). A general angular regression model for the analysis of data on animal movement in ecology. *Journal of the Royal Statistical Society, Series C, Applied Statistics*, *3*(65), 445–463.

Sarma, Y., & Jammalamadaka, S. (1993). *Circular regression*. In K. Matusita, M. L. Puri, & T. Hayakawa (Eds.), *Proceedings of the Third Pacific Area Statistical Conference: Statistical sciences and data analysis* (pp. 109–128). VSP.

Sjöberg, S., & Nilsson, C. (2015a). *Data from: Nocturnal migratory songbirds adjust their travelling direction aloft: evidence from a radiotelemetry and radar study*. Dryad, Dataset https://doi.org/10.5061/dryad.86020

Sjöberg, S., & Nilsson, C. (2015b). Nocturnal migratory songbirds adjust their travelling direction aloft: evidence from a radiotelemetry and radar study. *Biology Letters*, *11*(6), 20150337.

Watson, G. S. (1961). Goodness-of-fit tests on a circle. *Biometrika*, *48*(1/2), 109–114.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ameijeiras-Alonso, J., & Gijbels, I. (2026). Semiparametric regression for circular response with application in ecology. *Scandinavian Journal of Statistics*, *53*(1), 54–101. https://doi.org/10.1111/sjos.70027

APPENDIX A. FURTHER DETAILS ON THE NOTATIONS AND ASSUMPTIONS

We begin this section by stating some important remarks related to the quantities μ_k, τ_k, $\omega_{k,l}$, $\zeta_{j,k,\mathfrak{p}_r}$, and $\xi_{k,\mathfrak{p}_r}$ introduced in Section 4 of the main paper.

Remark A1. In the definitions of μ_k, τ_k, $\omega_{k,l}$, $\zeta_{j,k,\mathfrak{p}_r}$, and $\xi_{k,\mathfrak{p}_r}$; the limits of integration were not included on purpose. For the circular case, we have that the region of integration is $S_\Theta = [-\pi, \pi)$. Denote the marginal density function of the predictor as g_Δ. In the linear case, when the analyzed point x_0 is in the interior of the support of g_X, denoted by S_X, the region of integration will coincide with the support of the kernel, S_L. If it is in the boundary, then given a bandwidth h, the region of integration is $\{z : (x_0 - hz) \in S_X\} \cap S_L$. Further details regarding the modifications in the asymptotic results when x_0 is at the boundary can be found in Fan et al. (1995), or also in Fan and Gijbels (1996).

Remark A2. If L and K_ν are symmetric (and centered on 0), when integrating, respectively, over S_L and $[-\pi, \pi)$, $\mu_k = 0$, if k is odd, and $\omega_{k,l} = 0$, if $(k + l)$ is odd. When k is even, most of the following assumptions and results for circular regressors simplify when, as $n \to \infty$, $\mu_{k+2} = o(\mu_k)$. Under that condition, we can directly employ μ_{j+l} instead of $\omega_{j,l}$ since $\mu_{j+l} = \omega_{j,l}(1 + o(1))$. This is the case for two of the most common circular kernels, the von Mises and the wrapped normal (see Section 5.1). Otherwise, we need to employ also $\omega_{j,l}$ but we always have that, as $n \to \infty$, $\mu_{j+l} = \omega_{j,l}(1 + O(1))$.

We follow now by introducing the different matrices that appear on the asymptotic results: $\mathfrak{I}(\delta_0)$, $\boldsymbol{H}_{\mathfrak{p}_r}$, $\boldsymbol{A}_{\delta_0}$, $\boldsymbol{\Sigma}_{\delta_0}$, and $\boldsymbol{T}$. The Fisher information matrix $\mathfrak{I}(\delta_0)$ can be derived from the results in Ameijeiras-Alonso et al. (2022). For obtaining its expression, let us introduce the following quantities:

$$
\begin{aligned}
D^{k_1,k_2,k_3;-}_{c(\delta),p_L,p_R} &= \int_{-\pi}^{0} [f_{0,c}(\phi + p_L w(\phi))]^{-1} [w'(\phi)]^{k_1} [f'_{0,c}(\phi + p_L w(\phi))]^{k_2} \left[\frac{\partial}{\partial c} f_{0,c}(\phi + p_L w(\phi))\right]^{k_3} d\phi \Big|_{c=c(\delta)},\\
D^{k_1,k_2,k_3;+}_{c(\delta),p_L,p_R} &= \int_{0}^{\pi} [f_{0,c}(\phi + p_R w(\phi))]^{-1} [w'(\phi)]^{k_1} [f'_{0,c}(\phi + p_R w(\phi))]^{k_2} \left[\frac{\partial}{\partial c} f_{0,c}(\phi + p_R w(\phi))\right]^{k_3} d\phi \Big|_{c=c(\delta)},\\
D^{k_1,k_2,k_3;\pm}_{c(\delta),p_L,p_R} &= D^{k_1,k_2,k_3;-}_{c(\delta),p_L,p_R} + D^{k_1,k_2,k_3;+}_{c(\delta),p_L,p_R},\\
C^{c}_{c(\delta),p_L,p_R} &= \frac{\partial}{\partial c} C_{c,p_L,p_R}\Big|_{c=c(\delta)}.
\end{aligned}
$$

From these quantities and under some assumptions (see Conditions C3 and C4 below), the Fisher information matrix of $m(\delta)$ and $c(\delta)$ can be obtained as follows,

$$
\begin{aligned}
\mathfrak{I}(\delta) &= \begin{bmatrix} \mathfrak{I}_{11}(\delta) & 0 \\ 0 & \mathfrak{I}_{22}(\delta) \end{bmatrix}, \qquad \text{(A1)}\\
\mathfrak{I}_{11}(\delta) &= \mathfrak{I}_{11}(c(\delta)) = \frac{1}{C_{c(\delta),p_L,p_R}}\\
&\quad \times \left(D^{0,2,0;\pm}_{c(\delta),p_L,p_R} + p_L^2 D^{2,2,0;-}_{c(\delta),p_L,p_R} + 2p_L D^{1,2,0;-}_{c(\delta),p_L,p_R} + p_R^2 D^{2,2,0;+}_{c(\delta),p_L,p_R} + 2p_R D^{1,2,0;+}_{c(\delta),p_L,p_R} \right),\\
\mathfrak{I}_{22}(\delta) &= \mathfrak{I}_{22}(c(\delta)) = \frac{1}{C_{c(\delta),p_L,p_R}} \left(-\frac{(C^{c}_{c(\delta),p_L,p_R})^2}{C_{c(\delta),p_L,p_R}} + D^{0,0,2;\pm}_{c(\delta),p_L,p_R} \right).
\end{aligned}
$$

The other matrices that also appear in the asymptotic results are defined below. Herein $\otimes$ denotes a generalization of the Kronecker product, which, if $\boldsymbol{C} = (c_{kl})$ is a $(r \times s)$ matrix and $\boldsymbol{D}$ is a partitioned matrix $\boldsymbol{D}$ with submatrices $\boldsymbol{D}_{kl}$, it is defined as $\boldsymbol{C} \otimes \boldsymbol{D} = (c_{kl}\boldsymbol{D}_{kl})$, with $k \in \{1, \ldots, r\}$

and $l \in \{1,\ldots,r\}$.

$$\begin{aligned}
\boldsymbol{H}_{\mathfrak{p}_r} &= \operatorname{diag}\left(1, h, \ldots, h^{\mathfrak{p}_r}\right),\\
\boldsymbol{A}_{\delta_0} &= g_{\Delta}(\delta_0)\mathfrak{F}(\delta_0) \otimes \boldsymbol{U},\\
\boldsymbol{\Sigma}_{\delta_0} &= g_{\Delta}(\delta_0)\mathfrak{F}(\delta_0) \otimes \boldsymbol{T},\\
\boldsymbol{T} &= \begin{pmatrix} (\tau_{k+l-2})_{1\le k,l\le \mathfrak{p}_1+1} & (\tau_{k+l-2})_{1\le k\le \mathfrak{p}_1+1,1\le l\le \mathfrak{p}_2+1}\\ (\tau_{k+l-2})_{1\le k\le \mathfrak{p}_2+1,1\le l\le \mathfrak{p}_1+1} & (\tau_{k+l-2})_{1\le k,l\le \mathfrak{p}_2+1}\end{pmatrix}.
\end{aligned}$$

To establish the necessary conditions to obtain the asymptotic results, first assume that the following partial derivatives of the log-likelihood function exist, with $r, s, t \in \{1, 2\}$:

$$\begin{aligned}
\psi_r(v_1(\delta), v_2(\delta); \phi) &= \frac{\partial}{\partial u_r} \ln g_{\Phi|\Delta;p_L,p_R}(\phi; h_1^{-1}(u_1), h_2^{-1}(u_2)|\Delta = \delta)\Big|_{(u_1,u_2)=(v_1(\Delta),v_2(\Delta))},\\
\psi_{rs}(v_1(\delta), v_2(\delta); \phi) &= \frac{\partial^2}{\partial u_r \partial u_s} \ln g_{\Phi|\Delta;p_L,p_R}(\phi; h_1^{-1}(u_1), h_2^{-1}(u_2)|\Delta = \delta)\Big|_{(u_1,u_2)=(v_1(\Delta),v_2(\Delta))},\\
\psi_{rst}(v_1(\delta), v_2(\delta); \phi) &= \frac{\partial^3}{\partial u_r \partial u_s \partial u_t} \ln g_{\Phi|\Delta;p_L,p_R}(\phi; h_1^{-1}(u_1), h_2^{-1}(u_2)|\Delta = \delta)\Big|_{(u_1,u_2)=(v_1(\Delta),v_2(\Delta))}.
\end{aligned}$$

APPENDIX B. SIMULATION STUDY

In this section, we analyze the finite-sample performance of the proposed semiparametric estimators, following the guidelines described in Section 5. We also examine the behavior of the proposed goodness-of-fit test and the bootstrap confidence intervals used in the real data application in Section 6.

In the following, first, we study the finite sample behavior of the proposed semiparametric estimators and compare it with the nonparametric approach of Alonso-Pena and Crujeiras (2023) (see Section 2.2) under both well-specified and misspecified scenarios. For our proposal, we employed the conditional generalized Batschelet distribution, and we have followed the considerations described in Section 5, by employing the local-linear semiparametric model and the proposed rule of thumb smoothing parameter. In the linear regressor case, the Epanechnikov kernel is considered, while in the circular regressor case, the von Mises kernel is used.

For the nonparametric estimators (NP), we employed the functions available in the R package `NPCirc` (Oliveira et al., 2014), `modalreg.lin.circ` and `modalreg.circ.circ`. Those functions employ the Gaussian kernel, in the linear regressor case, and the von Mises kernel, in the circular regressor case. When employing the Gaussian kernel instead of the (optimal) Epanechnikov kernel in our semiparametric approach, we obtained results similar to those reported on. The reason for employing the compactly supported Epanechnikov kernel can be found in Condition (D5). In principle, the smoothing parameter in the NP method is chosen by modal cross-validation (see Alonso-Pena & Crujeiras, 2023, for details). As discussed in Section 2.2, if, for at least one value of δ, a conditional multimodal estimation is obtained, we decrease the value of ρ. In particular, for this simulation study, we iteratively multiply by 0.9 the estimated value of ρ until the conditional distribution is unimodal for all the analyzed regressor values.

In the simulation study, the predictors are generated by a uniform distribution over the interval $[-1, 1]$, in the linear case, and over the arc $[-\pi, \pi)$, in the circular case. For the conditional

densities, we consider nine models (Models A–I below). Some of these regression models were proposed in the circular literature, and their reference is mentioned in parentheses.

A. $g_{\Phi|X;0,0}(\phi; 2\tan^{-1}(X), 3|X = x)$ (Johnson & Wehrly, 1978).
B. $g_{\Phi|X;0,0}(\phi; 2\tan^{-1}(X), \exp(2X)|X = x)$ (Fisher & Lee, 1992).
C. $g_{\Phi|X;-0.5,0.7}(\phi; m(X), 1 + |2X||X = x)$, with $m(X) = 2\pi X$ wrapped on the interval $[-\pi, \pi)$.
D. $g_{\Phi|\Theta;0,0}(\phi; atan2(\sin(\Theta), 3\cos(\Theta)), 0.5|\Theta = \theta)$ (Sarma & Jammalamadaka, 1993).
E. $g_{\Phi|\Theta;0,0}(\phi; 2\tan^{-1}(\tan((\Theta - \pi)/2)/4), 0.5|\Theta = \theta)$ (Downs & Mardia, 2002).
F. $g_{\Phi|\Theta;-0.5,0.7}(\phi; \pi\cos(\Theta), 3 + 2\cos(\Theta)|\Theta = \theta)$.
G. $h_{\Phi|X;0.5,\pi/4}(\phi; 2\tan^{-1}(X), 0.5|X = x)$.
H. $h_{\Phi|X;0.5,\pi/4}(\phi; 2\tan^{-1}(X), (\tan^{-1}(X) + 2)/5|X = x)$.
I. $h_{\Phi|\Theta;0.5,\pi/4}(\phi; atan2(\sin(\Theta), 3\cos(\Theta)), (3 + 2\cos(X))/9|\Theta = \theta)$.

In the first six considered models (Models A–F), the dependent data are generated from the conditional generalized Batschelet distribution (see (4) and (2)), whose conditional density function is denoted as $g_{\Phi|\Delta;p_L,p_R}(\phi; m(\Delta), c(\Delta)|\Delta = \delta)$. In the three last models (Models G–I), the conditional (asymmetric) Kato and Jones (2015) model,

$$h_{\Phi|\Delta,\varsigma,\varpi}(\phi; \mathbb{m}(\delta), \mathbb{c}(\delta)|\delta) = \frac{1}{2\pi}\left(1 + 2\mathbb{c}(\delta)\frac{\cos(\theta - \mathbb{m}(\delta)) - \varsigma\cos\varpi}{1 + \varsigma^2 - 2\varsigma\cos(\theta - \mathbb{m}(\delta) - \varpi)}\right), \tag{B1}$$

is employed to generate the data. Herein $\mathbb{m}(\delta) \in [-\pi, \pi)$, $\mathbb{c}(\delta) \in [0, 1)$, $\varsigma \in [0, 1)$, and $\varpi \in [-\pi, \pi)$. The parameters must also satisfy that $(\varsigma\cos\varpi - \mathbb{c}(\delta))^2 + (\varsigma\sin\varpi)^2 \leq (1 - \mathbb{c}(\delta))^2$. For this simulation study, we have set $\varsigma = 0.5$ and $\varpi = \pi/4$, corresponding to a situation where the modal and antimodal directions (values at which the density function reaches its maximum and minimum) are "close" to each other. This aligns with the scenario where there is no parameter configuration for which the density in Equation (2) provides a good approximation of the true conditional density (see Ameijeiras-Alonso et al., 2022). Thus, Models G–I represent challenging scenarios where our conditional density is misspecified.

Table B1 summarizes some characteristics of Models A–I, such as whether the predictor is linear or circular, or if the concentration varies with the predictor or not. Additionally, it reports whether the conditional density is correctly imposed in the semiparametric approach. There are two misspecification cases in this simulation study. First, the conditional von Mises (SP_{vM}, i.e., $p_L = p_R = 0$) is employed in the estimators, while the true peakedness parameters are $p_L = -0.5$ and $p_R = 0.7$. Second, the generating model is the "left-skewed" conditional distribution of Kato and Jones (2015), instead of the condition density in Equation (4). Since for Models G–I, we cannot impose the correct values of p_L and p_R, we have studied the behavior of our left-skewed semiparametric estimator with $p_L = 0$ and $p_R = 1$ ($SP_{0,1}$). These values of p_L and p_R would be close to the maximum likelihood estimators of these two parameters when employing the generalized Batschelet distribution in data generated by the non-conditional Kato and Jones (2015) distribution, with $\varsigma = 0.5$ and $\varpi = \pi/4$.

An example of $n = 200$ data points generated by each of the nine models, together with the conditional modal direction functions, and the considered regression estimators are represented in Figure B1. For each sample, both the considered nonparametric (NP) and semiparametric (SP) regression estimators are shown. In cases where two peakedness parameter configurations were employed for the models, we present both SP_{vM} ($p_L = p_R = 0$) and $SP_{-0.5,0.7}$ ($p_L = -0.5$ and $p_R = 0.7$) or $SP_{0,1}$ ($p_L = 0$ and $p_R = 1$).

TABLE B1 Summary of some characteristics of the regression models utilized in the simulation study.

Model	Conditional density	Predictor (circular or linear)	Concentration (constant or not)	p_L, p_R (or ς, ϖ) values	Misspecification (yes or no)
A	$g_{\Phi\|\Delta;p_L,p_R}$ in (4)	Linear	Constant	$p_L = p_R = 0$	No, with SP_{vM}
B	$g_{\Phi\|\Delta;p_L,p_R}$	Linear	Varies	$p_L = p_R = 0$	No, with SP_{vM}
C	$g_{\Phi\|\Delta;p_L,p_R}$	Linear	Varies	$p_L = -0.5, p_R = 0.7$	Yes, with SP_{vM}
					No, with $SP_{-0.5,0.7}$
D	$g_{\Phi\|\Delta;p_L,p_R}$	Circular	Constant	$p_L = p_R = 0$	No, with SP_{vM}
E	$g_{\Phi\|\Delta;p_L,p_R}$	Circular	Constant	$p_L = p_R = 0$	No, with SP_{vM}
F	$g_{\Phi\|\Delta;p_L,p_R}$	Circular	Varies	$p_L = -0.5, p_R = 0.7$	Yes, with SP_{vM}
					No, with $SP_{-0.5,0.7}$
G	$h_{\Phi\|\Delta,\varsigma,\varpi}$ in (B1)	Linear	Constant	$\varsigma = 0.5, \varpi = \pi/4$	Yes, with SP_{vM}
					Yes, also with $SP_{0,1}$
H	$h_{\Phi\|\Delta,\varsigma,\varpi}$	Linear	Varies	$\varsigma = 0.5, \varpi = \pi/4$	Yes, with SP_{vM}
					Yes, also with $SP_{0,1}$
I	$h_{\Phi\|\Delta,\varsigma,\varpi}$	Circular	Varies	$\varsigma = 0.5, \varpi = \pi/4$	Yes, with SP_{vM}
					Yes, also with $SP_{0,1}$

From each model, 1000 bivariate samples of size $n = 50$ and $n = 200$ were generated. The semiparametric (SP_{vM}, $SP_{-0.5,0.7}$, and $SP_{0,1}$) and the nonparametric (NP) estimators are evaluated in a regular grid of $n_{ngrid} = 100$ points over the interval/arc at which the regression functions are estimated. To evaluate the performance of an estimator $\hat{m}$ compared with the true conditional modal direction, $m(\delta)$, we employed the median of the circular (geodesic) distances MCD, defined by

$$MCD(\hat{m}) = Median\Big(d(\hat{m}(\delta_1), m(\delta_1)), \ldots, d\Big(\hat{m}\Big(\delta_{n_{ngrid}}\Big), m\Big(\delta_{n_{ngrid}}\Big)\Big)\Big),$$

where d is the geodesic distance.

Results are presented as boxplots in Figure B2. Upon inspection of the quartiles of the MCD, we can see that, in general, the nonparametric approach (NP) and the semiparametric approach with a conditional von Mises density (SP_{vM}) yield comparable results. The only exceptions, where NP exhibits a slight superiority over SP_{vM}, are observed in Models D, E, G, and H (when $n = 50$). The noteworthy aspect of the equal (and sometimes slightly superior) performance of SP_{vM}, especially for $n = 200$, is that only in Models A, B, D, and E does the conditional von Mises align with the true underlying model.

The first case of misspecification in our semiparametric approach corresponds to situations where the peakedness parameters are not correctly specified. Specifically, in Models C and F, where data is generated by a conditional asymmetric generalized Batschelet distribution ($p_L = -0.5$ and $p_R = 0.7$). In this instance, we observe that NP and SP_{vM} yield similar results. In this case, we also observe that significantly better outcomes are achieved by correctly imposing the true peakedness parameters in our semiparametric approach ($SP_{-0.5,0.7}$).

The second case of misspecification involves situations where the conditional distribution is a conditional Kato and Jones (2015) distribution, as in Models G–I. In these cases, SP_{vM} remains

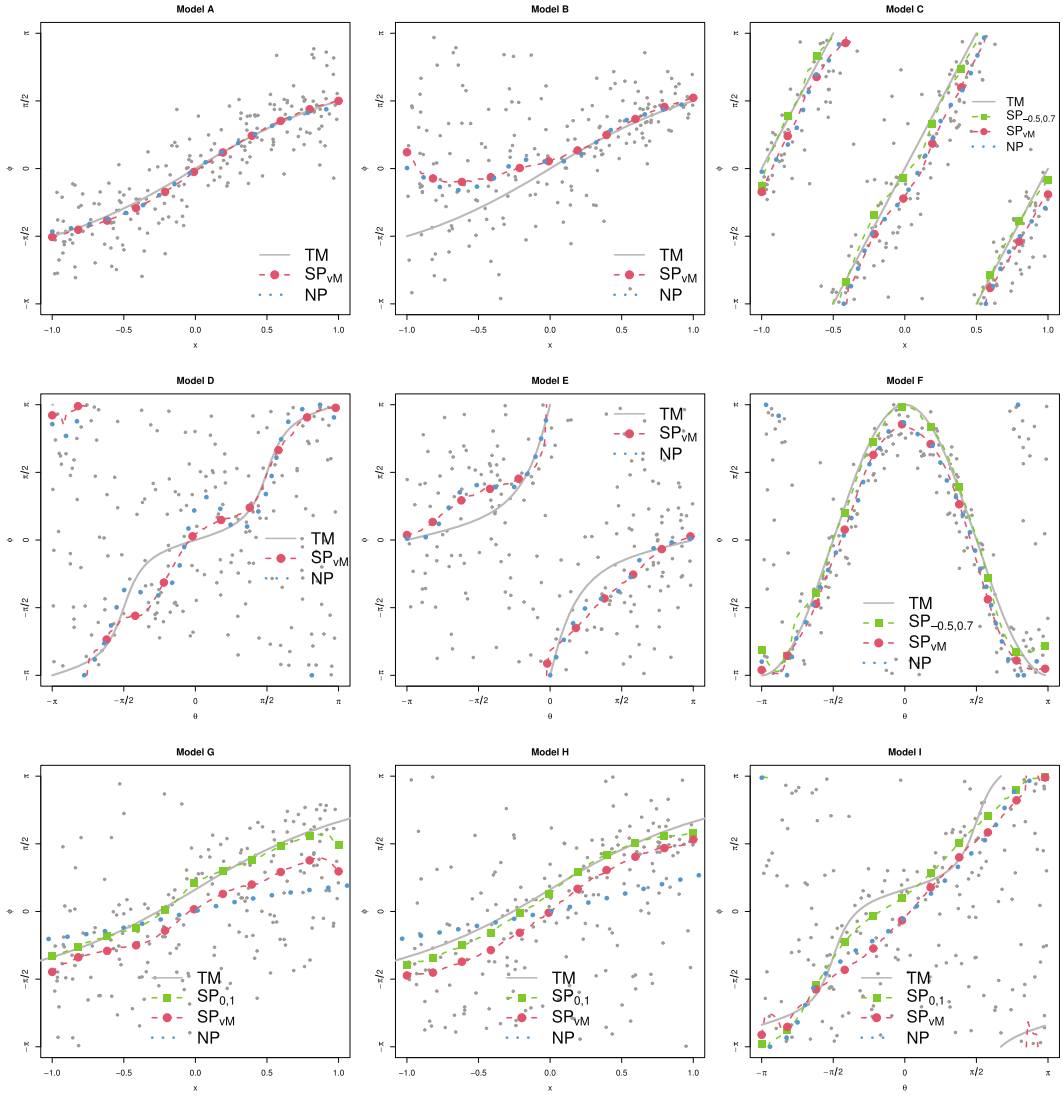

FIGURE B1 Gray dots: $n = 200$ sample points generated by Models A–I. Lines: Conditional modal direction functions, with the continuous line depicting the true modal (TM) function $m(\delta)$, dashed lines representing semiparametric (SP) estimates, and dotted lines indicating nonparametric (NP) estimates. In the semiparametric estimation, peakedness parameters are set as $p_L = p_R = 0$ (vM, circle-dashed lines), $p_L = -0.5$ and $p_R = 0.7$ (square-dashed lines in Models C and F), $p_L = 0$ and $p_R = 1$ (square-dashed lines in Models G–I).

competitive against NP. Moreover, when accounting for left-skewness, by imposing $p_L = 0$ and $p_R = 1$ in the semiparametric approach, significantly improved results are obtained with our semiparametric approach compared to NP, even in this misspecification scenario.

These findings support the use of the semiparametric approach over NP, especially when considering the additional advantage of having an estimator not only for the conditional modal direction but also for the conditional concentration, and in extension for the conditional mean direction.

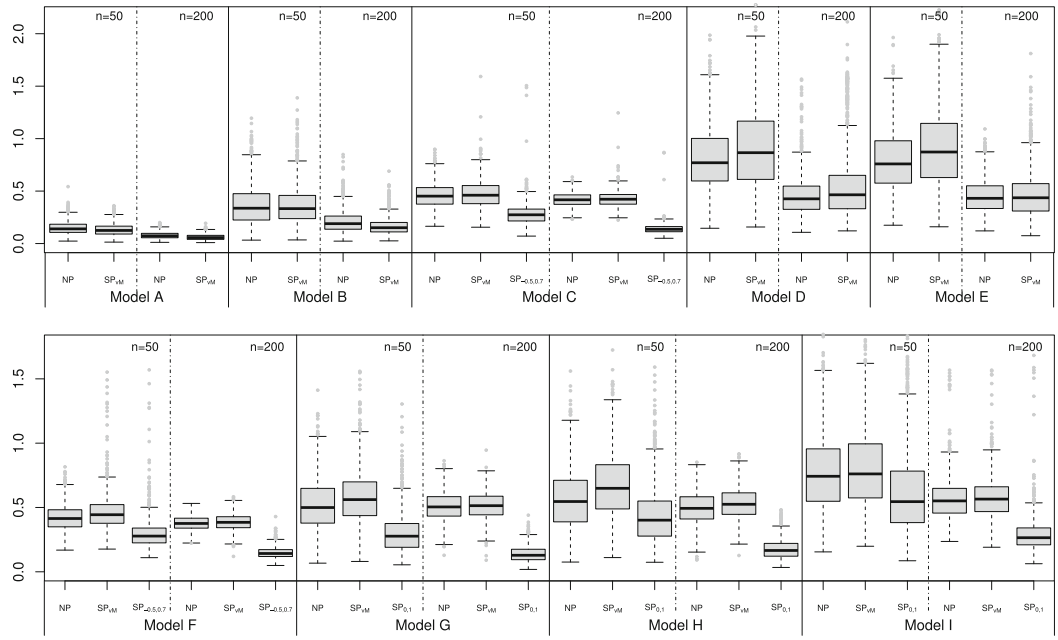

FIGURE B2 Boxplots of the MCD obtained from 1000 samples of size $n = 50$ and $n = 200$ generated by Models A–I. Three estimators are employed: Nonparametric approach (NP), semiparametric approach with conditional von Mises (i.e., $p_L = p_R = 0$; $\mathrm{SP}_{\mathrm{vM}}$), and semiparametric approach with conditional generalized Batschelet distribution ($p_L = -0.5$ and $p_R = 0.7$, $\mathrm{SP}_{-0.5,0.7}$, in Models C and F; $p_L = 0$ and $p_R = 1$, $\mathrm{SP}_{0,1}$, in Models G–I).

TABLE B2 Percentage of rejections at significance level $\alpha = 0.05$ for Models A, C, and F (which depend on a regressor) and NR1 and NR2 (which do not), for sample sizes $n = 50$ and $n = 200$. Results are shown under the null hypothesis, i.e., when using the true values of the peakedness parameters p_L and p_R from which the data were generated ($p_L = p_R = 0$ for Models A and NR1; $p_L = -0.5$ and $p_R = 0.7$ for Models C, F, and NR2), and under the alternative hypothesis, where misspecified values are imposed ($p_L = -1$ and $p_R = 1$ for Models A and NR1; $p_L = 1$ and $p_R = -1$ for Models C, F, and NR2). Results are obtained using the usual critical values of the test for circular uniformity (Asym) and employing the proposed bootstrap procedure (Boot).

		Semiparametric regression						**Parametric non-regression (NR) setting**			
		Under H_0			**Under H_1**			**Under H_0**		**Under H_1**	
	Model	**A**	**C**	**F**	**A**	**C**	**F**	**NR1**	**NR2**	**NR1**	**NR2**
$n = 50$	Asym	0.006	0.030	0.008	0.052	0.040	0.029	0.004	0.006	0.094	0.128
	Boot	0.004	0.038	0.006	0.046	0.042	0.022	0.002	0.008	0.082	0.096
$n = 200$	Asym	0.002	0.004	0.022	0.722	0.034	0.224	0.004	0.002	0.458	0.478
	Boot	0.000	0.004	0.028	0.690	0.044	0.226	0.000	0.004	0.442	0.468

In the following, we study the performance of the goodness-of-fit test based on the probability integral transformation and test statistic U^2 of Watson (1961), as suggested in Section 6. To that end, we evaluate the finite-sample performance of the test using two calibration approaches: The usual (asymptotic) critical values and the proposed bootstrap procedure.

In Table B2, we report the percentage of times the null hypothesis was rejected across 500 samples of size $n = 50$ and $n = 200$ generated from different models, at a significance level of

$\alpha = 0.05$. As seen from the table, comparable results are obtained when using the usual critical values and the bootstrap approach, regardless of the scenario considered. Therefore, we comment on both sets of results jointly in what follows.

In the first three columns (counting only colums with numeric results) of Table B2, we assess the calibration of the test by imposing the true peakedness parameters for Model A ($p_L = p_R = 0$), and for Models C and F ($p_L = -0.5$ and $p_R = 0.7$). The results show that the proposed test tends to be conservative in most cases, even when $n = 200$.

To determine whether the observed conservativeness is due to the semiparametric nature of the method or is also present in the standard parametric case, we examined the performance of a parametric version of the test that does not depend on the predictor. Specifically, we computed the classical goodness-of-fit test based on the probability integral transformation and the U^2 statistic of Watson (1961), using both the standard critical values for tests of circular uniformity and a parametric bootstrap version (see, e.g., Pewsey et al., 2013, Section 6.2.3). In this version, the conditional distribution function $G_{\Phi|\Delta;p_L,p_R}(\phi; \hat{m}(\Delta), \hat{c}(\Delta)|\Delta = \delta)$ used in the semiparametric test is replaced by the unconditional distribution $G_{\Phi;p_L,p_R}(\phi; \hat{m}, \hat{c})$ in the construction of the U^2 test statistic and its bootstrap counterpart $U^{2,j*}$. To compute rejection percentages, we generate samples from the generalized Batschelet model corresponding to $G_{\Phi;0,0}(\cdot; 0, 2)$ (NR1) and $G_{\Phi|\Delta;-0.5,0.7}(\cdot; 0, 2)$ (NR2). Columns 7 and 8 of Table B2 report the percentage of rejections under the null, i.e., when the correct peakedness parameters p_L and p_R are imposed. Once again, we observe the same conservative behavior as in the semiparametric case.

To analyze the power of both the semiparametric and parametric versions of the test, we compute p-values under alternative scenarios by imposing misspecified values for the peakedness parameters. Specifically, we set $p_L = -1$ and $p_R = 1$ for Models A and NR1, and $p_L = 1$ and $p_R = -1$ for Models C, F, and NR2. In Columns 4–6 (semiparametric case) and 9–10 (parametric case) of Table B2, we observe a lack of power when $n = 50$. For $n = 200$, the test is able to detect some alternatives: The rejection rates are around 0.7 (Model A) and 0.2 (Model F) in the semiparametric case, and around 0.4 (Model NR1) and 0.5 (Model NR2) in the parametric case. The test (using either of the two calibration methods) fails to detect the alternative under Model C, even for a sample size of $n = 200$.

These results indicate that, when the null hypothesis is rejected, we can be confident that the data do not follow the semiparametric generalized Batschelet distribution. However, the goodness-of-fit test based on the semiparametric approach is similarly conservative and exhibits comparable power limitations to its parametric counterpart. Therefore, further research is needed to develop a more powerful testing procedure.

To explore the performance of the bootstrap confidence intervals described in Section 6, we generated 500 samples of sizes $n = 50$ and $n = 200$ from Models A–F. We then computed the 95% confidence intervals for the conditional modal direction $m(\delta)$ and concentration $c(\delta)$ over a regular grid of 100 values of δ in the interval $[0, 1]$ (for models with linear predictors) and $[-\pi, \pi)$ (for models with circular predictors).

Figure B3 presents boxplots showing the percentage of times that the true values of $m(\delta_i)$ and $c(\delta_i)$, for $i \in \{1,\ldots,100\}$, fall within the corresponding confidence intervals. As the sample size increases, the coverage approaches the nominal 95%, especially for the conditional modal direction, where in most cases the median coverage is close to 0.95. In the case of the conditional concentration, the bottom panel of Figure B3 shows that larger sample sizes may be required to achieve nominal coverage; however, the percentage still tends to approach 0.95 as the sample size increases.

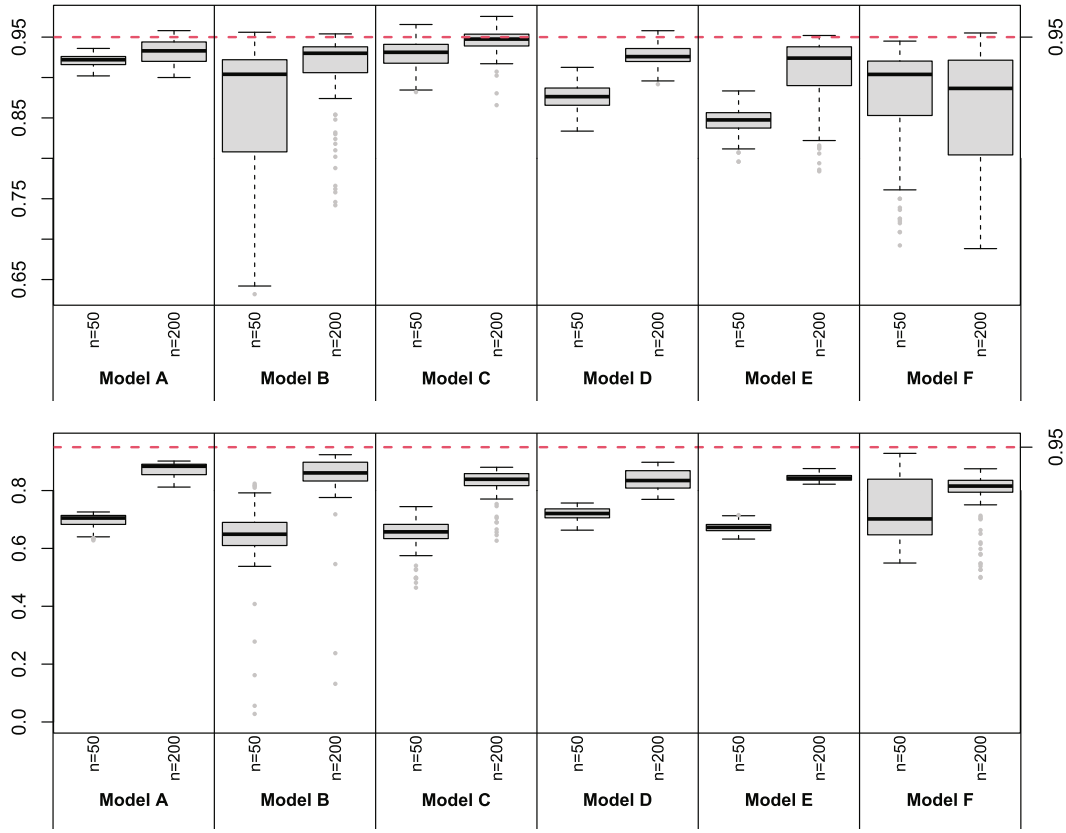

FIGURE B3 Boxplots showing the percentage of times that the true conditional modal direction $m(\delta)$ (top) and concentration $c(\delta)$ (bottom) fall within the 95% bootstrap confidence intervals. Results are based on 500 samples of sizes $n = 50$ and $n = 200$ generated from Models A–F.

APPENDIX C. FURTHER DETAILS ON THE DATA APPLICATION

In this section, we first provide additional analysis related to the data application presented in Section 6. Secondly, we explore in more detail the multivariate extension of the proposed model and apply it to the flight orientation data.

The first objective of this section is to investigate whether a parametric model allowing for conditional asymmetry can adequately fit the flight orientation data. We therefore estimated a parametric model with a conditional density given by (4). The peakedness parameters, p_L and p_R, are the same as those used in Section 6 for each scenario. To estimate the conditional modal direction and concentration, we have employed $m(\delta) = \check{h}_1^{-1}(\check{\eta}_{1,p_1}(\delta); l, 1) = l + 2\tan^{-1}(\check{\eta}_{1,p_1}(\delta))$ and $c(\delta) = h_2^{-1}(\check{\eta}_{2,p_2}(\delta)) = \exp(\check{\eta}_{2,p_2}(\delta))$. The parametric models $\check{\eta}_{r,p_r}$ are $\check{\eta}_{r,p_r}(x) = \check{\eta}_{r;0} + \check{\eta}_{r;1}x + \ldots + \check{\eta}_{r;p_r}x^{p_r}$ in the linear case, and $\check{\eta}^{\circ}_{r,p_r}(\theta) = \check{\eta}_{r;0} + \check{\eta}_{r;1}\sin(\theta) + \ldots + \check{\eta}_{r;p_r}\sin^{p_r}(\theta)$, in the circular case; for $r \in \{1, 2\}$. The parameters l, $\check{\boldsymbol{\eta}}_1$, and $\check{\boldsymbol{\eta}}_2$ are estimated via maximum likelihood using an optimization routine based on the Nelder-Mead algorithm.

We estimated the parametric model for different degrees of the polynomials, with p_1 and p_2 ranging from 1 to 4. To determine the "best" parametric fit, we computed the AIC

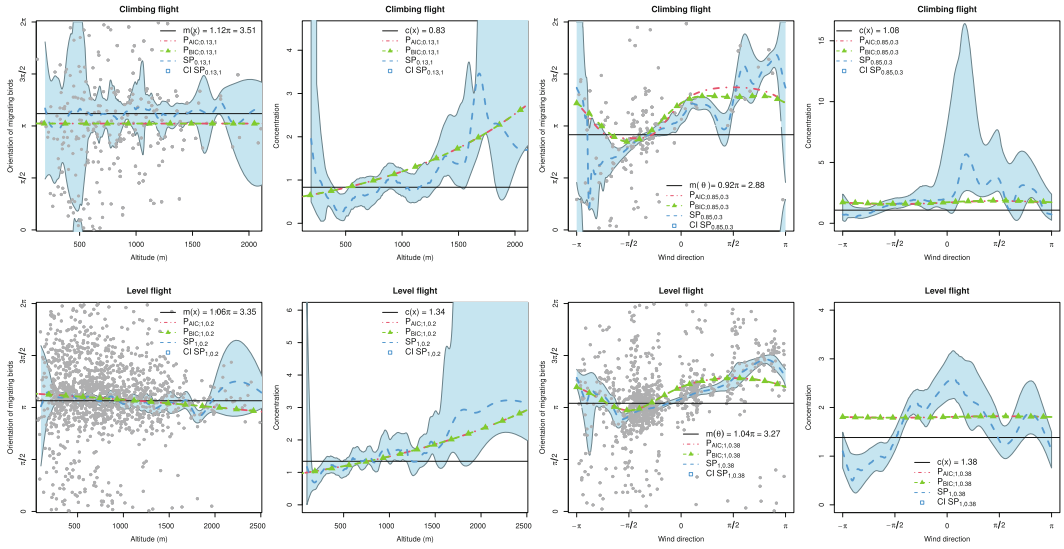

FIGURE C1 Gray dots: Flight orientation, altitude (first and second column), and wind direction (third and fourth column) for the migration birds in climbing (top) and level (bottom) flight. Modal direction (first and third column) and concentration (second and fourth column), assuming that they do (discontinuous lines) or not (continuous lines), depend on the regressor. Best parametric fits according to AIC (dot-dashed) and BIC (triangle-dashed), along with semiparametric (dashed) point estimates. Colored (light blue) area: 95% confidence bands based on the semiparametric approach.

and BIC values for all 16 model combinations. The best-fitting models according to each criterion are presented in Figure C1. In Figure C1, we observe that, for the altitude predictor, the parametric fits follow the shape of the semiparametric estimates, and most of the estimated curve falls within the confidence bands. This indicates that, for this regressor, even low-degree polynomials (AIC and BIC select $p_1 = 1$ and $p_2 = 1$ for level and climbing flights, respectively) are sufficient to fit the data well using the asymmetric parametric model. However, this is not the case for the wind predictor. For that regressor, the parametric fits fail to capture the oscillatory behavior of the conditional modal and concentration parameters.

Regarding the multivariate extension of the regression method, one can find only limited attempts in the circular regression literature to tackle this problem. In the parametric setting, among others, this problem was studied by Fisher and Lee (1992), for a multivariate linear regressor, by Kim and SenGupta (2017), for a multivariate circular regressor, and by Rivest et al. (2016), for a combination of linear and circular regressors. In the nonparametric setting, Meilán-Vila et al. (2021) employed a local polynomial estimator for the case of having a multivariate linear regressor.

The theory presented in this paper could also be extended to the multivariate regression problem. In that case, we propose to employ a similar approach to that considered in Fan et al. (1995) if all the covariates are of the same nature, all linear or circular. For that purpose, the product kernel $\mathcal{W}_1(\delta_1) \times \ldots \times \mathcal{W}_d(\delta_d)$ could be used. Then, the asymptotic results for the d-variate local linear kernel estimators of $m(\boldsymbol{\delta})$ and $c(\boldsymbol{\delta})$ can be derived from the proofs in Appendix D. The estimators would be $\hat{m}(\boldsymbol{\delta}) = h_1^{-1}(\hat{\beta}_{1;0}(\boldsymbol{\delta}))$ and $\hat{c}(\boldsymbol{\delta}) = h_2^{-1}(\hat{\beta}_{2;0}(\boldsymbol{\delta}))$, where

$$\left(\hat{\boldsymbol{\beta}}_1(\boldsymbol{x}), \hat{\boldsymbol{\beta}}_2(\boldsymbol{x})\right) \overset{\text{linear}}{=} \underset{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2}{\operatorname{argmax}} \sum_{i=1}^{n} \ell\left(h_1^{-1}\left(\beta_{1;0} + \sum_{j=1}^{d} \beta_{1;j}(X_{ij} - x_j) \right), \right.$$
$$\left. \times\, h_2^{-1}\left(\beta_{2;0} + \sum_{j=1}^{d} \beta_{2;j}(X_{ij} - x_j) \right); \Phi_i \right) \prod_{j=1}^{d} L_{j;h_j}(X_{ij} - x_j),$$

$$\left(\hat{\boldsymbol{\beta}}_1(\boldsymbol{\theta}), \hat{\boldsymbol{\beta}}_2(\boldsymbol{\theta})\right) \overset{\text{circular}}{=} \underset{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2}{\operatorname{argmax}} \sum_{i=1}^{n} \ell\left(h_1^{-1}\left(\beta_{1;0} + \sum_{j=1}^{d} \beta_{1;j} \sin(\Theta_{ij} - \theta_j) \right), \right.$$
$$\left. \times\, h_2^{-1}\left(\beta_{2;0} + \sum_{j=1}^{d} \beta_{2;j} \sin(\Theta_{ij} - \theta_j) \right); \Phi_i \right) \prod_{j=1}^{d} K_{j;\nu_j}(\Theta_{ij} - \Theta_j).$$

When the covariates are of different nature (circular and linear), similar ideas could be employed by considering the d-variate cylindrical product kernel. The main issue, in that case, is that getting the asymptotic results would be more involved as one needs to combine the linear and circular theory presented in Appendix D.

To motivate the applicability of the multivariate semiparametric estimator, in Figure C2, we show simultaneously how both altitude and wind direction influence the orientation of migratory birds in the data application of Section 6. As mentioned before, the main complication in obtaining the estimation is that deriving asymptotic results is not straightforward, and the rule-of-thumb smoothing parameter for a single predictor may not be related to the multivariate case. For this reason, we manually selected the smoothing parameters. Specifically, we set $h_1 = 2000$ for altitude (the sample standard deviation and interquartile range for climbing flights are $s_c = 387.76$ and $\text{IQR}_c = 511.95$, and for level flights $s_l = 448.32$ and $\text{IQR}_l = 577$), and $\nu_2 = 4$ for wind direction (the sample maximum likelihood estimate of the concentration parameter for a von Mises distribution is $\hat{\kappa}_c = 1.3$ for climbing flights and $\hat{\kappa}_l = 1.05$ for level flights).

Relying on these smoothing parameters, by observing Figure C2, similar to our previous findings, we can see that for both climbing and level flights, the conditional modal direction closely aligns with its non-conditional estimate for wind directions close to $-\pi/4$. The largest disparities between conditional and non-conditional modal directions occur for wind directions around $3\pi/4$. For a fixed wind direction, in general, the value of the conditional modal direction almost does not change with the altitude. Notably, certain combinations of altitude and wind direction yield results that differ from those discussed in Section 6. For instance, in climbing flights, Figure 5 showed that the largest disparities from the non-conditional modal direction occurred for wind directions close to π. When considering both regressors simultaneously, this remains true at low altitudes but not, e.g., at an altitude of 1400 m.

Regarding the conditional concentration, consistent with previous observations, we note an increase with altitude. The combination of high altitudes and a wind direction of $3\pi/4$ in climbing flights, and a direction of 0 in level flights, produces the highest concentration values. Conversely, low altitudes with a wind direction around $-3\pi/4$ result in lower concentration values.

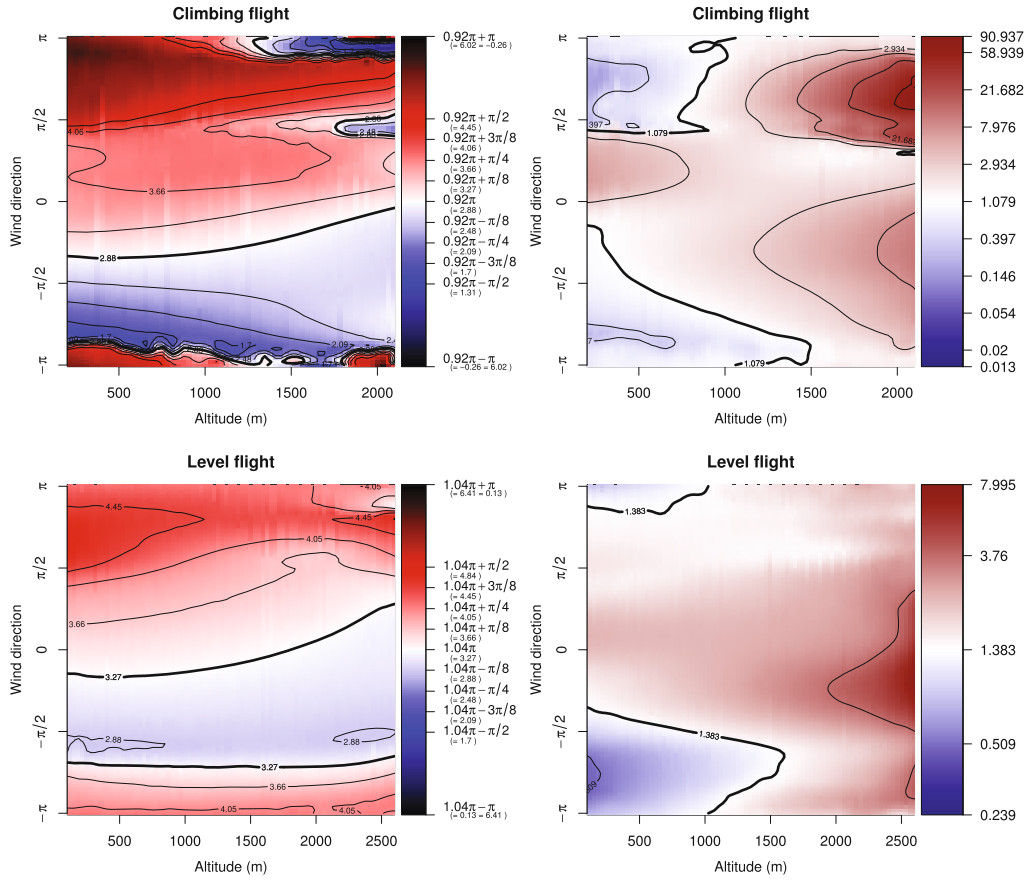

FIGURE C2 Contour plots depicting conditional parameters in the flight orientation of migrating birds concerning altitude (x-axis) and wind direction (y-axis). The employed smoothing parameters are $h_1 = 2000$ for altitude and $v_2 = 4$ for wind direction. Lighter colors on the plot indicate that the modal direction and concentration are closer to non-conditional values estimated from the entire sample. Top: Climbing flights (with peakedness parameters $p_L = 0.85$ and $p_R = 0.3$). Bottom: Level flights (with peakedness parameters $p_L = 1$ and $p_R = 0.38$). Left: Conditional modal direction estimation. Right: Conditional concentration estimation (colors are represented in log-scale).

APPENDIX D. PROOFS OF THE THEORETICAL RESULTS

D1 Proof of Theorem 1

The proof of Theorem 1 for the linear predictor case is done following the general asymptotic normality results for local quasi-likelihood estimators (see Fan et al., 1995; Fan & Gijbels, 1996, Ch. 5), taking into account that, in this case, two parameters, the modal direction and the concentration, depend on the covariate. For the circular predictor case, some particularities need to be considered.

Given a value δ_0, we first introduce the estimator $\hat{\boldsymbol{\beta}}_r^*$,

$$\hat{\boldsymbol{\beta}}_r^* \overset{\text{linear}}{=} a_n^{-1}\left(\hat{\beta}_{r;0} - \beta_{r;0}, h\left(\hat{\beta}_{r;1} - \beta_{r;1}\right), \ldots, h^{\mathfrak{p}_r}\left(\hat{\beta}_{r;\mathfrak{p}_r} - \beta_{r;\mathfrak{p}_r}\right)\right)^\top,$$
$$a_n \overset{\text{linear}}{=} (nh)^{-1/2},$$

$$\hat{\boldsymbol{\beta}}_r^* \overset{\text{circular}}{=} a_n^{-1}\left(\hat{\beta}_{r;0} - \beta_{r;0}, \hat{\beta}_{r;1} - \beta_{r;1}, \ldots, \hat{\beta}_{r;\mathfrak{p}_r} - \beta_{r;\mathfrak{p}_r}\right)^\top,$$
$$a_n \overset{\text{circular}}{=} n^{-1/2}.$$

Then, if $(\hat{\boldsymbol{\beta}}_1(\delta_0), \hat{\boldsymbol{\beta}}_2(\delta_0))$ maximizes $\mathcal{L}_n(\delta_0; \boldsymbol{\beta_1}, \boldsymbol{\beta_2})$, $(\hat{\boldsymbol{\beta}}_1^*(\delta_0), \hat{\boldsymbol{\beta}}_2^*(\delta_0))$ maximizes

$$\mathcal{L}_n(\delta_0; \boldsymbol{\beta}_1^*, \boldsymbol{\beta}_2^*) = \sum_{i=1}^n \ell\left(h_1^{-1}\left(\overline{\eta}_1(\delta_0, \Delta_i) + a_n \boldsymbol{\beta}_1^{*\top} \boldsymbol{Z}_{1,i}\right), h_2^{-1}\left(\overline{\eta}_2(\delta_0, \Delta_i) + a_n \boldsymbol{\beta}_2^{*\top} \boldsymbol{Z}_{2,i}\right); \Phi_i\right) \mathcal{W}(\Delta_i - \delta_0),$$

as a function of $(\boldsymbol{\beta}_1^*, \boldsymbol{\beta}_2^*)$, where

$$\begin{aligned}
\mathcal{W}(x) &\overset{\text{linear}}{=} L(x/h) \qquad\qquad \mathcal{W}(\theta) \overset{\text{circular}}{=} K_\nu(\theta),\\
\overline{\eta}_r(x_0, X_i) &\overset{\text{linear}}{=} \beta_{r;0}(x_0) + \beta_{r;1}(x_0)(X_i - x_0) + \ldots + \beta_{r;\mathfrak{p}_r}(x_0)(X_i - x_0)^{\mathfrak{p}_r},\\
\overline{\eta}_r(\theta_0, \Theta_i) &\overset{\text{circular}}{=} \beta_{r;0}(\theta_0) + \beta_{r;1}(\theta_0)\sin(\Theta_i - \theta_0) + \ldots + \beta_{r;\mathfrak{p}_r}(\theta_0)\sin^{\mathfrak{p}_r}(\Theta_i - \theta_0),\\
\boldsymbol{Z}_{r,i} &\overset{\text{linear}}{=} \{1, (X_i - x_0)/h, \ldots, (X_i - x_0)^{\mathfrak{p}_r}/h^{\mathfrak{p}_r}\}^\top,\\
\boldsymbol{Z}_{r,i} &\overset{\text{circular}}{=} \{1, \sin(\Theta_i - \theta_0), \ldots, \sin^{\mathfrak{p}_r}(\Theta_i - \theta_0)\}^\top.
\end{aligned}$$

Equivalently, $(\hat{\boldsymbol{\beta}}_1^*(\delta_0), \hat{\boldsymbol{\beta}}_2^*(\delta_0))$ maximizes

$$\begin{aligned}
\mathcal{L}_n^*(\delta_0; \boldsymbol{\beta}_1^*, \boldsymbol{\beta}_2^*) = \sum_{i=1}^n \Big\{ &\ell\left(h_1^{-1}\left(\overline{\eta}_1(\delta_0, \Delta_i) + a_n \boldsymbol{\beta}_1^{*\top} \boldsymbol{Z}_{1,i}\right), h_2^{-1}\left(\overline{\eta}_2(\delta_0, \Delta_i) + a_n \boldsymbol{\beta}_2^{*\top} \boldsymbol{Z}_{2,i}\right); \Phi_i\right)\\
&- \ell\left(h_1^{-1}\left(\overline{\eta}_1(\delta_0, \Delta_i)\right), h_2^{-1}\left(\overline{\eta}_2(\delta_0, \Delta_i)\right); \Phi_i\right) \Big\} \mathcal{W}(\Delta_i - \delta_0),
\end{aligned} \tag{D1}$$

as a function of $(\boldsymbol{\beta}_1^*, \boldsymbol{\beta}_2^*)$. Note that under the assumptions of the theorem, the function $\mathcal{L}_n^*(\delta_0; \boldsymbol{\beta}_1^*, \boldsymbol{\beta}_2^*)$ is almost always differentiable except in a finite number of points.

Using a Taylor expansion in Equation (D1), around $(\overline{\eta}_1(\delta_0, \Delta_i), \overline{\eta}_2(\delta_0, \Delta_i))$, we obtain

$$\begin{aligned}
\mathcal{L}_n^*(\delta_0; \boldsymbol{\beta}_1^*, \boldsymbol{\beta}_2^*) = &\sum_{r=1}^2 \boldsymbol{\beta}_r^{*\top} \boldsymbol{\mathcal{L}}_{r;n} + \frac{1}{2}\sum_{r=1}^2\sum_{s=1}^2 \boldsymbol{\beta}_r^{*\top} \boldsymbol{\mathcal{L}}_{rs;n} \boldsymbol{\beta}_s^*\\
&+ \frac{a_n^3}{6}\sum_{i=1}^n\sum_{r=1}^2\sum_{s=1}^2\sum_{t=1}^2 \psi_{rst}(\eta_{1,i}^*, \eta_{2,i}^*; \Phi_i)\boldsymbol{\beta}_r^{*\top} \boldsymbol{Z}_{r,i} \boldsymbol{\beta}_s^{*\top} \boldsymbol{Z}_{s,i} \boldsymbol{\beta}_t^{*\top} \boldsymbol{Z}_{t,i} \mathcal{W}(\Delta_i - \delta_0),
\end{aligned} \tag{D2}$$

where, for $r, s \in \{1, 2\}$,

$$\begin{aligned}
\eta_{r,i}^* &\in [\overline{\eta}_r(\delta_0, \Delta_i), \overline{\eta}_r(\delta_0, \Delta_i) + a_n \boldsymbol{\beta}_r^{*\top} \boldsymbol{Z}_{r,i}],\\
\boldsymbol{\mathcal{L}}_{r;n} &= a_n \sum_{i=1}^n \psi_r(\overline{\eta}_1(\delta_0, \Delta_i), \overline{\eta}_2(\delta_0, \Delta_i); \Phi_i) \boldsymbol{Z}_{r,i} \mathcal{W}(\Delta_i - \delta_0),\\
\boldsymbol{\mathcal{L}}_{rs;n} &= a_n^2 \sum_{i=1}^n \psi_{rs}(\overline{\eta}_1(\delta_0, \Delta_i), \overline{\eta}_2(\delta_0, \Delta_i); \Phi_i) \boldsymbol{Z}_{r,i} \boldsymbol{Z}_{s,i}^\top \mathcal{W}(\Delta_i - \delta_0).
\end{aligned}$$

In the following, we study the asymptotic behavior of the terms in Equation (D2). Regarding the last term in Equation (D2), note that $\boldsymbol{Z}_{r,i}$ is a bounded random vector since $\mathcal{W}$ has always a

compact support, and also that ψ_{rst} is a bounded function (see Assumption (D3)). Then, we obtain that the expected value of the last term in Equation (D2) is

$$
\begin{aligned}
& n\frac{a_n^3}{6}\mathbb{E}\left(\left|\sum_{r=1}^{2}\sum_{s=1}^{2}\sum_{t=1}^{2}\psi_{rst}(\eta_{1,1}^*,\eta_{2,1}^*;\Phi_1)\boldsymbol{\beta}_r^{*\top}\boldsymbol{Z}_{r,1}\boldsymbol{\beta}_s^{*\top}\boldsymbol{Z}_{s,1}\boldsymbol{\beta}_t^{*\top}\boldsymbol{Z}_{t,1}\mathcal{W}(\Delta_1-\delta_0)\right|\right) \\
& \quad = na_n^3 O(\mathbb{E}(\mathcal{W}(\Delta_1-\delta_0))) \\
& na_n^3 O(\mathbb{E}(\mathcal{W}(X_1-x_0))) \overset{\text{linear}}{=} na_n^3 O(h) = O(a_n). \\
& na_n^3 O(\mathbb{E}(\mathcal{W}(\Theta_1-\theta_0))) \overset{\text{circular}}{=} na_n^3 O(1) = O(a_n).
\end{aligned}
$$

Therefore, the last term in Equation (D2) is of order $O_P(a_n)$.

Regarding the second term in Equation (D2), we have that the entry (k,l) of the matrix $\boldsymbol{\mathcal{L}}_{rs;n}$ is equal to

$$
(\mathcal{L}_{rs;n})_{kl} = (\mathbb{E}[\boldsymbol{\mathcal{L}}_{rs;n}])_{kl} + O_P\big(((\mathrm{Var}[\boldsymbol{\mathcal{L}}_{rs;n}])_{kl})^{1/2}\big).
$$

Using the following notation

$$
\begin{aligned}
& \boldsymbol{\psi}_{rs,i}^* = \psi_{rs}(\overline{\eta}_1(\delta_0,\Delta_i),\overline{\eta}_2(\delta_0,\Delta_i);\Phi_i)\boldsymbol{Z}_{r,i}\boldsymbol{Z}_{s,i}^{\top}\mathcal{W}(\Delta_i-\delta_0), \\
& \lambda_r(\varphi_1(\delta),\varphi_2(\delta)) = \mathbb{E}_{\Phi|\Delta}[\psi_r(\varphi_1(\Delta),\varphi_2(\Delta);\Phi)|\Delta=\delta], \\
& \lambda_{rs}(\varphi_1(\delta),\varphi_2(\delta)) = \frac{\partial}{\partial u_s}\lambda_r(u_1,u_2)\Big|_{(u_1,u_2)=(\varphi_1(\delta),\varphi_2(\delta))},
\end{aligned}
$$

we have that $\mathbb{E}[\boldsymbol{\mathcal{L}}_{rs;n}] = na_n^2\mathbb{E}(\boldsymbol{\psi}_{rs,1}^*)$, for $r,s\in\{1,2\}$. The (k,l) component of $\mathbb{E}(\boldsymbol{\psi}_{rs,1}^*)$ can be obtained using a change of variables and the Taylor expansion as shown below. Note that in the following, we have employed that under Conditions (C1)–(C3), $\mathfrak{T}(\delta_0) = -\lambda_{rs}(\overline{\eta}_1(\delta_0,\delta_0),\overline{\eta}_2(\delta_0,\delta_0))$. For the linear case, we employ the bounded support of the kernel. For the circular case, we have also employed the periodicity of K_ν, the powers of sin being periodic, $\mu_{k+2} = \mu_k(1+O(1))$ and $\mu_{j+l} = \omega_{j,l}(1+O(1))$.

$$
\begin{aligned}
(\mathbb{E}[(\boldsymbol{\psi}_{rs,1}^*)])_{kl} &\overset{\text{linear}}{=} \int\int_{-\pi}^{\pi}\psi_{rs}(\overline{\eta}_1(x_0,x),\overline{\eta}_2(x_0,x),\phi)\Big(\frac{x-x_0}{h}\Big)^{k+l-2}L\Big(\frac{x-x_0}{h}\Big)g_{X,\Phi}(x,\phi)d\phi dx \\
&= h\int L(z)z^{k+l-2}g_X(x_0+hz)\lambda_{rs}\big(\overline{\eta}_1(x_0,x_0+hz),\overline{\eta}_2(x_0,x_0+hz)\big)dz \\
&= -h\mu_{k+l-2}g_X(x_0)\mathfrak{T}_{rs}(x_0) - h^2\mu_{k+l-1}\frac{\partial}{\partial x_0}\big[g_X(x_0)\mathfrak{T}_{rs}(x_0)\big] + o(h^2), \\
(\mathbb{E}[\boldsymbol{\mathcal{L}}_{rs;n}])_{kl} &\overset{\text{linear}}{=} -\mu_{k+l-2}g_X(x_0)\mathfrak{T}_{rs}(x_0) - h\mu_{k+l-1}\frac{\partial}{\partial x_0}\big[g_X(x_0)\mathfrak{T}_{rs}(x_0)\big] + o(h), \\
(\mathbb{E}[(\boldsymbol{\psi}_{rs,1}^*)])_{kl} &\overset{\text{circular}}{=} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\psi_{rs}(\overline{\eta}_1(\theta_0,\theta),\overline{\eta}_2(\theta_0,\theta),\phi)\sin^{k+l-2}(\theta-\theta_0)K_\nu(\theta-\theta_0)g_{\Theta,\Phi}(\theta,\phi)d\phi d\theta \\
&= \int_{-\pi+\theta_0}^{\pi+\theta_0}K_\nu(\vartheta)\sin(\vartheta)^{k+l-2}g_\Theta(\theta_0+\vartheta)\lambda_{rs}\big(\overline{\eta}_1(\theta_0,\theta_0+\vartheta),\overline{\eta}_2(\theta_0,\theta_0+\vartheta)\big)d\vartheta, \\
(\mathbb{E}[\boldsymbol{\mathcal{L}}_{rs;n}])_{kl} &\overset{\text{circular}}{=} -\mu_{k+l-2}g_\Theta(\theta_0)\mathfrak{T}_{rs}(\theta_0) - \omega_{k+l-2,1}\frac{\partial}{\partial\theta_0}\big[g_\Theta(\theta_0)\mathfrak{T}_{rs}(\theta_0)\big] + o(\omega_{k+l-2,2}) + o(\omega_{k+l-2,3}).
\end{aligned}
$$

Using similar arguments, it is straightforward to show that, in both the linear and the circular case, $(\mathrm{Var}[\boldsymbol{\mathcal{L}}_{rs;n}])_{kl} = O(a_n^2)$. Given $r, s \in \{1, 2\}$, considering that, under Conditions (C3) and (C4), $\mathfrak{T}_{rs}(\delta_0) = 0$ if $r \neq s$ we obtain,

$$\begin{aligned}
\boldsymbol{\mathcal{L}}_{rr;n} &\overset{\text{linear}}{=} -g_X(x_0)\mathfrak{T}_{rr}(x_0)(\mu_{k+l-2})_{1\leq k,l\leq \mathfrak{p}_r+1} - h\frac{\partial}{\partial x_0}\left[g_X(x_0)\mathfrak{T}_{rr}(x_0)\right] \\
&\quad \times (\mu_{k+l-1})_{1\leq k,l\leq \mathfrak{p}_r+1} + o_P(h) + O_P(a_n) \\
&= -g_X(x_0)\mathfrak{T}_{rr}(x_0)(\mu_{k+l-2})_{1\leq k,l\leq \mathfrak{p}_r+1} + o_P(1), \\
\boldsymbol{\mathcal{L}}_{rr;n} &\overset{\text{circular}}{=} -g_\Theta(\theta_0)\mathfrak{T}_{rr}(\theta_0)(\mu_{k+l-2})_{1\leq k,l\leq \mathfrak{p}_r+1} - \frac{\partial}{\partial \theta_0}\left[g_\Theta(\theta_0)\mathfrak{T}_{rr}(\theta_0)\right](\omega_{k+l-2,1})_{1\leq k,l\leq \mathfrak{p}_r+1} \\
&\quad + o_P\left(\max_{1\leq k,l\leq \mathfrak{p}_r+1}\omega_{k+l-2,1}\right) + O_P(a_n) \\
&= -g_\Theta(\theta_0)\mathfrak{T}_{rr}(\theta_0)(\mu_{k+l-2})_{1\leq k,l\leq \mathfrak{p}_r+1} + o_P(1), \\
\boldsymbol{\mathcal{L}}_{rs;n} &\overset{\text{linear}}{=} o_P(h) + O_P(a_n), \\
\boldsymbol{\mathcal{L}}_{rs;n} &\overset{\text{circular}}{=} o_P\left(\max_{1\leq k,l\leq \mathfrak{p}_r+1}\omega_{k+l-2,1}\right) + O_P(a_n), \\
\boldsymbol{\mathcal{L}}_{rs;n} &= o_P(1).
\end{aligned}$$

Denote by $\hat{\boldsymbol{\beta}}^* = (\hat{\boldsymbol{\beta}}_1^{*\top}, \hat{\boldsymbol{\beta}}_2^{*\top})^\top$. If $\boldsymbol{\mathcal{L}}_{r;n}$ is a sequence of stochastically bounded random vectors, by noting that $g_\Delta(\delta_0)\mathfrak{T}_{rr}(\theta_0) > 0$, from the Quadratic Approximation Lemma (Fan et al., 1995) we obtain the following result

$$\hat{\boldsymbol{\beta}}^* = \boldsymbol{A}_{\delta_0}^{-1}\mathcal{L}_n + o_P(1). \tag{D3}$$

In addition, assume that, as $n \to \infty$, for the linear case $nh^3 \to \infty$, and for the circular predictor case, $\omega_{k+l-2,1} = b_n(B_{\mathfrak{p}_r\mathfrak{p}_s})_{k,l}$, where $B_{\mathfrak{p}_r\mathfrak{p}_s}$ are non-random matrices and b_n a sequence of constants tending to zero and satisfying $b_n/a_n^2 \to \infty$. Then,

$$\hat{\boldsymbol{\beta}}^* = \left(\boldsymbol{A}_{\delta_0}^{-1} - b_n\boldsymbol{A}_{\delta_0}^{-1}(\boldsymbol{D}(\delta_0)\otimes\boldsymbol{B})\boldsymbol{A}_{\delta_0}^{-1}\right)\mathcal{L}_n + o_P(b_n), \tag{D4}$$

$$\boldsymbol{D}(\delta_0) = \begin{pmatrix} \frac{\partial}{\partial\delta_0}\left[g_\Delta(\delta_0)\mathfrak{T}_{11}(\delta_0)\right] & 0 \\ 0 & \frac{\partial}{\partial\delta_0}\left[g_\Delta(\delta_0)\mathfrak{T}_{22}(\delta_0)\right] \end{pmatrix},$$

$$\boldsymbol{B} = \begin{pmatrix} B_{\mathfrak{p}_1\mathfrak{p}_1} & B_{\mathfrak{p}_1\mathfrak{p}_2} \\ B_{\mathfrak{p}_2\mathfrak{p}_1} & B_{\mathfrak{p}_2\mathfrak{p}_2} \end{pmatrix} \overset{\text{linear}}{=} \begin{pmatrix} (\mu_{k+l-1})_{1\leq k,l\leq \mathfrak{p}_1+1} & (\mu_{k+l-1})_{1\leq k\leq \mathfrak{p}_1+1,1\leq l\leq \mathfrak{p}_2+1} \\ (\mu_{k+l-1})_{1\leq k\leq \mathfrak{p}_2+1,1\leq l\leq \mathfrak{p}_1+1} & (\mu_{k+l-1})_{1\leq k,l\leq \mathfrak{p}_2+1} \end{pmatrix},$$

where for the linear case $b_n = h$.

In the following, we obtain the asymptotic normality of $\mathcal{L}_n$ from which the asymptotic normality of $\hat{\boldsymbol{\beta}}^*$ is derived. For doing so, we first compute the expected value of $\boldsymbol{\mathcal{L}}_{r;n}$ using the following notation

$$\boldsymbol{\psi}_{r,i}^* = \psi_r(\overline{\eta}_1(\delta_0, \Delta_i), \overline{\eta}_2(\delta_0, \Delta_i); \Phi_i)\boldsymbol{Z}_{r,i}\mathcal{W}(\Delta_i - \delta_0).$$

Then, we obtain that $\mathbb{E}[\mathcal{L}_{r;n}] = na_n\mathbb{E}(\boldsymbol{\psi}^*_{r,1})$, for $r \in \{1,2\}$. Using similar arguments as before,

$$(\mathbb{E}[(\boldsymbol{\psi}^*_{r,1})])_k \overset{\text{linear}}{=} h\int L(z)z^{k-1}g_X(x_0+hz)\lambda_r\big(\overline{\eta}_1(x_0,x_0+hz),\overline{\eta}_2(x_0,x_0+hz)\big)dz,$$

$$(\mathbb{E}[(\boldsymbol{\psi}^*_{r,1})])_k \overset{\text{circular}}{=} \int_{-\pi+\theta_0}^{\pi+\theta_0} K_\nu(\vartheta)\sin(\vartheta)^{k-1}g_\Theta(\theta_0+\vartheta)\lambda_r\big(\overline{\eta}_1(\theta_0,\theta_0+\vartheta),\overline{\eta}_2(\theta_0,\theta_0+\vartheta)\big)d\vartheta.$$

In this case, considering the Taylor expansion of η_r and of the power of the sines, we obtain

$$\eta_r(\delta) = \eta_r(\delta_0) + \eta_r'(\delta_0)(\delta-\delta_0) + \ldots + \frac{\eta_r^{(\mathfrak{p}_r)}(\delta_0)}{\mathfrak{p}_r!}(\delta-\delta_0)^{\mathfrak{p}_r} + \frac{\eta_r^{(\mathfrak{p}_r+1)}(\delta_0)}{(\mathfrak{p}_r+1)!}(\delta-\delta_0)^{\mathfrak{p}_r+1} + \frac{\eta_r^{(\mathfrak{p}_r+2)}(\delta_0)}{(\mathfrak{p}_r+2)!}(\delta-\delta_0)^{\mathfrak{p}_r+2} + O\big((\delta-\delta_0)^{\mathfrak{p}_r+3}\big),$$

$$\eta_r(\delta) = \overline{\eta}_r(\delta,\delta_0) + \eta^*_{r,\mathfrak{p}_r+1}(\delta_0)(\delta-\delta_0)^{\mathfrak{p}_r+1} + \eta^*_{r,\mathfrak{p}_r+2}(\delta_0)(\delta-\delta_0)^{\mathfrak{p}_r+2} + O\big((\delta-\delta_0)^{\mathfrak{p}_r+3}\big),$$

$$\eta^*_{r,k}(\delta_0) = \frac{\eta_r^{(k)}(\delta_0)}{k!}, \text{ in the linear case or if } \mathfrak{p}_r \le 2 \text{ in the circular case.}$$

In the circular case, if $\mathfrak{p}_r > 2$, $\eta^*_{r,k}(\theta_0)$ is obtained from the above Taylor expansion and corresponds with a linear combination of the quantities $\eta_j^{(j)}(\theta_0)$, with $j \in \{0,\ldots,k\}$ (see Remark 1). Employing again a Taylor expansion and considering that $\lambda_r(\overline{\eta}_1(\delta_0,\delta_0),\overline{\eta}_2(\delta_0,\delta_0)) = 0$, we obtain that

$$\lambda_r\big(\overline{\eta}_1(x_0,x_0+hz),\overline{\eta}_2(x_0,x_0+hz)\big) \overset{\text{linear}}{=} -\sum_{s=1}^{2}\lambda_{rs}\big(\overline{\eta}_1(x_0,x_0+hz),\overline{\eta}_2(x_0,x_0+hz)\big) \times\left(\frac{\eta_r^{(\mathfrak{p}_r+1)}(x_0)}{(\mathfrak{p}_r+1)!}(hz)^{\mathfrak{p}_r+1} + \frac{\eta_r^{(\mathfrak{p}_r+2)}(x_0)}{(\mathfrak{p}_r+2)!}(hz)^{\mathfrak{p}_r+2}\right) + o(h^{\mathfrak{p}_r+2}),$$

$$\lambda_r\big(\overline{\eta}_1(\theta_0,\theta_0+\vartheta),\overline{\eta}_2(\theta_0,\theta_0+\vartheta)\big) \overset{\text{circular}}{=} -\sum_{s=1}^{2}\lambda_{rs}\big(\overline{\eta}_1(\theta_0,\theta_0+\vartheta),\overline{\eta}_2(\theta_0,\theta_0+\vartheta)\big) \times\left(\eta^*_{r,\mathfrak{p}_r+1}(\theta_0)\vartheta^{\mathfrak{p}_r+1} + \eta^*_{r,\mathfrak{p}_r+2}(\theta_0)\vartheta^{\mathfrak{p}_r+2}\right) + o(\vartheta^{\mathfrak{p}_r+2}) + o(\vartheta^{\mathfrak{p}_r+3}).$$

Thus, as done for $(\mathbb{E}[\mathcal{L}_{rs;n}])_{kl}$, considering that $\mathfrak{T}_{rs}(\delta_0) = 0$ if $r \neq s$, from the Taylor expansion we obtain,

$$\begin{aligned}(\mathbb{E}[\mathcal{L}_{r;n}])_k \overset{\text{linear}}{=}\ & \sqrt{nh}h^{\mathfrak{p}_r+1}\left(g_X(x_0)\mathfrak{T}_{rr}(x_0)\left(\frac{\eta_r^{(\mathfrak{p}_r+1)}(x_0)}{(\mathfrak{p}_r+1)!}\mu_{\mathfrak{p}_r+k} + h\frac{\eta_r^{(\mathfrak{p}_r+2)}(x_0)}{(\mathfrak{p}_r+2)!}\mu_{\mathfrak{p}_r+k+1}\right)\right.\\ &\left. + h\frac{\partial}{\partial x_0}\left[g_X(x_0)\mathfrak{T}_{rr}(x_0)\right]\frac{\eta_r^{(\mathfrak{p}_r+1)}(x_0)}{(\mathfrak{p}_r+1)!}\mu_{\mathfrak{p}_r+k+1}\right)\\ &+ o\left(\sqrt{nh}h^{\min(\mathfrak{p}_1,\mathfrak{p}_2)+2}\right),\end{aligned}$$

$$\begin{aligned}(\mathbb{E}[\mathcal{L}_{r;n}])_k \overset{\text{circular}}{=}\ & \sqrt{n}\left(g_\Theta(\theta_0)\mathfrak{T}_{rr}(\theta_0)\left(\eta^*_{r,\mathfrak{p}_r+1}(\theta_0)\omega_{\mathfrak{p}_r+1,k-1} + \eta^*_{r,\mathfrak{p}_r+2}(\theta_0)\omega_{\mathfrak{p}_r+2,k-1}\right)\right.\\ &\left. + \frac{\partial}{\partial\theta_0}\left[g_\Theta(\theta_0)\mathfrak{T}_{rr}(\theta_0)\right]\eta^*_{r,\mathfrak{p}_r+1}(\theta_0)\omega_{\mathfrak{p}_r+2,k-1}\right)\\ &+ o\left(\sqrt{n}\max_{p\in\{\mathfrak{p}_1+2,\mathfrak{p}_1+3,\mathfrak{p}_2+2,\mathfrak{p}_2+3\}}\omega_{p,k-1}\right).\end{aligned}$$

For computing the covariance between the kth component of $\boldsymbol{\mathcal{L}}_{r;n}$ and the lth component of $\boldsymbol{\mathcal{L}}_{s;n}$, we have that

$$\begin{aligned}
\mathrm{Cov}\left[(\mathcal{L}_{r;n})_k,(\mathcal{L}_{s;n})_l\right] &= \mathbb{E}\left[(\mathcal{L}_{r;n})_k(\mathcal{L}_{s;n})_l\right] - \mathbb{E}\left[(\mathcal{L}_{r;n})_k\right]\mathbb{E}\left[(\mathcal{L}_{s;n})_l\right] = na_n^2\mathbb{E}\left[(\psi^*_{r,1})_k(\psi^*_{s,1})_l\right] \\
&\quad - \mathbb{E}\left[(\mathcal{L}_{r;n})_k\right]\mathbb{E}\left[(\mathcal{L}_{s;n})_l\right] \\
\mathrm{Cov}\left[(\mathcal{L}_{r;n})_k,(\mathcal{L}_{s;n})_l\right] &\overset{\text{linear}}{=} \frac{1}{h}\mathbb{E}\left[(\psi^*_{r,1})_k(\psi^*_{s,1})_l\right] + O(h^{2\min(\mathfrak{p}_1,\mathfrak{p}_2)+3}) \\
\mathrm{Cov}\left[(\mathcal{L}_{r;n})_k,(\mathcal{L}_{s;n})_l\right] &\overset{\text{circular}}{=} \mathbb{E}\left[(\psi^*_{r,1})_k(\psi^*_{s,1})_l\right] + O\left(\max_{p\in\{\mathfrak{p}_1+1,\mathfrak{p}_1+2,\mathfrak{p}_2+1,\mathfrak{p}_2+2\}}\omega_{p,k-1}\right).
\end{aligned}$$

Now, $\mathbb{E}\left[(\psi^*_{r,1})_k(\psi^*_{s,1})_l\right]$ can be obtained using the Taylor expansion, in a similar way as before,

$$\begin{aligned}
\mathbb{E}\left[(\psi^*_{r,1})_k(\psi^*_{s,1})_l\right] &\overset{\text{linear}}{=} \int\int_{-\pi}^{\pi}\psi_r(\overline{\eta}_1(x_0,x),\overline{\eta}_2(x_0,x),\phi)\psi_s(\overline{\eta}_1(x_0,x),\overline{\eta}_2(x_0,x),\phi)\left(\frac{x-x_0}{h}\right)^{k+l-2} \\
&\quad \times L_h^2(x-x_0)g_{X,\Phi}(x,\phi)dxd\phi \\
&\overset{\text{linear}}{=} g_X(x_0)\mathfrak{T}_{rs}(x_0)\int z^{k+l-2}L^2(z)dz + O(h), \\
\mathbb{E}\left[(\psi^*_{r,1})_k(\psi^*_{s,1})_l\right] &\overset{\text{circular}}{=} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\psi_r(\overline{\eta}_1(\theta_0,\theta),\overline{\eta}_2(\theta_0,\theta),\phi)\psi_s(\overline{\eta}_1(\theta_0,\theta),\overline{\eta}_2(\theta_0,\theta),\phi)0\sin^{k+l-2} \\
&\quad \times(\theta-\theta_0)K_\nu^2(\theta-\theta_0)g_{\Theta,\Phi}(\theta,\phi)d\theta d\phi \\
&\overset{\text{circular}}{=} g_\Theta(\theta_0)\mathfrak{T}_{rs}(\theta_0)\int_{-\pi}^{\pi}\sin^{k+l-2}(\vartheta)K_\nu^2(\vartheta)d\vartheta \\
&\quad + O\left(\int_{-\pi}^{\pi}\vartheta\sin^{k+l-2}(\vartheta)K_\nu^2(\vartheta)d\vartheta\right).
\end{aligned}$$

Assume that, as $n\to\infty$, the following quantities tend to zero: h in the linear case; $\omega_{p,k-1}$, for $p\in\{\mathfrak{p}_1+1,\mathfrak{p}_1+2,\mathfrak{p}_2+1,\mathfrak{p}_2+2\}$, $\int_{-\pi}^{\pi}\vartheta\sin^{k+l-2}(\vartheta)K_\nu^2(\vartheta)d\vartheta$, for $1\le k,l\le\mathfrak{p}_r+1$, in the circular case. Then,

$$\mathrm{Cov}[\mathcal{L}_n]\to\boldsymbol{\Sigma}_{\delta_0}.$$

As mentioned before the asymptotic normality of $\hat{\boldsymbol{\beta}}^*$ is a consequence of the asymptotic normality of $\mathcal{L}_n$, which can be written as

$$(\mathrm{Cov}[\mathcal{L}_n])^{-1/2}(\mathcal{L}_n-\mathbb{E}[\mathcal{L}_n])\xrightarrow{D}N(\mathbf{0},\boldsymbol{I}_{\mathfrak{p}_1+\mathfrak{p}_2+2}).$$

This asymptotic result is obtained as in Fan and Gijbels (1996, Ch. 5) from the Cramér-Wold device, i.e., by seeing that, for any unit vector $\boldsymbol{a}$,

$$(\boldsymbol{a}^\top\mathrm{Cov}[\mathcal{L}_n]\boldsymbol{a})^{-1/2}(\boldsymbol{a}^\top\mathcal{L}_n-\boldsymbol{a}^\top\mathbb{E}[\mathcal{L}_n])\xrightarrow{D}N(0,1).$$

The last result is derived from Lyapounov's condition.

From the asymptotic normality of $\mathcal{L}_n$, as $n\to\infty$, we obtain that,

$$\left(\boldsymbol{A}_{\delta_0}^{-1}\boldsymbol{\Sigma}_{\delta_0}\boldsymbol{A}_{\delta_0}^{-1}\right)^{-1/2}\left(\hat{\boldsymbol{\beta}}^{*}-\boldsymbol{\mathcal{B}}_{\delta_0}\right)\xrightarrow{D}N(\mathbf{0},\boldsymbol{I}_{\mathfrak{p}_1+\mathfrak{p}_2+2}),\tag{D5}$$

$$\left(\sqrt{nh}\big(\boldsymbol{H}_{\mathfrak{p}_1;n}(\hat{\boldsymbol{\beta}}_1(x_0)-\boldsymbol{\beta}_1(x_0)),\boldsymbol{H}_{\mathfrak{p}_2;n}(\hat{\boldsymbol{\beta}}_2(x_0)-\boldsymbol{\beta}_2(x_0))\big)^{\top}-\boldsymbol{\mathcal{B}}_{x_0}\right)\xrightarrow[\text{linear}]{D}N\big(\mathbf{0},\boldsymbol{A}_{x_0}^{-1}\boldsymbol{\Sigma}_{x_0}\boldsymbol{A}_{x_0}^{-1}\big),$$

$$\left(\sqrt{n}\big(\hat{\boldsymbol{\beta}}_1(\theta_0)-\boldsymbol{\beta}_1(\theta_0),\hat{\boldsymbol{\beta}}_2(\theta_0)-\boldsymbol{\beta}_2(\theta_0)\big)^{\top}-\boldsymbol{\mathcal{B}}_{\theta_0}\right)\xrightarrow[\text{circular}]{D}N\Big(\mathbf{0},\boldsymbol{A}_{\theta_0}^{-1}\boldsymbol{\Sigma}_{\theta_0}\boldsymbol{A}_{\theta_0}^{-1}\Big).$$

Depending on if the result (D3) or (D4) is employed, the bias $\boldsymbol{\mathcal{B}}_{\delta_0}$ is (respectively) equal to

$$\boldsymbol{\mathcal{B}}_{\delta_0}\overset{(D3)}{=}\boldsymbol{A}_{\delta_0}^{-1}\mathbb{E}[\mathcal{L}_n](1+o(1)),$$

$$\boldsymbol{\mathcal{B}}_{\delta_0}\overset{(D4)}{=}\left(\boldsymbol{A}_{\delta_0}^{-1}-b_n\boldsymbol{A}_{\delta_0}^{-1}(\boldsymbol{D}(\delta_0)\otimes\boldsymbol{B})\boldsymbol{A}_{\delta_0}^{-1}\right)\mathbb{E}[\mathcal{L}_n](1+o(1)).$$

In the following, we provide some useful results to compute the asymptotic expression of $\boldsymbol{\mathcal{B}}_{\delta_0}$. Inspired in the proof of Lemma 3 in Fan et al. (1995), we have the following result. Denote as $c_{\mathfrak{p}_r;kl}$ to the cofactor of $(\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r})_{k,l}$, with $1\le k,l\le\mathfrak{p}_r$. Then, given a non-negative integer value of l, by expanding the determinant of $\boldsymbol{V}_{k,\mathfrak{p}_r}(x)$ along the $(k+1)$st column, we obtain

$$\begin{aligned}\zeta_{k,\mathfrak{p}_r+l+1,\mathfrak{p}_r}&\overset{\text{linear}}{=}\int x^{\mathfrak{p}_r+l+1}W_{k,\mathfrak{p}_r}(x)dx=\frac{k!}{|\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}|}\int\sum_{j=1}^{\mathfrak{p}_r+1}x^{\mathfrak{p}_r+j+l}c_{\mathfrak{p}_r;j(k+1)}L(x)dx\\&=k!\sum_{j=1}^{\mathfrak{p}_r+1}\left(\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}^{-1}\right)_{j(k+1)}\mu_{\mathfrak{p}_r+j+l},\\\zeta_{k,\mathfrak{p}_r+l+1,\mathfrak{p}_r}&\overset{\text{circular}}{=}\int\sin^{\mathfrak{p}_r+l+1}(\theta)W_{k,\mathfrak{p}_r}(\theta)d\theta=\frac{k!}{|\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}|}\int_{-\pi}^{\pi}\sum_{j=1}^{\mathfrak{p}_r+1}\sin^{\mathfrak{p}_r+l+1}(\theta)\theta^{j-1}c_{\mathfrak{p}_r;j(k+1)}K_\nu(\theta)d\theta\\&=k!\sum_{j=1}^{\mathfrak{p}_r+1}\left(\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}^{-1}\right)_{j(k+1)}\omega_{\mathfrak{p}_r+l+1,j-1}.\end{aligned}$$

If the convergence (D3) is employed, given $r\in\{1,2\}$ and $k\in\{0,\dots,\mathfrak{p}_r\}$, from the previous results, the components of the expression of $\boldsymbol{\mathcal{B}}_{\delta_0}$ can be obtained as follows,

$$\begin{aligned}\boldsymbol{A}_{\delta_0}^{-1}&=\begin{pmatrix}(A_{\delta_0}^{-1})_{11}&0\\0&(A_{\delta_0}^{-1})_{22}\end{pmatrix},\\(A_{\delta_0}^{-1})_{rr}&=g_\Delta(\delta_0)^{-1}\mathfrak{F}_{rr}^{-1}(\delta_0)\left(\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}\right)^{-1},\\\left((A_{x_0}^{-1})_{rr}\mathbb{E}\left[\boldsymbol{\mathcal{L}}_{r;n}\right]\right)_{k+1}&\overset{\text{linear}}{=}\frac{\sqrt{nh}h^{\mathfrak{p}_r+1}}{k!}\Bigg(\frac{\eta_r^{(\mathfrak{p}_r+1)}(x_0)}{(\mathfrak{p}_r+1)!}\zeta_{k,\mathfrak{p}_r+1,\mathfrak{p}_r}\\&\quad+h\frac{\eta_r^{(\mathfrak{p}_r+2)}(x_0)}{(\mathfrak{p}_r+2)!}\zeta_{k,\mathfrak{p}_r+2,\mathfrak{p}_r}+h\frac{\frac{\partial}{\partial x_0}\left[g_X(x_0)\mathfrak{F}_{rr}(x_0)\right]}{g_X(x_0)\mathfrak{F}_{rr}(x_0)}\frac{\eta_r^{(\mathfrak{p}_r+1)}(x_0)}{(\mathfrak{p}_r+1)!}\zeta_{k,\mathfrak{p}_r+2,\mathfrak{p}_r}\Bigg)\\&\quad+o\left(\sqrt{nh}h^{\min(\mathfrak{p}_1,\mathfrak{p}_2)+2}\right),\end{aligned}$$

$$\left((A_{\theta_0}^{-1})_{rr}\mathbb{E}\left[\mathcal{L}_{r;n}\right]\right)_{k+1} \overset{\text{circular}}{=} \frac{\sqrt{n}}{k!}\Bigg(\eta^*_{r,\mathfrak{p}_r+1}(\theta_0)\zeta_{k,\mathfrak{p}_r+1,\mathfrak{p}_r} + \eta^*_{r,\mathfrak{p}_r+2}(\theta_0)\zeta_{k,\mathfrak{p}_r+2,\mathfrak{p}_r} + \frac{\frac{\partial}{\partial\theta_0}\left[g_\Theta(\theta_0)\mathfrak{I}_{rr}(\theta_0)\right]}{g_\Theta(\theta_0)\mathfrak{I}_{rr}(\theta_0)}\eta^*_{r,\mathfrak{p}_r+1}(\theta_0)\zeta_{k,\mathfrak{p}_r+2,\mathfrak{p}_r}\Bigg) + o\left(\sqrt{n}\max_{p\in\{\mathfrak{p}_1+2,\mathfrak{p}_1+3,\mathfrak{p}_2+2,\mathfrak{p}_2+3\}}\omega_{p,k-1}\right).$$

Now, the remaining elements of the components of $\mathcal{B}_{\delta_0}$, when employing result (D3) are obtained by considering that,

$$\boldsymbol{A}_{\delta_0}^{-1}(\boldsymbol{D}(\delta_0)\otimes\boldsymbol{B})\boldsymbol{A}_{\delta_0}^{-1} = \begin{pmatrix}(\boldsymbol{A}_{\delta_0}^{-1}(\boldsymbol{D}(\delta_0)\otimes\boldsymbol{B})\boldsymbol{A}_{\delta_0}^{-1})_{11} & 0\\ 0 & (\boldsymbol{A}_{\delta_0}^{-1}(\boldsymbol{D}(\delta_0)\otimes\boldsymbol{B})\boldsymbol{A}_{\delta_0}^{-1})_{22}\end{pmatrix},$$

$$(\boldsymbol{A}_{\delta_0}^{-1}(\boldsymbol{D}(\delta_0)\otimes\boldsymbol{B})\boldsymbol{A}_{\delta_0}^{-1})_{rr} = \frac{\frac{\partial}{\partial\delta_0}\left[g_\Delta(\delta_0)\mathfrak{I}_{rr}(\delta_0)\right]}{g_\Delta(\delta_0)^2\mathfrak{I}_{rr}^2(\delta_0)}\left(\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}\right)^{-1}B_{\mathfrak{p}_r\mathfrak{p}_r}\left(\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}\right)^{-1}.$$

Note that, for $k\in\{1,\ldots,\mathfrak{p}_r+1\}$ and l a non-negative integer, in the linear case, we have that $(B_{\mathfrak{p}_r\mathfrak{p}_r})_{kl} = (\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r})_{k(l+1)}$. Under Condition (D7) since $\omega_{k+l-2,1} = b_n(B_{\mathfrak{p}_r p_s})_{k,l}$, using the Fourier approximation, in the circular case, we obtain that $\omega_{k+l-2,1} = \mu_{k+l-1}(1+o(1))$, and thus, $(B_{\mathfrak{p}_r\mathfrak{p}_r})_{kl} = (b_n)^{-1}(\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r})_{k(l+1)}(1+o(1))$. From the proof of Lemma 2 in Fan et al. (1995), for $j,k\in\{1,\ldots,\mathfrak{p}_r+1\}$, it can be shown that,

$$\left(\left(\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}\right)^{-1}B_{\mathfrak{p}_r\mathfrak{p}_r}\left(\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}\right)^{-1}\right)_{jk} \overset{\text{linear}}{=} \left(\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}\right)^{-1}_{(j-1)k} + \left(\sum_{l=1}^{\mathfrak{p}_r+1}\left(\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}\right)^{-1}_{jl}\mu_{\mathfrak{p}_r+l}\right)\left(\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}\right)^{-1}_{(\mathfrak{p}_r+1)k},$$

$$\left(\left(\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}\right)^{-1}B_{\mathfrak{p}_r\mathfrak{p}_r}\left(\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}\right)^{-1}\right)_{jk} \overset{\text{circular}}{=} b_n^{-1}\Bigg(\left(\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}\right)^{-1}_{(j-1)k} + \left(\sum_{l=1}^{\mathfrak{p}_r+1}\left(\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}\right)^{-1}_{jl}\omega_{\mathfrak{p}_r+1,l-1}\right) \times\left(\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}\right)^{-1}_{(\mathfrak{p}_r+1)k}\Bigg)(1+o(1)),$$

$$\left(\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}\right)^{-1}_{(j-1)k} = 0, \text{ when } j=1.$$

From the previous equalities, the second term of the bias obtained when employing the result (D4) can be derived as follows,

$$\left(b_n\left(\boldsymbol{A}_{x_0}^{-1}(\boldsymbol{D}(x_0)\otimes\boldsymbol{B})\boldsymbol{A}_{x_0}^{-1}\right)_{rr}\mathbb{E}\left[\mathcal{L}_{r;n}\right]\right)_{k+1} \overset{\text{linear}}{=} \sqrt{nh}h^{\mathfrak{p}_r+2}\frac{\frac{\partial}{\partial x_0}\left[g_X(x_0)\mathfrak{I}_{rr}(x_0)\right]}{g_X(x_0)\mathfrak{I}_{rr}(x_0)}\frac{\eta_r^{(\mathfrak{p}_r+1)}(x_0)}{(\mathfrak{p}_r+1)!} \times\left(\frac{\zeta_{k-1,\mathfrak{p}_r+1,\mathfrak{p}_r}}{(k-1)!} + \frac{\zeta_{\mathfrak{p}_r,\mathfrak{p}_r+1,\mathfrak{p}_r}\zeta_{k,\mathfrak{p}_r+1,\mathfrak{p}_r}}{k!\mathfrak{p}_r!}\right) + o\left(\sqrt{nh}h^{\min(\mathfrak{p}_1,\mathfrak{p}_2)+2}\right),$$

$$\left(b_n\left(\boldsymbol{A}_{\theta_0}^{-1}(\boldsymbol{D}(\theta_0)\otimes\boldsymbol{B})\boldsymbol{A}_{\theta_0}^{-1}\right)_{rr}\mathbb{E}\left[\mathcal{L}_{r;n}\right]\right)_{k+1} \overset{\text{circular}}{=} \sqrt{n}\Bigg(\frac{\frac{\partial}{\partial\theta_0}\left[g_\Theta(\theta_0)\mathfrak{I}_{rr}(\theta_0)\right]}{g_\Theta(\theta_0)\mathfrak{I}_{rr}(\theta_0)} \times\left(\eta^*_{r,\mathfrak{p}_r+1}(\theta_0)\left(\frac{\zeta_{k-1,\mathfrak{p}_r+1,\mathfrak{p}_r}}{(k-1)!} + \frac{\zeta_{\mathfrak{p}_r,\mathfrak{p}_r+1,\mathfrak{p}_r}\zeta_{k,\mathfrak{p}_r+1,\mathfrak{p}_r}}{k!\mathfrak{p}_r!}\right)\right.$$

$$+\eta^*_{r,\mathfrak{p}_r+2}(\theta_0)\left(\frac{\zeta_{k-1,\mathfrak{p}_r+2,\mathfrak{p}_r}}{(k-1)!}+\frac{\zeta_{\mathfrak{p}_r,\mathfrak{p}_r+2,\mathfrak{p}_r}\zeta_{k,\mathfrak{p}_r+2,\mathfrak{p}_r}}{k!\mathfrak{p}_r!}\right)\Bigg)$$
$$+\left(\frac{\frac{\partial}{\partial\theta_0}\left[g_\Theta(\theta_0)\mathfrak{I}_{rr}(\theta_0)\right]}{g_\Theta(\theta_0)\mathfrak{I}_{rr}(\theta_0)}\right)^2\eta^*_{r,\mathfrak{p}_r+1}(\theta_0)$$
$$\times\left(\frac{\zeta_{k-1,\mathfrak{p}_r+2,\mathfrak{p}_r}}{(k-1)!}+\frac{\zeta_{\mathfrak{p}_r,\mathfrak{p}_r+2,\mathfrak{p}_r}\zeta_{k,\mathfrak{p}_r+2,\mathfrak{p}_r}}{k!\mathfrak{p}_r!}\right)\Bigg)$$
$$+o\left(\sqrt{n}\max_{p\in\{\mathfrak{p}_1+2,\mathfrak{p}_1+3,\mathfrak{p}_2+2,\mathfrak{p}_2+3\}}\omega_{p,k-1}\right).$$

Now, since, in both the linear and the circular case, $\mu_k=0$ if k is odd, we obtain that $\zeta_{k,j,\mathfrak{p}_r}=0$ if $k+j$ is odd. Consider, also, that if $\mathfrak{p}_r\geq 0$, $r\in\{1,2\}$, for any non-negative integer k, $\zeta_{k,\mathfrak{p}_r+2,\mathfrak{p}_r}\to 0$, as $n\to\infty$, since $\zeta_{k,\mathfrak{p}_r+2,\mathfrak{p}_r}=O(\mu_{\mathfrak{p}_r+2})+O(\mu_{\mathfrak{p}_r+3})$. When (D4) holds, $k+\mathfrak{p}_r$ is even, with $k\in\{0,1,\ldots,\mathfrak{p}_r\}$, and x_0 is a point in the interior of S_X; these last results imply that the components of the asymptotic bias are

$$\left(\mathrm{A}\boldsymbol{\mathcal{B}}_{x_0}\right)_{r,k+1}\overset{\text{linear}}{=}\frac{\sqrt{nh}h^{\mathfrak{p}_r+2}}{k!}$$
$$\times\left(\frac{\eta_r^{(\mathfrak{p}_r+2)}(x_0)}{(\mathfrak{p}_r+2)!}\zeta_{k,\mathfrak{p}_r+2,\mathfrak{p}_r}+\frac{\frac{\partial}{\partial x_0}\left[g_X(x_0)\mathfrak{I}_{rr}(x_0)\right]}{g_X(x_0)\mathfrak{I}_{rr}(x_0)}\frac{\eta_r^{(\mathfrak{p}_r+1)}(x_0)}{(\mathfrak{p}_r+1)!}\left(\zeta_{k,\mathfrak{p}_r+2,\mathfrak{p}_r}-k\zeta_{k-1,\mathfrak{p}_r+1,\mathfrak{p}_r}\right)\right),$$

$$\left(\mathrm{A}\boldsymbol{\mathcal{B}}_{\theta_0}\right)_{r,k+1}\overset{\text{circular}}{=}\frac{\sqrt{n}}{k!}$$
$$\times\left(\eta^*_{r,\mathfrak{p}_r+2}(\theta_0)\zeta_{k,\mathfrak{p}_r+2,\mathfrak{p}_r}+\frac{\frac{\partial}{\partial\theta_0}\left[g_\Theta(\theta_0)\mathfrak{I}_{rr}(\theta_0)\right]}{g_\Theta(\theta_0)\mathfrak{I}_{rr}(\theta_0)}\eta^*_{r,\mathfrak{p}_r+1}(\theta_0)\left(\zeta_{k,\mathfrak{p}_r+2,\mathfrak{p}_r}-k\zeta_{k-1,\mathfrak{p}_r+1,\mathfrak{p}_r}\right)\right).$$

If $k+\mathfrak{p}_r$ is odd, with $k\in\{0,1,\ldots,\mathfrak{p}_r\}$, then, the asymptotic bias can be obtained from result (D3) and its components are equal to

$$\left(\mathrm{A}\boldsymbol{\mathcal{B}}_{x_0}\right)_{r,k+1}\overset{\text{linear}}{=}\frac{\sqrt{nh}h^{\mathfrak{p}_r+1}}{k!}\frac{\eta_r^{(\mathfrak{p}_r+1)}(x_0)}{(\mathfrak{p}_r+1)!}\zeta_{k,\mathfrak{p}_r+1,\mathfrak{p}_r},$$

$$\left(\mathrm{A}\boldsymbol{\mathcal{B}}_{\theta_0}\right)_{r,k+1}\overset{\text{circular}}{=}\frac{\sqrt{n}}{k!}\eta^*_{r,\mathfrak{p}_r+1}(\theta_0)\zeta_{k,\mathfrak{p}_r+1,\mathfrak{p}_r}.$$

Finally, we will study the elements of the asymptotic covariance matrix in Equation (D5). For doing so, first note that the covariance of $\hat{\boldsymbol{\beta}}^*_r$, $r\in\{1,2\}$, can be obtained from

$$\mathrm{Cov}\left[\hat{\boldsymbol{\beta}}^*_r\right]=\left(\boldsymbol{A}^{-1}_{\delta_0}\boldsymbol{\Sigma}_{\delta_0}\boldsymbol{A}^{-1}_{\delta_0}\right)_{rr}=(g_\Delta(\delta_0)\mathfrak{I}_{rr}(\delta_0))^{-1}(\boldsymbol{U}^{-1}_{\delta_0})_{rr}\boldsymbol{T}_{rr}(\boldsymbol{U}^{-1}_{\delta_0})_{rr}.$$

For computing the asymptotic variance of $\hat{\boldsymbol{\beta}}^*_{r(k+1)}$, $k\in\{0,\ldots,\mathfrak{p}_r\}$, arguing in the same way as for $\zeta_{k,\mathfrak{p}_r+l+1,\mathfrak{p}_r}$, we obtain that

$$\xi_{k,\mathfrak{p}_r} \overset{\text{linear}}{=} \frac{(k!)^2}{|\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}|^2} \int \sum_{j=1}^{\mathfrak{p}_r+1}\sum_{l=1}^{\mathfrak{p}_r+1} x^{j+l-2} c_{\mathfrak{p}_r;j(k+1)} c_{\mathfrak{p}_r;l(k+1)} L^2(x)dx,$$

$$\xi_{k,\mathfrak{p}_r} \overset{\text{circular}}{=} \frac{(k!)^2}{|\boldsymbol{U}_{\mathfrak{p}_r\mathfrak{p}_r}|^2} \int \sum_{j=1}^{\mathfrak{p}_r+1}\sum_{l=1}^{\mathfrak{p}_r+1} \sin^{j+l-2}(\theta) c_{\mathfrak{p}_r;j(k+1)} c_{\mathfrak{p}_r;l(k+1)} K_\nu^2(\theta)d\theta,$$

$$\xi_{k,\mathfrak{p}_r} = (k!)^2 \sum_{j=1}^{\mathfrak{p}_r+1}\sum_{l=1}^{\mathfrak{p}_r+1} \left(\boldsymbol{U}^{-1}_{\mathfrak{p}_r\mathfrak{p}_r}\right)_{j(k+1)} \tau_{j+l-2} \left(\boldsymbol{U}^{-1}_{\mathfrak{p}_r\mathfrak{p}_r}\right)_{l(k+1)}.$$

D2 Proof of Corollary 1

For deriving the asymptotic distributions of $\hat{m}(\delta_0)$ and $\hat{c}(\delta_0)$, first we need to stablish the asymptotic distributions of $\hat{\eta}_r^{(k)}(\delta_0)$. For doing so, first note that $\hat{\eta}_r^{(k)}(\delta_0) = \nu!\hat{\beta}_{r;k}(\delta_0)$, in the linear case for any non-negative integer, and in the circular case if also $k \leq 2$. Then, in those cases, the bias and variance of the derivatives of the estimator $\eta_r(\delta_0)$ are equal to $\mathbb{E}[\hat{\eta}_r^{(k)}(\delta_0)] = k!\mathbb{E}[\hat{\beta}_{r;k}(\delta_0)]$ and $\text{Var}[\hat{\eta}_r^{(k)}(\delta_0)] = (k!)^2\text{Var}[\hat{\beta}_{r;k}(\delta_0)]$. Thus, we obtain that $\hat{\eta}_r^{(k)}(\delta_0)$ asymptotically converges, as $n \to \infty$, in distribution to a univariate Gaussian with the following mean and variance,

$$\mathbb{E}[\hat{\eta}_r^{(k)}(x_0)] \overset{\text{linear}}{=} \eta_r^{(k)}(\delta_0) + \frac{k!}{\sqrt{nh}h^k}\left(\text{A}\boldsymbol{\mathcal{B}}_{x_0}\right)_{r,k+1},$$

$$\mathbb{E}[\hat{\eta}_r^{(k)}(\theta_0)] \overset{\text{circular}}{=} \eta_r^{(k)}(\delta_0) + \frac{k!}{\sqrt{n}}\left(\text{A}\boldsymbol{\mathcal{B}}_{\theta_0}\right)_{r,k+1},$$

$$\left(\text{Cov}\left[\hat{\boldsymbol{\beta}}_r^*\right]\right)_{(k+1)(k+1)} = \left((k!)^2 g_\Delta(\delta_0)\mathfrak{T}_{rr}(\delta_0)\right)^{-1}\xi_{k,\mathfrak{p}_r},$$

$$\text{Var}[\hat{\eta}_r^{(k)}(x_0)] \overset{\text{linear}}{=} \frac{(k!)^2}{nh^{2k+1}}\text{Var}[\hat{\beta}^*_{r;k}(x_0)] = \left(nh^{2k+1}g_X(x_0)\mathfrak{T}_{rr}(x_0)\right)^{-1}\xi_{k,\mathfrak{p}_r},$$

$$\text{Var}[\hat{\eta}_r^{(k)}(\theta_0)] \overset{\text{circular}}{=} \frac{(k!)^2}{n}\text{Var}[\hat{\beta}^*_{r;k}(\theta_0)] = (ng_\Theta(\theta_0)\mathfrak{T}_{rr}(\theta_0))^{-1}\xi_{k,\mathfrak{p}_r}.$$

Using the previous result, the asymptotic distribution of the conditional modal direction and concentration, can be derived, using the delta method, by considering that $\hat{m}(\delta_0) = h_1^{-1}(\hat{\beta}_{1;0}(\delta_0))$, $\hat{c}(\delta_0) = h_2^{-1}(\hat{\beta}_{2;0}(\delta_0))$. If the derivative of h_1, in $m(\delta_0)$, and h_2, in $c(\delta_0)$, exist and are not zero-valued, then we obtain the results in Corollary 1. In the circular case, for a derivative value of $r > 2$, the bias and variance of $\hat{\eta}_r^{(k)}(\theta_0)$, can be obtained similarly considering the linear combinations stated in Remark 1.

D3 Proof of Corollary 2

Results regarding the optimal smoothing parameter of $m(\delta)$ and $c(\delta)$ can be derived from the expressions of $\mathbb{E}[\hat{\eta}_r^{(k)}(\delta_0)]$ and $\text{Var}[\hat{\eta}_r^{(k)}(\delta_0)]$ in Section D2. When both parameters are unknown, $m(\delta)$ and $c(\delta)$, the performance of the estimation method can be measured via the (weighted) sum of the mean squared errors $\text{MSE}(\hat{m}(\delta), \hat{c}(\delta)) = \text{MSE}(\hat{m}(\delta))w_1(\delta) + \text{MSE}(\hat{c}(\delta))w_2(\delta)$. For simplicity, we assume that $\mathfrak{p}_1 = \mathfrak{p}_2 = \mathfrak{p}$, with $\mathfrak{p}$ a positive odd number. Then, given the weight functions w_1 and w_2, the AMSE is equal to

$$\begin{aligned}\text{AMSE}(\hat{m}(\delta_0), \hat{c}(\delta_0)) = &\left((\text{ABias}(\hat{m}(\delta_0)))^2 + \text{AVar}(\hat{m}(\delta_0))\right)w_1(\delta_0)\\ &+ \left((\text{ABias}(\hat{c}(\delta_0)))^2 + \text{AVar}(\hat{c}(\delta_0))\right)w_2(\delta_0),\end{aligned}$$

$$\begin{aligned}\mathrm{AMSE}(\hat{m}(x_0),\hat{c}(x_0)) &\overset{\text{linear}}{=} h^{2\mathfrak{p}+2}\left(\frac{\zeta_{0,\mathfrak{p}+1,\mathfrak{p}}}{(\mathfrak{p}+1)!}\right)^2\left(w_1(x_0)\left(\frac{\eta_1^{(\mathfrak{p}+1)}(x_0)}{h_1'(m(x_0))}\right)^2+w_2(x_0)\left(\frac{\eta_2^{(\mathfrak{p}+1)}(x_0)}{h_2'(c(x_0))}\right)^2\right)\\ &\quad+\frac{\xi_{0,\mathfrak{p}}}{nhg_X(x_0)}\left(w_1(x_0)\left(\mathfrak{I}_{11}(x_0)\left(h_1'(m(x_0))\right)^2\right)^{-1}\right.\\ &\quad\left.+w_2(x_0)\left(\mathfrak{I}_{22}(x_0)\left(h_2'(c(x_0))\right)^2\right)^{-1}\right),\\ \mathrm{AMSE}(\hat{m}(\theta_0),\hat{c}(\theta_0)) &\overset{\text{circular}}{=} \zeta_{0,\mathfrak{p}+1,\mathfrak{p}}^2\left(w_1(\theta_0)\left(\frac{\eta_{1,\mathfrak{p}+1}^*(\theta_0)}{h_1'(m(\theta_0))}\right)^2+w_2(\theta_0)\left(\frac{\eta_{2,\mathfrak{p}+1}^*(\theta_0)}{h_2'(c(\theta_0))}\right)^2\right)\\ &\quad+\frac{\xi_{0,\mathfrak{p}}}{ng_\Theta(\theta_0)}\left(w_1(\theta_0)\left(\mathfrak{I}_{11}(\theta_0)\left(h_1'(m(\theta_0))\right)^2\right)^{-1}\right.\\ &\quad\left.+w_2(\theta_0)\left(\mathfrak{I}_{22}(\theta_0)\left(h_2'(c(\theta_0))\right)^2\right)^{-1}\right).\end{aligned}$$

From the previous expressions, in the linear case, the bandwidth minimizing the AMSE expression, and its integrated version, can be obtained by differentiating the corresponding expression with respect to h and equating it to zero. For any positive odd value of $\mathfrak{p}$, the optimal bandwidths are equal to

$$\begin{aligned}&h_{p_L,p_R;\mathrm{opt}}^{\mathrm{AMSE}}(x_0)\\ &\overset{\text{linear}}{=} \mathfrak{c}_{\mathfrak{p}}\left(\frac{w_1(x_0)(\mathfrak{I}_{11}(x_0))^{-1}\left(h_1'(m(x_0))\right)^{-2}+w_2(x_0)(\mathfrak{I}_{22}(x_0))^{-1}\left(h_2'(c(x_0))\right)^{-2}}{g_X(x_0)\left(w_1(x_0)\left(\frac{\eta_1^{(\mathfrak{p}+1)}(x_0)}{h_1'(m(x_0))}\right)^2+w_2(x_0)\left(\frac{\eta_2^{(\mathfrak{p}+1)}(x_0)}{h_2'(c(x_0))}\right)^2\right)}\right)^{\frac{1}{2\mathfrak{p}+3}}n^{-\frac{1}{2\mathfrak{p}+3}},\\ &h_{p_L,p_R;\mathrm{opt}}^{\mathrm{AMISE}}\\ &\overset{\text{linear}}{=} \mathfrak{c}_{\mathfrak{p}}\left(\frac{\int\frac{w_1(x_0)\left(\mathfrak{I}_{11}(x)\right)^{-1}\left(h_1'(m(x))\right)^{-2}+w_2(x_0)\left(\mathfrak{I}_{22}(x)\right)^{-1}\left(h_2'(c(x))\right)^{-2}}{g_X(x)}dx}{\int\left(w_1(x_0)\left(\frac{\eta_1^{(\mathfrak{p}+1)}(x)}{h_1'(m(x))}\right)^2+w_2(x_0)\left(\frac{\eta_2^{(\mathfrak{p}+1)}(x)}{h_2'(c(x))}\right)^2\right)dx}\right)^{\frac{1}{2\mathfrak{p}+3}}n^{-\frac{1}{2\mathfrak{p}+3}},\\ &\mathfrak{c}_{\mathfrak{p}}\overset{\text{linear}}{=}\left(\frac{((\mathfrak{p}+1)!)^2\xi_{0,\mathfrak{p}}}{(2\mathfrak{p}+1)\zeta_{0,\mathfrak{p}+1,\mathfrak{p}}}\right)^{\frac{1}{2\mathfrak{p}+3}}.\end{aligned}$$

By substituting the optimal bandwidth in the previous AMSE expression, we can see that the order of the AMSE and the AMISE is $O(n^{-(2\mathfrak{p}+2)/(2\mathfrak{p}+3)})$.

The circular case is, in general, more complicated as deriving the expression of the optimal concentration parameter is not straightforward. One important exception is the kernels satisfying Condition (D9) when $\mathfrak{p}=1$. Using that $\zeta_{0,2,1}=\mu_2$ and $\xi_{0,1}=\tau_0$, we obtain that the optimal smoothing parameters can be derived from

$$\mu^{\text{AMSE}}_{2;p_L,p_R;\text{opt}}(\theta_0)$$
$$\underset{=}{\text{circular}} \left(\frac{w_1(\theta_0)(\mathfrak{I}_{11}(\theta_0))^{-1}\left(h_1'(m(\theta_0))\right)^{-2} + w_2(\theta_0)(\mathfrak{I}_{22}(\theta_0))^{-1}\left(h_2'(c(\theta_0))\right)^{-2}}{2\sqrt{\pi}g_\Theta(\theta_0)\left(w_1(\theta_0)\left(\frac{\eta_1''(\theta_0)}{h_1'(m(\theta_0))}\right)^2 + w_2(\theta_0)\left(\frac{\eta_2''(\theta_0)}{h_2'(c(\theta_0))}\right)^2\right)} \right)^{2/5} n^{-2/5}.$$

$$\mu^{\text{AMISE}}_{2;p_L,p_R;\text{opt}}$$
$$\underset{=}{\text{circular}} \left(\frac{\int \frac{w_1(\theta)\left(\mathfrak{I}_{11}(\theta)\right)^{-1}\left(h_1'(m(\theta))\right)^{-2} + w_2(\theta)\left(\mathfrak{I}_{22}(\theta)\right)^{-1}\left(h_2'(c(\theta))\right)^{-2}}{g_\Theta(\theta)} d\theta}{2\sqrt{\pi}\int \left(w_1(\theta)\left(\frac{\eta_1''(\theta)}{h_1'(m(\theta))}\right)^2 + w_2(\theta)\left(\frac{\eta_2''(\theta)}{h_2'(c(\theta))}\right)^2\right) d\theta} \right)^{2/5} n^{-2/5}.$$

By substituting the optimal concentration in the AMSE and AMISE expressions, we can see that their order is $O(n^{-4/5})$.

Similar steps could be followed to derive the optimal concentration parameter for some general values of $\mathfrak{p}_1 = \mathfrak{p}_2 = \mathfrak{p}$, with $\mathfrak{p} > 1$. In that case, the added difficulty is that the AMISE depends on the value of ν not only through μ_2 and τ_0, but also through other values of μ_k, τ_k and $\omega_{k,l}$.

For the case $\mathfrak{p}_1 = \mathfrak{p}_2 = 1$, if Condition (D9) is relaxed, the value of $\mu_{2;p_L,p_R;\text{opt}}$ can be obtained in a similar way if, as $n \to \infty$, $\tau_{0;n} = \mathcal{K}\mu_{2;n}^{-\gamma}$, for some known values of $\mathcal{K}$ and γ.

When the explicit expression of the optimal concentration cannot be obtained, an alternative is to use numerical methods to calculate the value of $\mu_{2;p_L,p_R;\text{opt}}$ from the given AMSE expression (or its integrated version).