Analyzing animal escape data with circular nonparametric multimodal regression

M. Alonso-Pena and R.M. Crujeiras

Version: This is the published version of the manuscript, which is available in Annals of Applied Statistics

https://doi.org/10.1214/22-AOAS1619

The Annals of Applied Statistics
2023, Vol. 17, No. 1, 130–152
https://doi.org/10.1214/22-AOAS1619

ANALYZING ANIMAL ESCAPE DATA WITH CIRCULAR NONPARAMETRIC MULTIMODAL REGRESSION

BY MARÍA ALONSO-PENA[a] AND ROSA M. CRUJEIRAS[b]

CITMAga, Universidade de Santiago de Compostela, [a]*mariaalonso.pena@usc.es,* [b]*rosa.crujeiras@usc.es*

Analyzing the escape direction of animals subject to covariates is a problem that requires statistical techniques beyond classical regression methods. Apart from the periodicity of the angle of direction, which demands the use of circular statistics, animal escape data usually call for the exploration of the preferred orientations rather than the expected orientation. In this paper we propose the use of a nonparametric method to estimate the conditional local modes of the escape directions of animals from a regression perspective. We present the estimation algorithms and study the asymptotic properties of the estimators as well as its finite sample performance through some simulation experiments. Our proposal is used to model the escape behavior of a group of larval zebrafish escaping from a robot predator. More broadly, the approach presented in this paper can be applied to many existing problems related to animal behavior or other fields.

1. Introduction. There is a broad literature in the biological field dealing with the study of orientation choices or escape behavior in animals after being startled by a stimulus. Escape responses have been studied both theoretically and through behavioral experiments for many different animal species (Domenici, Blagburn and Bacon (2011a, 2011b)). Of especial interest is the investigation of how certain covariates can influence the escape direction (Card and Dickinson (2008), Marchetti and Scapini (2003), Obleser et al. (2016), Sato, Shidara and Ogawa (2019), Scapini et al. (2002)). In particular, the motivation of our research is to study the escape responses of larval zebrafish (*Danio rerio*) when sensing the visual stimulus of a chasing predator. Although logic would suggest that the animals should turn immediately away from predators, this is not always the case, and the study of their escape directions becomes necessary to understand the behavior of the fish and validate theoretical escape strategies. In this paper we analyze a data set obtained from Nair et al. (2017a), concerning an experimental study on the escaping behavior of larval zebrafish, where the situation of a predator chasing the larvae was simulated.

1.1. *Zebrafish larve data.* In our case study, a collection of $n = 502$ larval zebrafish were kept in an aquarium which disposed of high-quality moving cameras used to collect the data regarding the movements of the fish. A robot was disguised with the body of a dead adult zebrafish and was used to imitate a real predator. This robot was kept still in the middle of the aquarium until the beginning of the experiment. At this point and with the goal of studying the escape responses of the larval zebrafish when startled by a visual stimulus, the robot was moved at a constant speed, emulating a real predator. The variable determining the direction of escape from the robot predator is the variable of interest in our study, and we aim to analyze how this escape direction changes with the angle from which the fish were approached by the predator. Thus, we analyze the zebrafish dataset from a regression perspective, where the angle of approach or direction of stimulus is the predictor variable and the escape direction is the response variable.

Received May 2021; revised January 2022.
Key words and phrases. Multimodal regression, circular data, animal escape.

FIG. 1. *Diagrams of the zebrafish experiment, where the gray shape represents the robot predator, the dark colored larvae represents the initial position of the fish and the light colored larvae represents its position after the escape response. Own elaboration using the image from Wikimedia Commons (2014). Left: Contralateral escape. Right: Ipsilateral escape.*

Figure 1 contains illustrative diagrams of the experiment showing how the variables of the study were defined, where the gray shape represents the robot predator, the dark larval zebrafish represents the original position of the fish and the light colored larvae represents the direction of the fish after initiating the escape response. The angle of stimulus, Θ, which is the angle in which the fish were approached by the robot predator, was measured (anticlockwise) with respect to the initial orientation of the zebrafish's body, as indicated in the left panel of Figure 1. The support of this variable is determined by the animal's vision field. Values of the stimulus direction close to $-\pi/4$ indicate that the robot approached the fish through the rostral (near the rostrum or front end) side of the body. On the other hand, stimulus directions close to π show that the robot approached the fish through the caudal side (near the tail). Approaching the fish through the rostral or caudal sides (front or tail) conveys that the robot appeared in the fish's peripherical vision. The direction of escape was considered as the response variable, Φ. This variable was measured (anticlockwise) with respect to the heading direction of the predator at the moment of the visual contact, as it can be seen in Figure 1. An escape response in the range $(-\pi, 0)$, such as the one depicted in the left panel of Figure 1, is considered to be a contralateral escape. On the contrary, an escape direction in the range $(0, \pi)$ is known as an ipsilateral escape, since the larvae would be escaping to the same side where its predator lies, as represented in the right panel of Figure 1. More technical details about the experiment can be found in Nair et al. (2017b), where the zebrafish dataset was initially analyzed and the authors tried to model the relationship between the escape direction and the angle of stimulus through simple linear regression. However, when trying to analyze this dataset with regular regression methods, we encounter two problems: (i) the periodicity of the variable determining the direction of escape, which is, by definition, a circular variable and (ii) the need to estimate the preferred directions (which may not be unique), conditionally on the stimulus direction, instead of the (conditional) expected or mean direction.

1.2. *Circular data.* Classical theory on circular data (observations defined on the circumference of the unit circle) has been around for decades (Fisher (1993b), Jammalamadaka and SenGupta (2001), Mardia and Jupp (2000)), but its usage in practice has been limited, maybe due to the lack of accurate records of circular random variables (circular observations). Advances on technology have made it possible to precisely register this type of data, increasing the interest of circular statistics in recent years (Ley and Verdebout (2017), Pewsey, Neuhäuser and Ruxton (2013)) from a more applied perspective. The particular nature of this kind of observations calls for the need of specific modeling and inferential methods beyond those tailored for data on the real line.

Three different regression contexts can be considered when dealing with circular variables: when the response is circular and the predictor is real-valued, when both variables are circular, as in our zebrafish experiment, and the case where the response is real-valued and the

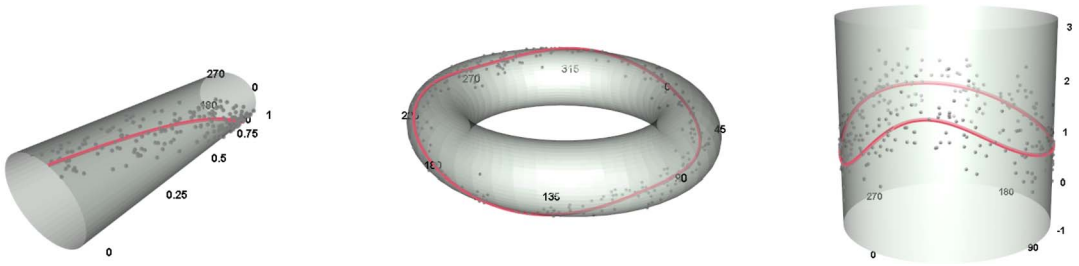


FIG. 2. *Representations on the cylinder and on the torus of simulated data from regression models with a real-valued predictor and a circular response* (*left*), *a circular predictor and a circular response* (*middle*) *and a circular predictor and a real-valued response* (*right*).

predictor is circular. The regression function can be represented on a horizontal cylinder in the first scenario, where the height of the cylinder represents the magnitude of the predictor variable and the angle represents the value of the response; when both variables are circular, the regression function can be represented on a torus and, in the real-valued response case, the regression function may be represented on the surface of a vertical oriented cylinder, as shown in Figure 2. Several proposals for parametric regression models in the three different contexts can be found in Jammalamadaka and SenGupta (2001). However, parametric models may not be sufficiently flexible to capture more complex relationships between variables, so nonparametric methods stand as a useful alternative. Di Marzio, Panzera and Taylor (2009) use a local-polynomial regression method for circular predictors and a real-valued response by including periodic kernels. For the cases where the response is a circular variable, Di Marzio, Panzera and Taylor (2013) propose a kernel-type regression estimator by smoothing the cosine and sine components of the response.

1.3. *Multimodal regression*. All the aforementioned parametric and nonparametic methods consider the conditional mean as the target function to estimate. However, as we stated in point (ii), the classical approach of regression *to the mean* might not be suitable in situations where the conditional density is multimodal. In order to explain why classical mean regression methods might fail for this type of data, let us present, as a toy example, the classical periwinkle dataset given in Fisher and Lee (1992), consisting of measurements of distance (in centimeters) and directions moved by small blue periwinkles after being relocated. The interest lies on studying the influence of the distance (real-valued predictor variable) on the escape direction (circular response variable). In the top-left panel of Figure 3, we have represented the data in a horizontal cylinder, where the height of the cylinder represents the magnitude of the distance. For clarity purposes, the top-right panel in Figure 3 shows a planar representation of the same dataset. The curve in continuous trace represents a nonparametric estimation of the regression function, regarded as the expected circular mean direction given the covariate, computed with the method of Di Marzio, Panzera and Taylor (2013) which accounts for the circular behavior of the response. We can see that, for small distances, this estimation is located in a region without data, where there seems to exist two different groups or two different preferred directions. Something similar happens in our case study with the zebrafish dataset which is represented both on the torus and on the plane in the bottom part of Figure 3. The nonparametric estimator of Di Marzio, Panzera and Taylor (2013) for the mean regression curve is also computed for this dataset (continuous trace). It is distinctly observed that the data presents a multimodal structure, and the conditional circular mean direction fails to effectively capture this feature. Thus, estimating the conditional local modes, instead of the conditional mean, would be a more adequate approach in this case, always taking into account the circular nature of the response variable.

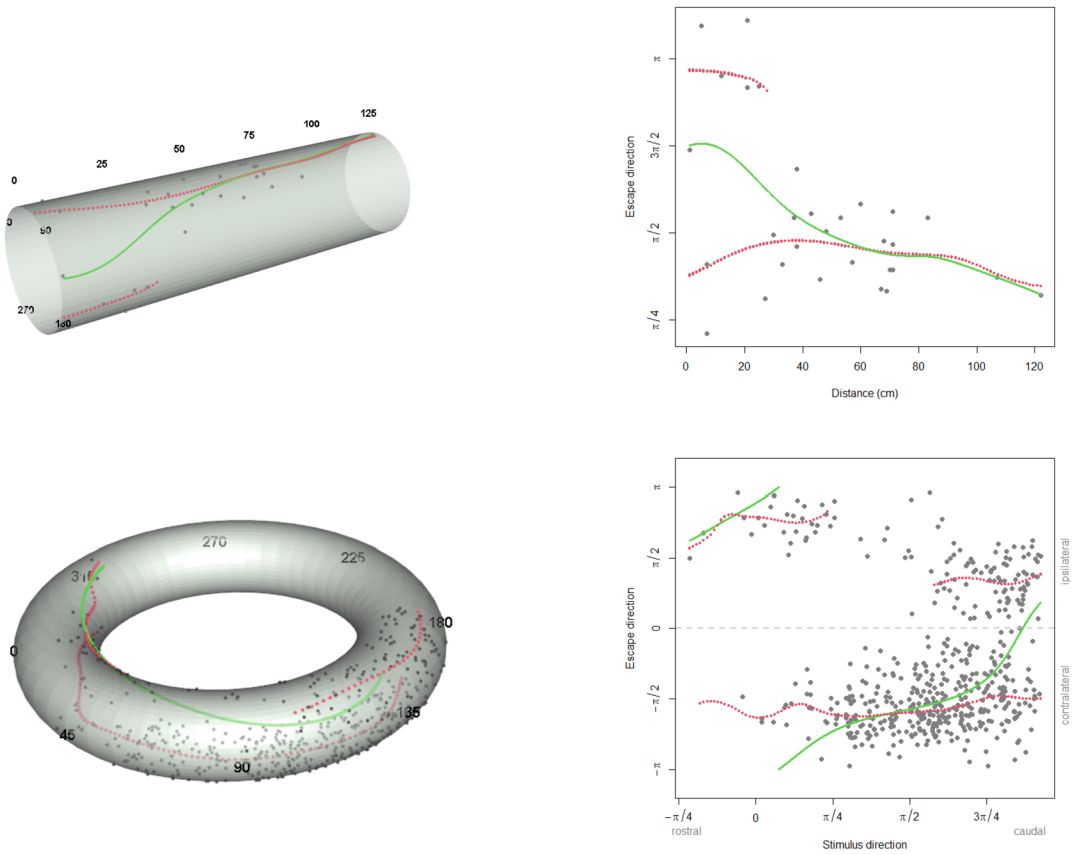


FIG. 3. *Top*: *Representations on cylinder* (*left*) *and on the plane* (*right*) *of the periwinke data with the kernel estimator of the mean regression function* (*continuous trace*) *and the proposed multimodal estimator* (*dotted trace*). *Bottom*: *Representations on the torus* (*left*) *and on the plane* (*right*) *of the zebrafish data with the estimated regression multifunction* (*dotted trace*) *and the estimated mean regression function* (*continuous trace*).

The idea of estimating the local modes of the conditional density in the Euclidean context was first introduced by Scott (1992), who obtained the conditional modes by computing a linear blend of the bivariate averaged shifted histogram estimate. Einbeck and Tutz (2006) proposed using a conditional version of the mean shift algorithm (Cheng (1995), Comaniciu and Meer (2002), Fukunaga and Hostetler (1975)) to estimate the regression multifunction based on a kernel-type conditional density estimator. The theoretical properties of this approach were studied by Chen et al. (2016) and extended to measurement error problems by Zhou and Huang (2016). A recent review of multimodal regression for Euclidean variables can be found in Chen (2018). The mean shift algorithm was extended to directional variables by Oba, Kato and Ishii (2005) in the context of nonparametric modal clustering, and it was also studied by Kobayashi and Otsu (2010) in the same setting. The convergence of the algorithm and some other theoretical results were recently obtained by Zhang and Chen (2021a), and the extension to directional product spaces was studied by Zhang and Chen (2021b). However, modal regression for circular variables, where the circular nature may characterize the response, the covariate or both, has not been addressed in the literature.

1.4. *Outline of the paper.* The goal of this manuscript is to introduce statistical tools to analyze the zebrafish dataset by taking into account the multimodal structure of the data. In addition, the methodology proposed in this work allows to estimate modal trends in regression contexts not only when both variables are circular as in the zebrafish experiment but

also when one of the variables (covariate or response) is real-valued. The potential of our approach is highlighted in Figure 3, where nonparametric multimodal regression estimators are represented both for the periwinkle example and for the zebrafish dataset. Clearly, the new approach allows to identify different trends in the data which are not captured by more classical circular regression methods.

The organization of the manuscript is as follows: Section 2 presents the regression scenarios and details the algorithms that allow to estimate the circular regression multifunctions. Section 3 deals with the error measurement in the circular multimodal regression context, while the problem of bandwidth selection is discussed in Section 4. The finite sample performance of the estimators is studied in Section 5 through simulation experiments, and Section 6 contains a more complete analysis of the zebrafish data with the proposed methodology. Some final comments are given in Section 7.

2. On the use of mean shift for circular modal regression. Nonparametric multimodal regression aims for estimating the conditional density of a response over a covariate with a flexible nonparametric estimator and, afterward, compute the conditional local modes with the so-called mean shift algorithm (Cheng (1995), Comaniciu and Meer (2002), Fukunaga and Hostetler (1975)). In this section we will detail the algorithms designed for estimating the multimodal regression multifunctions in the different contexts involving circular variables depicted in Figure 2. Although we aim to use this methodology to study the zebrafish escape behavior (and, more generally, to animal escape problems with a circular response variable), we start by contemplating the case of a real-valued response and a circular predictor, since the methodology in this case is closer to the real-valued scenario (Chen et al. (2016), Einbeck and Tutz (2006)) and is a good starting point for understanding our proposal. In addition, multimodal regression for a circular covariate and a real-valued predictor has many potential applications in the animal behavior field, for example, when studying the size of a school of fish as a function of a periodic variable such as the flood/ebb index (Williamson et al. (1962)).

Consequently, we will distinguish the case where the response is a real-valued variable in Section 2.1, where we make use of the usual conditional mean shift algorithm with suitably adapted weights and the cases where the response is circular in Section 2.2, where the circular conditional mean shift is proposed and applied to the regression context. Throughout the manuscript the circumference will be represented by $\mathbb{T} = (-\pi, \pi]$, and all of the arithmetic in the circumference will be performed modulo 2π; that is, for $\theta_1, \theta_2 \in \mathbb{T}$, we will write $\theta_1 + \theta_2$ instead of $(\theta_1 + \theta_2) (\mathrm{mod}\, 2\pi)$.

2.1. *Real-valued response.* Consider a circular predictor, Θ, with support on the circumference of the unit circle, $\mathbb{T}$ and a real-valued response, denoted by Y, with support on $\Omega \subset \mathbb{R}$. Let $\{(\Theta_j, Y_j)\}_{j=1}^n$ be a bivariate sample from (Θ, Y) with conditional density function $f(y|\theta)$ satisfying assumption (A1) in Supplementary Material 1 (Alonso-Pena and Crujeiras (2023a)). In order to model the relationship between the predictor and the response, we consider the modal regression multifunction (Einbeck and Tutz (2006)) which, for each $\theta \in \mathbb{T}$, is defined as the set of local modes (or local maxima) of the conditional density function,

$$M(\theta) = \left\{ y \in \Omega : \frac{\partial}{\partial y} f(y|\theta) = 0, \frac{\partial^2}{\partial y^2} f(y|\theta) < 0 \right\}. \tag{1}$$

The estimation of the set $M(\theta)$ in (1) is carried out through an indirect approach: first, the conditional density is estimated, and next, the conditional local modes are computed. For

the estimation of the conditional density of the response over the covariate, we will use a kernel-type estimator,

$$\hat{f}(y|\theta) = \frac{\sum_{j=1}^{n} K_\kappa(\theta - \Theta_j) L_h(y - Y_j)}{\sum_{j=1}^{n} K_\kappa(\theta - \Theta_j)}, \tag{2}$$

where $L_h(\cdot) = \frac{1}{h} L(\cdot/h)$ is a *linear* kernel with bandwidth h, that is, a symmetric around zero density where $L(\cdot)$ satisfies condition (A2) in Supplementary Material 1 (Alonso-Pena and Crujeiras (2023a)), and $K_\kappa(\cdot)$ is a circular kernel with concentration κ,

$$K_\kappa(\cdot) = c_\kappa K\big[\kappa\big(1 - \cos(\cdot)\big)\big], \tag{3}$$

where c_κ is a normalizing constant depending on κ and $K(\cdot)$ is a function satisfying assumption (A4-1) in Supplementary Material 1 (Alonso-Pena and Crujeiras (2023a)). This matches the definition of directional kernel given in Bai, Rao and Zhao (1988) for the particular case of circular data. It must be noted that κ plays a role opposite to h in the smoothing of the kernel density estimator: if κ is small, an oversmoothed estimator is obtained, and if the value of κ is large, the kernel density estimator will be undersmooothed. A circular kernel usually employed, in practice, is the von Mises density $f(\theta; \mu, \kappa) = \exp\{\kappa \cos(\theta - \mu)\}/I_0(\kappa)$, where $I_0(\kappa)$ denotes the modified Bessel function of the first kind and order zero. This is a unimodal circular density, centered at μ with concentration κ. Note that the numerator in (2) is proportional to the product kernel estimator of the joint density in the cylinder (García-Portugués, Crujeiras and González-Manteiga (2013)) and the denominator is proportional to the circular kernel density estimator (Fisher (1989)). Thus, the kernel density estimator of the conditional density depends on two smoothing parameters, κ and h which, as we will comment later, play a different role in the estimation of the regression multifunction.

In order to estimate the regression multifunction, we plug the expression in (2) into (1), obtaining

$$\widehat{M}(\theta) = \left\{ y \in \Omega : \frac{\partial}{\partial y} \hat{f}(y|\theta) = 0, \frac{\partial^2}{\partial y^2} \hat{f}(y|\theta) < 0 \right\}. \tag{4}$$

The computation of the conditional local modes is not straightforward, and thus, the conditional mean shift algorithm is used. We will assume $L_h(\cdot)$ is a radially symmetric kernel with profile l so that $L(\cdot) = c_L \times l\{(\cdot)^2\}$, with $c_L \in \mathbb{R}^+$. A local maximum of $\hat{f}(y|\theta)$ must satisfy

$$\frac{\partial}{\partial y} \hat{f}(y|\theta) = \frac{\frac{2c_L}{h^3} \sum_{j=1}^{n} K_\kappa(\theta - \Theta_j) l'\{\big(\frac{y - Y_j}{h}\big)^2\}(y - Y_j)}{\sum_{j=1}^{n} K_\kappa(\theta - \Theta_j)} = 0.$$

By taking $g(\cdot) = -l(\cdot)'$, we have

$$\frac{\partial}{\partial y} \hat{f}(y|\theta) = \frac{-2c_L}{nh^3 \hat{f}(\theta)} \sum_{j=1}^{n} K_\kappa(\theta - \Theta_j) g\left\{\left(\frac{y - Y_j}{h}\right)^2\right\}(y - Y_j) = 0, \tag{5}$$

where $\hat{f}(\theta) = \sum_{j=1}^{n} K_\kappa(\theta - \Theta_j)/n$. Note that the last factor in the sum of the first equality in (5), $(y - Y_j)$, is the shift of each datum Y_j to the point y. Therefore, the partial derivative in (5) is a weighted sum of the shifts from the observations to the point y. Now, by assuming that $G(\cdot)$ is a function proportional to $g\{(\cdot)^2\}$, we have

$$y = \frac{\sum_{j=1}^{n} K_\kappa(\theta - \Theta_j) G\big(\frac{y - Y_j}{h}\big) Y_j}{\sum_{j=1}^{n} K_\kappa(\theta - \Theta_j) G\big(\frac{y - Y_j}{h}\big)}.$$

Algorithm 1: Conditional mean shift

Data: Sample $\{(\Theta_j, Y_j)\}_{j=1}^n$, smoothing parameters κ and h.

1. Initialize mesh points $\mathcal{T} \subset \mathbb{T}$.
2. For each $\theta \in \mathcal{T}$, select starting points $y_0^{(1)}(\theta), \ldots, y_0^{(p)}(\theta)$.
3. For each $\theta \in \mathcal{T}$ and for $k = 1, \ldots, p$ iterate until convergence:

$$y_{l+1}^{(k)} = \frac{\sum_{j=1}^n K_\kappa(\theta - \Theta_j) G\left(\frac{y_l^{(k)} - Y_j}{h}\right) Y_j}{\sum_{i=1}^n K_\kappa(\theta - \Theta_j) G\left(\frac{y_l^{(k)} - Y_j}{h}\right)} \quad \text{with } l = 0, 1, \ldots.$$

However, this equation cannot be solved for y since it appears in both equation terms, that is, we have a fixed-point equation $y = \omega(y)$, where

$$\omega(\cdot) = \frac{\sum_{j=1}^n K_\kappa(\theta - \Theta_j) G\left(\frac{\cdot - Y_j}{h}\right) Y_j}{\sum_{j=1}^n K_\kappa(\theta - \Theta_j) G\left(\frac{\cdot - Y_j}{h}\right)}.$$

Note that, when $L(\cdot)$ is a Gaussian kernel, $G(\cdot)$ is also a Gaussian kernel and that the function $\omega(y)$ is actually a weighted mean of the observations with weights depending on the proximity of the data to the point (θ, y). In order to obtain the solutions of the equation $y - \omega(y) = 0$, we apply the conditional mean shift algorithm. The unconditional version of the mean shift for Euclidean variables was proposed by Fukunaga and Hostetler (1975) in the context of gradient density estimation and subsequently studied by Cheng (1995), Comaniciu and Meer (2002), Li, Hu and Wu (2007). The convergence of the algorithm was also investigated by Aliyari Ghassabeh, Linder and Takahara (2015), Aliyari Ghassabeh (2013, 2015). The algorithm was generalized to a conditional version by Einbeck and Tutz (2006) in the regression setting, and Chen, Genovese and and Wasserman (2014) considered general weights in the context of mode and ridge estimation. Now, we apply the conditional mean shift where the weights are given by circular kernels. For each value of θ, we consider the so-called mean shift function, defined as

$$m(y) = \omega(y) - y, \tag{6}$$

which represents the shift we get when moving from y to the weighted mean of the observations with weights depending on y, where the function is evaluated. The key of the mean shift algorithm is that, for a local maximum of the conditional density, the mean shift function takes the value zero. The conditional version of the mean shift algorithm with weights given by circular kernels is described in Algorithm 1.

It can be easily seen that the conditional mean shift is actually a gradient ascent method on $\hat{f}(y|\theta)$ for a fixed θ, where the step size is implicitly chosen, since $m(y)$ in (6) is proportional to and has the same direction as $\frac{\partial}{\partial y}\hat{f}(y|\theta)$. The convergence of the algorithm is guaranteed by the convergence of the unconditional version of the algorithm and, following Comaniciu and Meer (2002), once $y_{l+1}^{(k)}$ gets sufficiently close to a local mode of the conditional density, it will converge to it. Note that the number of starting points in Algorithm 1, namely p, may be different for each value of θ. Einbeck and Tutz (2006) use a fixed set of starting points for all the values of the predictor variable and Chen et al. (2016) take the whole sample as the starting points, as it is usually done in nonparametric modal clustering. However, in order to initialize the algorithm at regions close to the data, we recommend a *local initialization* where, for each value of θ, the starting values are the quartiles of the sample responses closer to θ.

2.2. *Circular response.* Consider now a circular response variable, Φ, with support on $\mathbb{T}$ and either a real-valued predictor X with support on $\Omega \subset \mathbb{R}$ or a circular predictor Θ with support on $\mathbb{T}$. We will use Δ to denote a generic predictor, with support either on the real line or on the unit circumference. Let $\{(\Delta_j, \Phi_j)\}_{j=1}^n$ be a random sample from (Δ, Φ), with conditional density function $f(\phi|\delta)$ satisfying assumption (A3) in Supplementary Material 1 (Alonso-Pena and Crujeiras (2023a)). Similarly to Section 2.1, for each value of the predictor variable, δ, the modal regression multifunction is defined as

$$M(\delta) = \left\{\phi : \frac{\partial}{\partial \phi} f(\phi|\delta) = 0, \frac{\partial^2}{\partial \phi^2} f(\phi|\delta) < 0\right\}, \tag{7}$$

with $f(\phi|\delta)$ being the conditional density of Φ given the value of Δ. Note that an interesting feature of multimodal regression for circular responses is that the parameter of interest is analogous to the Euclidean setting and it is uniquely defined. This is not the case when considering mean regression methods, since not only can one define the mean direction through the circular mean (see Fisher (1993b)) but also through the Fréchet mean, and they are not always defined, making the problem of circular mean regression not easy to handle in some situations. Conversely, the local modes are always well defined as long as the conditional density is not uniform.

In order to estimate the multifunction in (7), we take the same indirect approach as in the previous section: estimate the conditional density, and then compute its local maxima for each value of the predictor. We consider the kernel density estimator of the conditional density for a circular response (Di Marzio et al. (2016)). If the predictor variable is real-valued ($\Delta = X$), the estimator is given by

$$\hat{f}(\phi|x) = \frac{\sum_{j=1}^n L_h(x - X_j) K_\kappa(\phi - \Phi_j)}{\sum_{j=1}^n L_h(x - X_j)}.$$

In this case, $L_h(\cdot)$ is a *linear* kernel with smoothing parameter h, and $K_\kappa(\cdot)$ is a circular kernel with concentration parameter κ. If the predictor is circular ($\Delta = \Phi$), the conditional density is estimated as

$$\hat{f}(\phi|\theta) = \frac{\sum_{j=1}^n K_\nu(\theta - \Theta_j) K_\kappa(\phi - \Phi_j)}{\sum_{j=1}^n K_\nu(\theta - \Theta_j)},$$

where the expression in the numerator is proportional to an estimation of the joint density with a product of circular kernels (Di Marzio, Panzera and Taylor (2011)). Here, the kernel associated to the predictor has concentration parameter ν, and the kernel associated to Φ has concentration κ. From now on, we will denote the weights corresponding to the predictor variable (X or Θ) at the point δ as $w_\delta(\Delta_j)$, $j = 1, \ldots, n$. Note that such weights depend on the smoothing parameter h or ν (depending on Δ being circular or real-valued). Thus,

$$\hat{f}(\phi|\delta) = \frac{1}{n\hat{f}(\delta)} \sum_{j=1}^n w_\delta(\Delta_j) K_\kappa(\phi - \Phi_j), \quad \hat{f}(\delta) = \frac{1}{n} \sum_{j=1}^n w_\delta(\Delta_j).$$

Consequently, the estimator of the multimodal regression function (7) is given by

$$\widehat{M}(\delta) = \left\{\phi : \frac{\partial}{\partial \phi} \hat{f}(\phi|\delta) = 0, \frac{\partial^2}{\partial \phi^2} \hat{f}(\phi|\delta) < 0\right\}. \tag{8}$$

To obtain the local maxima of $\hat{f}(\delta|\phi)$, we establish the necessary condition for a critical point: $\frac{\partial}{\partial \phi} \hat{f}(\phi|\delta) = 0$. By applying (3), it holds

$$\frac{\partial}{\partial \phi} \hat{f}(\phi|\delta) = \frac{\kappa c_\kappa}{n\hat{f}(\delta)} \sum_{j=1}^n w_\delta(\Delta_j) K'\big[\kappa\big(1 - \cos(\phi - \Phi_j)\big)\big] \sin(\phi - \Phi_j). \tag{9}$$

Consequently, the derivative of the estimated conditional density with respect to ϕ is a weighted sum of the sines of the differences from each observation to the point ϕ. In this case the sine function is used for measuring the shifts from the observations to ϕ. This is very intuitive since, if $\phi = \Phi_j$, then $\sin(\phi - \Phi_j) = 0$. In addition, if the difference $\phi - \Phi_j$ is very small, then $\sin(\phi - \Phi_j) \approx \phi - \Phi_j$. Therefore, the right side of equation (9) can be interpreted as a weighted sum of the shifts from the observations to ϕ. By expanding the last factor in the right side of (9), we get

$$\frac{\partial}{\partial \phi}\hat{f}(\phi|\delta) = \frac{\kappa c_\kappa}{n\hat{f}(\delta)}\sum_{j=1}^{n} w_\delta(\Delta_j)K'\big[\kappa\big(1-\cos(\phi-\Phi_j)\big)\big](\sin\phi\cos\Phi_j - \cos\phi\sin\Phi_j),$$

and by equating it to zero, we have

$$\sin\phi\sum_{j=1}^{n} w_\delta(\Delta_j)T(\phi-\Phi_j)\cos\Phi_j = \cos\phi\sum_{j=1}^{n} w_\delta(\Delta_j)T(\phi-\Phi_j)\sin\Phi_j,$$

where $T(\cdot)$ is proportional to $K'[\kappa(1-\cos(\cdot))]$. Therefore, if we denote

$$S_\delta(\phi) = \sum_{j=1}^{n} w_\delta(\Delta_j)T(\phi-\Phi_j)\sin\Phi_j \quad \text{and} \quad C_\delta(\phi) = \sum_{j=1}^{n} w_\delta(\Delta_j)T(\phi-\Phi_j)\cos\Phi_j,$$

we have that if $S_\delta(\phi) \neq 0$ or $C_\delta(\phi) \neq 0$, then $\phi = \operatorname{atan2}(S_\delta(\phi), C_\delta(\phi))$, where the operator $\operatorname{atan2}(a,b)$ returns the angle between the x-axis and the vector from the origin to (b,a) (see Jammalamadaka and SenGupta (2001), Chapter 1). Thus, we obtain that

$$\phi = \tilde{\omega}(\phi) = \operatorname{atan2}\big(S_\delta(\phi), C_\delta(\phi)\big). \tag{10}$$

Note that the function $\tilde{\omega}(\phi)$ actually returns a weighted circular mean direction of the observations, since the quantity $S_\delta(\phi)$ is a weighted sum of $\sin\Phi_j$ and $C_\delta(\phi)$ is a weighted sum of $\cos\Phi_j$ (where the weights depend on the point (δ,ϕ)). Since in the previous expression we have a fixed-point equation, we use a mean shift-type algorithm in order to obtain the conditional mode estimator. As an analogy with the mean shift function in equation (6), we define the circular mean shift function as

$$\tilde{m}(\phi) = \sin\big(\tilde{\omega}(\phi) - \phi\big).$$

As discussed before, given that for small values of $\tilde{\omega}(\phi)-\phi$ we have $\sin(\tilde{\omega}(\phi)-\phi) \approx \tilde{\omega}(\phi)-\phi$, the sine function is used to measure the shift from ϕ to $\tilde{\omega}(\phi)$. In addition, for a local mode of the conditional density estimator, the circular mean shift function takes the value zero. Consequently, the estimated regression multifunction (8) is obtained by using the conditional circular mean shift algorithm, which is described in Algorithm 2. As in Algorithm 1, the number of starting points, p, can be different for each value of δ, and we recommend the same initialization, as in in Algorithm 1, but using circular quartiles (see Fisher (1993b)).

REMARK. The unconditional version of Algorithm 2 actually coincides with the directional mean shift proposed by Oba, Kato and Ishii (2005) and studied by Kobayashi and Otsu (2010) and Zhang and Chen (2021a) in the particular case of data in the unit circumference. Note that, however, the derivations of the algorithms are different. The results in Zhang and Chen (2021a) guarantee the convergence of Algorithm 2 and the fact that the algorithm converges to a mode of the estimated conditional density.

Algorithm 2: Circular conditional mean shift

Data: Sample $\{(\Delta_j, \Phi_j\}_{j=1}^n$, smoothing parameters κ and h/ν.

1. Initialize mesh points $\mathcal{S} \subset \Omega$ if $\Delta = X$ or $\mathcal{T} \subset \mathbb{T}$ if $\Delta = \Theta$.
2. For each $\delta \in \mathcal{S}$ (or $\delta \in \mathcal{T}$), select starting points $\phi_0^{(1)}(\delta), \ldots, \phi_0^{(p)}(\delta)$.
3. For $k = 1, \ldots, p$ iterate until convergence:

$$\phi_{l+1}^{(k)} = \operatorname{atan2}\left(\sum_{j=1}^{n} w_\delta(\Delta_j) T\big(\phi_l^{(k)} - \Phi_j\big) \sin \Phi_j, \sum_{j=1}^{n} w_\delta(\Delta_j) T\big(\phi_l^{(k)} - \Phi_j\big) \cos \Phi_j\right),$$

with $l = 0, 1, \ldots$

3. Error measurement. The aim of this section is to introduce a way to measure the quality of the estimators described in Section 2. Note that usual error metrics in kernel regression (such as the mean integrated squared error or the integrated squared error) are not adequate to measure the quality of the estimators since, in the multimodal regression context, our objective function is actually a multifunction, that is, for each value of the predictor, we seek for a set of values of the response. Following Chen et al. (2016), we consider pointwise and mean integrated errors based on a distance between sets.

We first introduce the Hausdorff distance which, for two sets $A, B \subset \mathbb{R}^q$ is defined as

$$\operatorname{Haus}(A, B) = \max\left\{\sup_{x \in A} d(x, B), \sup_{x \in B} d(x, A)\right\}, \tag{11}$$

where, $d(x, A) = \inf_{z \in A} \|x - z\|$. The Hausdorff distance measures how close two Euclidean sets are from each other. In the real-valued response case exposed in Section 2.1, the Hausdorff distance is an adequate measure of the distance from the true regression multifunction to the estimated multifunction, because $M(\theta)$ in equation (1) and its estimated version $\widehat{M}(\theta)$ are subsets of $\mathbb{R}$. However, in the circular response scenario, our multifunctions $M(\delta)$, $\widehat{M}(\delta)$ are subsets of $\mathbb{T}$, and thus, it is necessary to generalize the definition of the Hausdorff distance by considering, for $A, B \subset \mathbb{T}$,

$$\widetilde{\operatorname{Haus}}(A, B) = \max\left\{\sup_{x \in A} \tilde{d}(x, B), \sup_{x \in B} \tilde{d}(x, A)\right\}, \tag{12}$$

with $\tilde{d}(x, A) = \inf_{z \in A} 1 - \cos(x - z)$. The Hausdorff distances defined in (11) and (12) are used to construct the pointwise error metrics, defined as

$$\Lambda(\theta) = \operatorname{Haus}\big(M(\theta), \widehat{M}(\theta)\big), \qquad \tilde{\Lambda}(\delta) = \widetilde{\operatorname{Haus}}\big(M(\delta), \widehat{M}(\delta)\big), \tag{13}$$

where the first pointwise error is applied to the multifunctions with a circular predictor and real-valued response and the second is used for multifunctions with a circular response. The pointwise errors measure how close the estimated multifunction is from the true regression multivalued function for each value of the predictor, θ or δ, respectively. In order to obtain a global error measure, we consider the modal mean integrated squared error (MISE_m), as in Chen et al. (2016), and the modal circular mean integrated error (CMIE_m), defined as

$$\text{MISE}_m(\widehat{M}) = \mathbb{E}\left[\int_{\theta \in \mathbb{T}} \Lambda^2(\theta)\, d\theta\right], \qquad \text{CMIE}_m(\widehat{M}) = \mathbb{E}\left[\int_{\delta \in \text{Supp}(\Delta)} \tilde{\Lambda}(\delta)\, d\delta\right],$$

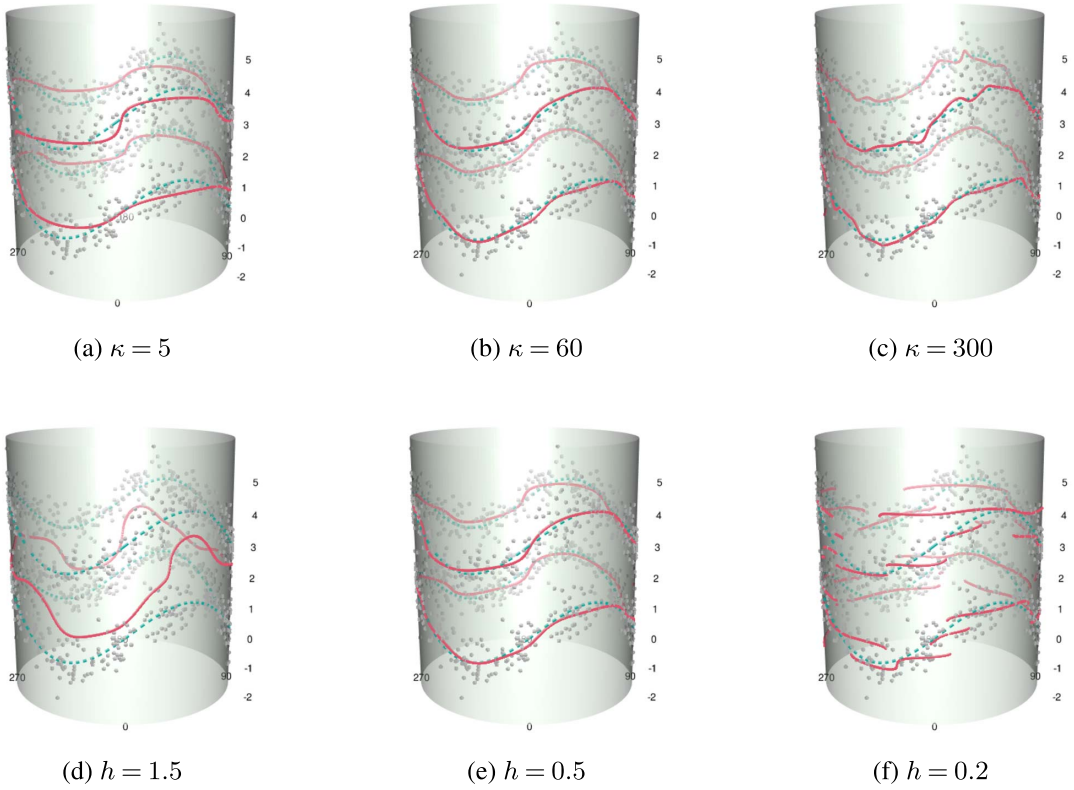
where $\text{Supp}(\Delta)$ denotes the support of Δ which, recall, can denote either a circular or real-valued predictor. The MISE_m is obtained by taking the expectation of the integrated squared pointwise error and is a global error measure for the multimodal regression estimator in (4). The CMIE_m, constructed to measure the global error of the estimator in (8), is the expectation

of the integrated pointwise error. The reason why the circular error is not squared is that it is constructed as an analogy with the Euclidean case: for real-valued variables, the mean squared error of an estimator is the expectation of the squared distance between the estimator and the parameter while, for circular variables, the so-called *circular mean squared error*, defined by Kim and SenGupta (2017), is the expectation of the cosine distance between the estimator and the parameter.

Theoretical results showing the asymptotic error rates of the estimators in terms of the pointwise and global errors are given in Supplementary Material 1 (Alonso-Pena and Crujeiras (2023a)) along with outlines of the proofs. Roughly speaking, the derived rates show that the pointwise errors $\Delta(\theta)$, $\tilde{\Delta}(x)$ and $\tilde{\Delta}(\theta)$ converge to zero at the same rate as the first partial derivative, with respect to the response variable, of the joint kernel density estimators, namely, $\frac{\partial}{\partial y}\hat{f}(\theta, y)$, $\frac{\partial}{\partial \phi}\hat{f}(x, \phi)$ and $\frac{\partial}{\partial \phi}\hat{f}(\theta, \phi)$. Note that all the obtained rates are the same as in the Euclidean case if one thinks of $\kappa \asymp h^{-2}$ and $\nu \asymp h^{-2}$.

4. Bandwidth selection. As it usually happens in kernel estimation, the selection of the smoothing parameters is a crucial issue in multimodal regression. In classical kernel methods for regression in the real line (Fan and Gijbels (1996)), a large value of the bandwidth h leads to an oversmoothed estimator, while a small value of h produces an undersmoothed estimation of the regression function. Similarly to the context of circular kernel density estimation exposed in Section 2.1, the behavior of the concentration parameter in circular kernel regression (Di Marzio, Panzera and Taylor (2009, 2013)) is reversed: when κ is large, an undersmoothed estimation of the regression function is obtained, and a small value of κ leads to an oversmoothed estimator. However, in the context of multimodal regression two smoothing parameters are required, one associated to the predictor variable and one associated to the response, and the role of the two parameters is very different. The parameter associated to the predictor variable controls the smoothing of the regression multifunction, having a similar role to the smoothing parameter in classical kernel regression. On the contrary, the parameter associated to the response variable influences the number of estimated modes. The reason behind this behavior is that an undersmoothed estimator of the conditional density will lead to many local modes, producing a large number of estimated branches in the regression multifunction. We illustrate this behavior in Figure 4 with simulated data from a multimodal regression model with a circular predictor and a real-valued response (so the data cloud lies on a cylinder), although similar interpretations are obtained for the scenarios involving a circular response variable. In the top row of Figure 4, we fix the parameter h, associated to the response variable, while κ (the parameter associated to the covariate) takes different values. As the value of the concentration increases, the multimodal estimator turns more undersmoothed. On the other hand, in the bottom row the value of κ was fixed, and we change the value of h. If h is too large, the two branches of the regression multifunction are merged into one, but if h is too small, many local modes are estimated for each value of the predictor, resulting into a too fragmented estimator.

The statistical literature on the selection of the smoothing parameters for multimodal regression in the Euclidean setting is quite scarce. Since the estimation is carried out by obtaining the local maxima of the conditional density estimator, Einbeck and Tutz (2006) recommend using methods for bandwidth selection tailored for conditional density estimation. However, such methods may not be adequate in practice, because mode estimation is related but not equivalent to density estimation. As noted by Casa, Chacón and Menardi (2020), an estimated density might be close to the true density in terms of ISE but have many estimated local modes, which in the case of modal regression would lead to many estimated branches of the regression multifunction. Zhou and Huang (2019) proposed two different methods for obtaining smoothing parameters in multimodal regression for real-valued response and covariate. The first of them, called modal cross-validation, aims to balance the number of estimated



(a) $\kappa = 5$ (b) $\kappa = 60$ (c) $\kappa = 300$

(d) $h = 1.5$ (e) $h = 0.5$ (f) $h = 0.2$

FIG. 4. *Simulated data from a multimodal regression model with a circular predictor and a real-valued response. The dashed curves represent the true regression multifunctions, and the continuous trace shows the estimated multifunction. In the top row the smoothing parameter is* $h = 0.6$, *and the values of* κ *are specified below each panel. In the bottom row the value of* κ *is* 30, *and the values of h are specified below each panel.*

local modes and the distance from the estimation to the data. Although the method shows a good performance, in practice, nothing assures that minimizing the modal cross-validation function will minimize the MISE_m of the estimator or any other error criterion. The second approach in Zhou and Huang (2019) is to minimize the ISE_m of the estimator (the integrated Hausdorff distance between the estimator and the true multifunction) by using a parametric bootstrap based on a mixture of parametric regressions. Another approach was proposed by Chen et al. (2016), which consists on constructing a prediction band for the regression multifunction and, afterward, selecting the parameters minimizing a loss function defined as the volume of the prediction band. However, the authors assume that the smoothing parameter is the same for both variables, which is not well justified, especially given the different role of the two parameters. In addition, the choice of the smoothing parameters depends on the previously selected prediction level.

4.1. *Modal cross-validation for circular regression.* The modal cross-validation method in Zhou and Huang (2019) is easily adaptable to the context of circular multimodal regression. Although the theoretical properties of the modal cross-validation method are not well studied, the behavior of the method in practice is quite satisfactory. Thus, here we modify the cross-validation method to deal with circular variables. For the scenario presented in Section 2.1 (circular predictor and real-valued response), the modal CV approach consists on selecting κ

and h by minimizing

$$CV_1(\kappa, h) = \frac{1}{n}\sum_{j=1}^{n} d^2\big(\widehat{M}_{-j}^{\kappa,h}(\Theta_j), Y_j\big)N_{-j}^2(\Theta_j), \tag{14}$$

where $d(x, A) = \inf_{z\in A}\|x - z\|$, $\widehat{M}_{-j}^{\kappa,h}(\theta)$ is the estimator of $M(\theta)$ using data $\{(\Theta_i, Y_i) : i \neq j\}$ constructed with smoothing parameters κ and h and $N_{-j}(\Theta_j)$ is the number of estimated modes of $\widehat{M}_{-j}^{\kappa,h}$, at $\theta = \Theta_j$.

For the circular response case exposed in Section 2.2, the parameters ρ and κ (where ρ represents either h or ν, depending on the nature of the predictor) are selected by minimizing

$$CV_2(\rho, \kappa) = \frac{1}{n}\sum_{j=1}^{n} \tilde{d}\big(\widehat{M}_{-j}^{\rho,\kappa}(\Delta_j), \Phi_j\big)N_{-j}(\Delta_j), \tag{15}$$

where $\tilde{d}(x, A) = \inf_{z\in A} 1 - \cos(x - z)$ and $\widehat{M}_{-j}^{\rho,\kappa}(\delta)$ is the estimator of $M(\delta)$ using data $\{(\Delta_i, \Phi_i) : i \neq j\}$ and constructed with smoothing parameters ρ and κ. This approach does not rely on a theoretical basis and computational trials are needed in order to assess the efficacy of the method.

In this work we use a modification of this method in order to accomplish more computational efficiency. Since the role of the two smoothing parameters is quite different, we first select a pair of reference parameters, then update the parameter corresponding to the predictor variable by minimizing the cross-validation function and, lastly, update the second smoothing parameter. The detailed modified criteria can be found in Supplementary Material 2 (Alonso-Pena and Crujeiras (2023b)). Through simulation experiments (see Supplementary Material 2 (Alonso-Pena and Crujeiras (2023b))) we have verified that results using this modification of the modal cross-validation criteria are very similar to the ones obtained with the original procedures described in this section, but the modification is more efficient in computational terms.

5. Simulation experiments. In this section we analyze the finite sample performance of the circular multimodal regression estimators in the three different contexts (circular predictor and real-valued response, real-valued predictor and circular response and circular predictor and circular response). First, the simulation scenarios are presented, and the obtained results are subsequently shown.

5.1. *Simulation scenarios.* In our simulated examples, the sample is divided into two groups, with each group corresponding to a branch of the regression multifunction. For each observation we use the subscript ij, where i denotes the group or branch number and j denotes the observation number within each group. In addition, n_i denotes the sample size for the ith group. Note that these underlying groups are not known in practice, and there is no information about them. The simulated models are presented in Table 1. The first model in each scenario corresponds to two parallel regression curves or, seen from another perspective, to a regression curve with a bimodal error. As for the second model of each scenario, the regression curves are not parallel. The same happens with the third model, although this time the support of the first group is different (and smaller) from the support of the second group. For all scenarios and models, the sample sizes are $(n_1, n_2) \in \{(100, 100), (100, 200), (200, 200), (200, 300), (300, 300)\}$. Examples of data from all simulated models are shown in Figure 5.

In order to assess the performance of the estimators, the modal mean integrated error (MISE_m) and the circular mean integrated error (CMIE_m) are approximated by generating

TABLE 1

Simulated models. CL stands for circular predictor and real-valued response, LC stands for real-valued predictor and circular response and CC stands for circular predictor and circular response. For CL models, $\varepsilon_{1j}, \varepsilon_{2j} \sim N(0, \sigma^2)$ *with* $\sigma \in \{0.75, 1, 1.25\}$. *For LC and CC models,* $\varepsilon_{1j}, \varepsilon_{2j} \sim vM(0, \tau)$ *with* $\tau \in \{6, 8, 10\}$ *for LC and* $\tau \in \{10, 12, 14\}$ *for CC*

Model	Data generation	Regression multifunction
CL-1	$Y_{1j} = 2\sin(2\Theta_{1j}) + \varepsilon_{1j}$ $Y_{2j} = 8 + 2\sin(2\Theta_{2j}) + \varepsilon_{2j}$ $\Theta_1, \Theta_2 \sim$ Circular Uniform	$M(\theta) = \{2\sin(2\theta), 8 + 2\sin(2\theta)\}$
CL-2	$Y_{1j} = -2\cos(\Theta_{1j}) + \varepsilon_{1j}$ $Y_{2j} = 2\sin(2\Theta_{2j}) - 10 + \varepsilon_{2j}$ $\Theta_1, \Theta_2 \sim$ Circular Uniform	$M(\theta) = \{2\sin(2\theta), 8 + 2\sin(2\theta)\}$
CL-3	$Y_{1j} = 2\cos(\Theta_{1j}) + \varepsilon_{1j}$ $Y_{2j} = 8 + 2\sin(2\Theta_{2j}) + \varepsilon_{2j}$ $\Theta_1 \sim U(0, \pi), \Theta_2 \sim$ Circular Uniform	$M(\theta) = \{2\sin(2\theta)\}$ if $\theta \in [0, \pi)$, $M(\theta) = \{8 + 2\sin(2\theta)\}$ otherwise,
LC-1	$\Phi_{1j} = 6\tan^{-1}(2.5X_{1j} - 3) + \varepsilon_{1j}$ $\Phi_{2j} = \pi + 6\tan^{-1}(2.5X_{2j} - 3) + \varepsilon_{2j}$ $X_1, X_2 \sim U(0, 1)$	$M(x) = \{6\tan^{-1}(2.5x - 3),$ $\pi + 6\tan^{-1}(2.5x - 3)\}$
LC-2	$\Phi_{1j} = \operatorname{atan2}(\sin(3X_{1j}^2), \cos(3X_{1j}^2)) + \varepsilon_{1j}$ $\Phi_{2j} = \pi/2 + 2\tan^{-1}(10X_{2j} - 1/2) + \varepsilon_{2j}$ $X_1, X_2 \sim U(0, 1)$	$M(x) = \{\operatorname{atan2}(\sin(3x^2), \cos(3x^2)),$ $\pi/2 + 2\tan^{-1}(10x + 1/2)\}$
LC-3	$\Phi_{1j} = \pi/4 - 3\tan^{-1}(2.5X_{1j} - 3) + \varepsilon_{1j}$ $\Phi_{2j} = \pi/2 + 6\tan^{-1}(2.5X_{2j} - 3) + \varepsilon_{2j}$ $X_1 \sim U(0, 0.5), X_2 \sim U(0, 1)$	$M(x) = \{\pi/4 - 3\tan^{-1}(2.5x - 3), \pi/2 +$ $6\tan^{-1}(2.5x - 3)\}$ if $x \in [0, 0.5]$ $M(x) = \{\pi/2 + 6\tan^{-1}(2.5x - 3)\}$ otherwise
CC-1	$\Phi_{1j} = 2\cos\Theta_{1j} + \varepsilon_{1j}$ $\Phi_{2j} = 3\pi/4 + 2\cos\Theta_{2j} + \varepsilon_{2j}$ $\Theta_1, \Theta_2 \sim$ Circular Uniform	$M(\theta) = \{2\cos\theta, 3\pi/4 + 2\cos\theta\}$
CC-2	$\Phi_{1j} = 3/4\cos\Theta_{1j} - \pi/2 + \varepsilon_{1j}$ $\Phi_{2j} = \pi/2 - \cos\Theta_{2j} + \varepsilon_{2j}$ $\Theta_1, \Theta_2 \sim$ Circular Uniform	$M(\theta) = \{3/4\cos\theta - \pi/2, \pi/2 - 2\cos\theta\}$
CC-3	$\Phi_{1j} = 2\pi/3 + 3\sin(\Theta_{1j}/3) + \varepsilon_{1j}$ $\Phi_{2j} = 3\pi/4 + 1.75\cos\Theta_{2j} + \varepsilon_{2j}$ $\Theta_1 \sim U(2\pi/3, 3\pi/2), \Theta_2 \sim$ Circular Uniform	$M(\theta) = \{2\pi/3 + 3\sin(\theta/3), 3\pi/4 + 1.75\cos\theta\}$ if $\theta \in [2\pi/3, 3\pi/2]$ $M(\theta) = \{3\pi/4 + 1.75\cos\theta\}$ otherwise

500 Monte Carlo replicates of the simulated data and computing the average of the modal integrated error/modal circular integrated error (ISE_m or CIE_m) of the multimodal estimator,

$$\mathrm{ISE}_m(\widehat{M}) = \int_{\theta \in \mathbb{T}} \Lambda^2(\theta)\, d\theta \quad \text{for Euclidean response,}$$

$$\widetilde{\mathrm{ISE}}_m(\widehat{M}) = \int_{\delta \in \mathrm{Supp}(\Delta)} \widetilde{\Lambda}(\delta)\, d\delta \quad \text{for circular response,}$$

where the integrals are numerically approximated by Simpson's rule and $\Lambda, \widetilde{\Lambda}$ are defined in (13). In the simulation experiments the smoothing parameters were selected by the modified version of the modal cross-validation method described in Supplementary Material 2 (Alonso-Pena and Crujeiras (2023b)).

5.2. *Simulation results.* The simulation results are presented in this section. Comments about the first and second models in each scenario are given, whereas the results of the third models (CL-3, LC-3 and CC-3) are discussed later as a special case.

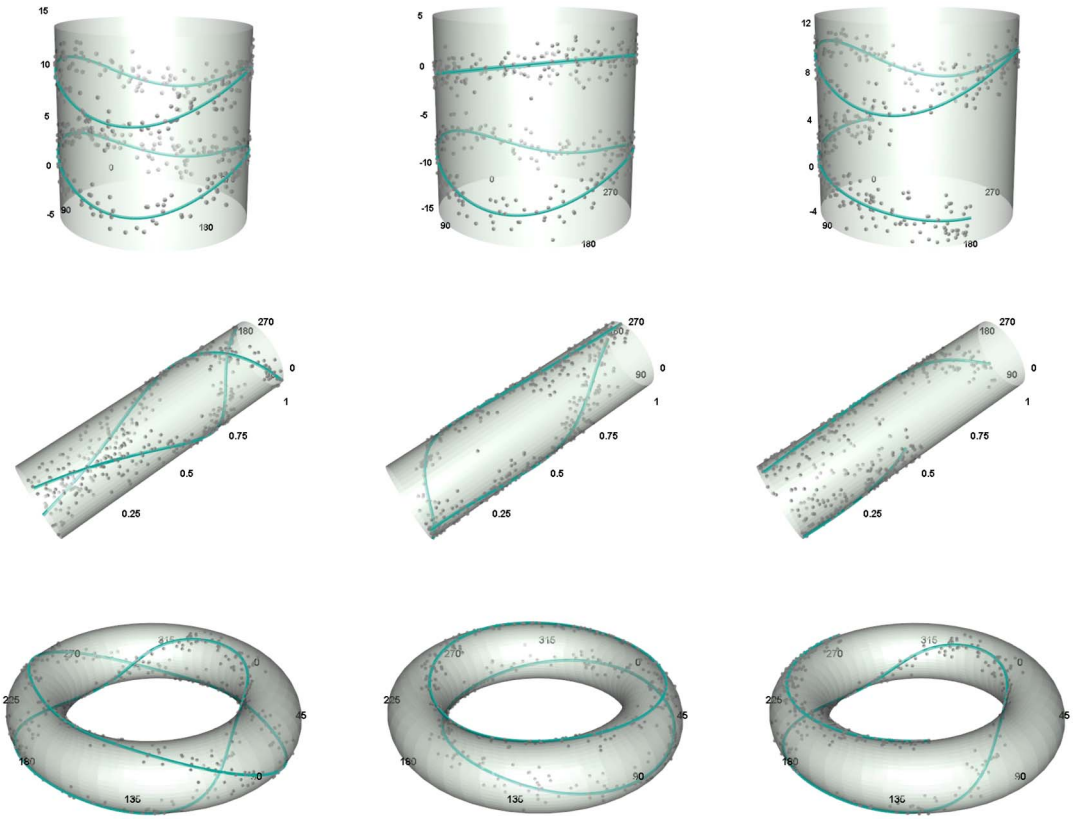


FIG. 5. *Representations on the cylinder and torus of simulated data and true regression multifunctions from models CL*-1 (*top left*), *CL*-2 (*top center*) *and CL*-3 (*top right*) *with standard deviation* $\sigma = 1$; *models LC*-1 (*middle left*), *LC*-2 (*middle center*) *and LC*-3 (*middle right*) *with concentration* $\tau = 8$; *and models CC*-1 (*bottom left*), *CC*-2 (*bottom center*) *and CC*-3 *with concentration* $\tau = 12$. *Sample sizes are* $(n_1, n_2) = (200, 200)$ *for all models*.

The estimated MISE_m for models CL-1 and CL-2 is reported in Table 2. Results are obtained with the smoothing parameters selected by the modified modal cross-validation and also with the *optimal* parameters minimizing ISE_m, which are used as benchmarks. As expected, the estimated MISE_m generally decreases both when increasing sample size and decreasing standard deviation. In addition, the performance of the modal cross-validation criterion to select the smoothing parameters is quite satisfactory, as the estimated MISE_m obtained using this method is reasonably close to the benchmark MISE_m. The larger differences are obtained for the smallest sample sizes, especially when $\sigma = 1.25$. Note that, although the estimations of MISE_m tend to decrease when increasing the sample size, a much remarkable decrease is obtained when the sample size is the same in both groups, that is, $(n_1, n_2) \in \{(100, 100), (200, 200), (300, 300)\}$. When the sample size in each group is different, we observe that the estimated MISE_m does not decrease so fast and is sometimes larger than with the previous sample size when using the modal cross-validation criterion. This behavior is due to the need of different degrees of smoothing in each branch of the multifunction when the sample sizes are highly unbalanced.

Results for models LC-1 and LC-2 are given in Table 3, while Table 4 contains the results for models CC-1 and CC-2. Results are similar to those of the real-valued response scenario: the estimated value of CMIE_m generally diminishes as the sample size is larger. The few cases where this behavior is not observed is when the sample sizes are unbalanced, that is, $(n_1, n_2) \in \{(100, 200), (200, 300)\}$. In addition, a large value of the error concentration also

TABLE 2

Monte Carlo estimates of MISE_m *for models CL-*1, *CL-*2 *and CL-*3 *with different values of the standard deviation* σ *and the sample sizes. Results under B (benchmark) correspond to the ones obtained with smoothing parameters minimizing* ISE_m, *and results under* CV_1 *refer to the ones obtained with smoothing parameters selected by modified modal cross-validation*

		$\sigma = 1.25$		$\sigma = 1$		$\sigma = 0.75$	
Model	(n_1, n_2)	B	CV_1	B	CV_1	B	CV_1
CL-1	(100, 100)	2.308	3.644	1.456	1.877	0.928	1.012
	(100, 200)	1.935	3.382	1.191	1.653	0.729	1.021
	(200, 200)	1.172	1.321	0.776	0.851	0.480	0.509
	(200, 300)	1.031	1.215	0.650	0.706	0.411	0.438
	(300, 300)	0.835	0.940	0.527	0.574	0.340	0.359
CL-2	(100, 100)	1.578	3.194	1.076	1.777	0.697	0.951
	(100, 200)	1.270	3.252	0.836	1.439	0.528	0.903
	(200, 200)	0.880	1.143	0.596	0.686	0.381	0.432
	(200, 300)	0.735	0.912	0.489	0.583	0.309	0.345
	(300, 300)	0.613	0.710	0.412	0.473	0.261	0.290
CL-3	(100, 100)	1.489	3.673	0.921	2.080	0.749	1.144
	(100, 200)	0.959	2.321	0.590	1.244	0.358	0.731
	(200, 200)	0.785	1.917	0.491	0.849	0.299	0.448
	(200, 300)	0.570	1.179	0.359	0.607	0.219	0.356
	(300, 300)	0.534	1.307	0.338	0.515	0.206	0.278

leads to a smaller estimated CMIE_m. The performance of the cross-validation criterion is also more than acceptable, as for large values of the sample size, the value of the estimated CMIE_m obtained with smoothing parameters selected by the modified modal cross-validation is very close to the values obtained with the *optimal* smoothing parameters. The poorest performance is achieved in model LC-2 when $\tau = 6$. As shown in the middle-center panel of Figure 5, when the value of the predictor variable is around 1, the two branches are very close, which makes it difficult to discern the two groups when the error concentration is small.

5.3. *The unknown boundary condition*. In this section we analyze results for models CL-3, LC-3 and CC-3. We study these models as a special case since, unlike in the other models, here the support of the first group is different from the support of the other group. In fact, the support of the first group is always contained in the support of the second group. This issue makes the analysis of the performance of the estimators rather tricky, as it is difficult to obtain the *optimal* smoothing parameters. We will illustrate this situation with model CL-3, but it also applies for the other two scenarios.

Figure 6 shows a representation on the plane of simulated data from model CL-3 with the true and estimated regression multifunctions. In the left panel the smoothing parameters were selected by modal cross-validation, while the right panel shows the estimation where the smoothing parameters are the *optimal* ones minimizing ISE_m. Focusing on the left panel, we can see that the lower branch of the estimated regression multifunction includes a part falling outside of its support (represented with triangles). This is a general and expected behavior for all kernel methods, both for density and regression. Usually, one just restricts the estimation to the support of the variable, but in multimodal regression the support of each group is unknown (recall that these groups are unknown in practice, and we have no information about them). On the early stages of modal regression, Scott (1992) had already noticed this behavior, referring to it as the *unknown boundary condition*. Thus, the yellow part in Figure 6 leads to a huge ISE_m. Since the part represented with triangles is the main source

TABLE 3

Monte Carlo estimates of CMIE_m *for models LC-1, LC-2 and LC-3 with different values of the concentration* τ *and the sample sizes. Results under B (benchmark) correspond to the ones obtained with smoothing parameters minimizing* CIE_m*, and results under* CV_2 *refer to the ones obtained with smoothing parameters selected by modified modal cross-validation*

		$\tau = 6$		$\tau = 8$		$\tau = 10$	
Model	(n_1, n_2)	B	CV_2	B	CV_2	B	CV_2
LC-1	(100, 100)	0.021	0.034	0.016	0.023	0.013	0.018
	(100, 200)	0.016	0.039	0.012	0.024	0.011	0.018
	(200, 200)	0.011	0.019	0.008	0.012	0.007	0.010
	(200, 300)	0.010	0.016	0.007	0.011	0.006	0.009
	(300, 300)	0.007	0.012	0.006	0.008	0.005	0.006
LC-2	(100, 100)	0.016	0.035	0.012	0.025	0.010	0.018
	(100, 200)	0.016	0.070	0.010	0.039	0.008	0.024
	(200, 200)	0.010	0.023	0.007	0.012	0.005	0.009
	(200, 300)	0.009	0.038	0.006	0.017	0.005	0.011
	(300, 300)	0.007	0.015	0.005	0.009	0.004	0.007
LC-3	(100, 100)	0.014	0.030	0.010	0.017	0.008	0.013
	(100, 200)	0.008	0.012	0.006	0.008	0.005	0.007
	(200, 200)	0.008	0.016	0.005	0.008	0.004	0.006
	(200, 300)	0.006	0.009	0.004	0.006	0.003	0.005
	(300, 300)	0.006	0.010	0.004	0.006	0.003	0.005

of the error in these situations, when computing the *optimal* parameters, which minimize ISE_m, we obtain very large concentrations (for a circular predictor) or very small values of the bandwidth (for a real-valued predictor), leading to a heavily undersmoothed estimator, as in the right panel of Figure 6. In order to show that the modal cross-validation criterion

TABLE 4

Monte Carlo estimates of CMIE_m *for models CC-1, CC-2 and CC-3 with different values of the concentration* τ *and the sample sizes. Results under B (benchmark) correspond to the ones obtained with smoothing parameters minimizing* CIE_m*, and results under* CV_2 *refer to the ones obtained with smoothing parameters selected by modified modal cross-validation*

		$\tau = 10$		$\tau = 12$		$\tau = 14$	
Model	(n_1, n_2)	B	CV_2	B	CV_2	B	CV_2
CC-1	(100, 100)	0.144	0.395	0.117	0.213	0.102	0.148
	(100, 200)	0.120	0.182	0.097	0.147	0.087	0.120
	(200, 200)	0.066	0.089	0.054	0.066	0.047	0.057
	(200, 300)	0.058	0.068	0.046	0.054	0.040	0.045
	(300, 300)	0.042	0.056	0.035	0.045	0.030	0.038
CC-2	(100, 100)	0.116	0.227	0.092	0.173	0.080	0.150
	(100, 200)	0.085	0.232	0.068	0.159	0.057	0.120
	(200, 200)	0.055	0.091	0.045	0.069	0.039	0.059
	(200, 300)	0.046	0.092	0.037	0.066	0.032	0.049
	(300, 300)	0.037	0.055	0.030	0.043	0.026	0.033
CC-3	(100, 100)	0.063	0.094	0.054	0.076	0.048	0.065
	(100, 200)	0.032	0.041	0.027	0.033	0.024	0.029
	(200, 200)	0.031	0.037	0.027	0.031	0.024	0.027
	(200, 300)	0.022	0.027	0.018	0.022	0.016	0.019
	(300, 300)	0.021	0.024	0.018	0.020	0.016	0.018

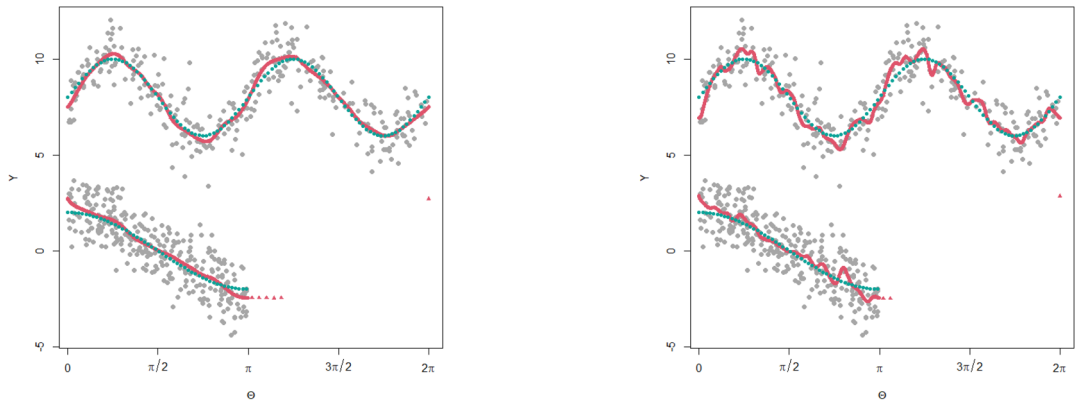


FIG. 6. *Simulated data from model CL-3 with the multimodal regression curve (dotted trace) and multimodal estimator (continuous trace and triangles) with smoothing parameters selected by cross-validation (left) and optimal smoothing parameters minimizing* ISE_m *(right).*

achieves a good performance within the support of the variables, in our simulation study we have computed the error after having removed the estimations outside the support of each group (which cannot be done with real data where this is unknown). In addition, the *optimal* parameters were also selected by removing these extra parts of the estimators (such as the triangle parts in Figure 6).

The bottom rows in Tables 2, 3 and 4 report the estimated MISE_m and CIE_m for models CL-3, LC-3 and CC-3, respectively. For model CL-3 the estimated MISE_m generally decreases by increasing the sample size and diminishing the standard deviation. For $\sigma = 1$ and $\sigma = 0.75$, when the sample size is large, the results obtained with the modal cross-validation criterion are very close to the ones obtain with the *optimal* parameters. More difficulties are encountered when $\sigma = 1.25$, as the large dispersion makes the two branches really close to each other in some areas. Unlike in models CL-1 and CL-2, a better performance is achieved when the sample sizes within groups are unbalanced, which is also expected since the support of the first group is half of the support of the second group, so unbalanced sample sizes are needed to obtain a similar density of data points in each branch. For models LC-3 and CC-3, similar results are obtained, with the estimated CMIE_m generally decreasing with the sample size, although sometimes increasing when going from unbalanced sample sizes to equal sample sizes. Again, a larger error concentration leads to a decrease of the estimated CMIE_m.

Removing the parts of the estimation outside of the support of each group is only possible in this simulation scenario where we know the information about the underlying groups. For practical purposes, a potential approach in this case would be to estimate the α-shape of the data (Edelsbrunner, Kirkpatrick and Seidel (1983), Edelsbrunner and Mucke (1984)) accounting for the cylindrical/toroidal nature of the observations and then restrict the estimation to the interior of the estimated α-shape.

6. Analysis of the escape responses in larval zebrafish. In this section we analyze the larval zebrafish data described in the Introduction. The angle of stimulus and the direction of escape (Θ, Φ) are represented in the bottom part of Figure 3, both on the surface of the torus and on the plane. While the representation on the plane seems clearer and easier to understand, it makes more difficult to visualize the periodicity of the data, and one must take into account that the top and bottom parts of the plot are actually united, as in the toroidal plot.

In order to study how the escape response Φ changes with the stimulus direction Θ, one can consider the more classical approach of estimating the conditional mean of the escape

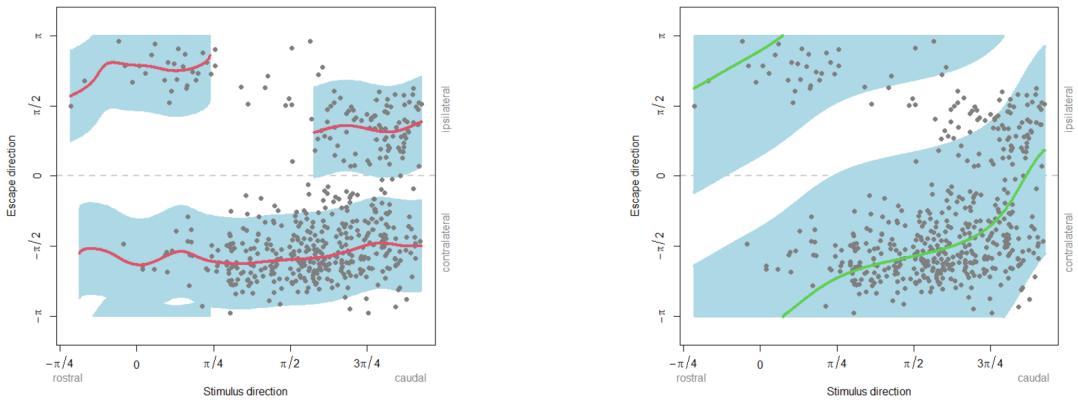


FIG. 7. *Left*: *Planar representation of the zebrafish data with the estimated regression multifunction and uniform prediction sets for a prediction level of* 0.90. *Right*: *Same representation with the nonparametric estimator of the mean regression function and uniform prediction bands with prediction level* 0.90.

direction, for example, with the kernel regression estimator for circular predictors and circular responses of Di Marzio, Panzera and Taylor (2013). This estimator was applied to the data, selecting the smoothing (concentration) parameter by cross-validation (see Oliveira, Crujeiras and Rodríguez–Casal (2014)), and the estimation is represented in the bottom part of Figure 3 with a continuous trace. According to this estimator, the mean escape direction was ipsilateral when the robot approached the rostral side of the fish, contralateral when approaching the fish's left or right side of the bodies and around zero when approaching the caudal side of the bodies. This translates into the larval zebrafish escaping, on average, in a direction toward the predator when recieving the visual stimulus peripherally around the rostrum, escaping away from the predator if the stimulus was received laterally, and escaping in the direction the predator is coming if the stimulus is caught peripherally around the tail.

As argued in the Introduction, estimating the conditional circular mean direction might not be the best approach for this kind of data, and multimodal regression might help to understand different features of the data. Thus, in order to obtain more insight into the relationship between the escape direction and the stimulus angle, we apply the multimodal estimator for circular predictors and a circular response, represented in a dotted trace in the bottom part of Figure 3. The smoothing parameters were obtained by applying the modal cross-validation criterion. Clearly from the bottom right panel of the figure, we can observe three different major trends in the data, features which the more classical regression to the mean approach fails to capture. We can observe that, when the robot appeared in the peripherical vision of the fish (both rostral and caudal sides), there were two preferred escape directions: one ipsilateral (toward the robot predator) and one contralateral (away from the threat). On the other hand, only one mode is estimated when the robot did not emerge through the peripherical vision of the fish, indicating a contralateral escape in this case.

Another advantage of the multimodal estimator is that it enables the construction of inferential tools for regression from a multimodal point of view, such as prediction sets. We briefly describe how to construct sets, which predict, for a given prediction level $1-\alpha$, the regions where new data will fall. Adapting the methodology of Chen et al. (2016) to the scenario with two circular variables, we can define a theoretical uniform prediction set as

$$\mathcal{P}_{1-\alpha} = \{(\theta, \phi) : \theta \in \mathbb{T}, \phi \in \mathcal{A}_{\varepsilon_{1-\alpha}}(\theta)\},$$

where

$$\mathcal{A}_{\varepsilon_{1-\alpha}}(\theta) = \{\phi \in \mathbb{T} : \tilde{d}(\phi, M(\theta)) \leq \varepsilon_{1-\alpha}\}$$

and

$$\varepsilon_{1-\alpha} = \inf\{\varepsilon > 0 : \mathbb{P}(\tilde{d}(\Phi, M(\Theta)) > \varepsilon) \leq \alpha\}.$$

With this definition we have $\mathbb{P}(\Phi \in \mathcal{P}_{1-\alpha}) \geq 1 - \alpha$. At sample level, following Chen et al. (2016), we can estimate

$$\widehat{\varepsilon}_{1-\alpha} = \text{Quantile}(\{\tilde{d}(\Phi_i, \widehat{M}(\Theta_i)), i = 1, \ldots, n\}, 1 - \alpha),$$
$$\widehat{\mathcal{A}}_{\widehat{\varepsilon}_{1-\alpha}}(\theta) = \{\phi \in \mathbb{T} : \tilde{d}(\phi, \widehat{M}(\theta)) \leq \widehat{\varepsilon}_{1-\alpha}\}$$

and, then,

$$\widehat{\mathcal{P}}_{1-\alpha} = \{(\theta, \phi) : \theta \in \mathbb{T}, \phi \in \widehat{\mathcal{A}}_{\varepsilon_{1-\alpha}}(\theta)\}.$$

For a prediction level of $1 - \alpha = 0.90$, we have estimated this uniform prediction set, which is represented in the left panel of Figure 7. We observe that the prediction set consists on three different prediction *bands*, each one corresponding to one of the estimated trends. The large dispersion of the data makes these prediction *bands* rather wide. The unknown boundary condition, discussed in Section 5.3, is clearly visible in the bottom-left part of the plot, where a modal estimation and a prediction set are located in a region without any data. The resulting prediction set is even more satisfactory if we compare it to the uniform prediction band, obtained with the circular mean regression estimator of Di Marzio, Panzera and Taylor (2013), which is shown in the right panel of Figure 7 for a prediction level of 0.90. The band was computed by using 1000 bootstrap resamples of the data using bootstrapping by pairs. The plot clearly shows that the uniform band constructed with the estimator of the conditional circular mean is considerably wider than the set constructed with the modal estimator.

In summary, the multimodal regression estimator allows us to estimate the two preferred directions of escape when a fish receives the visual stimulus of a predator through its peripherical vision, that is, coming toward the front or tail ends of its body. It also shows how, if the predator approaches the fish through the left or right sides of the body (where the fish's eyes are located), the direction of escape is in an opposite direction to the threat.

7. Discussion. When studying the escape response of fish after being startled by the visual stimulus of a predator, the variable determining the escape response is of circular nature. Thus, the analysis of animal escape data subject to covariates should be carried out with regression methods tailored for circular data. However, the existing literature dealing with regression involving circular observations has focused mainly on the estimation of the conditional circular mean. Animal escape data often shows a multimodal structure, as it happens with the escape responses of the zebrafish we have considered in this work, and in such cases, the estimation of the mean may yield to misleading results.

In this paper we have introduced a nonparametric method to study the local modes of the escape response conditioned on the predictor variable. Our multimodal regression method allows to estimate the most likely values of the response, given the covariate, providing a more suitable insight on the variables relationship beyond mean regression settings. The properties of our proposal have been studied both theoretically (see Supplementary Material 1 (Alonso-Pena and Crujeiras (2023a))) and through a simulation study in different scenarios. The kernel smoothing approach taken in our proposal requires the selection of a suitable smoothing parameter, a task that has been also accomplished in this work.

With the new methodology we have shown that the larval zebrafish have one preferred direction of escape when being chased from a direction lateral to their bodies. However, when the zebrafish are pursued from the caudal or rostral sides of their bodies, there are two

preferred directions of escape. In addition, the multimodal estimator allows to graphically perceive the differences between the escape behavior of the animals when being chased from the caudal and rostral sides. Our proposal also allows to construct prediction sets, which are much more informative than the ones obtained with more classical mean regression methods, as shown in Section 6.

Regarding future work, it would be interesting to consider variable bandwidths and variable concentration parameters in the conditional density estimators (see, e.g., Zelnik-Manor and Perona (2004)). This consideration could lead to a better performance of the modal estimator, although it would increase the computational cost of the method.

R functions have been programmed for the proposed methodology and are currently available from the authors upon request.

Acknowledgments. The authors thank the Editor, Associate Editor and the anonymous reviewers for their helpful comments, which considerably improved the quality of the manuscript. The authors also acknowledge the Supercomputing Center of Galicia (CESGA) for the computational resources.

Funding. This work was supported by grant PID2020-116587GB-I00 funded by MCIN/AEI/10.13039/501100011033 and the Competitive Reference Groups 2021/2024 (ED431C 2021/24) from the Xunta de Galicia through the ERDF. Research of M. Alonso-Pena was supported by the Xunta de Galicia through the grant ED481A-2019/139 from the Consellería de Educación, Universidade e Formación Profesional.

SUPPLEMENTARY MATERIAL

1. Assumptions, results and proofs (DOI: 10.1214/22-AOAS1619SUPPA; .pdf). This supplement contains the assumptions needed for the consistency of the estimators and the propositions stating such consistency, including outlines of the proofs.

2. Modified modal cross-validation (DOI: 10.1214/22-AOAS1619SUPPB; .pdf). This supplement gives the detailed algorithms containing the modification of the modal cross-validation criterion presented in Section 4, which was used in the simulations experiments in Section 5, along with the simulation results showing the performance of the modified method.

REFERENCES

ALIYARI GHASSABEH, Y. (2015). A sufficient condition for the convergence of the mean shift algorithm with Gaussian kernel. *J. Multivariate Anal.* **135** 1–10. MR3306422 https://doi.org/10.1016/j.jmva.2014.11.009

ALIYARI GHASSABEH, Y. (2013). On the convergence of the mean shift algorithm in the one-dimensional space. *Pattern Recogn. Lett.* **34** 1423–1427.

ALIYARI GHASSABEH, Y., LINDER, T. and TAKAHARA, G. (2015). On some convergence properties of the subspace constrained mean shift. *Pattern Recognit.* **46** 3140–3147.

ALONSO-PENA, M. and CRUJEIRAS, R. M. (2023a). Supplement to "Analyzing animal escape data with circular nonparametric multimodal regression: Assumptions, results and proofs." https://doi.org/10.1214/22-AOAS1619SUPPA

ALONSO-PENA, M. and CRUJEIRAS, R. M. (2023b). Supplement to "Analyzing animal escape data with circular nonparametric multimodal regression: Modifed modal cross-validation." https://doi.org/10.1214/22-AOAS1619SUPPB

BAI, Z. D., RAO, C. R. and ZHAO, L. C. (1988). Kernel estimators of density function of directional data. *J. Multivariate Anal.* **27** 24–39. MR0971170 https://doi.org/10.1016/0047-259X(88)90113-3

CARD, G. and DICKINSON, M. H. (2008). Visually mediated motor planning in the escape response of Drosophila. *Curr. Biol.* **18** 1300–1307.

CASA, A., CHACÓN, J. E. and MENARDI, G. (2020). Modal clustering asymptotics with applications to bandwidth selection. *Electron. J. Stat.* **14** 835–856. MR4062160 https://doi.org/10.1214/20-EJS1679

CHEN, Y.-C. (2018). Modal regression using kernel density estimation: A review. *Wiley Interdiscip. Rev.: Comput. Stat.* **10** e1431. MR3826098 https://doi.org/10.1002/wics.1431

CHEN, Y. C., GENOVESE, C. R. and WASSERMAN, L. (2014). Generalized mode and ridge estimation. ArXiv E-prints. Available at arXiv:1406.1803.

CHEN, Y.-C., GENOVESE, C. R., TIBSHIRANI, R. J. and WASSERMAN, L. (2016). Nonparametric modal regression. *Ann. Statist.* **44** 489–514. MR3476607 https://doi.org/10.1214/15-AOS1373

CHENG, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **17** 790–799.

COMANICIU, D. and MEER, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **24** 603–619.

DI MARZIO, M., PANZERA, A. and TAYLOR, C. C. (2009). Local polynomial regression for circular predictors. *Statist. Probab. Lett.* **79** 2066–2075. MR2571770 https://doi.org/10.1016/j.spl.2009.06.014

DI MARZIO, M., PANZERA, A. and TAYLOR, C. C. (2011). Kernel density estimation on the torus. *J. Statist. Plann. Inference* **141** 2156–2173. MR2772221 https://doi.org/10.1016/j.jspi.2011.01.002

DI MARZIO, M., PANZERA, A. and TAYLOR, C. C. (2013). Non-parametric regression for circular responses. *Scand. J. Stat.* **40** 238–255. MR3066413 https://doi.org/10.1111/j.1467-9469.2012.00809.x

DI MARZIO, M., FENSORE, S., PANZERA, A. and TAYLOR, C. C. (2016). A note on nonparametric estimation of circular conditional densities. *J. Stat. Comput. Simul.* **86** 2573–2582. MR3511014 https://doi.org/10.1080/00949655.2016.1146279

DOMENICI, P., BLAGBURN, J. M. and BACON, J. P. (2011a). Animal escapology I: Theoretical issues and emerging trends in escape trajectories. *J. Exp. Biol.* **214** 2463–2473.

DOMENICI, P., BLAGBURN, J. M. and BACON, J. P. (2011b). Animal escapology II: Escape trajectory case studies. *J. Exp. Biol.* **214** 2474–2494.

EDELSBRUNNER, H., KIRKPATRICK, D. G. and SEIDEL, R. (1983). On the shape of a set of points in the plane. *IEEE Trans. Inf. Theory* **29** 551–559. MR0713690 https://doi.org/10.1109/TIT.1983.1056714

EDELSBRUNNER, H. and MUCKE, E. P. (1984). Three-dimensional alpha shapes. *ACM Trans. Graph.* **13** 43–72.

EINBECK, J. and TUTZ, G. (2006). Modelling beyond regression functions: An application of multimodal regression to speed-flow data. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **55** 461–475. MR2242274 https://doi.org/10.1111/j.1467-9876.2006.00547.x

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications. Monographs on Statistics and Applied Probability* **66**. CRC Press, London. MR1383587

FISHER, N. I. (1989). Smoothing a sample of circular data. *J. Struct. Geol.* **11** 775–778.

FISHER, N. I. (1993b). *Statistical Analysis of Circular Data*. Cambridge Univ. Press, Cambridge. MR1251957 https://doi.org/10.1017/CBO9780511564345

FISHER, N. I. and LEE, A. J. (1992). Regression models for an angular response. *Biometrics* **48** 665–677. MR1187598 https://doi.org/10.2307/2532334

FUKUNAGA, K. and HOSTETLER, L. D. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans. Inf. Theory* **21** 32–40. MR0388638 https://doi.org/10.1109/tit.1975.1055330

GARCÍA-PORTUGUÉS, E., CRUJEIRAS, R. M. and GONZÁLEZ-MANTEIGA, W. (2013). Kernel density estimation for directional-linear data. *J. Multivariate Anal.* **121** 152–175. MR3090475 https://doi.org/10.1016/j.jmva.2013.06.009

JAMMALAMADAKA, S. R. and SENGUPTA, A. (2001). *Topics in Circular Statistics. Series on Multivariate Analysis* **5**. World Scientific Co., Inc., River Edge, NJ. MR1836122 https://doi.org/10.1142/9789812779267

KIM, S. and SENGUPTA, A. (2017). Multivariate-multiple circular regression. *J. Stat. Comput. Simul.* **87** 1277–1291. MR3621948 https://doi.org/10.1080/00949655.2016.1261292

KOBAYASHI, T. and OTSU, N. (2010). von Mises-Fisher mean shift for clustering on a hypersphere. In *Proceedings of the* 20*th International Conference on Pattern Recognition* 2130–2133. IEEE, New York.

LEY, C. and VERDEBOUT, T. (2017). *Modern Directional Statistics. Chapman & Hall/CRC Interdisciplinary Statistics Series*. CRC Press, Boca Raton, FL. MR3752655

LI, X., HU, Z. and WU, F. (2007). A note on the convergence of the mean shift. *Pattern Recognit.* **40** 1756–1762.

MARCHETTI, G. and SCAPINI, F. (2003). Use of multiple regression models in the study of sandhopper orientation under natural conditions. *Estuar. Coast. Shelf Sci.* **58** 207–215.

MARDIA, K. V. and JUPP, P. E. (2000). *Directional Statistics. Wiley Series in Probability and Statistics*. Wiley, Chichester. MR1828667

NAIR, A., CHANGSING, K., STEWART, W. J. and MCHENRY, M. J. (2017a). Data from: Fish prey change strategy with the direction of a threat. https://doi.org/10.5061/dryad.47mq9.

NAIR, A., CHANGSING, K., STEWART, W. J. and MCHENRY, M. J. (2017b). Fish prey change strategy with the direction of a threat. *Proc. Biol. Sci.* **284**. https://doi.org/10.1098/rspb.2017.0393

OBA, S., KATO, K. and ISHII, S. (2005). Multi-scale clustering for gene expression profiling data. In 5*th IEEE Symposium on Bioinformatics and Bioengineering* (*BIBE*'05) 210–217. IEEE, New York.

OBLESER, P., HART, V., MALKEMPER, E. P., BEGALL, S., HOLÁ, M., PAINTER, M. S., CERVENÝ, J. and BURDA, H. (2016). Compass-controlled escape behavior in Roe deer. *Behav. Ecol. Sociobiol.* **70** 1345–1355.

OLIVEIRA, M., CRUJEIRAS, R. M. and RODRÍGUEZ–CASAL, A. (2014). NPCirc: An R package for nonparametric circular methods. *J. Stat. Softw.* **61**.

PEWSEY, A., NEUHÄUSER, M. and RUXTON, G. D. (2013). *Circular Statistics in R*. Oxford Univ. Press, Oxford. MR3156170

SATO, N., SHIDARA, H. and OGAWA, H. (2019). Trade-off between motor performance and behavioural flexibility in the action selection of cricket escape behaviour. *Sci. Rep.* **9** 18112. https://doi.org/10.1038/s41598-019-54555-7

SCAPINI, F., ALOIA, A., BOUSLAMA, M., CHELAZZI, L., COLOMBINI, I., ELGTARI, M., FALLACI, M. and MARCHETTI, G. M. (2002). Multiple regression analysis of the sources of variation in orientation of two sympatric sandhoppers, talitrus saltator and talorchestia brito, from an exposed Mediterranean beach. *Behav. Ecol.* **52** 403–414.

SCOTT, D. W. (1992). *Multivariate Density Estimation. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, New York. MR1191168 https://doi.org/10.1002/9780470316849

WIKIMEDIA COMMONS (2014). Larval zebrafish. [Online; accessed September 29, 2021].

WILLIAMSON, B., FRASER, S., WILLIAMSON, L., NIKORA, V. and SCOTT, B. (2019). Predictable changes in fish school characteristics due to a tidal turbine support structure. *Renew. Energy* **141** 1092–1102.

ZELNIK-MANOR, L. and PERONA, P. (2004). Self-tuning spectral clustering. In *Advances in Neural Information Processing Systems* (*NIPS*).

ZHANG, Y. and CHEN, Y.-C. (2021a). Kernel smoothing, mean shift, and their learning theory with directional data. *J. Mach. Learn. Res.* **22** Paper No. 154. MR4318510

ZHANG, Y. and CHEN, Y. C. (2021b). Mode and ridge estimation in euclidean and directional product spaces: A mean shift approach. ArXiv E-prints. Available at arXiv:2110.08505.

ZHOU, H. and HUANG, X. (2016). Nonparametric modal regression in the presence of measurement error. *Electron. J. Stat.* **10** 3579–3620. MR3575565 https://doi.org/10.1214/16-EJS1210

ZHOU, H. and HUANG, X. (2019). Bandwidth selection for nonparametric modal regression. *Comm. Statist. Simulation Comput.* **48** 968–984. MR3957561 https://doi.org/10.1080/03610918.2017.1402044

Supplement to "Analyzing animal escape data with circular nonparametric multimodal regression": assumptions, results and proofs

María Alonso-Pena[1] and Rosa M. Crujeiras[1]

[1] CITMAga, Universidade de Santiago de Compostela

This supplementary material for the paper "Analyzing animal escape data with circular nonparametric multimodal regression" provides further results complementing the main text. Section 1 contains the assumptions needed for the consistency of the estimators. Section 2 gives the propositions stating the consistency, including outlines of the proofs.

1 Assumptions

We will first state the assumptions needed for obtaining the theoretical results stating the consistency of the estimators. We will divide these assumptions in three different blocks, according to what they are needed for. In the first place, we will assume:

(A1) The joint density function $f(\theta, y)$ is at least three times continuously differentiable and its first two partial derivatives with respect of y are square integrable on $\mathbb{R}$.

(A2) The linear kernel $L(\cdot)$ is a symmetric around zero linear density and

(A2-1)

$$0 < \int_{\mathbb{R}} L^2(u)du < \infty, \quad 0 < \int_{\mathbb{R}} u^2 L^2(u)du < \infty.$$

(A2-2)

$$0 < \int_{\mathbb{R}} {L^{(q)}}^2(u)du < \infty, \quad 0 < \int_{\mathbb{R}} u^2 {L^{(q)}}^2(u)du < \infty, \quad \text{for} \quad q = 1, 2.$$

(A3) The joint density function $f(\delta, \phi)$ is at least three times continuously differentiable and its first two partial derivatives with respect of ϕ are square integrable on $\mathbb{T}$.

(A4) The circular kernel $K_\kappa(\cdot)$ satisfies $K_\kappa(\cdot) = c_\kappa K[\kappa(1-\cos(\cdot))]$ where c_κ is a normalization constant, $K : [0,\infty) \to [0,\infty)$ with

(A4-1)

$$0 < \int_0^\infty r^{\frac{j-1}{2}} K^l(r)dr < \infty, \quad \text{for } l = 1,2,\ j = 0,2.$$

(A4-2)

$$0 < \int_0^\infty r^{\frac{j-1}{2}} {K^{(q)}}^2(r)dr < \infty \quad \text{and} \quad 0 < \int_0^\infty r^{\frac{j-1}{2}} K^{(2)}(r)K^{(1)}(r)dr < \infty,$$

for $l = 1,2, \quad q = 1,2,\ j = 0,2,4.$

In this first block of assumptions, we handle conditions on the joint densities and its partial derivatives, since the estimation of the regression multifunction is related to the estimation of the partial derivatives of the joint density. Note that the regression multifunction in equation (1) of the main text is equivalent to

$$M(\theta) = \left\{ y : \frac{\partial}{\partial y} f(\theta, y) = 0, \frac{\partial^2}{\partial y^2} f(\theta, y) < 0 \right\}$$

and the multifunction in equation (7) of the main text is equivalent to

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

Note that conditions **(A1)** and **(A3)** are standard smoothing conditions on the true density required in order to obtain the convergence rates of the partial derivatives of the joint density estimators. Condition **(A2)** is also a usual condition on kernel density estimation literature and is satisfied by many kernels in practice, such as the normal kernel. The last condition, **(A4)**, was required by Zhang and Chen (2021) in order to obtain the asymptotic convergence rates of the derivatives of the kernel density estimator in the unit hypersphere and, as noted by the authors, integrating by parts shows that the von Mises kernel satisfies **(A4)**.

For the next block of assumptions, we define the following function classes:

$$\mathcal{KL} = \left\{ K\{\kappa[1-\cos(\cdot-\theta)]\} \frac{\partial^q}{\partial y^q} L\left(\frac{\cdot - y}{h}\right) : (\theta, y) \in \mathbb{T} \times \mathbb{R}, \kappa, h > 0,\ q = 0,1,2 \right\},$$

$$\mathcal{LK} = \left\{ L\left(\frac{\cdot - x}{h}\right) \frac{\partial^q}{\partial \phi^q} K\{\kappa[1-\cos(\cdot-\phi)]\} : (x, \phi) \in \mathbb{R} \times \mathbb{T}, h, \kappa > 0,\ q = 0,1,2 \right\},$$

$$\mathcal{KK} = \left\{ K\{\nu[1-\cos(\cdot-\theta)]\} \frac{\partial^q}{\partial \phi^q} K\{\kappa[1-\cos(\cdot-\phi)]\} : (\theta, \phi) \in \mathbb{T}\times\mathbb{T}, \upsilon, \kappa > 0,\ q = 0,1,2 \right\}.$$

We assume the following conditions:

(B1) The function $K\{\kappa[1-\cos(\cdot-\theta)]\}\frac{\partial^q}{\partial y^q}L\left(\frac{\cdot-y}{h}\right)$ is bounded in absolute value by C_{KL} for $q=0,1,2$ and the collection $\mathcal{KL}$ is a VC-type class, i.e., there exist $A,u>0$ such that for any $0<\varepsilon<1$,

$$\sup_Q N(\mathcal{KL}, L_2(Q), C_{KL}\varepsilon) \leq \left(\frac{A}{\varepsilon}\right)^u,$$

where $N(T,d,\varepsilon)$ is the ε-covering number for a semimetric space (T,d) and Q is any probability measure.

(B2) The function $L\left(\frac{\cdot-x}{h}\right)\frac{\partial^q}{\partial\phi^q}K\{\kappa[1-\cos(\cdot-\phi)]\}$ is bounded in absolute value by C_{LK} for $q=0,1,2$ and the collection $\mathcal{LK}$ is a VC-type class: there exist $A',u'>0$ such that for any $0<\varepsilon<1$,

$$\sup_Q N(\mathcal{LK}, L_2(Q), C_K\varepsilon) \leq \left(\frac{A'}{\varepsilon}\right)^{u'}.$$

(B3) The function $K\{\nu[1-\cos(\cdot-\theta)]\}\frac{\partial^q}{\partial\phi^q}K\{\kappa[1-\cos(\cdot-\phi)]\}$ is bounded in absolute value by C_{KK} for $q=0,1,2$ and the collection $\mathcal{KK}$ is a VC-type class: there exist $A'',u''>0$ such that for any $0<\varepsilon<1$,

$$\sup_Q N(\mathcal{KK}, L_2(Q), C_{KK}\varepsilon) \leq \left(\frac{A''}{\varepsilon}\right)^{u''}.$$

Conditions **(B1)**, **(B2)** and **(B3)** correspond to standard conditions on the kernels in order to obtain the uniform convergence rates (Giné and Guillou, 2002; Einmahl and Mason, 2005). Such conditions can be satisfied in practice by, for example, the Gaussian kernel in the linear case and the von Mises kernel in the circular case.

With the objective of obtaining the asymptotic error rates for the multimodal estimators we also need to assume the following conditions on the joint densities:

(C1) There exists $\lambda_1>0$ such that for any $(\theta,y)\in\mathbb{T}\times\Omega$ meeting $\frac{\partial}{\partial y}f(\theta,y)=0$, $\left|\frac{\partial^2}{\partial y^2}f(\theta,y)\right|<\lambda_1$. In addition, for each $\theta\in\mathbb{T}$ there exist $B_1,\rho_1>0$ such that

$$\left\{y\in\Omega:\max_{y\in\Omega}\left|\frac{\partial}{\partial y}f(\theta,y)\right|\leq B_1, \frac{\partial^2}{\partial^2 y}f(\theta,y)\leq-\lambda_1<0\right\}\subset M(\theta)\oplus\rho_1,$$

where $A\oplus r=\{x\in\mathbb{R}:\inf_{z\in A}|z-x|<r\}$.

(C2) There exists $\lambda_2>0$ such that for any $(\delta,\phi)\in\text{Supp}(\Delta)\times\mathbb{T}$ satisfying $\frac{\partial}{\partial\phi}f(\delta,\phi)=0$, $\left|\frac{\partial^2}{\partial\phi^2}f(\delta,\phi)\right|<\lambda_2$. In addition, for each $\delta\in\text{Supp}(\Delta)$ there exist $B_2,\rho_2>0$ such that

$$\left\{\phi\in\mathbb{T}:\max_{\phi\in\mathbb{T}}\left|\frac{\partial}{\partial\phi}f(\delta,\phi)\right|\leq B_2, \frac{\partial^2}{\partial^2\phi}f(\delta,\phi)\leq-\lambda_2<0\right\}\subset M(\delta)\oplus\rho_2,$$

where $A\oplus r=\{\gamma\in\mathbb{T}:\inf_{z\in A}(1-\cos(z-\gamma))<r\}$.

Conditions **(C1)** and **(C2)** are imposed so that every conditional local mode is isolated from other critical values and so that points with near zero first partial derivative and negative second partial derivative are close to a conditional local mode. This assumptions eliminate the cases where the branches of the regression multifunction cross each other (Chen et al., 2016).

2 Theoretical results

Proposition 1. *Let Θ be a circular random variable and Y a real-valued random variable. Consider the multimodal regression estimator in equation (3) of the main text. Under assumptions **(A1)**, **(A2)**, **(A4-1)**, **(B1)** and **(C1)**, it holds that,*

$$\Lambda(\theta) = O(\kappa^{-1} + h^2) + O_P\left(\sqrt{\frac{\kappa^{1/2}}{nh^3}}\right)$$

and

$$MISE_m(\widehat{M}) = O(\kappa^{-2} + h^4) + O\left(\frac{\kappa^{1/2}}{nh^3}\right),$$

as $h \to 0$, $\kappa \to \infty$ and $\frac{nh^{1+2q}}{\kappa^{1/2}\log n} \to \infty$, for $q = 0, 1, 2$.

Outline of the proof. In the first place, by using the same approach as in Propositions 1 and 2 in García-Portugués et al. (2013), under conditions **(A1)**, **(A2)** and **(A4-1)** we obtain that, for fixed $(\theta, y) \in \mathbb{T} \times \mathbb{R}$,

$$\frac{\partial^q}{\partial y^q r}\hat{f}(\theta, y) - \frac{\partial^q}{\partial y^q} f(\theta, y) = O(\kappa^{-1} + h^2) + O_P\left(\sqrt{\frac{\kappa^{1/2}}{nh^{1+2q}}}\right) \quad \text{as } nh^{1+2q}\kappa^{-1/2} \to \infty,$$

for $q = 0, 1, 2$. Now, by also taking into account assumption **(B1)**, we apply the method of Giné and Guillou (2002) and Einmahl and Mason (2005), noting that the circular kernel is bounded, and obtain

$$\sup_{\theta, y}\left|\left|\frac{\partial^q}{\partial y^q}\hat{f}(\theta, y) - \frac{\partial^q}{\partial y^q} f(\theta, y)\right|\right| = O(\kappa^{-1} + h^2) + O_P\left(\sqrt{\frac{\kappa^{1/2}\log n}{nh^{1+2q}}}\right)$$

as $h \to 0$, $\kappa \to \infty$ and $\frac{nh^{1+2q}}{\kappa^{1/2}\log n} \to \infty$, for $q = 0, 1, 2$.

Lastly, by also using assumption **(C1)** and applying the arguments of Theorem 3 in Chen et al. (2016), we have

$$\Lambda(\theta) = O(\kappa^{-1} + h^2) + O_P\left(\sqrt{\frac{\kappa^{1/2}}{nh^3}}\right).$$

The second assertion is obtained by applying the arguments in Chacón et al. (2011) and Chacón and Duong (2013), leading to

$$\mathrm{MISE}_m(\widehat{M}) = O(\kappa^{-2} + h^4) + O\left(\frac{\kappa^{1/2}}{nh^3}\right).$$

□

Proposition 2. *Let X be a real-valued random variable and Φ a circular random variable. Consider the multimodal regression estimator in equation (7) of the main text. Under assumptions **(A2-1)**, **(A3)**, **(A4)**, **(B2)** and **(C2)**, it holds that*

$$\tilde{\Lambda}(x) = O(h^2 + \kappa^{-1}) + O_P\left(\sqrt{\frac{\kappa^{3/2}}{nh}}\right)$$

and

$$CMIE_m(\widehat{M}) = O(h^2 + \kappa^{-1}) + O\left(\sqrt{\frac{\kappa^{3/2}}{nh}}\right),$$

as $\kappa \to \infty$, $h \to 0$ and $\frac{nh}{\kappa^{(1+2q)/2}\log n} \to \infty$, for $q = 0, 1, 2$.

Outline of the proof. First, under conditions **(A2-1)**, **(A3)** and **(A4)**, we use the approach in Propositions 1 and 2 in García-Portugués et al. (2013) and Theorem 2 in Zhang and Chen (2021) and obtain that, for fixed $(x, \phi) \in \mathbb{R} \times \mathbb{T}$,

$$\frac{\partial^q}{\partial \phi^q}\hat{f}(x,\phi) - \frac{\partial^q}{\partial \phi^q} f(x,\phi) = O(\kappa^{-1} + h^2) + O_P\left(\sqrt{\frac{\kappa^{(1+2q)/2}}{nh}}\right) \quad \text{as } nh\kappa^{\frac{-(1+2q)}{2}} \to \infty,$$

for $q = 0, 1, 2$. Now, by also assuming, **(B2)** and by analogy with the proof of Theorem 4 in Zhang and Chen (2021) (noting that the linear kernel is bounded), we obtain

$$\sup_{x,\phi}\left|\left|\frac{\partial^q}{\partial \phi^q}\hat{f}(x,\phi) - \frac{\partial^q}{\partial \phi^q} f(x,\phi)\right|\right| = O(\kappa^{-1} + h^2) + O_P\left(\sqrt{\frac{\kappa^{(1+2q)/2}\log n}{nh}}\right)$$

as $h \to 0$, $\kappa \to \infty$ and $\frac{nh}{\kappa^{(1+2q)/2}\log n} \to \infty$, for $q = 0, 1, 2$.

Lastly, under assumption **(C2)** and applying the arguments of Theorem 6 in Zhang and Chen (2021), we have

$$\tilde{\Lambda}(x) = O(h^2 + \kappa^{-1}) + O_P\left(\sqrt{\frac{\kappa^{3/2}}{nh}}\right).$$

Now, the CMIE_m yields the rate of convergence

$$\mathrm{CMIE}_m(\widehat{M}) = O(h^2 + \kappa^{-1}) + O\left(\sqrt{\frac{\kappa^{3/2}}{nh}}\right).$$

□

Proposition 3. *Let Θ and Φ be circular random variables. Consider the multimodal regression estimator in equation (7) of the main text. Under assumptions* ***(A2-1)****,* ***(A3)****,* ***(A4)****,* ***(B3)*** *and* ***(C2)****, it holds that*

$$\tilde{\Lambda}(\theta) = O(\nu^{-1} + \kappa^{-1}) + O_P\left(\sqrt{\frac{\kappa^{3/2}\nu^{1/2}}{n}}\right)$$

and

$$CMIE_m(\widehat{M}) = O(\nu^{-1} + \kappa^{-1}) + O\left(\sqrt{\frac{\kappa^{3/2}\nu^{1/2}}{n}}\right),$$

as $\kappa \to \infty$, $\nu \to \infty$ and $\frac{n}{\kappa^{(1+2q)/2}\nu^{1/2}\log n} \to \infty$, for $q = 0, 1, 2$.

Outline of the proof. The proof is analogous to that of Proposition 2 by using a circular kernel. □

References

Chacón, J. E. and Duong, T. (2013). Data-driven density derivative estimation, with applications to nonparametric clustering and bump hunting. *Electron. J. Stat.*, 7:499–532.

Chacón, J. E., Duong, T., and Wand, M. (2011). Asymptotics for general multivariate kernel density derivative estimators. *Stat. Sin.*, 21:807–840.

Chen, Y.-C., Genovese, C. R., Tibshirani, R. J., and Wasserman, L. (2016). Nonparametric modal regression. *Ann. Stat.*, 44:489–514.

Einmahl, U. and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Stat.*, 33:1380–1403.

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2013). Kernel density estimation for directional-linear data. *J. Multivar. Anal.*, 121:152–275.

Giné, E. and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. I. H. Poincare-Pr*, 38:907–921.

Zhang, Y. and Chen, Y.-C. (2021). Kernel smoothing, mean shift, and their learning theory with directional data. *J. Mach. Learn. Res.*, pages 1–91.

Supplement to "Analyzing animal escape data with circular nonparametric multimodal regression": modified modal cross-validation

María Alonso-Pena[1] and Rosa M. Crujeiras[1]

[1] CITMAga, Universidade de Santiago de Compostela

This supplementary material for the paper "Analyzing animal escape data with circular nonparametric multimodal regression" provides further results complementing the main text. In this supplement we give the detailed algorithms containing the modification of the modal cross-validation criterion presented in Section 4 of the paper and some simulation results comparing their performance against the original method. This modified criterion was used in the simulations experiments in Section 5 of the main text.

1 Modification of the modal cross-validation criterion

In the simulation experiments of Section 5 in the main text, we use a modification of the modal cross-validation criterion in order to attain more computational efficiency. The motivation behind this modification is the different role the two smoothing parameters have in the estimation of the regression multifunction.

Circular predictor and real-valued response Consider first the case with a real-valued response, Y, and a circular predictor, Θ. We denote the multimodal regression estimator in equation (3) of the main text as $\widehat{M}^{\kappa,h}(\theta)$, where κ is the concentration parameter and h is the bandwidth. The cross-validation criterion presented in Section 4.1 of the main text can be modified as follows:

1. Select a pair of reference smoothing parameters by minimizing the Mean Integrated Squared Error of the joint density estimator $\hat{f}(\theta, y)$ by assuming that the joint density is a mixture of independent von Mises and Gaussian densities (García-Portugués et al., 2013). We denote these reference parameters as $(\kappa_{\text{ref}}, h_{\text{ref}})$.

2. Set $\kappa = \kappa_{\text{ref}}$ and select h_{cv} minimizing

$$CV_1(h) = \frac{1}{n}\sum_{j=1}^{n} d^2\left(\widehat{M}_{-j}^{\kappa_{\text{ref}},h}(\Theta_j), Y_j\right) N_{-j}^2(\Theta_j),$$

where $\widehat{M}_{-j}^{\kappa_{\text{ref}},h}$ represents the estimator computing without observations (Θ_j, Y_j), $N_{-j}(\Theta_j)$ represents the number of estimated local modes at $\theta = \Theta_j$ with estimator $\widehat{M}_{-j}^{\kappa_{\text{ref}},h}$ and $d(x, A) = \inf_{z\in A} ||x - z||$.

3. Set $h = h_{\text{cv}}$ and select κ minimizing

$$CV_1(\kappa) = \frac{1}{n}\sum_{j=1}^{n} d^2\left(\widehat{M}_{-j}^{\kappa,h_{\text{cv}}}(\Theta_j), Y_j\right) N_{-j}^2(\Theta_j).$$

Real-valued predictor and circular response Consider now the case with a circular response, Φ, and a real-valued predictor, X. Denote by $\widehat{M}^{h,\kappa}(x)$ the multimodal estimator described in equation (7) of the main text, computed with smoothing parameters h and κ. In order to automatically select the smoothing parameters, we use the following modification of the cross-validation criterion:

1. Select a pair of reference smoothing parameters by minimizing the Mean Integrated Squared Error of the joint density estimator $\hat{f}(x, \phi)$ by assuming that the joint density is a mixture of independent von Mises and Gaussian densities (García-Portugués et al., 2013). We denote these reference parameters as $(h_{\text{ref}}, \kappa_{\text{ref}})$.

2. Set $h = h_{\text{ref}}$ and select κ_{cv} minimizing

$$CV_2(\kappa) = \frac{1}{n}\sum_{j=1}^{n} \tilde{d}\left(\widehat{M}_{-j}^{h_{\text{ref}},\kappa}(\Delta_j), Y_j\right) N_{-j}(\Delta_j),$$

where $\tilde{d}(x, A) = \inf_{z\in A} 1 - \cos(x - z)$.

3. Set $\kappa = \kappa_{\text{cv}}$ and select h minimizing

$$CV_2(h) = \frac{1}{n}\sum_{j=1}^{n} \tilde{d}\left(\widehat{M}_{-j}^{h,\kappa_{\text{cv}}}(\Delta_j), Y_j\right) N_{-j}(\Delta_j).$$

Circular predictor and circular response Finally, when we have a circular response, Φ, and a circular predictor, Θ, denote the estimated regression multifunction as $\widehat{M}^{\nu,\kappa}(\theta)$ when using concentration parameters ν and κ. We employ the following modification of the algorithm:

1. Select a pair of reference smoothing parameters by minimizing the Mean Integrated Squared Error of the joint density estimator $\hat{f}(\theta, \phi)$ by assuming that the joint density is a mixture of products of von Mises densities. We denote these reference parameters as $(\nu_{\text{ref}}, \kappa_{\text{ref}})$.

2. Set ν_{ref} and select κ_{cv} minimizing

$$CV_2(\kappa) = \frac{1}{n}\sum_{j=1}^{n} \tilde{d}\left(\widehat{M}_{-j}^{\nu_{\text{ref}},\kappa}(\Delta_j), Y_j\right) N_{-j}(\Delta_j).$$

3. Set $\kappa = \kappa_{\text{cv}}$ and select ν minimizing

$$CV_2(g) = \frac{1}{n}\sum_{i=j}^{n} \tilde{d}\left(\widehat{M}_{-j}^{\nu,\kappa_{\text{cv}}}(\Delta_j), Y_j\right) N_{-j}(\Delta_j).$$

2 Performance of the modified modal cross-validation

In this section we compare, through a brief simulation study, the performances of the modal cross-validation method described in Section 4 of the main text and the modification introduced in the previous section. For that aim, we simulated data from models CL-1, LC-1 and CC-1 of Section 5 of the manuscript, where the dispersion/concentration parameters were set to $\sigma = 1$ for CL-1, $\tau = 8$ for LC-1 and $\tau = 12$ for CC-1. We estimated the MISE_m for model CL-1 and the CMIE_m for models LC-1 and CC-1, using the same sample sizes and number of replicates as in the simulation study of the main text. Results when selecting the parameters by the original cross-validation procedure (O) and with its modified version (M) are shown in Table 1. Both methods yield a very similar performance in terms of the estimated MISE_m and estimated CMIE_m. Regarding the computational efficiency, it is clear that the modified version requires less computational time, since it performs an optimization over one variable (twice) while the modal cross-validation procedure described in Section 4 of the main text implies optimization over two variables.

References

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2013). Kernel density estimation for directional-linear data. *J. Multivar. Anal.*, 121:152–275.

Table 1: Monte Carlo estimates of MISE_m for model CL-1 ($\sigma = 1$) and CMIE_m for models LC-1 ($\tau = 8$) and CC-1 ($\tau = 12$), with different values of the sample sizes. Results under O correspond to the ones obtained with smoothing parameters selected by the original modal cross-validation and results under M refer to the modified modal cross-validation method.

	CL-1		LC-1		CC-1	
(n_1, n_2)	O	M	O	M	O	M
$(100, 100)$	1.877	1.865	0.023	0.024	0.213	0.201
$(100, 200)$	1.653	1.653	0.024	0.024	0.147	0.150
$(200, 200)$	0.851	0.840	0.012	0.012	0.066	0.065
$(200, 300)$	0.706	0.702	0.011	0.011	0.054	0.056
$(300, 300)$	0.574	0.574	0.008	0.008	0.045	0.043