



---

*Research article*

## **A method for determining groups in nonparametric regression curves: Application to prefrontal cortex neural activity analysis**

**Javier Roca-Pardiñas<sup>1</sup>, Celestino Ordóñez<sup>2,\*</sup> and Luís Meira Machado<sup>3</sup>**

<sup>1</sup> Department of Statistics and Operational Research, Vigo University, Vigo 36310, Spain

<sup>2</sup> Department of Mining Exploitation and Prospecting, Geomatics and Computer Graphics Lab, Oviedo University, Mieres 33600, Spain

<sup>3</sup> Center of Mathematics, Minho University, Braga 4704-553, Portugal

\* **Correspondence:** Email: [ordonezcelestino@uniovi.es](mailto:ordonezcelestino@uniovi.es); Tel: +34985458027.

**Abstract:** Generalized additive models provide a flexible and easily-interpretable method for uncovering a nonlinear relationship between response and covariates. In many situations, the effect of a continuous covariate on the response varies across groups defined by the levels of a categorical variable. When confronted with a considerable number of groups defined by the levels of the categorical variable and a factor-by-curve interaction is detected in the model, it then becomes important to compare these regression curves. When the null hypothesis of equality of curves is rejected, leading to the clear conclusion that at least one curve is different, we may assume that individuals can be grouped into a number of classes whose members all share the same regression function. We propose a method that allows determining such groups with an automatic selection of their number by means of bootstrapping. The validity and behavior of the proposed method were evaluated through simulation studies. The applicability of the proposed method is illustrated using real data from an experimental study in neurology.

**Keywords:** clustering of regression curves; factor-by-curve interaction; generalized additive model; multiple regression curves; nonlinear regression; number of groups

---

### **1. Introduction**

One of the main goals of statistical modeling is to understand the effect of some explanatory variables,  $X_i, i = 1, \dots, p$ , on a dependent variable,  $Y$ , also known as the response. This type of dependence is often modeled using a generalized linear regression model (GLM) [1] that imposes a linear relationship between explanatory and dependent variables without specifying in advance the function that links

them. Generalized additive models (GAMs) [2, 3] offer an extension of the GLM through the incorporation of nonlinear forms for the explanatory variables. GAM has been widely used as an effective technique for conducting nonlinear regression analysis in various fields such as economics, finance, the environment, medicine, and biology. In a wide range of these applications, problems can be observed which it might be useful to compare two or more regression curves. This often occurs when it is necessary to check if the effect of a continuous covariate on the response varies across groups defined by levels of a categorical variable. Testing the hypothesis of equality of the regression functions is a topic of statistical inference that has been widely researched in the literature. One important review on this topic may be found in [4] (see their Section 7). Examples of methods for clustering regression curves using  $L^2$  distance and other metrics are in [5–7]. For nonparametric models, most approaches rely on smoothing techniques, such as splines or Nadaraya-Watson [8], in the construction of the nonparametric estimators of regression functions. The general idea is to use these nonparametric estimators directly, either by contrasting all individual estimators with the pooled estimator, or by performing pairwise comparisons. Other authors considered the use of empirical processes to avoid the necessity of selecting a smoothing parameter required in the construction of the nonparametric estimators of regression functions [9–14]. The idea behind these tests is to use the distribution of the residuals and to compare them via Kolmogorov-Smirnov or Cramér-von Mises type statistics. ANOVA-type test statistics have also been used to compare regression functions (see for example [15] and [16]). Reference [17] illustrates how the SiZer exploratory tool is capable of comparing multiple curves based on the residuals. In [18] a kernel-based nonparametric approach is considered while [19] proposed a testing procedure for single index models.

A second, though related question is how to determine groups among a series of regression curves when the null hypothesis of equality of the regression functions is rejected. Though the aforementioned methods can be used to compare regression curves, to the best of our knowledge, those methods cannot be used to determine groups among a series of regression curves, for example, among groups defined by the levels of a categorical variable. Some of them are not recommended when confronted with a considerable number of curves. If the null hypothesis of equality of curves is rejected, then this leads to the clear conclusion that at least one regression curve is different. However, these methods cannot be used to ascertain whether individuals can be grouped into a reduced number of classes whose members all share the same regression function or if all the regression curves are different from each other. One naïve approach would be to perform pairwise comparisons. However, this approach would lead to a large number of comparisons (e.g., 7 groups would lead to 21 pairwise comparisons). This could be done but without the possibility of determining groups with similar regression curves.

A statistical procedure to estimate the unknown group structure was proposed in [20] and later in [21]. The first paper of these authors was about financial time series whereas the second deals with environmental statistics. More recently, the comparison of survival curves was addressed [22]. Similarly, in this paper, we propose an approach that allows determining groups that share the same regression function with an automatic selection of their number. The proposed method can be used, for instance, to establish groups with similar monotonic trends in the regression curves (e.g., over the levels of categorical variables). The proposed methodology will be shown in the framework of the generalized additive model with a binary outcome and a logit link function. The generalization of the proposed methods to other link functions is possible.

The remainder of the paper is organized as follows: The following section provides the notation

and the methodological background. Then, the performance of the proposed methods is investigated through simulations, and their usage is illustrated through the analysis of a real data set from neuronal activity in the prefrontal cortex of monkeys during a discrimination task. Finally, the last section contains a discussion and the main conclusions of the work.

## 2. Materials and methods

### 2.1. Mathematical model

Let  $Y$  denote the binary (0/1) target response and  $\mathbf{X} = (X_1, \dots, X_p)$  a set of  $p$  continuous covariates. In this regression context, the logistic generalized additive model expresses the conditional probability  $P(\mathbf{X}) = P(Y = 1|\mathbf{X})$  as

$$\log \frac{P(\mathbf{X})}{1 - P(\mathbf{X})} = \alpha + \sum_{j=1}^p f_j(X_j) \quad (2.1)$$

where  $\alpha$  is a constant and each partial function  $f_j$  represents the effect of the covariate  $X_j$ . These unknown smooth partial functions are modeled from the data without specifying in advance any parametric structure. Therefore, the GAM in (2.1) combines flexibility with interpretability, in which each of the additive components describes the influence of each covariate separately.

In practice, the effect of a continuous covariate  $X_j$  can vary across the levels  $1, \dots, K$  of a categorical factor  $F$ . For simplicity of notation, here we consider that only the effect of the last covariate,  $X_p$ , depends on the levels of  $F$ , so the pure GAM in (2.1) can incorporate this type of factor-by-curve relationship as follows:

$$\text{logit}(\mathbf{F}, \mathbf{X}) = \log \frac{P(Y = 1|\mathbf{F}, \mathbf{X})}{1 - P(Y = 1|\mathbf{F}, \mathbf{X})} = \alpha + \sum_{j=1}^{p-1} f_j(X_j) + \begin{cases} g_1(X_p) & \text{if } F = 1 \\ g_2(X_p) & \text{if } F = 2 \\ \vdots & \vdots \\ g_K(X_p) & \text{if } F = K \end{cases} \quad (2.2)$$

where  $g_1, \dots, g_K$  represent the specific effect of  $X_p$  for each of the possible  $K$  levels established by the factor  $F$ .

### 2.2. Estimating partitions

The study of the partial functions  $g_1, \dots, g_K$  can be useful in the comparison of two or more groups, which is an important problem associated with statistical inference. Interest centers on the null hypothesis  $H_0 : g_1 = \dots = g_K$ , namely, that the effect of  $X_p$  does not depend on the levels of the factor  $F$ . When the equality of the  $K$  curves is rejected, leading to the clear conclusion that at least one curve is different, it can be interesting to ascertain whether groups can be performed or, by contrast, that all these curves are different from each other. More clearly, if  $H_0$  is rejected, it could be interesting to determine: a) if any pair of curves are different from each other, that is  $g_i \neq g_j$ , for all  $i \neq j \in \{1, \dots, K\}$ ,

or b) if the indices in  $\{1, \dots, K\}$  can be grouped in a partition of  $J (J < K)$  groups  $\mathbf{I} = (\mathbf{I}_1, \dots, \mathbf{I}_J)$  with  $J < K$ , so that  $g_i = g_j$  for each  $i, j \in I_k$ , for some  $k$ . Note that  $\mathbf{I}$  must be a partition of  $\{1, \dots, K\}$  and, therefore, must satisfy  $\mathbf{I}_1 \cup \dots \cup \mathbf{I}_J = \{1, \dots, K\}$  and  $\mathbf{I}_i \cap \mathbf{I}_j = \emptyset$  for all  $i \neq j \in \mathbf{I}$ .

In some situations, the possible partition  $\mathbf{I} = (\mathbf{I}_1, \dots, \mathbf{I}_J)$ , can be established in advance, and the interest is to validate such partition by testing the null model

$$\text{logit}(\mathbf{F}, \mathbf{X}) = \alpha + \sum_{j=1}^{p-1} f_j(X_j) + \begin{cases} g_1(X_p) & \text{if } F \in \mathbf{I}_1 \\ g_2(X_p) & \text{if } F \in \mathbf{I}_2 \\ \vdots & \vdots \\ g_K(X_p) & \text{if } F \in \mathbf{I}_J \end{cases} \quad (2.3)$$

versus the general model given in (2.2). However, in general, the partitions are unknown and will have to be estimated from the data.

Before explaining the procedure for determining the groups, let us introduce some notation. Given a partition  $\mathbf{I} = (\mathbf{I}_1, \dots, \mathbf{I}_J)$ , we define a function  $G_{\mathbf{I}} : \{1, \dots, K\} \rightarrow \{1, \dots, J\}$  so  $G_{\mathbf{I}}(k) = j$  if  $k \in \mathbf{I}_j$ . Now, given a sample  $\{F_i, \mathbf{X}_i, Y_i\}_{i=1}^n$  and using the estimates  $\hat{g}_k$  ( $k = 1, \dots, K$ ) of the model in (2.2), the estimated partition  $\hat{\mathbf{I}} = (\hat{\mathbf{I}}_1, \dots, \hat{\mathbf{I}}_J)$  can be obtained by minimizing the following Cramér-von Mises type distance

$$\min_{\mathbf{I}=(\mathbf{I}_1, \dots, \mathbf{I}_J)} \sum_{k=1}^K \int_x (\hat{g}_k(x) - \hat{C}_{G_{\mathbf{I}}(k)}(x))^2 dx \quad (2.4)$$

Alternatively, the estimated partition can be obtained using the Kolmogorov-Smirnov type distance

$$\min_{\mathbf{I}=(\mathbf{I}_1, \dots, \mathbf{I}_J)} \sum_{k=1}^K \int_x |\hat{g}_k(x) - \hat{C}_{G_{\mathbf{I}}(k)}(x)| dx \quad (2.5)$$

In both cases the estimated centroids for  $j = 1, \dots, J$  are defined as

$$\hat{C}_j(x) = \frac{\sum_{k=1}^K \hat{g}_k(x) \cdot I\{G_{\mathbf{I}}(k) = j\}}{\sum_{k=1}^K I\{G_{\mathbf{I}}(k) = j\}} \quad (2.6)$$

The above minimization problems can have a high computational cost for large values of  $K$ , because they require the evaluation of all the different combinations of the  $K$  curves into  $J$  groups. To solve the quadratic minimization problem in (2.4) and L1-norm minimization problem in (2.5) we propose the use of the k-means or k-medians algorithms. In both cases, the partial functions have to be estimated in a common grid of  $X_p, x_1^* < \dots < x_N^*$ , of size  $N$ , leading to a matrix

$$\begin{pmatrix} \hat{g}_1(x_1^*) & \dots & \hat{g}_K(x_1^*) \\ \vdots & & \vdots \\ \hat{g}_1(x_N^*) & \dots & \hat{g}_K(x_N^*) \end{pmatrix}_{N \times K}$$

This matrix will be the input of both heuristic methods, k-means and k-medians, and from these, the estimated partition  $\hat{\mathbf{I}} = (\hat{\mathbf{I}}_1, \dots, \hat{\mathbf{I}}_J)$  is obtained.

### 2.3. Determining the number of groups

The procedure to obtain the estimation partition  $\hat{\mathbf{I}} = (\hat{\mathbf{I}}_1, \dots, \hat{\mathbf{I}}_J)$  for a given  $J$  ( $J < K$ ) has been explained in the previous section. This section focuses on determining the number of groups for the regression functions. With this goal in mind, one possible approach consists in fitting the model in (2.3) for each value of  $j = 1, \dots, J$ , and selecting the number of groups  $K$  that minimizes the Bayesian Information Criterion (BIC). BIC is one of the most widely used tools in statistical model selection. Its popularity is derived from its computational simplicity and effective performance in many modeling approaches. Our simulations on curve clustering demonstrate that the BIC criterion can be considered a valid option, performing quite satisfactorily choosing the correct number of clusters in simulation scenarios with a large number of curves and complex data sets. Another possibility would be to test, for a given  $J$ , the null hypothesis  $H_0(J)$  that at least one partition  $\mathbf{I}$  of length  $J$  exists, so that the model in (2.3) is fulfilled, against the alternative that for any  $J$ -partition at least a group  $\mathbf{I}_j$  exists in which  $g_m \neq g_n$  for some  $m, n \in \mathbf{I}_j$ . In order to test  $H_0(J)$ , for a given sample  $\{F_i, \mathbf{X}_i, Y_i\}_{i=1}^n$ , we consider the following two statistics

$$D_{\text{CM}} = \sum_{k=1}^K \sum_{i=1}^n \left( \hat{g}_k(X_{ip}) - \hat{g}_{G_{\hat{\mathbf{I}}}(k)}^\circ(X_{ip}) \right)^2 \quad (2.7)$$

$$D_{\text{KS}} = \sum_{k=1}^K \sum_{i=1}^n \left| \hat{g}_k(X_{ip}) - \hat{g}_{G_{\hat{\mathbf{I}}}(k)}^\circ(X_{ip}) \right| \quad (2.8)$$

where the estimated partition  $\hat{\mathbf{I}} = (\hat{\mathbf{I}}_1, \dots, \hat{\mathbf{I}}_J)$  is obtained by solving the minimization problems in (2.4) or (2.5), and  $\hat{g}_j^\circ$  functions are obtained from model (2.3) using the same partition. Alternatively, we can use the deviance as an appropriate measure of discrepancy between observed and fitted values and consider the following likelihood ratio test

$$D_{\text{LR}} = \frac{\sum_{i=1}^n \text{Dev}_i(Y_i, \hat{P}_i) - \sum_{i=1}^n \text{Dev}_i(Y_i, \hat{P}_i^\circ)}{\sum_{i=1}^n \text{Dev}_i(Y_i, \hat{P}_i)} \quad (2.9)$$

which compares the deviance of the null model (2.3) with the deviance of the general model in (2.2), where  $\hat{P}_i^\circ$  and  $\hat{P}_i$  are the estimates values of the true probabilities  $P_i = P(Y_i = 1 | \mathbf{F}_i, \mathbf{X}_i)$  under the null and general model, respectively. The individual deviance  $\text{Dev}_i(Y_i, \hat{P}_i)$  is defined as  $\text{Dev}_i(Y_i, \hat{P}_i) = -2(Y_i \log \hat{P}_i + (1 - Y_i) \log(1 - \hat{P}_i))$ .

Note that if  $H_0(J)$  is verified, the value of the test statistic  $D$  should be close to zero. The decision rule based on each of the three statistics,  $D$ , consists in rejecting the null hypothesis if  $D$  is greater than  $(1 - \alpha)$ -percentile obtained under the null hypothesis. Nevertheless, the theory for determining such percentiles is not closed. To approximate the distributions of the test statistics, resampling methods such as the bootstrap method can be applied [23–26]. The binary bootstrap used in our paper is a particular case of the bootstrap techniques suggested by [27] and [28] for inference in nonparametric models with response belonging to the binary family. The binary bootstrap involves the following steps:

**Step 1.** Using the original sample data  $\{F_i, \mathbf{X}_i, Y_i\}_{i=1}^n$ , compute the test statistics  $D$  as explained above. Then, obtain for  $i = 1, \dots, n$  the estimated  $\hat{P}_i^\circ$  of the true probabilities  $P_i = P(Y_i = 1 | \mathbf{F}_i, \mathbf{X}_i)$  obtained from the null model in (2.3).

**Step 2.** For  $b = 1, \dots, B$ , generate the bootstrap sample  $\{F_i, \mathbf{X}_i, Y_i^*\}_{i=1}^n$ , with  $Y_i^* \in \text{Bernoulli}(\hat{P}_i^c)$ , and compute the bootstrap statistics  $D^{\bullet,b}$ . Note that the estimation of the partition is required in this Step.

Finally, the decision rule, based on each test statistic  $D$ , consists in rejecting the null hypothesis if  $D > \hat{D}^{1-\alpha}$ , where  $\hat{D}^{1-\alpha}$  is the empirical  $(1-\alpha)$ -percentile of the values  $D^{\bullet,1}, \dots, D^{\bullet,B}$  obtained previously.

The bootstrap-based test introduced here can be very useful to automatically determine the number of groups  $J$ . Specifically, the rule proposed is as follows: the procedure begins by testing  $H_0(1)$ . If this hypothesis is true, we decide that  $J = 1$  (all regression curves are equal). When the previous hypothesis is rejected, it will be necessary to test  $H_0(2)$ . If this new hypothesis is not rejected, we decide that  $J = 2$ . When  $H_0(2)$  is rejected, it will be required to test  $H_0(3)$  and so on until a certain  $H_0(J)$  is not rejected. Finally, note that we are aware that this approach could deal with the problem of multiple hypothesis testing where a set of  $J$  p-values corresponding to the  $J$  null hypotheses,  $H_0(1), H_0(2), \dots, H_0(J)$ , are given. Even though several methods have been proposed to deal with this problem (see e.g. [29] for an introduction to this area), there is still a challenge because there is no information about the minimum number of tests needed to apply these techniques. Accordingly, and considering that our rule finishes with a low number of them, we have not considered this problem.

#### 2.4. Simulation study

This section reports the results of a simulation study conducted to evaluate the practical performance of the proposed methodology. The response  $Y$  was generated under model (2.3) with

$$\log \frac{P(Y = 1|F, X)}{1 - P(Y = 1|F, X)} = \begin{cases} g_1(X) = 1.5X & \text{if } F \in \mathbf{I}_1 = \{1, 2, 3\} \\ g_2(X) = 2X^2 - 3 & \text{if } F \in \mathbf{I}_2 = \{4, 5\} \\ g_3(X) = 2 \sin(2X) - 2 & \text{if } F \in \mathbf{I}_3 = \{6, 7\} \\ g_4(X) = 2 \cos(2X) & \text{if } F \in \mathbf{I}_4 = \{8, 9\} \\ g_5(X) = 2 \cos(2X) + a \exp(0.5X) & \text{if } F \in \mathbf{I}_5 = \{10, 11, 12, 13\} \\ g_6(X) = 4 - 2X & \text{if } F \in \mathbf{I}_6 = \{14, 15\} \end{cases} \quad (2.10)$$

The categorical covariate  $X$  was drawn from an uniform distribution  $U[-2, 2]$ , and its factors,  $F$ , were generated by taking a random value in  $\{1, \dots, 15\}$  with associated probabilities  $\mathbf{p} = (p_1, \dots, p_{15}) = \mathbf{n} / \sum n_k, k = 1, \dots, 15$ , where:

$$\mathbf{n} = (n_1, \dots, n_{15}) = (1.0, 2.0, 1.0, 1.5, 1.0, 2.0, 1.5, 2.0, 2.0, 1.0, 1.0, 1.0, 1.5, 1.5, 1.0)$$

In this way, and as can be seen in Table 1, we have considered unequal sample sizes for each ( $k = 1, \dots, 15$ ) curve.

We explore the validity of the proposed tests assuming that we aim to test if the  $K = 15$  regression curves can be grouped in five ( $J = 5$ ) groups. To study the size and power of the proposed tests, different values were considered for  $a$ , ranging from 0 to 2. It should be noted that  $g_5(X) = g_4(X) + a \exp(0.5X)$ , so a value of  $a = 0$  corresponds to a null hypothesis of five groups, while a value of  $a > 0$  corresponds to a null hypothesis of six groups.

#### 2.5. Application to neural activity in the prefrontal cortex during a discrimination task

The data analyzed in this section come from laboratory experiments of the extra-cellular single unit activity in the prefrontal cortex of a monkey. These experiments were conducted in the Laboratory of

**Table 1.** Sample size  $n_k$  for each regression curve  $k$  ( $k = 1, \dots, 15$ ) for different total sample sizes  $n = \sum n_k$ .

$n$	$n_1$	$n_2$	$n_3$	$n_4$	$n_5$	$n_6$	$n_7$	$n_8$	$n_9$	$n_{10}$	$n_{11}$	$n_{12}$	$n_{13}$	$n_{14}$	$n_{15}$
300	14	28	14	21	14	28	21	28	28	14	14	14	21	21	14
500	23	47	23	35	23	47	35	47	47	23	23	23	35	35	23
1000	47	95	47	71	47	95	71	95	95	47	47	47	71	71	47
2000	95	190	95	142	95	190	142	190	190	95	95	95	142	142	95
4000	190	380	190	285	190	380	285	380	380	190	190	190	285	285	190

Neurophysiology of the University of Vigo (Spain). The monkey was trained to discriminate between different stimuli. The stimuli consisted of stationary bright line segments presented on a monitor screen in front of the monkey that changed their orientation over time.

A trial was initiated when the monkey pressed a lever key with its right hand and then two stimuli (reference and test), each of 500 ms duration, were presented in sequence with a fixed inter-stimulus interval (ISI: 1000 ms). At the end of the second stimulus, the subject released the key in a 1200 ms time window and pressed one of the two switches (left or right), indicating whether the orientation of the second stimulus was clockwise (right) or counter-clockwise (left) to the reference stimulus. While the monkey worked on the task, its extracellular unit activity was recorded. The monkey was rewarded for correct discrimination.

To gather enough data and to account for the cell response variability, the neuron was recorded over a number of  $T = 80$  trials. For each of the trials, we consider the angle (trial specific),  $TA$ , corresponding to the test stimulus. The reference angle is  $90^\circ$ . The following  $TAs$  were considered in the experiment:

- $TA \in \{78^\circ, 102^\circ\}$ : test stimuli more separated from the reference, and therefore very easy to discriminate
- $TA \in \{81^\circ, 99^\circ\}$ : test stimuli is easy to discriminate
- $TA \in \{84^\circ, 96^\circ\}$ : test stimuli is difficult to discriminate
- $TA \in \{87^\circ, 93^\circ\}$ : test stimuli closest to the reference, therefore more difficult to discriminate

In this experiment, the outcome of interest is the neuronal activity in the interval  $t \in [t_{min}, t_{max}] = [-500, 3000]$  in ms. At each instant  $t$  and trial  $j = 1, \dots, T$ , this outcome may then be represented by a temporal binary sequence,  $Y_t^j$ , where  $Y_t^j = 1$  if there is a spike in  $[t, t + 1)$  ms and 0 otherwise. The explanatory variables are the time  $t$  when a spike was detected and the interval of time  $\Delta t$  since the last spike. The last one is an adjustment variable that takes into account that the neuron might fire more easily if it has been fired recently. Accordingly, for each trial, the data set consists of the following information

$$\{TA, (t, \Delta t), Y_t\}_{t=t_{min}}^{t_{max}}$$

We use generalized additive modeling to assess whether the association between neural activity and decision-making depends on the difficulty of the discrimination task related to the angle of the test stimulus. We consider the following GAM:

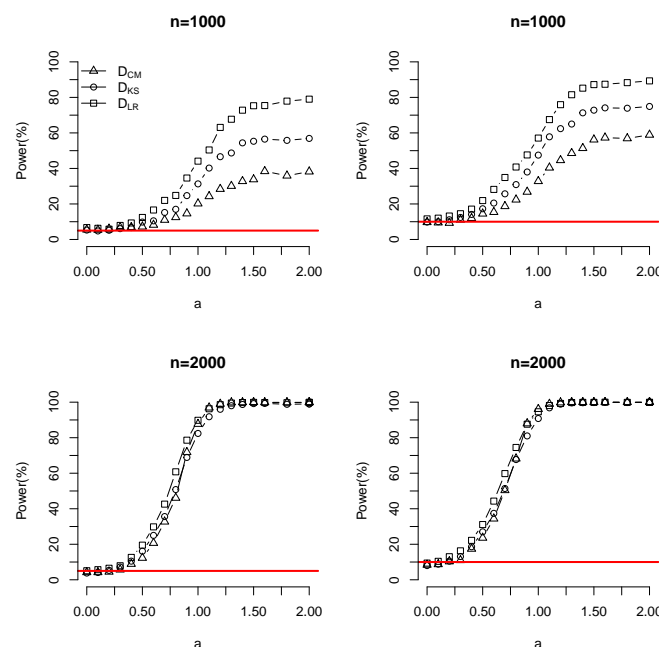
$$\log\left(\frac{P(TA, t, \Delta t)}{1 - P(TA, t, \Delta t)}\right) = \alpha + f(\Delta t) + \begin{cases} g_1(t) & \text{if } TA = 78^\circ \\ g_2(t) & \text{if } TA = 102^\circ \\ g_3(t) & \text{if } TA = 81^\circ \\ g_4(t) & \text{if } TA = 99^\circ \\ g_5(t) & \text{if } TA = 84^\circ \\ g_6(t) & \text{if } TA = 96^\circ \\ g_7(t) & \text{if } TA = 87^\circ \\ g_8(t) & \text{if } TA = 93^\circ \end{cases} \quad (2.11)$$

where  $P(TA, t, \Delta t) = P(Y_t = 1|F, t, \Delta t)$ ,  $\alpha$  is a fixed parameter, and  $g_k$  for  $k = 1, \dots, 8$  are the specific time functions associated to each of the test angles considered in the study.

### 3. Results

#### 3.1. Results on simulation data set

Type I error rates and power values of the proposed tests as a function of  $a$  are shown in Figure 1. They were calculated from 1000 simulation runs using  $B = 400$  bootstrap samples at each repetition. We compare the results of the three test statistics under sample sizes  $n = 1000$  and  $2000$ , at the significance levels of 0.05 and 0.10. As can be seen in Figure 1, the three curves show the expected behavior pattern, with an increase in the power as  $a$  increases, and an improvement in it as the sample size grows.



**Figure 1.** Rejection probabilities (power of the hypothesis test) of the proposed test statistics as a function of  $a$ , for sample size  $n = 1000$  and  $n = 2000$  (upper and lower panels, respectively) at a 0.05 and 0.10 (left and right panels, respectively) significant levels (red line).



The test statistic based on the likelihood ratio test (labeled as  $D_{LR}$ ) is that with the highest values of power for  $n = 1000$ , but the differences between the three test statistics are minimal for  $n = 2000$ . The Cramér-von Mises-type statistic test (labeled as  $D_{CM}$ ) revealed a poorer behavior with lower rejection probabilities for a sample size of 1000. Values for type I error, for the three test statistics and for different significance levels and different sample sizes ( $n = 1000, 2000$ , and  $3000$ ), are reported in Table 2. The three test statistics work satisfactorily according to the type I error, coming quite close to the nominal level regardless of the sample size.

**Table 2.** Estimated type I error (in %) and nominal level percentages (1, 5, 10, 15, and 20) for different sample sizes.

	level	$D_{CM}$	$D_{KS}$	$D_{LR}$
n = 1000	1	1.4	1.3	1.0
	5	5.5	5.3	5.6
	10	9.7	9.8	11.5
	15	14.5	15.3	16.5
	20	18.7	20.0	21.1
n = 2000	1	0.9	0.7	0.8
	5	4.2	3.8	5.7
	10	8.4	8.0	9.4
	15	12.8	12.4	15.6
	20	16.5	18.0	21.8
n = 3000	1	0.9	0.7	1.2
	5	4.2	4.0	6.5
	10	8.8	7.6	11.1
	15	13.0	12.1	16.2
	20	16.2	17.0	22.3

In Table 3, we report results that can be used to evaluate the accuracy of the bootstrap-based algorithm introduced in Section 2.3. Again, different values of  $a$  were considered, ranging from 0 to 2. Recall that the value  $a = 0$  corresponds to the null hypothesis, which assumes that the fifteen regression functions can be classified into five groups, while, when  $a \neq 0$ , the number of groups is six. Note that to select the correct number  $J$  of groups of regression functions, the bootstrap-based algorithm must first reject the first null hypothesis,  $H_0(1)$ , then reject the second hypothesis,  $H_0(2)$ , and so on until it accepts  $H_0(5)$  if  $a = 0$ , or until  $H_0(6)$  when  $a > 0$ . Results shown in Table 3 for the three test statistics, using a nominal level of 5%, display the number of times that the procedure selects the number of groups  $J$ . Results in bold denote the correct classifications according to model (2.3), revealing the high accuracy of the proposed bootstrap-based algorithm for  $a = 0$ , showing the correct number of groups in percentages quite close to the nominal level. As the value of  $a$  increases, so does the percentage of cases in which the proposed method suggests six groups. When comparing the three test statistics, it can be observed that all have a similar performance with a small advantage for the test statistic based on the likelihood ratio test (labeled as  $D_{LR}$ ).

**Table 3.** Percentage of simulations (over 1000 repetitions) that the bootstrap-based algorithm estimated the number of groups  $J$ . Results for the three test statistics were obtained for sample size  $n = 2000$  and different values of  $a$ , with a nominal level of 5%.

$a$	$D_{CM}$			$D_{KS}$			$D_{LR}$		
	$J = 5$	$J = 6$	$J = 7$	$J = 5$	$J = 6$	$J = 7$	$J = 5$	$J = 6$	$J = 7$
0.0	<b>95.8</b>	3.1	1.2	<b>96.2</b>	2.5	1.3	<b>94.3</b>	3.4	2.3
0.2	95.5	<b>3.2</b>	1.3	94.7	<b>3.8</b>	1.6	94.2	<b>3.2</b>	2.6
0.4	91.1	<b>6.1</b>	2.8	89.5	<b>7.6</b>	2.9	87.6	<b>7.9</b>	4.5
0.6	79.3	<b>17.6</b>	3.1	75	<b>21.2</b>	3.8	72.4	<b>22.2</b>	5.4
0.8	53.9	<b>42.2</b>	3.9	49.2	<b>46.7</b>	4.1	40.7	<b>53.9</b>	5.4
1.0	12.3	<b>83</b>	4.7	17.6	<b>78.5</b>	3.8	10.2	<b>84.5</b>	5.3
1.2	1	<b>94.5</b>	4.5	4	<b>91.8</b>	4.2	1.4	<b>93.6</b>	4.9
1.4	0.1	<b>95.3</b>	4.6	1.3	<b>95.2</b>	3.5	0.2	<b>94.6</b>	5.2
1.6	0.1	<b>96.1</b>	3.8	0.7	<b>95.4</b>	3.9	0.2	<b>94.8</b>	5.1
1.8	0.1	<b>95.7</b>	4.2	1.2	<b>94.7</b>	4.1	0.3	<b>94.9</b>	4.8
2.0	0.1	<b>96.5</b>	3.4	1	<b>95.4</b>	3.7	0.2	<b>94.6</b>	5.2

According to our simulation scenario, the  $K = 15$  regression curves should be grouped into five groups if  $a = 0$  and six groups otherwise. In addition to this, the assignment of each curve should follow model (2.3). Results reported in Table 4 can be used to check if the proposed algorithm is performing well by assigning each of the fifteen regression curves to the correct group. Specifically, the results correspond to a value of  $a = 0.7$ , for different sample sizes. The entries in the first column specify the number of regression curves with an incorrect classification, whereas the remaining columns show the corresponding percentages for different sample sizes. Therefore, the results reported in the first line correspond to the ideal situation in which none of the fifteen curves has an incorrect classification. As we can see, the percentages in this line increase with an increase in the sample size, up to rates of success of around 95%, coming quite close to that established. In addition, we can also see that the number of incorrect classifications decreases with an increase in the sample size. For  $n = 2000$ , for example, the percentages of incorrect classifications have a sharp decrease whereas for  $n = 4000$  only 1 or 2 curves have incorrect classifications with low percentage values of 4.9 and 0.6%, respectively.

To measure how well the procedure assigns each of the  $k = 1, \dots, 15$  to their correct group  $I_1, \dots, I_6$ , we report in Table 4 the percentage of incorrect classifications for each regression curve, for different sample sizes, considering  $a = 0.7$ . While it can be seen that the procedure performs quite well for large sample sizes, it also reveals some difficulties in the case of small samples.

Note that, according to Table 1, for a sample size of  $n = 300$  the number of observations in each group is between 14 and 28. This can explain the high misclassification rates for this sample size, in particular in the case of regression curves with fewer observations. As the sample size increases, the percentage of incorrect classifications for each regression curve decreases.

In Table 5 we show which group each of the  $k = 1, \dots, 15$  regression curves is assigned to. Results reported in this table were obtained for  $a = 0.7$  and four sample sizes. The numbers in bold denote the correct classifications according to model (2.3). The results are in agreement with the previous findings, revealing the accuracy of the proposed method for higher sample sizes. We can also see that

for all the samples, the assignment is, in most cases, correct. As expected, higher misclassification rates can be observed for those regression curves ( $k = 10, 11, 12, 13$ ) that belong to group  $\mathbf{I}_5$  but were classified in group  $\mathbf{I}_4$ . Results not reported here show that this misclassification is less evident as the value of  $a$  increases.

**Table 4.** Percentages of misclassifications over the total of the 15 regression curves. Scenario with  $a = 0.7$  for different sample sizes.

n.misclass	$n = 300$	$n = 500$	$n = 1000$	$n = 2000$	$n = 4000$
0	0.3	0.9	20.4	67.3	94.6
1	2.7	6.1	26.4	22.7	4.9
2	5.1	12.0	15.0	4.3	0.6
3	12.9	32.9	26.4	4.9	0.0
4	15.4	16.0	2.6	0.1	0.0
5	18.1	14.0	4.1	0.3	0.0
6	17.0	8.1	3.1	0.1	0.0
7	10.4	4.9	0.6	0.0	0.0
8	8.9	3.6	1.1	0.3	0.0
9	5.6	1.0	0.1	0.0	0.0
10	2.3	0.4	0.0	0.0	0.0
11	1.1	0.1	0.0	0.0	0.0
12	0.1	0.0	0.0	0.0	0.0
13	0.0	0.0	0.0	0.0	0.0
14	0.0	0.0	0.0	0.0	0.0
15	0.0	0.0	0.0	0.0	0.0

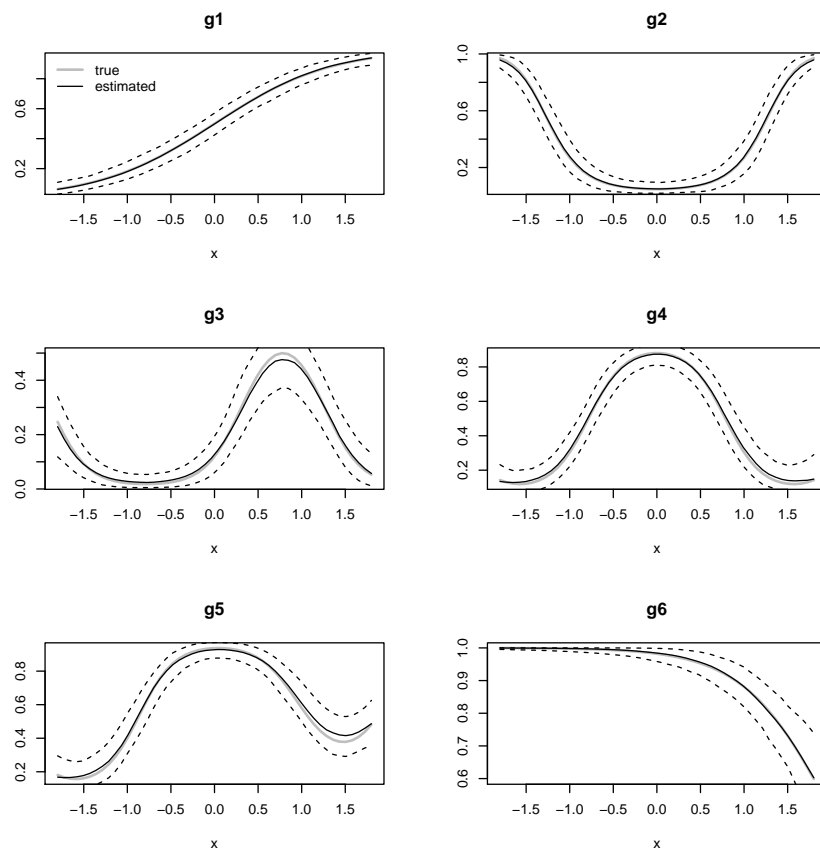
The success of the proposed methods depends on recovering the true functional forms of the corresponding six curves  $\hat{g}_k$ ,  $k = 1, \dots, 6$ , introduced in model (2.3). In this regard, we present in Figure 2 the estimated curves with the corresponding 95% confidence limits for a sample size of  $n = 2000$ . This figure reveals that the resulting estimates recover the functional form of the corresponding true curves very successfully.

As mentioned above, the Bayesian Information Criterion (BIC) can be used to select the number of groups  $J$ . This is a method of computational simplicity that consists of choosing the model (2.3) that minimizes the BIC statistics. Table 6 can be used to compare the performance of the BIC against the results reported in Table 3, based on the bootstrap method introduced in Section 3. Both tables report the percentage of simulations (over 1000 repetitions) that the BIC procedure and the bootstrap-based algorithm estimated to have a number of groups  $J$  for different values of  $a$ . BIC is able to identify the correct number of groups (five when  $a = 0$  and six otherwise) for sample sizes  $n \geq 1000$  (correct classifications represented in bold). Even though the results shown in the two tables indicate good agreement between the two procedures, one can see that the BIC procedure is more conservative in detecting important differences between the regression curves, whereas, the bootstrap method is more effective in detecting less important differences. Since the BIC procedure has obvious computational advantages, we suggest that it should be used to contrast and thoroughly examine conclusions obtained by applying the proposed bootstrap method. We should not consider the BIC procedure as merely an

alternative to the bootstrap but rather as a supplement that offers additional information. In this sense, BIC could be used to obtain an initial approximation to the number of groups, then our bootstrapping method will be used to tune this value.

**Table 5.** Assignment of curves to groups. The entries in the table specify the percentage of times that curve  $K$  ( $k = 1, \dots, 15$ ) is classified in cluster  $\mathbf{I}_j$  ( $j = 1, \dots, 6$ ). Results for four different sample sizes  $n$  with  $a = 0.7$ .

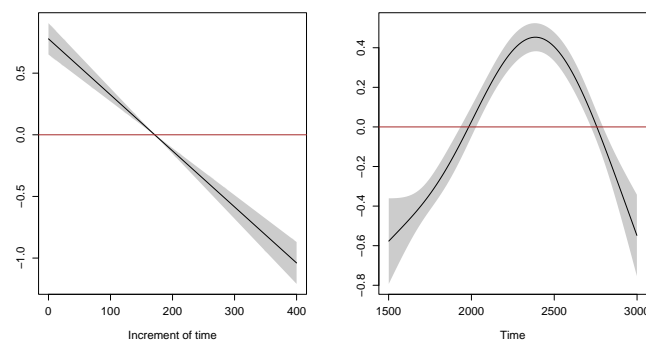
$K$	$n = 500$						$n = 1000$					
	$\mathbf{I}_1$	$\mathbf{I}_2$	$\mathbf{I}_3$	$\mathbf{I}_4$	$\mathbf{I}_5$	$\mathbf{I}_6$	$\mathbf{I}_1$	$\mathbf{I}_2$	$\mathbf{I}_3$	$\mathbf{I}_4$	$\mathbf{I}_5$	$\mathbf{I}_6$
1	<b>80.9</b>	1.8	5.2	4.1	7.6	0.3	<b>95.3</b>	0.0	1.0	0.8	2.9	0.0
2	<b>82.4</b>	1.0	4.0	4.2	8.4	0.0	<b>94.6</b>	0.0	0.4	1.0	4.1	0.0
3	<b>79.7</b>	3.5	4.8	4.5	7.1	0.3	<b>94.5</b>	0.2	0.7	1.0	3.6	0.0
4	2.0	<b>89.9</b>	5.8	1.4	0.8	0.0	0.5	<b>97.0</b>	0.5	0.7	1.3	0.0
5	1.6	<b>89.2</b>	5.8	2.3	0.7	0.4	0.6	<b>97.8</b>	0.7	0.6	0.2	0.0
6	3.4	1.8	<b>89.8</b>	3.4	1.6	0.0	0.2	0.1	<b>97.7</b>	1.4	0.5	0.0
7	3.0	2.3	<b>89.9</b>	2.8	2.0	0.0	0.0	0.1	<b>97.7</b>	1.6	0.6	0.0
8	0.3	0.3	4.1	<b>66.9</b>	28.5	0.0	0.0	0.0	0.0	<b>83.5</b>	16.5	0.0
9	0.4	0.1	1.8	<b>70.3</b>	27.3	0.0	0.0	0.0	0.1	<b>84.4</b>	15.5	0.0
10	12.6	0.4	0.7	35.1	<b>46.0</b>	5.1	2.6	0.0	0.1	32.6	<b>64.1</b>	0.5
11	10.9	0.1	1.6	38.8	<b>43.2</b>	5.4	2.4	0.0	0.0	34.4	<b>62.8</b>	0.4
12	10.8	0.4	2.0	37.7	<b>44.8</b>	4.4	3.0	0.0	0.0	32.0	<b>64.7</b>	0.2
13	9.1	0.1	0.4	41.9	<b>44.9</b>	3.5	0.8	0.0	0.0	35.4	<b>63.8</b>	0.0
14	0.3	0.1	0.0	0.8	2.1	<b>96.6</b>	0.0	0.0	0.0	0.0	0.4	<b>99.6</b>
15	0.1	0.1	0.0	0.6	2.1	<b>97.0</b>	0.0	0.0	0.0	0.1	0.5	<b>99.4</b>
$K$	$n = 2000$						$n = 4000$					
	$\mathbf{I}_1$	$\mathbf{I}_2$	$\mathbf{I}_3$	$\mathbf{I}_4$	$\mathbf{I}_5$	$\mathbf{I}_6$	$\mathbf{I}_1$	$\mathbf{I}_2$	$\mathbf{I}_3$	$\mathbf{I}_4$	$\mathbf{I}_5$	$\mathbf{I}_6$
1	<b>99.6</b>	0.0	0.1	0.1	0.1	0.0	<b>100.0</b>	0.0	0.0	0.0	0.0	0.0
2	<b>99.5</b>	0.0	0.1	0.1	0.3	0.0	<b>100.0</b>	0.0	0.0	0.0	0.0	0.0
3	<b>99.5</b>	0.1	0.1	0.1	0.1	0.0	<b>100.0</b>	0.0	0.0	0.0	0.0	0.0
4	0.1	<b>99.7</b>	0.0	0.0	0.1	0.0	0.0	<b>100.0</b>	0.0	0.0	0.0	0.0
5	0.1	<b>99.7</b>	0.0	0.0	0.1	0.0	0.0	<b>100.0</b>	0.0	0.0	0.0	0.0
6	0.0	0.0	<b>99.9</b>	0.1	0.0	0.0	0.0	0.0	<b>100.0</b>	0.0	0.0	0.0
7	0.0	0.0	<b>99.9</b>	0.1	0.0	0.0	0.0	0.0	<b>100.0</b>	0.0	0.0	0.0
8	0.0	0.0	0.0	<b>98.2</b>	1.8	0.0	0.0	0.0	0.0	<b>100.0</b>	0.0	0.0
9	0.0	0.0	0.0	<b>98.1</b>	1.9	0.0	0.0	0.0	0.0	<b>100.0</b>	0.0	0.0
10	0.0	0.0	0.0	12.0	<b>88.0</b>	0.0	0.0	0.0	0.0	1.3	<b>98.7</b>	0.0
11	0.3	0.0	0.0	10.4	<b>89.4</b>	0.0	0.0	0.0	0.0	1.7	<b>98.3</b>	0.0
12	0.4	0.0	0.0	12.1	<b>87.4</b>	0.1	0.0	0.0	0.0	1.9	<b>98.1</b>	0.0
13	0.0	0.0	0.0	10.2	<b>89.8</b>	0.0	0.0	0.0	0.0	1.1	<b>98.9</b>	0.0
14	0.0	0.0	0.0	0.0	0.0	<b>100.0</b>	0.0	0.0	0.0	0.0	0.0	<b>100.0</b>
15	0.0	0.0	0.0	0.0	0.1	<b>99.9</b>	0.0	0.0	0.0	0.0	0.0	<b>100.0</b>



**Figure 2.** Estimated centroids  $g_j$  ( $j = 1, \dots, 6$ ) for the six clusters with the 95% confidence limits. Results obtained from  $n = 2000$  and  $a = 0.7$ .

### 3.2. Results on real data set

Figure 3 shows the effects of both explanatory variables,  $t$  and  $\Delta t$ , on the response variable, applying model 2.11 to the experimental data.



**Figure 3.** Effect of each explanatory variable on the probability of spike.

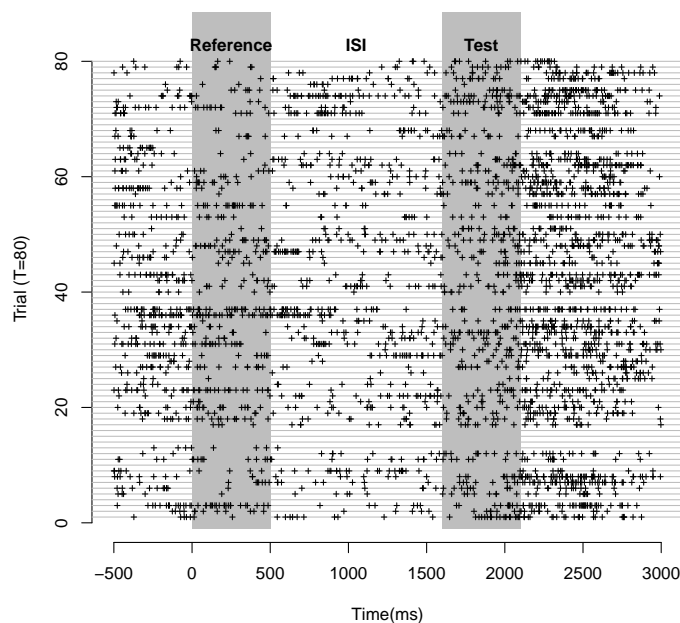
**Table 6.** Percentage of simulations (over 1000 repetitions) that the BIC procedure estimated the number of groups  $J$ . Results for different values of  $a$ .

	$a$	Number of selected groups							
		$J = 1$	$J = 2$	$J = 3$	$J = 4$	$J = 5$	$J = 6$	$J = 7$	$J > 7$
$n = 500$	0.0	0.0	0.0	3.4	33.3	<b>59.1</b>	4.2	0.0	0.0
	0.2	0.0	0.0	3.9	28.8	62.1	<b>4.8</b>	0.3	0.1
	0.4	0.0	0.1	4.7	28.7	61.2	<b>5.1</b>	0.2	0.0
	0.6	0.0	0.1	5.0	28.8	58.5	<b>7.4</b>	0.2	0.0
	0.8	0.0	0.3	7.9	31.4	50.8	<b>8.8</b>	0.7	0.1
	1.0	0.0	0.3	9.0	30.1	45.0	<b>13.7</b>	1.8	0.1
	1.2	0.0	0.5	8.0	31.9	39.1	<b>18.2</b>	2.1	0.2
	1.4	0.0	0.6	9.6	27.0	34.8	<b>24.8</b>	2.8	0.4
	1.6	0.0	0.2	7.8	24.8	35.2	<b>28.0</b>	3.6	0.4
	1.8	0.0	0.4	6.8	24.0	35.8	<b>29.2</b>	3.2	0.6
	2.0	0.0	0.8	4.8	21.6	36.8	<b>32.4</b>	3.6	0.0
$n = 1000$	0.0	0.0	0.0	0.0	1.9	<b>97.3</b>	0.7	0.1	0.0
	0.2	0.0	0.0	0.0	1.8	97.3	<b>0.9</b>	0.0	0.0
	0.4	0.0	0.0	0.0	1.3	97.2	<b>1.5</b>	0.0	0.0
	0.6	0.0	0.0	0.0	1.2	94.8	<b>3.8</b>	0.2	0.0
	0.8	0.0	0.0	0.0	2.1	84.0	<b>13.5</b>	0.4	0.0
	1.0	0.0	0.0	0.0	1.2	60.4	<b>37.4</b>	1.0	0.0
	1.2	0.0	0.0	0.0	1.7	29.0	<b>67.9</b>	1.4	0.0
	1.4	0.0	0.0	0.0	1.4	16.4	<b>79.8</b>	2.4	0.0
	1.6	0.0	0.0	0.1	0.7	10.4	<b>85.3</b>	3.4	0.1
	1.8	0.0	0.0	0.0	0.2	11.0	<b>85.3</b>	3.2	0.3
	2.0	0.0	0.0	0.0	0.6	10.3	<b>85.2</b>	3.4	0.5
$n = 4000$	0.0	0.0	0.0	0.0	0.0	<b>99.9</b>	0.1	0.0	0.0
	0.2	0.0	0.0	0.0	0.0	99.7	<b>0.3</b>	0.0	0.0
	0.4	0.0	0.0	0.0	0.0	99.3	<b>0.7</b>	0.0	0.0
	0.6	0.0	0.0	0.0	0.0	80.2	<b>19.8</b>	0.0	0.0
	0.8	0.0	0.0	0.0	0.0	13.9	<b>86.1</b>	0.0	0.0
	1.0	0.0	0.0	0.0	0.0	0.2	<b>99.8</b>	0.0	0.0
	1.2	0.0	0.0	0.0	0.0	0.0	<b>100.0</b>	0.0	0.0
	1.4	0.0	0.0	0.0	0.0	0.0	<b>100.0</b>	0.0	0.0
	1.6	0.0	0.0	0.0	0.0	0.0	<b>100.0</b>	0.0	0.0
	1.8	0.0	0.0	0.0	0.0	0.0	<b>100.0</b>	0.0	0.0
	2.0	0.0	0.0	0.0	0.0	0.0	<b>100.0</b>	0.0	0.0

As can be appreciated, the neuron activation decreases with the time interval  $\Delta t$ . The activation time  $t$ , for its part, tends to increase during the period when the angle of the rod is changed, reaching a maximum at about 2200 ms and then decreasing. This means that it takes time for the neuron to fire

from the moment the monkey sees the rod.

Figure 4 shows the raster plot of the response of the neuron recorded in the prefrontal cortex of a monkey while it is carrying out the visual discrimination task. Each row represents one of the total of 80 trials, and each tick represents a potential action (spike). The spontaneous cell activity, i.e., in the absence of any stimuli, is very irregular, as can be seen in the interval from -500 to 0 ms, where there was no stimulus for the neuron.

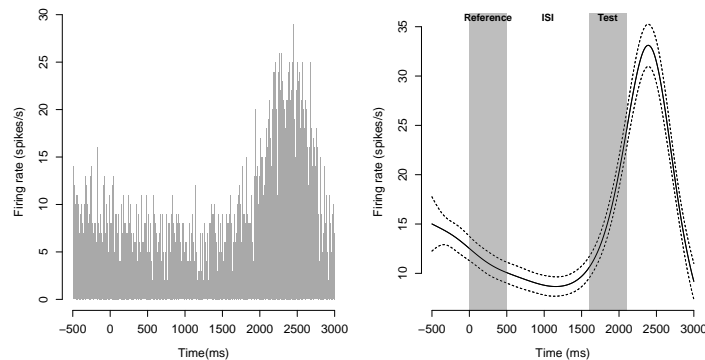


**Figure 4.** Response of a neuron during the performance of the discrimination task. The observed spikes for each of the 80 trials are on separate lines. The shaded regions correspond to reference stimulus (from 0 to 500 ms) and to the interval to test the response (1600–2100 ms).

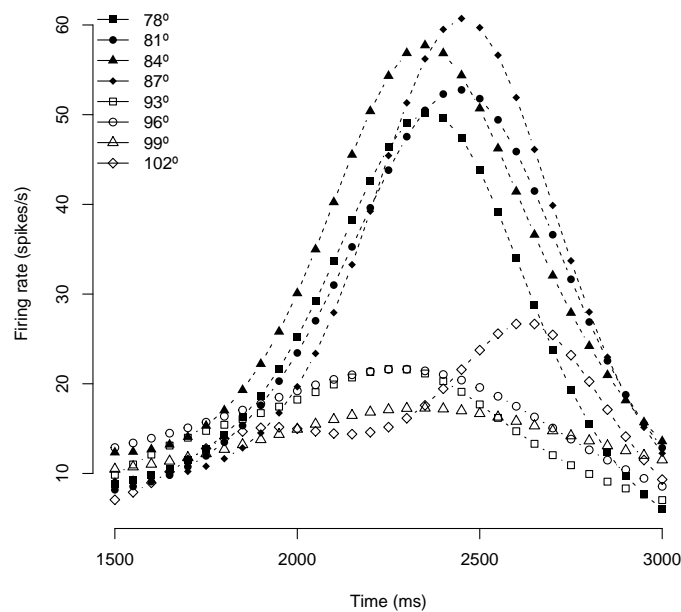
The panel on the left of Figure 5 shows the data pooled across trials represented as counts of spikes within 10 ms intervals. Units of mean firing rate are the number of spikes per trial per second. The fitted curve, obtained using a generalized additive model, is shown in the panel on the right. As can be seen in both plots, the temporal evolution of the cell firing rate during the task indicates that there is an increase in the firing rate between 1750 and 2750 ms, which corresponds to the presentation of the test stimulus and the reaction time, before the monkey motor response. Therefore, we restricted the data to the time interval between 1500 and 3000 ms. One hundred milliseconds before the test stimuli were taken as control because there were no stimuli present and the firing rate variability from trial to trial was very low. Accordingly, for each trial, 1500 measurements of the neuron activity were made.

Figure 6 shows the estimated regression curves of the firing rate (spikes) for the eight different experimental conditions, i.e., for each of the eight levels of the trial angle (TA) as defined in the model (2.11). Two groups stand out, particularly those with a trial angle below  $90^\circ$ , with higher firing rates per second in the time interval between 1900 and 2700 ms, and those with a trial angle above  $90^\circ$ , with lower firing rate per second. Although it may appear strange, higher neuron firing rates for angles less

than  $90^\circ$  (clockwise rotation) were previously reported in a study [26].



**Figure 5.** Representation of neural activity over time. Observed spikes pool counts of spikes within 10 ms intervals (left) and the corresponding smoothing version (right plot).



**Figure 6.** Fitted regression curves for the eight different experimental conditions.

Results for the bootstrap method using the three test statistics defined in Section 3 are shown in Table 7. As can be seen, the three test statistics indicate that there are two groups.  $D_{KD}$ , and slightly less so  $D_{CM}$ , are close to tipping the balance in favor of three groups.

We have also applied the BIC criterion to determine the number of groups. The results of this cluster solution also lead to the same conclusion, namely,  $J = 2$  groups. These findings are in agreement with those reported in our simulations that indicate that BIC is more conservative, but capable of detecting

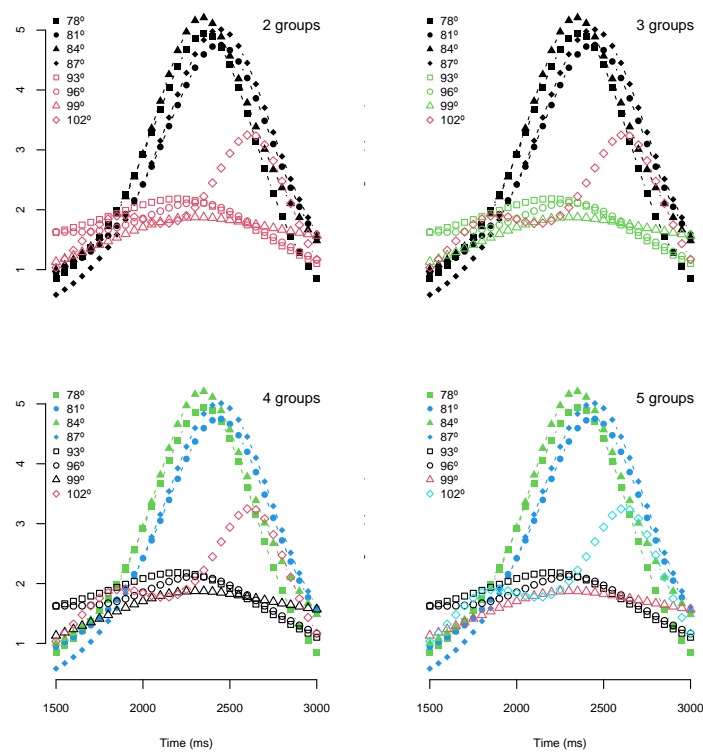


important differences.

**Table 7.** Probability values for testing the null hypothesis  $H_0(J)$ . Results based on the bootstrap method with different test statistics.

$J$	$D_{CM}$	$D_{KS}$	$D_{LR}$
1	< 0.001	< 0.001	< 0.001
2	< 0.06	< 0.05	< 0.29
3	< 0.19	< 0.13	< 0.53
4	< 0.09	< 0.76	< 0.48

Clusters for a fixed number of groups between  $J = 2$  and  $J = 5$  are shown in Figure 7 to facilitate the comprehension of the problem, although the results obtained indicate that there are only two groups.



**Figure 7.** Estimated regression curves according to the groups to which they belong. The curves assigned into two ( $J = 2$ ), until five groups ( $J = 5$ ) are represented in each of the four panels. The statistics used to estimate the number of groups found only two.

#### 4. Conclusions

In this work we propose a bootstrap-based method to determine the number of groups in generalized additive models when the effect of a continuous covariate on the response varies across groups defined by levels of a categorical variable. The simulation analysis confirms the capacity of the three proposed statistical tests to reproduce the theoretical values of type I and type II errors, even in the presence of a

large number of curves. However, the power turned out to be higher for the test based on the likelihood test ratio. As expected, the percentage of misclassifications decreases with the sample size,  $n$ , being almost null for  $n = 4000$ .

The simulated data were also used to compare our method with BIC, a well-known approach that has been largely used previously in clustering analysis to estimate the number of clusters in an artificial data set. The number of detected groups was the same for both methods, but the simulation study showed that BIC has a lower statistical power than the bootstrap method, although they are close for larger sample sizes. Accordingly, BIC could be used to obtain an initial estimation of the number of clusters that could be tuned using our method.

The application of the proposed method to an experiment to determine the activity of a neuron of a monkey subject to visual stimuli led to the same conclusions when we compared our approach with BIC. In both cases the neuron only reacts to two of the eight stimuli, corresponding to the most evident differences with respect to a reference state. Regarding the three statistics tested, they provide the same result, although the evidence for two clusters is stronger for the statistic based on the likelihood ratio test.

Finally, it can be said that although the proposed method was designed to detect groups of regression curves in generalized additive models with a binary response, it can be extended without much difficulty to determine groups in models with another kind of response.

## Acknowledgments

This work was partially supported by project 2017/00001/006/001/097: *Ayudas para el mantenimiento de actividades de investigación de institutos universitarios de investigación y grupos de investigación de la Universidad de Oviedo para el ejercicio 2021*.

Luís Meira-Machado acknowledges financial support from Portuguese Funds through FCT - "Fundação para a Ciência e a Tecnologia", within the projects UIDB/00013/2020, UIDP/00013/2020.

Javier Roca-Pardiñas acknowledges financial support from Grant PID2020-118101GB-I00, Ministerio de Ciencia e Innovación (MCIN/ AEI /10.13039/501100011033).

## Conflict of interest

The authors declare there is no conflict of interest.

## References

1. P. McCullagh, J. Nelder, *Generalized Linear Models*, 2nd edition, Chapman and Hall/CRC, Boca Raton, 1989. <https://doi.org/10.1201/9780203753736>
2. T. J. Hastie, R. J. Tibshirani, *Generalized Additive Models*, Chapman & Hall/CRC, New York, 1990.
3. S. Wood, *Generalized Additive Models: An Introduction with R*, Chapman & Hall/CRC, 2006. <https://doi.org/10.1201/9781420010404>
4. W. González-Manteiga, R. M. Crujeiras, An updated review of Goodness-of-Fit tests for regression models, *Test*, **22** (2013), 361–411. <https://doi.org/10.1007/s11749-013-0327-5>

5. H. Dette, A. Munk, Testing heterocedasticity in nonparametric regression, *J. R. Stat. Soc. B*, **60** (1998), 693–708. <https://doi.org/10.1111/1467-9868.00149>
6. H. Dette, N. Neumeier, Nonparametric analysis of covariance, *Ann. Stat.*, **29** (2001), 1361–1400. <https://doi.org/10.1214/aos/1013203458>
7. L. García-Escudero, A. Gordaliza, A proposal for robust curve clustering, *J. Classif.*, **22** (2005), 185–201. <https://doi.org/10.1007/s00357-005-0013-8>
8. E. A. Nadaraya, On estimating regression, *Theory Probab. Its Appl.*, **9** (1964), 141–142. <https://doi.org/10.1137/1109020>
9. M. A. Delgado, Testing the equality of nonparametric regression curves, *Stat. Probab. Lett.*, **17** (1993), 199–204. [https://doi.org/10.1016/0167-7152\(93\)90167-H](https://doi.org/10.1016/0167-7152(93)90167-H)
10. K. B. Kulasekera, Comparison of regression curves using quasi-residuals, *J. Am. Stat. Assoc.*, **90** (1995), 1085–1093. <https://doi.org/10.1080/01621459.1995.10476611>
11. K. B. Kulasekera, J. Wang, Smoothing parameter selection for power optimality in testing of regression curves, *J. Am. Stat. Assoc.*, **92** (1997), 500–511. <https://doi.org/10.1080/01621459.1997.10474003>
12. K. B. Kulasekera, J. Wang, Bandwidth selection for power optimality in a test of equality of regression curves, *Stat. Probab. Lett.*, **37** (1998), 287–293. [https://doi.org/10.1016/S0167-7152\(97\)84155-7](https://doi.org/10.1016/S0167-7152(97)84155-7)
13. N. Neumeier, H. Dette, Nonparametric comparison of regression curves: An empirical process approach, *Ann. Stat.*, **31** (2003), 31880–31920.
14. J. C. Pardo-Fernández, I. Keilegom, W. González-Manteiga, Testing for the equality of k regression curves, *Stat. Sin.*, **17** (2007), 1115–1137.
15. S. G. Young, A. W. Bowman, Non-parametric analysis of covariance, *Biometrics*, **51** (1995), 920–931. <https://doi.org/10.2307/2532993>
16. J. C. Pardo-Fernández, M. D. Jiménez-Gamero, A. Ghouch, A non-parametric ANOVA-type test for regression curves based on characteristic functions. *Scand. J. Stat.*, **42** (2015), 197–213. <https://doi.org/10.1111/sjos.12102>
17. C. Park, K. Kang, Sizer analysis for the comparison of regression curves, *Comput. Stat. Data. Anal.*, **52** (2008), 3954–3970. <https://doi.org/10.1016/j.csda.2008.01.006>
18. C. Park, J. Hannig, K. Kang, Nonparametric comparison of multiple regression curves in scale-space, *J. Comput. Graphical Stat.*, **23** (2014), 657–677. <https://doi.org/10.1080/10618600.2013.822816>
19. W. Lin, K. B. Kulasekera, Testing the equality of linear single-index models, *J. Multivar. Anal.*, **101** (2010), 1156–1167.
20. M. Vogt, O. Linton, Classification of non-parametric regression functions in longitudinal data models, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **79** (2017), 5–27. <https://doi.org/10.1111/rssb.12155>
21. M. Vogt, O. Linton, Multiscale clustering of nonparametric regression curves, *J. Econometrics*, **216** (2020), 305–325.

22. N. M. Villanueva, M. Sestelo, L. Meira-Machado, A method for determining groups in multiple survival curves, *Stat. Med.*, **38** (2019), 866–877. <https://doi.org/10.1002/sim.8016>
23. P. Hall, J. D. Hart. Bootstrap test for difference between means in nonparametric regression, *J. Am. Stat. Assoc.*, **85** (412), 1039–1049.
24. M. C. Rodríguez-Campos, W. González-Manteiga, R. Cao, Testing the hypothesis of a generalized linear regression model using nonparametric regression estimation, *J. Stat. Plan. Infer.*, **67** (1998), 99–122. [https://doi.org/10.1016/S0378-3758\(97\)00098-0](https://doi.org/10.1016/S0378-3758(97)00098-0)
25. J. Roca-Pardiñas, C. Cadarso-Suárez, V. Nácher, C. Acuña, Bootstrap-based methods for testing factor-by-curve interactions in Generalized Additive Models: assessing pre-frontal cortex neural activity related to decision-making, *Stat. Med.*, **25**(2006), 2483–2501. <https://doi.org/10.1002/sim.2415>
26. C. Cadarso-Suárez, J. Roca-Pardiñas, G. Molenberghs, F. Faes, V. Nácher, S. Ojeda, et al., Flexible modelling of neuron firing rates across different experimental conditions. An application to neural activity in the prefrontal cortex during a discrimination task, *J. R. Stat. Soc. Ser. C*, **55** (2006), 431–447.
27. S. Sperlich, D. Tjøstheim, L. Yang, Nonparametric estimation and testing of interaction in additive models, *Econom. Theory*, **18** (2002), 197–251. <https://doi.org/10.1017/S0266466602182016>
28. L. Yang, S. Sperlich, W. Härdle, Derivative estimation and testing in generalized additive models, *J. Stat. Plan. Infer.*, **115** (2003), 521–542. [https://doi.org/10.1016/S0378-3758\(02\)00163-5](https://doi.org/10.1016/S0378-3758(02)00163-5)
29. S. Dudoit, M. J. Van Der Laan, *Multiple Testing Procedures with Applications to Genomics*, Springer, Springer Series in Statistics, New York, 2007.



AIMS Press

©2022 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)