



# A Nonparametric Graphical Tests of Significance in Functional GLM

Tomáš Mrkvička<sup>1</sup>  · Tomáš Roskovec<sup>1</sup> · Michael Rost<sup>1</sup>

Received: 1 February 2019 / Revised: 15 October 2019 / Accepted: 20 November 2019 /

Published online: 16 December 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

A new nonparametric graphical test of significance of a covariate in functional GLM is proposed. Our approach is especially interesting due to its functional graphical interpretation of the results. As such, it is able to find not only if the factor of interest is significant but also which functional domain is responsible for the potential rejection. In the case of functional multi-way main effect ANOVA or functional main effect ANCOVA models it is able to find which groups differ (and where they differ), in the case of functional factorial ANOVA or functional factorial ANCOVA models it is able to find which combination of levels (which interactions) differ (and where they differ). The described tests are extensions of global envelope tests in the GLM models. It applies Freedman-Lane algorithm for the permutation of functions, and as such, it approximately achieves the desired significance level.

**Keywords** Functional ANCOVA · Freedman-Lane algorithm · Global envelope test · Groups comparison · Permutation test

**Mathematics Subject Classification (2010)** 62H15 · 62G10

## 1 Introduction

Functional general linear models (GLM) appear in various scientific fields, where the observed data are in the form of function, e.g. in medicine, finance, biology, etc. In this paper, we will consider a  $d$ -dimensional functions and study their dependence on various factors - continuous, categorical and also interactions, through a GLM.

The problem of functional GLM is widely studied in the literature. For example Ram-say and Silverman (2006) described a bootstrap procedure based on pointwise  $F$ -tests, Abramovich and Angelini (2006) used wavelet smoothing techniques, and Ferraty et al. (2007) used dimension reduction approach. Further, Cuesta-Albertos and Febrero-Bande

---

✉ Tomáš Mrkvička  
mrkvicka.toma@gmail.com

<sup>1</sup> Department of Applied Mathematics and Informatics, Faculty of Economics, University of South Bohemia, Studentská 13, 37005 České Budejovice, Czech Republic

(2010) applied the  $F$ -test on several random univariate projections and bound the tests together through false discovery rate.

Since we deal with functional data the assumptions for parametric methods are more complex and in practice not guaranteed to be fulfilled, therefore the nonparametric methods, which have much fewer assumptions, are very popular in this area, e.g. Nichols and Holmes (2001), Winkler et al. (2014) and Pantazis et al. (2005) concentrate on certain pointwise statistics, such as the  $F$ -statistic, and find the distribution of its maxima by permutation. Hahn (2012) used a univariate integral deviation statistic to summarise the deviances between groups in one-way ANOVA. All these permutation methods find a pure maximum of a statistic or compute the integral from a statistic over the study domain. It requires that the statistic has homogeneous distribution across the functional domain, which is not necessarily the case in the practice. On the other hand, our nonparametric procedure solves this problem by applying nonparametric rank envelope test (Myllymäki et al. 2017; Mrkvička et al. 2017) instead. The rank envelope test was introduced in the field of spatial statistic in order to solve the problem of goodness-of-fit for the functional test statistic. The test functions are generated independently under the null model, which guarantee the desired exchangeability in the underlying Monte Carlo test. And the test functions are sorted from the most extreme to the least extreme by extreme rank length measure, which allows for the desired graphical interpretation.

Mrkvička et al. (2018) proposed to generate the functions under the null model by permutation instead, as it is usual in functional permutation methods (Nichols and Holmes 2001; Winkler et al. 2014; Pantazis et al. 2005), and apply the rank envelope test on the permuted functional test statistics. Mrkvička et al. (2018) proposed this approach for functional one-way ANOVA since for this setting the simple permutation of functions guarantees the exchangeability of the test statistics, and it implies the exactness of the test, i.e. the true significance level of the test is the pre-set significance level of the test. When adding the nuisance factors, the exchangeability is not straightforward. Certain permutation strategies have to be applied in order to achieve the exact test (Winkler et al. 2014). In our approach, we follow the recommendation accepted in the univariate permutation tests (Anderson and Ter Braak 2003), i.e. to apply the permutation of residuals under the reduced model, which was first described in Freedman and Lane (1983). This permutation strategy is not exact but comes the closest to the conceptually exact level (Anderson and Robinson 2001) and performed the best in various circumstances (Legendre and Anderson 1999). This permutation strategy is also recommended in the functional permutation test described in details in Winkler et al. (2014).

Thus the proposed method is an extension of the one-way functional ANOVA described in Mrkvička et al. (2018) to functional general linear models by changing the permutation strategy. Then the proposed approach is not exact, but it achieves acceptable nominal levels as can be seen from the simulation study. The main contribution of this paper is showing how to apply the rank envelope test in various general linear models and how to reach the interesting graphical interpretation of the results. This graphical interpretation is able to show which two groups differ together with showing in what area of the functions these groups differ without application of any posthoc test in case of a categorical factor of interest. It is able to indicate which area is responsible for the rejection of a continuous factor of interest. It is also able to show which combinations of levels in multi-way ANOVA design or multi-way ANCOVA design are responsible for the rejection of the interaction factor. By this, the interpretation of analysis is much easier. The graphical visualisation of such a generality was not introduced to this field before by our knowledge.

The proposed procedure is implemented in the R package GET, which contains all methods connected with the global envelope tests. The package is available at CRAN.

The organisation of the paper is as follows. In Section 2 we specify the mathematical setting for our method and how every step of algorithm works. First, we discuss the exchangeability property and usage of Freedman–Lane algorithm. Then we describe the Global extreme rank length envelope test and how it is applied on test vectors.

In Section 3, we perform a simulation study to compare the powers of our graphical procedures with the powers of the procedures which are already available. We have chosen the random projection method (RPM) which is available through the software R (Febrero-Bande and Oviedo de la Fuente 2012) and  $F$ -max method (Nichols and Holmes 2001). Also, we show the graphical output for the categorical predictor, continuous predictor and interactions. The Section 4 is left for conclusions and discussion.

## 2 Theory

### 2.1 Mathematical Specification

We study a set of functions (possibly multidimensional)  $\mathbf{Y} = (y_1(t), \dots, y_n(t))^T$ ,  $t \in D \subset \mathbb{R}^d$ . We consider several exploratory factors of these functions and set a general linear model by two matrices  $\mathbf{X} = (x_{ij}(t))_{i=1,\dots,n,j=1,\dots,J}$  (defining the factor of interest) and  $\mathbf{Z} = (z_{ij}(t))_{i=1,\dots,n,j=1,\dots,Z}$  (defining nuisance factors). Generally, we assume that the factors can be functional, but usually the factors are constants (e.g. age, gender, ... ).

By *full model* we understand

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\gamma + \varepsilon,$$

where  $\beta$  and  $\gamma$  are parameters of GLM calculated in every  $t \in D$ . Thus  $\beta = (\beta_j(t))_{j=1,\dots,J}$  and  $\gamma = (\gamma_j(t))_{j=1,\dots,Z}$  are  $J$  and  $Z$  dimensional functions representing an effect of factors.  $\varepsilon = (\varepsilon_1(t), \dots, \varepsilon_n(t))^T$  is the set of residuals.

By *null model* we understand

$$\mathbf{Y} = \mathbf{Z}\gamma + \varepsilon,$$

i.e. we ignore  $\mathbf{X}$  and regress considering only the nuisance factors.

The elements of  $\mathbf{Y}$  represent functions  $y(t)$  depending on variable  $t$ , possibly more than one dimensional. But since the data have to be discretised anyway, we consider only finitely many values of  $t$  and therefore  $y \in \mathbf{Y}$  is a vector of values of the function in question. For simplicity, we refer to the variable  $t$  as “time” further in this paper, regardless of specific data meaning and the dimension of this variable.

We assume that the discretization is done for same  $K$  values of  $t$  for every function  $y(t)$ . That means  $\mathbf{Y}$  is a multi-vector of size  $n \times K$ ,  $\beta$  is a multi-vector of size  $J \times K$ ,  $\gamma(t)$  of size  $Z \times K$  and  $\varepsilon$  of size  $n \times K$ .

Our study aims to test null hypothesis of the significance of the factor contained in the matrix  $\mathbf{X}$ , i.e.

$$H_0 : \beta_{jk} = 0, \text{ for all } j = 1, \dots, J, k = 1, \dots, K.$$

Remark that both  $\mathbf{X}$  and  $\mathbf{Z}$  may include all kinds of factors, both continuous and categorical ones and even the interactions of effects.

## 2.2 Choice of Methods

Our goal is to test  $H_0$  by nonparametric permutation method with graphical interpretation. We based it on two advanced results. First, the permutation scheme with the presence of nuisance effects is based on *Freedman-Lane algorithm* (Freedman and Lane 1983). Second, we visualise the relationship between the original and the permuted data by the *global extreme rank length envelope test* (Mrkvička et al. 2017; Myllymäki et al. 2017).

## 2.3 ANOVA Effect Study

Let us begin with univariate one way ANOVA. Then the data consist of several groups of samples, and we decide if there is a difference between the groups or if the grouping has no effect. The general nonparametric permutation method, which is usually applied for the non-normal data, suggests permuting the samples between groups randomly. Then we compare the initial sorting to the sorting of new permuted data, for example by  $F$ -test statistics. The result should be the same if the belonging to a group does not affect the sample within. The statistical comparison of the test statistics is in permutation tests made by Monte Carlo test, which requires the *exchangeability* of the test statistic. We say that a test statistic  $\mathbf{T}_1$  is exchangeable if the observed and simulated (permuted) test statistic  $\mathbf{T}_2, \dots, \mathbf{T}_I$  are exchangeable, i.e. the joint distribution of  $\mathbf{T}_1, \dots, \mathbf{T}_I$  is not affected by permutation. The property of *exchangeability* is lost if we add a nuisance effect. For example, if we want to test medicine results of several treatments and the result depends on the age of a patient, the permutation may sort all old patients data into one group and all young patients data into the other. Then we may wrongly conclude the result based on the age and not on the used medicine. There are several ways how to avoid this situation, for example, we may restrict the family of admissible permutations or bound the samples into vectors of samples in order to keep information about nuisance effect despite permuting. We avoid this mistake by the Freedman-Lane algorithm described further, recommended in the univariate GLM permutation studies (Anderson and Ter Braak 2003) but also in functional GLM (Winkler et al. 2014). Suggested Freedman-Lane algorithm permutes the residuals under the null model instead of permuting data itself. The exchangeability property is not satisfied with this scheme, but the significance level for such permutation procedures is close to the nominal one while retaining good power of the tests as it will be shown by the simulation study.

## 2.4 Freedman-Lane Algorithm

1. Regress data  $\mathbf{Y}$  against the full model containing both the effect of interest  $\beta$  and the nuisance effect  $\gamma$  as

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\gamma + \varepsilon.$$

2. Regress data  $\mathbf{Y}$  against the reduced model containing only the nuisance effect  $\gamma$  as

$$\mathbf{Y} = \mathbf{Z}\gamma + \varepsilon_{\mathbf{Z}}.$$

3. Compute the permuted data  $\mathbf{Y}_j^*$ . We get this data by permuting the residuals of reduced model  $\varepsilon_{\mathbf{Z}}$  by permutation  $\pi_j$  into  $\varepsilon_{\mathbf{Z},j} = \pi_j(\varepsilon_{\mathbf{Z}})$ . We get

$$\mathbf{Y}_j^* = \mathbf{Z}\gamma + \varepsilon_{\mathbf{Z},j}.$$

4. Regress the permuted data  $\mathbf{Y}_j^*$  against the full model and get a new effect of interest  $\beta_j^*$  from formula

$$\mathbf{Y}_j^* = \mathbf{X}\beta_j^* + \mathbf{Z}\gamma_j^* + \varepsilon_j.$$

## 2.5 Global Extreme Rank Length Envelope Test

The other specification of our work is the graphical interpretation and finding the differing groups without the need of post-hoc test in case of the categorical factor of interest or interactions. If there is a nonzero effect of interest for a categorical factor, there is the need of post-hoc test to find the responsible party. But the post-hoc procedure may slightly differ from the result of the ANOVA test. Our method provides a  $(1 - \alpha)100\%$  global envelope, which contains simultaneously all parameters  $\beta_{jk}$ ,  $j = 1, \dots, J, k = 1, \dots, K$ . So if the test vector is completely contained in the envelope for every time, the null hypothesis is not rejected. If the test vector leaves the envelope in any value, the null hypothesis is rejected, and this group is identified as the group responsible for rejecting the null hypothesis. Moreover, since we study functions, we see the values of time for whose this group does not fit into the envelope and may use this information for interpretation of data and conclusions. The positions of the envelopes and the test vectors can be visualised, and the results are easy to be interpreted.

### 2.5.1 Test Vector Choice

To apply the rank envelope test, we have first to select a test vector. The first possible choice is, based on the values of effects

$$\mathbf{T} = (\beta_{jk}), j = 1, \dots, J, k = 1, \dots, K.$$

That is a multi-vector  $J \times K$  for  $J$  different groups of the factor of interest and  $K$  values of time. I.e. for continuous factor of interest  $J = 1$  and  $\mathbf{T} = (\beta_{1k}), k = 1, \dots, K$ . For the categorical factor of interest,  $J$  is equal to the number of groups of the categorical factor, with each vector  $(\beta_{j1}, \beta_{j2}, \dots, \beta_{jK})$  having the same role in the univariate model, adding the additional condition of  $\sum_j \beta_{jk} = 0$  for all  $k \in \{1, \dots, K\}$ . For interaction of continuous and categorical factor,  $J$  is also equal to the number of groups of the categorical factor, adding the same additional condition of  $\sum_j \beta_{jk} = 0$  for all  $k \in \{1, \dots, K\}$ . For the interaction of two categorical factors,  $J$  is equal to the product of the numbers of groups of both categorical factors, adding the same additional condition of  $\sum_j \beta_{jk} = 0$  for all  $k \in \{1, \dots, K\}$ . To construct the envelope, we consider test vectors of the same type based on permuted data.

The second possible choice of the test vector, applicable for at least one categorical factor, consists of differences between two group parameters of  $\beta$ . Since we do not have to check the couples  $i = j$  and  $i > j$ , we check only the couples where the first index is lower than the second one. We get a different test vector

$$\mathbf{T}' = (\beta_{11} - \beta_{21}, \dots, \beta_{1K} - \beta_{2K}, \beta_{11} - \beta_{31}, \dots, \dots, \beta_{1K} - \beta_{3K}, \dots, \beta_{(J-1)1} - \beta_{J1}, \dots, \beta_{(J-1)K} - \beta_{JK}).$$

The number of coordinates of  $\mathbf{T}'$  is  $(J(J - 1)/2) \times K$ . For the first test vector  $\mathbf{T}$  we get the information about the groups causing possible rejection of the null hypothesis. For the second test vector  $\mathbf{T}'$  we get the information about the couples of groups which are different. Although we ask the same question about the data, either test is sensitive to the different kind of misfits.

In case of unequal variances among groups of functions, in case of categorical factors either nuisance or one of interest or categorical interactions, we may use the group variance normalising transformation before the analysis as described in Mrkvička et al. (2018). The check for homoscedasticity of the groups in the presence of nuisance factors can be done in the same way as in Mrkvička et al. (2018), where it is described without nuisance factors.

## 2.5.2 Global Extreme Rank Length Envelope Test

We briefly introduce the method developed in Myllymäki et al. (2017) and Mrkvička et al. (2017). We suppose now that the type of test vector has been chosen. Let us suppose we have  $I$  permutations produced by Freedman-Lane algorithm. Let us denote  $\mathbf{T}_i$  the test vector based on  $i$ -th permutation, especially  $\mathbf{T}_1$  is the vector based on identical permutation or the original data. Using the permuted data, we aim to define boundary vectors  $\mathbf{T}_{upp}$  and  $\mathbf{T}_{low}$ . These vectors create a natural envelope including the typical values and excluding the extreme values. We do not reject the null hypothesis  $H_0$  if for every element  $k$  it holds

$$\mathbf{T}_1(k) \in (\mathbf{T}_{low}(k), \mathbf{T}_{upp}(k)).$$

We reject the hypothesis  $H_0$ , if there exists an element  $k$  such that

$$\mathbf{T}_1(k) \notin (\mathbf{T}_{low}(k), \mathbf{T}_{upp}(k)).$$

We proceed to show how to set  $T_{upp}$  and  $T_{low}$  and a  $p$ -value of the test. First, let us calculate for every vector  $\mathbf{T}_i(k)$  the ranks of its values for each element  $k$  separately and denote this value  $S_i(k)$ . As we use two side envelope, we search for the extremeness from both sides, i.e.  $R_i(k) = \min(S_i(k), n - S_i(k) + 1)$ . As an *extreme rank* we denote

$$R_i = \min_{k \in \{1, 2, \dots, K\}} R_i(k).$$

There is a huge risk of ties, so we have to decide which of two vectors with the same *extreme rank* is more centred. We break these ties and order the vectors by so-called *extreme rank length*.

First, we order the ranks for every test vector  $T_i(k)$  into nondecreasing sequence of  $K$  numbers  $\mathbf{R}_i$ , starting with *extreme rank*  $R_i$ . More precisely we set

$$\mathbf{R}_i = (R_i(k_1), R_i(k_2) \dots R_i(k_K)), \text{ such that } R_i(k_l) \leq R_i(k_{l+1}) \text{ for } l \in \{1, 2 \dots K - 1\}.$$

Then we define the *extreme rank length* relation  $<$  by:

$$\mathbf{T}_i < \mathbf{T}_{i'},$$

if there exists  $N > 0$  such that the ranks  $R_i(k_l) = R_{i'}(k_l)$  for the first  $N - 1$  elements and  $R_i(k_N) < R_{i'}(k_N)$ . Roughly speaking, we compare the most extreme rank of vectors, and in case of a tie, we compare how many values are so extreme for said vectors. If there is another tie in the length of the most extreme rank, we compare the second most extreme rank, then we compare its length and so on until there is the difference.

Due to the above ordering, we may set the  $p$ -value as

$$p = 1 - (I^{-1} \sum_{i=1}^I \mathbf{1}(\mathbf{T}_1 < \mathbf{T}_i)),$$

where  $I$  is the number of test vectors produced by permuting introduced before.

Now we are ready to construct the  $(1 - \alpha)$  envelope, where  $\alpha$  is the pre-set significance level of the test of the null hypothesis. Remark here that due to the nuisance factors and the

Freedman and Lane permutation strategy, the probability that  $T_1(k)$  leaves this envelope is only approximately equal to  $\alpha$ . Let  $\mathcal{I}_{ex} \subset \{1, \dots, I\}$ ,  $|\mathcal{I}_{ex}| = I\alpha$  denotes the set indexes which corresponds to  $I\alpha$  the most extreme vectors. Then we define the *global extreme rank length envelope* for all  $k$  as

$$(T_{low}(k), T_{upp}(k)) = (\min_{i \in \{1, \dots, I\} \setminus \mathcal{I}_{ex}} \{T_i(k)\}, \max_{i \in \{1, \dots, I\} \setminus \mathcal{I}_{ex}} \{T_i(k)\}), \quad k \in \{1, \dots, K\}.$$

### 3 Examples with Simulation Study

This Section aims to show several different functional GLM where our methods are applicable and to compare the powers with other non-graphical methods available for functional GLM. Namely with random projection method (RPM) of Cuesta-Albertos and Febrero-Bande (2010) and with  $F$ -max test (Nichols and Holmes 2001), which is often used in neuroimage data analysis.

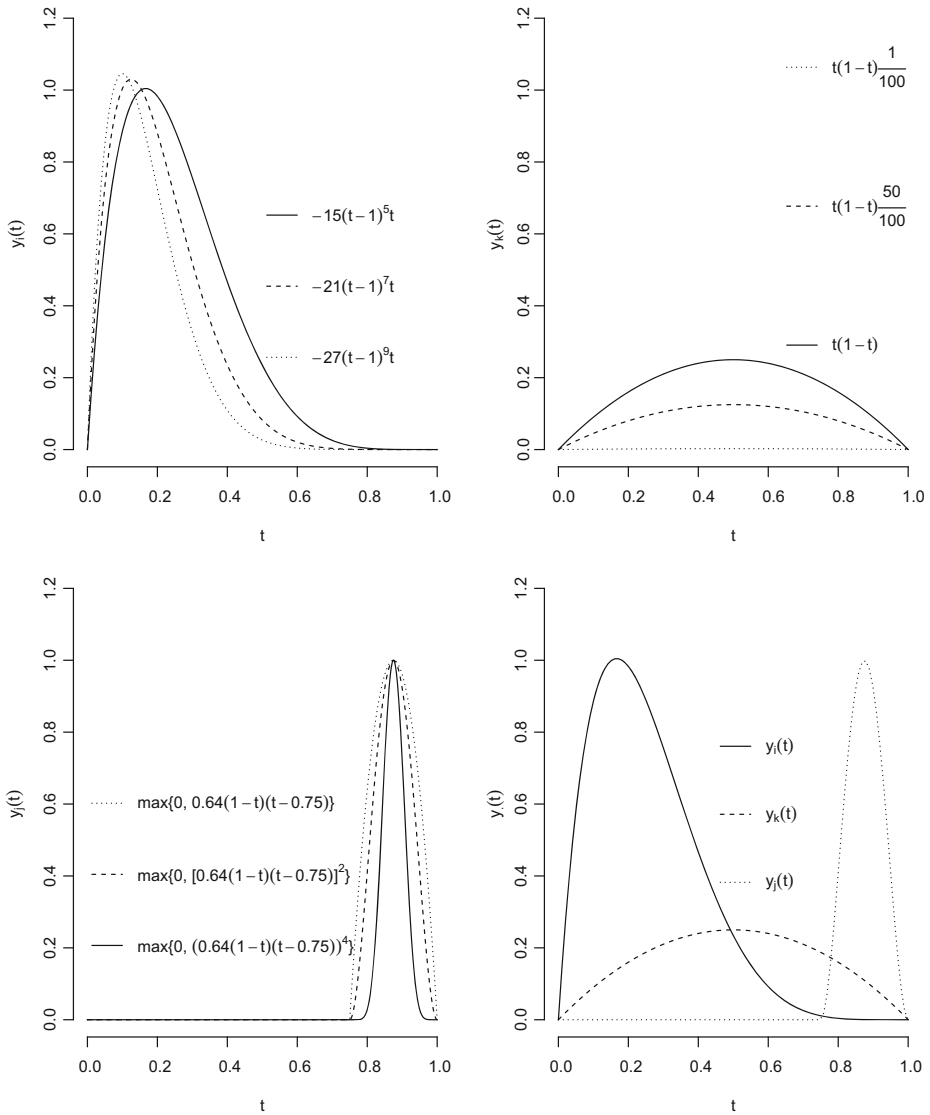
For this purpose we use a model function which consists of three summands. By the use of this model function, we will simulate all our examples.

$$\begin{aligned} y_{i,j,k}(t) &= 3(5 + 2i)t(1 - t)^{5+2i} + (\max\{0; 64(1 - t)(t - 0.75)\})^j + t(1 - t)k/100 \\ &= y_i(t) + y_j(t) + y_k(t). \end{aligned}$$

We describe the design of our model function now. There are three parameters  $i, j, k$  that may represent the effects of three different factors. In every of our example we define the data function by picking a triplet  $i, j, k$  and adding the error term  $e(t)$  to function  $y_{i,j,k}(t)$ . The domain of functions is set to  $t \in [0, 1]$  and it is discretized into 100 equidistant values for every function. If the parameter represent categorical factor, the choice of value for parameter represents an involvement of data function to some category. For example, we pick values of  $i = 0, 1$  or  $2$ , which identify the three levels of the categorical factor, and we may sort all data function into three groups based on the value of  $i$  used in  $y_{i,j,k}$ . On the other hand, if the parameter represent a continuous factor, we pick an interval and the value of the parameter would be randomly drawn from this interval. For example we pick  $k \in [0, 100]$  and for every data function we use random  $k$  from this interval to evaluate  $y_{i,j,k}$ , the corresponding continuous factor is then equal to  $k$ .

For better visualisation, we present Fig. 1. In the first three quadrants, we pick a factor and three typical values for the parameter corresponding to the factor in question. The function representing the effect of the parameter is shown. Function  $y_i$  based on parameter  $i$  effects much more the values for  $t < 0.5$  and its effect is fading as we enclose  $t = 1$ , function  $y_j$  based on the value of  $j$  effects only the interval  $t \in [0.75, 1]$  and  $y_k$  based on  $k$  effects all values of  $t$ , but the most significant influence is always in the middle close to  $t = 0.5$ . We would recall these observations later, as we present the sensitivity of our method to these properties. In the last, fourth, quadrant, we put all the factors in one figure to show the comparison of values and the supports of the functions representing the factors.

The error term  $e(t)$  is chosen to be an i.i.d. error of zero mean and standard deviation equal to  $\sigma$ . In the Appendix, we also give the results of the simulation study for the Brownian motion error term. Each example consists of 60 simulated data functions, and the significance levels and powers of studied tests are computed from 1000 repetition of the experiment on the pre-set significance level  $\alpha = 0.05$ . The  $p$ -value of every permutation test is calculated from 1000 permutations. The RPM method is computed from 30 random projections. Four tests are applied on testing of significance of the factor of interest. The



**Fig. 1** Three factors of simulation: In first three quadrants the effects of parameters  $i$ ,  $j$ , and  $k$  on corresponding factors are visualised; in fourth quadrant the differences between the supports of effects in dependence on  $t$  are shown

studied tests are our graphical test based on the regression parameters  $\mathbf{T}$  (GETP), our graphical test based on differences of regression parameters  $\mathbf{T}'$  (GETDP), Random projection method (RPM) and  $F$ -max test.

### 3.1 Categorical Factor of Interest in Main Effect GLM

Let us consider two categorical factors, the first one with three levels being of interest and the other one with two levels being a nuisance factor. This setting produces six groups of



functions, each one defined by a unique triplet of parameters  $i, j, k$  and consisting of 10 functions created by adding an error terms  $e(t)$ . The estimated probability of rejection is shown in Table 1 for four tests and six different models. In the left column, we show the setting of parameters  $i, j, k$  in six triplets generating six groups of functions. The factor of interest is amplified by bold font, and we note by  $\mathbf{F1} = 0$  that the factor of interest has no effect on the model and by  $\mathbf{F1} \neq 0$  that the factor of interest have some effect in the model. Similarly, we use this notation for nuisance factors  $F2$ . In the second column, we mark the row with the method which result is presented in the following columns. In the top row, we show the choice of the standard deviation of i.i.d. error.

Let us mention that the effects of first and second factor influence the value of  $y(t)$  on the same functional domain in case 4. whereas on the different functional domain for the last setting, since supports of functions  $y_i$  and  $y_j$  do not overlap.

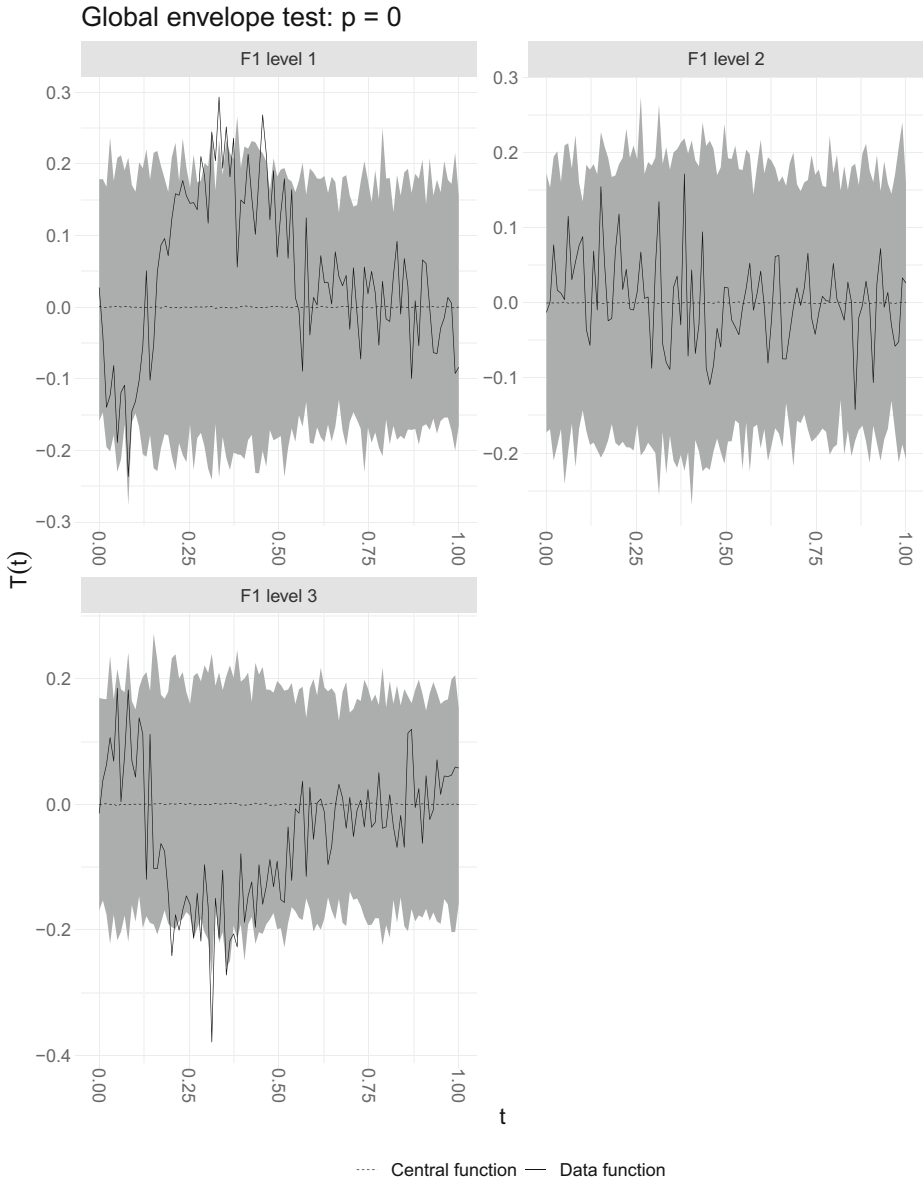
Table 1 shows that the estimated significance levels (case 1. and 2.) are slightly bigger than the desired 0.05 level in tests using Freedman-Lane permutation scheme. On the other hand, the estimated power (cases 3., 4., 5., 6.) is much higher for our two tests than for the other two.

We also show the graphical output of our tests for one realisation of model 4 from Table 1, for  $\sigma(e) = 0.3$ . The global 95% envelope is drawn in grey, and the test vector is drawn in

**Table 1** The estimated probabilities of rejecting the factor of interest in main effect FGLM with two categorical factors with significance level  $\alpha = 0.05$

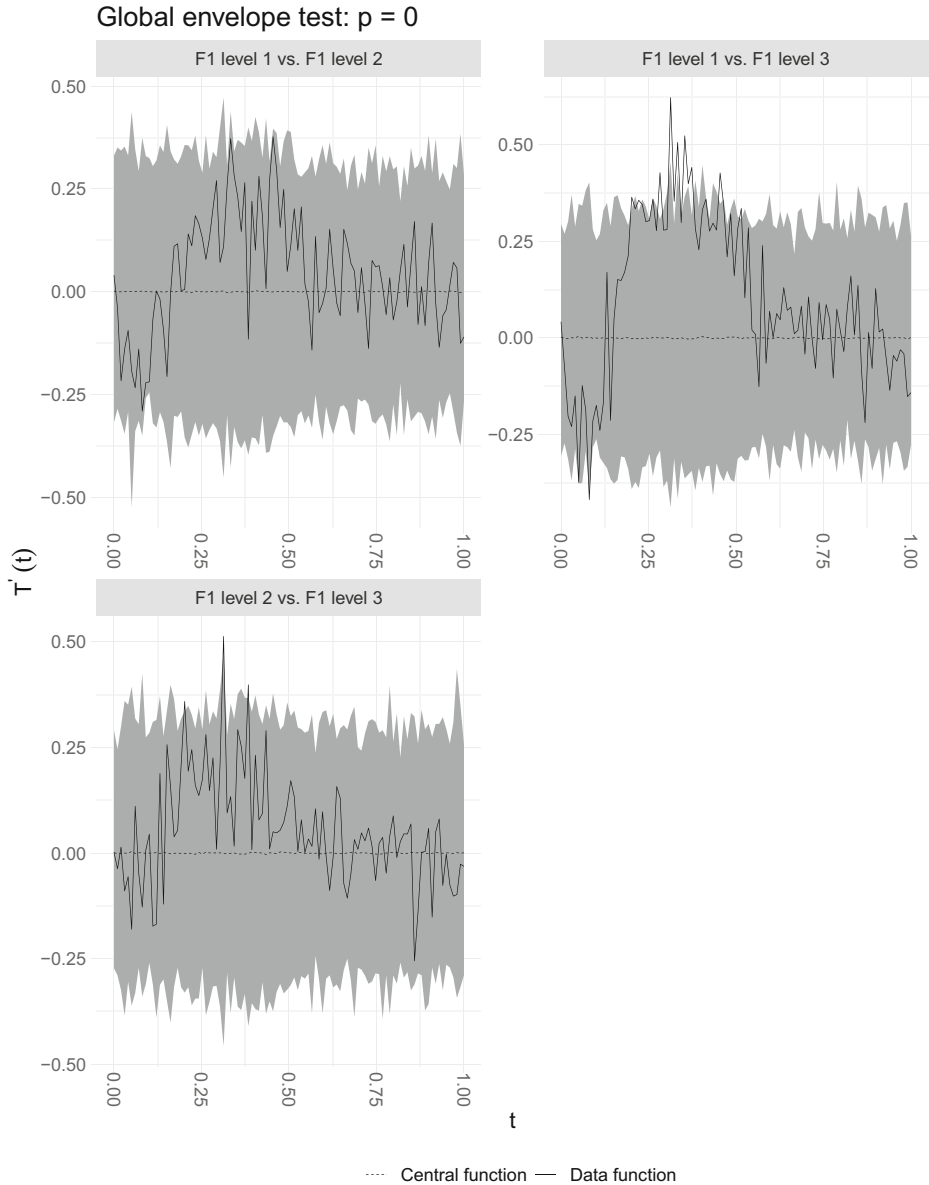
	Method	$\sigma(e) = 0.3$	$\sigma(e) = 0.5$	$\sigma(e) = 0.8$	
1.	$\mathbf{i}=(1,1,1,1,1,1)$	GETP	0.059	0.057	0.070
	$\mathbf{j}=(1,1,1,1,1,1)$	GETDP	0.071	0.059	0.059
	$\mathbf{k}=(1,1,1,1,1,1)$	RPM	0.058	0.062	0.073
	$\mathbf{F1} = 0, F2 = 0$	F-max	0.064	0.055	0.044
2.	$\mathbf{i}=(1,1,1,1,1,1)$	GETP	0.052	0.050	0.063
	$\mathbf{j}=(1,1,1,1,1,1)$	GETDP	0.069	0.056	0.068
	$\mathbf{k}=(1,1,1,50,50,50)$	RPM	0.069	0.061	0.087
	$\mathbf{F1} = 0, F2 \neq 0$	F-max	0.037	0.041	0.050
3.	$\mathbf{i}=(0,1,2,0,1,2)$	GETP	1.000	0.996	0.594
	$\mathbf{j}=(1,1,1,1,1,1)$	GETDP	1.000	0.997	0.602
	$\mathbf{k}=(1,1,1,1,1,1)$	RPM	1.000	0.766	0.322
	$\mathbf{F1} \neq 0, F2 = 0$	F-max	1.000	0.893	0.329
4.	$\mathbf{i}=(0,1,2,0,1,2)$	GETP	1.000	0.992	0.623
	$\mathbf{j}=(1,1,1,1,1,1)$	GETDP	1.000	0.996	0.609
	$\mathbf{k}=(1,1,1,50,50,50)$	RPM	0.999	0.778	0.340
	$\mathbf{F1} \neq 0, F2 \neq 0$	F-max	1.000	0.894	0.323
5.	$\mathbf{i}=(1,1,1,1,1,1)$	GETP	1.000	0.932	0.366
	$\mathbf{j}=(1,2,4,1,2,4)$	GETDP	1.000	0.940	0.370
	$\mathbf{k}=(1,1,1,1,1,1)$	RPM	0.907	0.427	0.183
	$\mathbf{F1} \neq 0, F2 = 0$	F-max	1.000	0.772	0.224
6.	$\mathbf{i}=(1,1,1,1,1,1)$	GETP	1.000	0.920	0.356
	$\mathbf{j}=(1,2,4,1,2,4)$	GETDP	1.000	0.928	0.374
	$\mathbf{k}=(1,1,1,50,50,50)$	RPM	0.905	0.389	0.177
	$\mathbf{F1} \neq 0, F2 \neq 0$	F-max	1.000	0.781	0.222

black. We recall the definitions of test vectors from Section 2.5.1, first we present graphical outcome of test vectors  $\mathbf{T}$  in Fig. 2 for three different groups determined by choice of parameter of interest  $i = 0, i = 1$  or  $i = 2$ . Figure 3 pictures test vector  $\mathbf{T}'$  for comparison between three groups of data function determined by  $i = 0, i = 1$  or  $i = 2$ . We should mention that all figures are plotted for 5000 permutations instead of 1000 permutations result presented in Tables, the reason is that the envelope is more smooth for more permutation used.



**Fig. 2** Graphical output for Model 4,  $\sigma(e) = 0.3$  test vector  $\mathbf{T}$ ; The first envelope plots the estimated parameters  $\beta_{1k}$  based on the choice  $i = 0$ , second envelope plots  $i = 1$ , third envelope plots  $i = 2$ . The values where data function exits the grey area of the 95% global envelope are responsible for the rejection of null hypothesis

The both presented tests reject the null hypothesis with  $p$ -value  $< 0.001$ . In addition our method gives us information about the values of  $t$  and groups responsible for rejection. As we observe in Fig. 2, there are values  $t \in (0.2, 0.5)$  in the group  $i = 0$  and group  $i = 2$  responsible for rejection. We may expect based on figure, that for values  $i = 0$  the values of functions  $y_i$  is bigger than average in responsible interval  $t \in (0.2, 0.5)$  and for functions



**Fig. 3** Graphical output for Model 4,  $\sigma(e) = 0.3$  test vector  $\mathbf{T}'$ ; The first envelope plots the differences of the estimated parameters  $\beta_{1k}$  and  $\beta_{2k}$ , second envelope plots the difference between the  $\beta_{1k}$  and  $\beta_{3k}$ , third envelope plots the difference between  $\beta_{2k}$  and  $\beta_{3k}$ . The values where data function exits the grey area of the 95% global envelope are responsible for the rejection of null hypothesis

in group  $i = 2$  we expect  $y_i$  to be lower than average for  $t \in (0.2, 0.5)$ . As we compare this observation to Fig. 1, we see that we catch the main difference between the shape of  $y_i$  for  $i = 0, i = 1$  or  $i = 2$ , that is the shift of the maximum of the function that cause the difference between the functions approximately in interval  $t \in (0.2, 0.5)$ . The similar outcome can be read from Fig. 3, where most significant differences are between the groups corresponding to choices  $i = 0$  and  $i = 2$ , and we see that the difference between them is obvious in interval  $(0.2, 0.5)$  and also possibly around  $t = 0.1$ . That is exactly where the distance between the values of  $y_i$  for choices  $i = 0$  and  $i = 2$  are the biggest. We may also observe some differences between the pairs of groups  $i = 0, i = 1$  or  $i = 1, i = 2$ , but the difference there is not that obvious. The big variability of the tested functions is caused by the nature of i.i.d. error term.

### 3.2 Continuous Factor of Interest in Main Effect GLM

Let us consider now the continuous factor as the factor of interest. We may consider a nuisance factor to be a categorical factor or another continuous factor. The estimated probability of rejection is shown in Table 2 for four tests and six different models in a similar way as above. In models 1., 2. the tested continuous factor should not influence the functions, so the parameter  $k$  is kept constant. In models 3., 4., 5. and 6. we draw  $k$  randomly from the

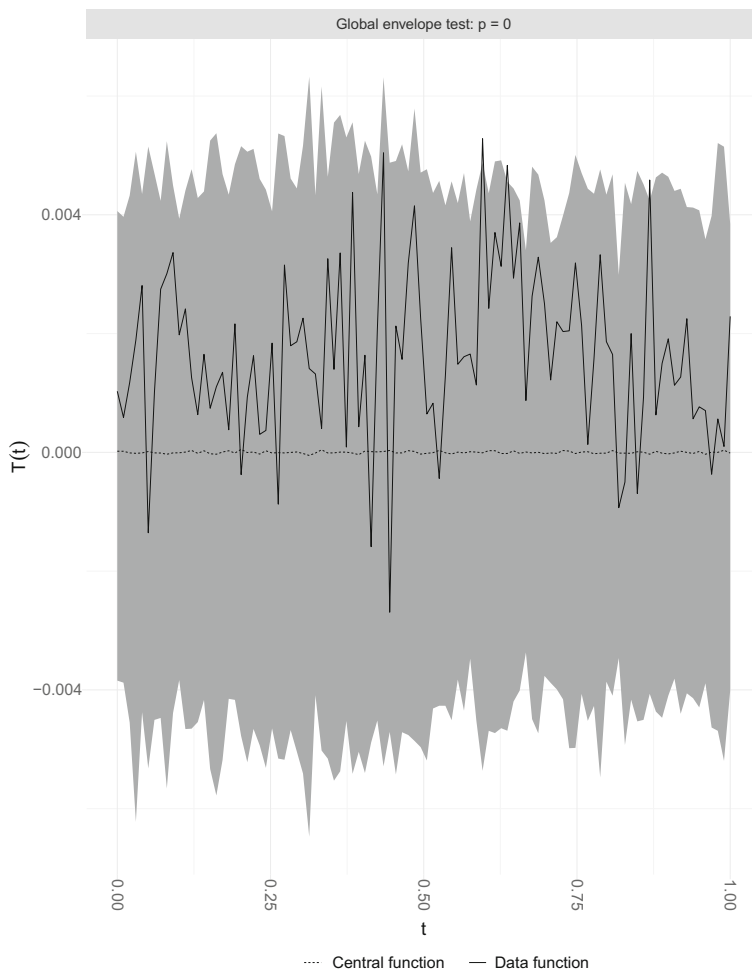
**Table 2** The estimated probabilities of rejecting of factor of interest in main effect two factor FGLM with continuous factor of interest

	Method	$\sigma(e) = 0.3$	$\sigma(e) = 0.5$	$\sigma(e) = 0.8$	
1.	$i=(1,1,1,1,1)$	GETP	0.045	0.057	0.056
	$j=(1,1,1,1,1)$	GETDP	–	–	–
	$k=(1,1,1,1,1)$	RPM	–	–	–
	$F1 = 0, F2 = 0$	F-max	0.049	0.058	0.059
2.	$i=(0,1,2,0,1,2)$	GETP	0.028	0.076	0.072
	$j=(1,1,1,1,1)$	GETDP	–	–	–
	$k=(1,1,1,1,1)$	RPM	–	–	–
	$F1 = 0, F2 \neq 0$	F-max	0.047	0.064	0.042
3.	$i=(1,1,1,1,1)$	GETP	0.970	0.626	0.202
	$j=(1,1,1,1,1)$	GETDP	–	–	–
	$k \in [0,100]$	RPM	–	–	–
	$F1 \neq 0, F2 = 0$	F-max	0.800	0.384	0.152
4.	$i=(0,1,2,0,1,2)$	GETP	0.704	0.355	0.208
	$j=(1,1,1,1,1)$	GETDP	–	–	–
	$k \in [0,100]$	RPM	–	–	–
	$F1 \neq 0, F2 \neq 0$	F-max	0.435	0.233	0.126
5.	$i \in [0,2]$	GETP	0.981	0.493	0.199
	$j=(1,1,1,1,1)$	GETDP	–	–	–
	$k \in [0,100]$	RPM	–	–	–
	$F1 \neq 0, F2 = 0$	F-max	0.857	0.262	0.132
6.	$i=(1,1,1,1,1)$	GETP	0.986	0.53	0.237
	$j=(1,2,4,1,2,4)$	GETDP	–	–	–
	$k \in [0,100]$	RPM	–	–	–
	$F1 \neq 0, F2 \neq 0$	F-max	0.841	0.325	0.175

interval  $k \in [0, 100]$  for each function separately in order to achieve a continuous effect. In model 5, we have a continuous nuisance factor, so we also draw  $i \in [0, 2]$  randomly. Model 4, has a nuisance effect in the same domain as the effect of interest, model 6, is analogous, but the nuisance effect is in a different domain than the factor of interest.

The GETDP method is not applicable in this case since we have only one parameter for the continuous effect. The R implementation of RPM is not designed for the continuous factors. Therefore Table 2 shows results only for the other two methods. Similarly like in the previous study the estimated significant level could be a bit higher than the nominal level for the GETP, but the power is much higher for the GETP than for the  $F$ -max procedure.

Once again, we demonstrate the visual outcome of our method by plotting the envelope in Fig. 4 for one realisation of model 4 from Table 2. We do not search for a significant group as previous, but we try to find the values of  $t$  responsible for rejection, it means the values where the continuous effect has the most influence on the functions. In Fig. 4, these are the ones that exit the envelope, and we see that it is the values in the middle of the



**Fig. 4** Graphical output results for one realisation of model 4 from Table 2,  $\sigma(e) = 0.3$ . The values where data function exits the grey area of the 95% global envelope are responsible for the rejection of null hypothesis

interval,  $t = 0.5$  and values close to it. We compare this result of envelope test to the model function plot in Fig. 1, and clearly, we observe that  $y_k$  has the most extreme values in the middle of the interval, as we expect based on visualisation given by envelope test.

### 3.3 Interactions

Let us consider now the factorial model of two categorical factors, the first one with three levels and the other one with two levels. The factor of interest are now interactions between these two factors. We get six groups of functions again, each of them consisting of 10 functions and defined by a unique triplet of parameters  $i, j, k$  and a random error term. The estimated probability of rejection is shown in Table 3 for four tests and four different models. The factor of interest is amplified by bold font, and we note by  $\mathbf{I} = 0$  that the interactions are not present in the model and by  $\mathbf{I} \neq 0$  that the interactions are present in the model.

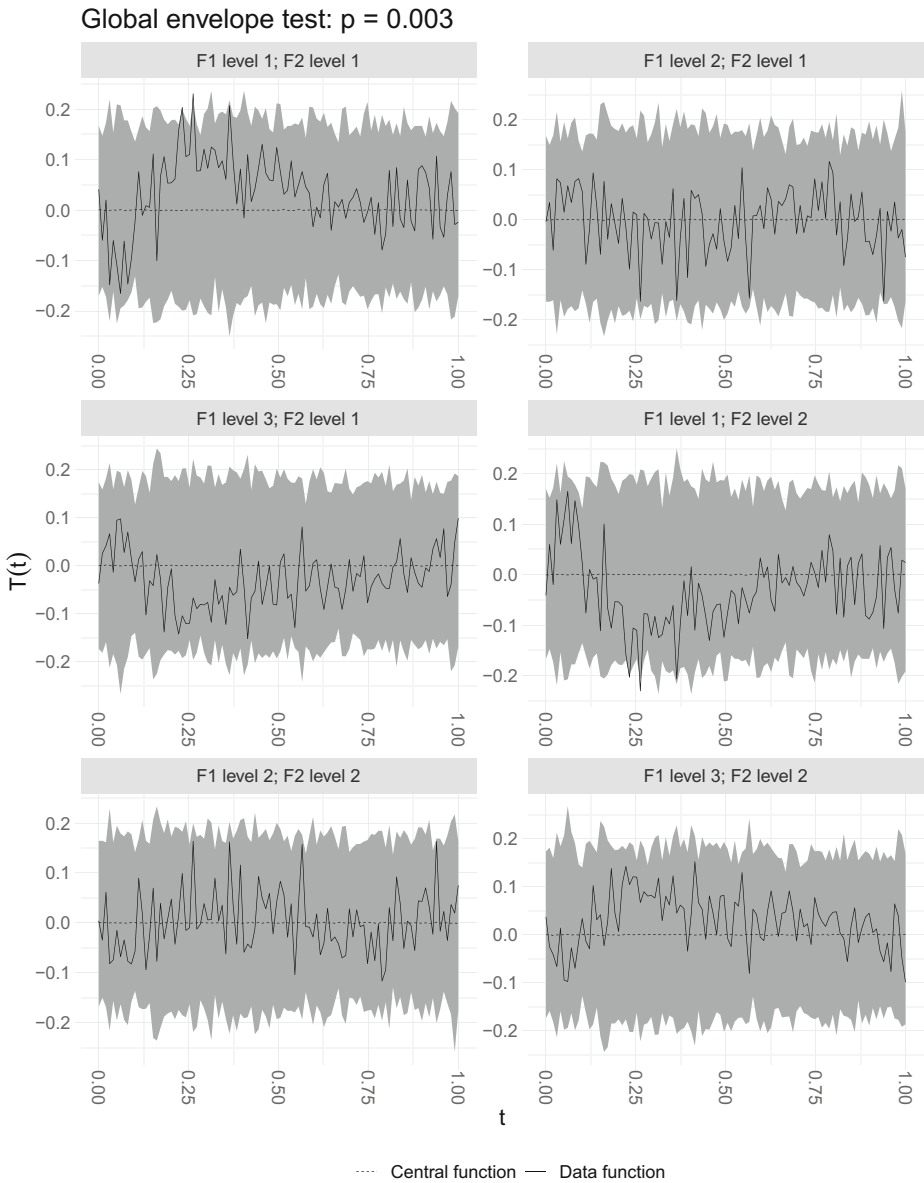
Here we observe the slight liberality of the methods GETP, GETDP and RPM and again much higher estimated powers for our methods than the other two methods.

Figure 5 shows the graphical results of the GETP test for one realisation of the model 4 from Table 3. The estimated parameter functions leave the envelope for factor 1 level 1 groups only. This outcome identifies the differences between the groups of factor 2 level 1 and factor 2 level 2 in level 1 of factor 1. Looking in Table 3 we should see also the differences in factor 1 level 3 but since the differences between functions with  $i = 1$  and  $i = 2$  are smaller than the differences between functions with  $i = 0$  and  $i = 1$  we do not identify this difference.

Figure 6 shows the graphical results of the GETDP test for the same realisation as above. Here we identify the differences between groups of functions identified by factor 1 level

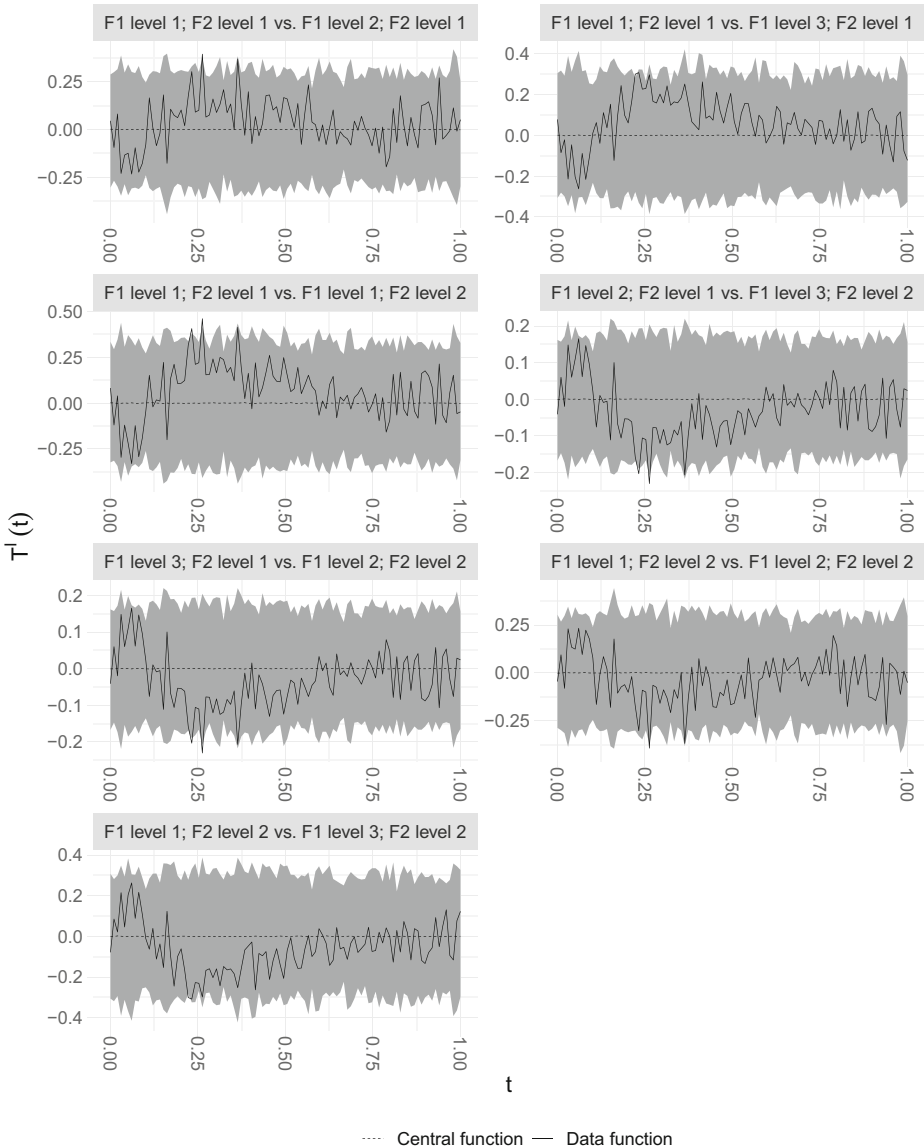
**Table 3** The estimated probabilities of rejecting of effect of interactions in factorial two factor FGLM

	Method	$\sigma(e) = 0.3$	$\sigma(e) = 0.5$	$\sigma(e) = 0.8$
1.	$i=(0,1,2,0,1,2)$ GETP	0.084	0.077	0.095
	$j=(1,1,1,1,1,1)$ GETDP	0.084	0.078	0.084
	$k=(1,1,1,50,50,50)$ RPM	0.076	0.069	0.078
	$\mathbf{I} = 0$ F-max	0.060	0.049	0.052
2.	$i=(0,1,2,1,1,1)$ GETP	0.961	0.457	0.184
	$j=(1,1,1,1,1,1)$ GETDP	0.962	0.440	0.182
	$k=(1,1,1,50,50,50)$ RPM	0.588	0.222	0.118
	$\mathbf{I} \neq 0$ F-max	0.652	0.192	0.096
3.	$i=(0,1,2,0,1,2)$ GETP	0.093	0.078	0.075
	$j=(1,1,1,2,2,2)$ GETDP	0.087	0.081	0.070
	$k=(1,1,1,1,1,1)$ RPM	0.079	0.070	0.070
	$\mathbf{I} = 0$ F-max	0.049	0.052	0.043
4.	$i=(0,1,2,1,1,1)$ GETP	0.959	0.425	0.188
	$j=(1,1,1,2,2,2)$ GETDP	0.951	0.425	0.189
	$k=(1,1,1,1,1,1)$ RPM	0.597	0.218	0.110
	$\mathbf{I} \neq 0$ F-max	0.647	0.199	0.086



**Fig. 5** Graphical output for one realisation of model 4 from Table 3,  $\sigma(\epsilon) = 0.3$ , with test vector  $\mathbf{T}$

1 and several other groups, as in the previous image. Furthermore, there is identified the difference between factor 1 level 3 groups and a few other groups, which was not identified by GETP test.



**Fig. 6** Graphical output for one realisation of model 4 from Table 3,  $\sigma(e) = 0.3$ , with test vector  $\mathbf{T}'$ . Due to the high number of comparisons, only the significant are presented

### 4 Discussion and Conclusions

In this paper, we introduced a nonparametric test of significance in the functional general linear model with either categorical and continuous factors and also with interactions. The simulation studies performed in cases of interactions, categorical and continuous factors of interest show that our proposed tests are very slightly liberal, due to the application of Freedman-Lane permutation scheme in the presence of nuisance factors. On the other hand, they show much bigger



power than random projection method and the  $F$ -max method. The differences between powers of the tests can be even bigger when the random error is not i.i.d., see [Appendix](#).

The advantage of our method in comparison to other methods is the graphical interpretation which it provides, especially it is able to detect the functional domain which is responsible for the potential rejection. The graphical interpretation is the primary purpose of why we propose this method. Also, its base in rank statistics makes it robust with respect to changes of the distribution of the test statistics across the domain of the functions. E.g.  $F$  statistic does not change its first and second moments across the domain but if the original functional data are not Gaussian than the other moments and especially the quantiles of the  $F$  statistic changes across the domain. In such cases, the  $F$ -max statistic can be blind to deviation from the null model in some parts of the domain. The size of the robustness of rank based method concerning  $F$ -max method will be studied in the future.

Finally, the post hoc nature of our test based on the differences of regression parameters is another advantage by providing the better interpretation capabilities of our method.

**Acknowledgements** The project has been financially supported by the Grant Agency of Czech Republic (Project No. 19-04412S).

## Appendix

**Table 4** The estimated probabilities of rejecting of factor of interest in main effect FGLM with two categorical factors and Brownian motion error

	Method	$\sigma(e(1)) = 3$	$\sigma(e(1)) = 5$	$\sigma(e(1)) = 8$
1. $i=(1,1,1,1,1,1)$	GETP	0.03	0.08	0.05
$j=(1,1,1,1,1,1)$	GETDP	0.06	0.06	0.02
$k=(1,1,1,1,1,1)$	RPM	0.03	0.09	0.07
$F1 = 0, F2 = 0$	F-max	0.03	0.06	0.02
2. $i=(1,1,1,1,1,1)$	GETP	0.08	0.08	0.03
$j=(1,1,1,1,1,1)$	GETDP	0.06	0.09	0.00
$k=(1,1,1,50,50,50)$	RPM	0.06	0.07	0.08
$F1 = 0, F2 \neq 0$	F-max	0.07	0.09	0.06
3. $i=(0,1,2,0,1,2)$	GETP	1.00	1.00	0.76
$j=(1,1,1,1,1,1)$	GETDP	1.00	1.00	0.73
$k=(1,1,1,1,1,1)$	RPM	1.00	0.77	0.33
$F1 \neq 0, F2 = 0$	F-max	1.00	0.90	0.23
4. $i=(0,1,2,0,1,2)$	GETP	1.00	1.00	0.71
$j=(1,1,1,1,1,1)$	GETDP	1.00	1.00	0.67
$k=(1,1,1,50,50,50)$	RPM	1.00	0.82	0.28
$F1 \neq 0, F2 \neq 0$	F-max	1.00	0.83	0.38
5. $i=(1,1,1,1,1,1)$	GETP	1.00	0.90	0.34
$j=(1,2,4,1,2,4)$	GETDP	1.00	0.88	0.29
$k=(1,1,1,1,1,1)$	RPM	0.89	0.46	0.14
$F1 \neq 0, F2 = 0$	F-max	1.00	0.75	0.22
6. $i=(1,1,1,1,1,1)$	GETP	1.00	0.87	0.34
$j=(1,2,4,1,2,4)$	GETDP	1.00	0.85	0.35
$k=(1,1,1,50,50,50)$	RPM	0.92	0.34	0.18
$F1 \neq 0, F2 \neq 0$	F-max	1.00	0.79	0.25

**Table 5** The estimated probabilities of rejecting of factor of interest in main effect two factor FGLM with continuous factor of interest and Brownian motion error

	Method	$\sigma(e(1)) = 3$	$\sigma(e(1)) = 5$	$\sigma(e(1)) = 8$
1. $i=(1,1,1,1,1,1)$	GETP	0.04	0.07	0.07
$j=(1,1,1,1,1,1)$	GETDP	–	–	–
$k=(1,1,1,1,1,1)$	RPM	–	–	–
$F1 = 0, F2 = 0$	F-max	0.07	0.08	0.02
2. $i=(0,1,2,0,1,2)$	GETP	0.05	0.07	0.046
$j=(1,1,1,1,1,1)$	GETDP	–	–	–
$k=(1,1,1,1,1,1)$	RPM	–	–	–
$F1 = 0, F2 \neq 0$	F-max	0.01	0.05	0.040
3. $i=(1,1,1,1,1,1)$	GETP	0.993	0.495	0.269
$j=(1,1,1,1,1,1)$	GETDP	–	–	–
$k \in [0,100]$	RPM	–	–	–
$F1 \neq 0, F2 = 0$	F-max	0.904	0.295	0.168
4. $i=(0,1,2,0,1,2)$	GETP	0.905	0.314	0.123
$j=(1,1,1,1,1,1)$	GETDP	–	–	–
$k \in [0,100]$	RPM	–	–	–
$F1 \neq 0, F2 \neq 0$	F-max	0.667	0.174	0.091
5. $i \in [0,2]$	GETP	0.991	0.497	0.206
$j=(1,1,1,1,1,1)$	GETDP	–	–	–
$k \in [0,100]$	RPM	–	–	–
$F1 \neq 0, F2 = 0$	F-max	0.874	0.266	0.139
6. $i=(1,1,1,1,1,1)$	GETP	0.955	0.426	0.179
$j=(1,2,4,1,2,4)$	GETDP	–	–	–
$k \in [0,100]$	RPM	–	–	–
$F1 \neq 0, F2 \neq 0$	F-max	0.747	0.264	0.127

Let us consider the previous simulation design, where the i.i.d. error term  $e(t)$  would be replaced by the Brownian motion. The difference is that with i.i.d. error used in previous sections the variance is constant, but with the Brownian motion, it is increasing in dependence on  $t$ . This may cause some trouble, since the bigger variance for bigger  $t$  means different sensitivity for effects influencing values close to  $t = 0$ , such as parameter  $i$  and effects that influence the values close to  $t = 1$  such as parameter  $j$ , see Fig. 1.

The standard deviation of the Brownian motion  $e(1)$  was kept ten times bigger than the standard deviation of the i.i.d. error, since then the increments in our discrete Brownian motion has the same standard deviation as the i.i.d. error and we get comparable results.

We present three tables in the same spirit as in the main text. The results here are calculated from 100 simulations only since we did not have enough time to finish the whole study. The full study will appear in the final version.

The estimated levels of significance are slightly liberal for the procedures using the Freedman-Lane algorithm. The powers of our tests are again much bigger than the powers of the other two tests. Even more, in some cases, the difference between these tests is more significant than for the i.i.d. error rate and in other cases, the difference between these tests is similar as for the i.i.d. error rate.

**Table 6** The estimated probabilities of rejecting of effect of interactions in factorial two factor FGLM with Brownian motion error

	Method	$\sigma(e(1)) = 3$	$\sigma(e(1)) = 5$	$\sigma(e(1)) = 8$
1. $\mathbf{i}=(0,1,2,0,1,2)$	GETP	0.09	0.09	0.15
$\mathbf{j}=(1,1,1,1,1,1)$	GETDP	0.11	0.11	0.09
$\mathbf{k}=(1,1,1,50,50,50)$	RPM	0.12	0.07	0.06
$\mathbf{I} = 0$	F-max	0.05	0.05	0.04
2. $\mathbf{i}=(0,1,2,1,1,1)$	GETP	0.97	0.45	0.22
$\mathbf{j}=(1,1,1,1,1,1)$	GETDP	0.95	0.46	0.17
$\mathbf{k}=(1,1,1,50,50,50)$	RPM	0.62	0.19	0.13
$\mathbf{I} \neq 0$	F-max	0.66	0.24	0.13
3. $\mathbf{i}=(0,1,2,0,1,2)$	GETP	0.10	0.06	0.05
$\mathbf{j}=(1,1,1,2,2,2)$	GETDP	0.10	0.07	0.05
$\mathbf{k}=(1,1,1,1,1,1)$	RPM	0.08	0.10	0.05
$\mathbf{I} = 0$	F-max	0.02	0.06	0.08
4. $\mathbf{i}=(0,1,2,1,1,1)$	GETP	1.00	0.54	0.28
$\mathbf{j}=(1,1,1,2,2,2)$	GETDP	1.00	0.58	0.27
$\mathbf{k}=(1,1,1,1,1,1)$	RPM	0.55	0.22	0.07
$\mathbf{I} \neq 0$	F-max	0.64	0.24	0.07

## References

- Abramovich F, Angelini C (2006) Testing in mixed-effects fanova models. *J Stat Plan Infer* 136(12):4326–4348
- Anderson MJ, Robinson J (2001) Permutation tests for linear models. *Austr New Zealand J Stat* 43(1):75–88. <https://doi.org/10.1111/1467-842X.00156>
- Anderson MJ, Ter Braak CJ (2003) Permutation tests for multi-factorial analysis of variance. *J Stat Comput Simul* 73(2):85–113
- Cuesta-Albertos JA, Febrero-Bande M (2010) A simple multiway anova for functional data. *TEST* 19(3):537–557. <https://doi.org/10.1007/s11749-010-0185-3>
- Febrero-Bande M, Oviedo de la Fuente M (2012) Statistical computing in functional data analysis: the R package fda.usc. *J Stat Softw* 51(4):1–28. <http://www.jstatsoft.org/v51/i04/>
- Ferraty F, Vieu P, Viguier-Pla S (2007) Factor-based comparison of groups of curves 51:4903–4910
- Freedman D, Lane D (1983) A nonstochastic interpretation of reported significance levels 1:292–98
- Hahn U (2012) A studentized permutation test for the comparison of spatial point patterns. *Am Stat Assoc J* 107(498):754–764
- Legendre P, Anderson MJ (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecol Monogr* 69(1):1–24
- Mrkvička T, Myllymäki M, Hahn U (2017) Multiple monte carlo testing, with applications in spatial point processes. *Stat Comput* 27(5):1239–1255. <https://doi.org/10.1007/s11222-016-9683-9>
- Mrkvička T, Myllymäki M, Jilek M, Hahn U (2018) A one-way anova test for functional data with graphical interpretation, arXiv:1612.03608 [stat.ME]
- Myllymäki M, Mrkvička T, Grabarnik P, Seijo H, Hahn U (2017) Global envelope tests for spatial processes. *J R Stat Soc: Series B (Stat Methodol)* 79(2):381–404. <https://doi.org/10.1111/rssb.12172>
- Nichols TE, Holmes AP (2001) Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping*
- Pantazis D, Nichols TE, Baillet S, Leahy RM (2005) A comparison of random field theory and permutation methods for the statistical analysis of meg data. *NeuroImage* 25(2):383–394. <http://www.sciencedirect.com/science/article/pii/S1053811904005671>
- Ramsay J, Silverman B (2006) *Functional data analysis*, 2nd edn. Springer Series in Statistics, Springer

Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE (2014) Permutation inference for the general linear model. *NeuroImage* 92:381–397. <http://www.sciencedirect.com/science/article/pii/S1053811914000913>

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