A one-way ANOVA test for functional data with graphical interpretation

Tomáš Mrkvička
Dpt. of Applied Mathematics and Informatics, Faculty of Economics, University of South Bohemia, Studentská 13, 37005 České Budějovice, Czech Republic e-mail: mrkvicka.toma@gmail.com
and

Ute Hahn
Centre for Stochastic Geometry and Advanced Bioimaging, Department of Mathematics, University of Aarhus, 8000 Aarhus C, Denmark e-mail: ute@imf.au.dk
and

Mari Myllymäki
Natural Resources Institute Finland (Luke), Latokartanonkaari 9, FI-00790 Helsinki, Finland e-mail: mari.myllymaki@luke.fi

Abstract: A new functional ANOVA test, with a graphical interpretation of the result, is presented. The test is an extension of the global envelope test introduced by Myllymäki et al. (2017, Global envelope tests for spatial processes, J. R. Statist. Soc. B 79, 381–404, doi: 10.1111/rssb.12172). The graphical interpretation is realized by a global envelope which is drawn jointly for all samples of functions. If an average function, computed over a sample, is out of the given envelope, the null hypothesis is rejected with the predetermined significance level $\alpha$. The advantages of the proposed one-way functional ANOVA are that it identifies the domains of the functions which are responsible for the potential rejection. We introduce two versions of this test: the first gives a graphical interpretation of the test results in the original space of the functions and the second immediately offers a post-hoc test by identifying the significant pair-wise differences between groups. The proposed tests rely on discretization of the functions, therefore the tests are also applicable in the multidimensional ANOVA problem.

MSC 2010 subject classifications: Primary 62H15; secondary 62G10.
Keywords and phrases: Functional ANOVA, Global envelope test, Groups comparison, Permutation test, Nonparametrical methods, Point processes, Rank envelope test.

1. Introduction

Functional data appear in a number of scientific fields, where the process of interest is continuously monitored. Those include e.g. monitoring of the share price, the temperature in a given location or monitoring of a body characteristic. A classical statistical problem is to decide, if there exist differences between the groups of functions (e.g. the control group and treatment group). This problem
is usually solved by determining the differences among the mean group functions and then we deal with a one-way functional ANOVA problem.

The functional ANOVA problem, both one-way and more complex designs, was previously studied by many authors. For example Cuevas et al. (2004) introduced asymptotic version of the ANOVA $F$-test, Zhang (2014) considers asymptotic or bootstrapped versions of $L_2$ norm based test, $F$-type statistic test and globalizing pointwise $F$-test. Górecki and Smaga (2015) introduced a method based on a basis function representation. Ramsay 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 (2010) applied the $F$-test on several random univariate projections and bound the tests together through false discovery rate.

There is also a possibility to transform the function into one number and use the classical ANOVA but such a procedure can be blind against some alternatives.

Also the nonparametric permutation procedures have been used to address this problem. Hahn (2012) used an one-dimensional integral deviation statistic to summarize the deviance between groups. Its distribution had been obtained by permuting of the functions. Nichols and Holmes (2001) concentrated either on certain pointwise statistics, such as the $F$-statistic, and found the distribution of its maxima by permutation, or on the size of the area restricted by some given threshold. Since these statistics need to satisfy the homogeneity across the functional domain, Pantazis et al. (2005) recommended to concentrate on the $p$-values which are implicitly homogeneous across the domain and find the distribution of its minima by permutation. This $p$-min and also $F$-max methods are able to identify the regions of rejections by identifying a threshold of the statistics of the interest.

However, none of the available methods is able to give a graphical interpretation of the test results in the original space of the functions which can help the user to understand what are the reasons of potential rejections, when or where the potential differences appear. Our new proposed method is based on the global rank envelope test and the extreme rank length measure introduced in Myllymäki et al. (2017). In particular, we introduce here graphical interpretation of the test based on the extreme rank length directly. The proposed tests are Monte Carlo tests where the simulations are obtained by permuting the functions. In this work, we concentrate on the one-way functional ANOVA problem only, because in this case the proposed Monte Carlo test is exact, i.e. its type I error is exactly the prescribed significance level $\alpha$.

In Section 2 we introduce two versions of our completely nonparametric method for the one-way functional ANOVA problem. The correction for heteroscedasticity is also described. The graphical interpretation of the second version of the test also gives an immediate post-hoc test, i.e. a test for which of the groups are different. Interestingly, this post-hoc comparison is done simultaneously with the ANOVA test, thus at the exact significance level $\alpha$. Therefore, no second comparison is needed to find which groups differ.
In Section 2.2 we also describe the global rank envelope test applied on the pointwise $F$-tests. This test does not have its graphical representation in the space of functions, but we introduce it as another possibility of applying the global envelope test.

In Section 3, we perform a simulation study in order to compare powers of our graphical procedures with the powers of the procedures which are already available. We have chosen only such procedures which are available through the software R (R Core Team, 2016). All our proposed methods can be found in the R package GET, which is available at github (https://github.com/myllym/GET).

In Section 4, we apply our methods to year temperature curves data and we study which days of the year are significantly warmer nowadays than in recent years.

Finally, in Section 5 we apply our methods to point pattern data, where the aim is to decide whether there are differences in the structure of the point patterns in the three groups. In this case, the typical approach is to choose a one-dimensional summary function, which summarizes the structure of the pattern, and compute it for each pattern. The summary function is typically a function of the one-dimensional distance. In the point patterns literature the group comparison is done either using functional ANOVA as in Aliy et al. (2013) or using a bootstrap procedure as in Diggle et al. (1991), Diggle et al. (2000) or Schladitz et al. (2003). Furthermore, Hahn (2012) proposed a pure permutation procedure to correct the inaccurate significance level of the bootstrap procedure. In these works univariate statistics summarizing the structure of the point patterns were used and permuted or bootstrapped. On the other hand in our approach whole summary functions are analyzed and we propose a way to take into account the differing variances of the used functions.

Section 6 is for further discussion.

### 2. Graphical functional ANOVA

Let us assume that we have $J$ groups which contain $n_1, \ldots, n_J$ functions and denote the functions by $T_{ij}, j = 1, \ldots, J, i = 1, \ldots, n_j$ observed on the finite interval $R = [a, b]$. Assume that $\{T_{ij}, i = 1, \ldots, n_i\}$ is an i.i.d. sample from a stochastic process $SP(\mu_j, \gamma_j)$ with a mean function $\mu_j$ and a covariance function $\gamma_j(s, t), s, t \in R$ for $j = 1, \ldots, J$. We want to test the hypothesis

$$H_0 : \mu_1(r) = \ldots = \mu_J(r), r \in R.$$  

We do not need to specify the stochastic process $SP$ in order to define our method and thus our method can be taken as completely nonparametric comparison of groups of functions.

The hypothesis $H_0$ is equivalent to the hypothesis

$$H'_0 : \mu_{ij'}(r) - \mu_j(r) = 0, r \in R, j' = 1, \ldots, J - 1, j = j', \ldots, J.$$  

This hypothesis corresponds to the post-hoc test done usually after the ANOVA test is significant.
In the following we introduce the test statistics either for hypothesis $H_0$ and $H'_0$ first for the case of equal covariance functions (i.e. for the case of $\gamma_1(s,t) = \ldots = \gamma_J(s,t)$, $s,t \in \mathbb{R}$) and then for the case of unequal variance functions (i.e. for the case of $\gamma_1(s,t)/\gamma_1(s,s) = \ldots = \gamma_J(s,t)/\gamma_J(s,s)$, $s,t \in \mathbb{R}$) (Sections 2.1 and 2.2). Then we describe how the permutations are performed under the null hypothesis (Section 2.3) and show how the combined rank envelope test can be used for these test statistics in order to obtain exact tests (Section 2.4).

In order to implement our method we rely on the discretization of functions. We assume that all functions are discretized in the same points $(r_1, \ldots, r_K)$. If this is not the case, we have to apply a smoothing techniques (e.g. those described in Zhang, 2014) and then make such a necessary discretization.

### 2.1. Test vectors

The hypothesis $H_0$ can be tested by the test vector consisting of the average of functions in the first group followed by the average of test functions in the second group, etc. We can shortly write that

$$ T = (\overline{T}_1(r), \overline{T}_2(r), \ldots, \overline{T}_J(r)), $$

where $\overline{T}_j(r) = (\overline{T}_j(r_1), \ldots, \overline{T}_j(r_K))$. Thus, the length of the test vector becomes $J \times K$, where $K$ stands for the number of $r$ values to which the functions are discretized.

The hypothesis $H'_0$ can be tested by test vector consisting of the differences of the group averages of functions. We can shortly write that

$$ T' = (\overline{T}_1(r) - \overline{T}_2(r), \overline{T}_1(r) - \overline{T}_3(r), \ldots, \overline{T}_{J-1}(r) - \overline{T}_J(r)). $$

Here the length of the test vector becomes $J(J - 1)/2 \times K$.

#### 2.1.1. Correction for an unequal variances

To deal with different variances of functions in different groups, we consider the rescaled functions $S_{ij}$ instead of the original functions $T_{ij}$,

$$ S_{ij}(r) = \frac{T_{ij}(r) - \overline{T}(r)}{\sqrt{\text{Var}(T(r))}} \cdot \sqrt{\text{Var}(T(r))} + T(r), \quad (1) $$

where the overall sample mean $\overline{T}(r)$ and overall sample variance $\text{Var}(T(r))$ are involved to keep the mean and variability of the functions at the original scale. The group sample variance $\text{Var}(T_j(r))$ corrects the unequal variances.

For small samples, the sample variance estimators can have big variance, which may influence the test procedure undesirably. In order to deal with this problem the variances can be smoothed by applying moving average (MA)
the estimated variance with a chosen window size $b$. Thus, the rescaled functions take the form

$$S_{ij}(r) = \frac{T_{ij}(r) - \bar{T}(r)}{\sqrt{MA_b(\text{Var}(T(r)))}} \cdot \sqrt{MA_b(\text{Var}(T(r)))} + T(r).$$

(2)

Thereafter, the test vectors are composed in the same way as in the case of the equal variances but with rescaled functions:

$$T_1 = (\bar{S}_1(r), \bar{S}_2(r), \ldots, \bar{S}_J(r)).$$

(3)

where $\bar{S}_j(r) = (\bar{S}_j(r_1), \ldots, \bar{S}_j(r_K))$, and

$$T'_1 = (\bar{S}_1(r) - \bar{S}_2(r), \bar{S}_1(r) - \bar{S}_3(r), \ldots, \bar{S}_{J-1}(r) - \bar{S}_J(r)).$$

(4)

Obviously, for the test vector $T'_1$ the last term in (1) or (2) plays no role and could be neglected.

**2.2. Rank envelope $F$-test**

When a graphical interpretation is not of the interest but the area of rejection is of the interest, one can utilize the $F$-test for each $r \in R$ separately and form the test vector from the $r$-wise $F$-statistics. In this case the correction for unequal variances can be done by choosing the variance corrected $F$-statistic.

The test vector is then

$$T_2 = (F(r_1), F(r_2), \ldots, F(r_K)),$$

where $F(r_k)$ stands for the $F$-statistic for $r_k$. This test vector is not able to detect which groups are different and therefore a post-hoc test would have to be performed afterwards.

**2.3. Permutations and exchangability of the test vectors**

The most important aspect of the permutation tests is the manner in which data are shuffled under the null hypothesis. In all our one-way ANOVA tests, we perform the simple permutation of raw functions among the groups. That is, if $G$ is a vector of group indices of length $N = \sum_{j=1}^{J} n_j$, then the permutation matrix $P$ is $N \times N$ matrix that has all elements being either 0 or 1, each row and column having exactly one 1. Pre-multiplicating the group indices $G$ by the matrix $P$ permutes the group indices. Note that the possible correction for unequal variances is performed prior to the permutation and the permutations are consequently done for the rescaled functions (1) or (2).

We say that a test vector $T$ is exchangeable if the observed and simulated (permuted) test vectors $T^1, \ldots, T^*$ are exchangeable, i.e. the joint distribution of $T^1, \ldots, T^*$ is not affected by permutation.
Proposition 2.1. The test vectors $T$, $T'$ and $T_2$ are exchangeable for permutations $P$ under the correspondent null hypotheses in the case of equal variances. The test vectors $T_1$, $T'_1$ and $T_2$ are asymptotically exchangeable for permutations $P$ under correspondent null hypotheses in the case of unequal variances. The asymptotics is taken over $\min_{j=1}^{J} n_j$.

Proof. Under null hypotheses and equal variances the set of all functions is an i.i.d. sample from a stochastic process. Since the permutations are performed on the whole functions (i.e. the block permutation scheme is used) the joint distribution of $T$ is equal to the joint distribution of $T'$ for permuted groups $P_G$.

In the case of unequal variances the functions are first scaled by the sample group variance for computation of $T_1$ and $T'_1$. The sample group variance $\text{Var}(T_j(r))$ converges a.s. to the true group variance. Similar holds for overall sample mean and overall sample variance. Thus the stochastic process $S_{ij}$ converges in distribution to $SP(\mu, \gamma_j(s,t) \gamma(s,s)/\gamma_j(s,s))$, where $\gamma(s,s)$ is the overall variance. Under the null hypotheses and unequal variances these stochastic processes are equal and thus the test vectors are asymptotically exchangeable. Similar proof can be made for $T_2$ in the case of unequal variances.

2.4. Combined global rank envelope test

The idea of the global rank envelope was introduced in Myllymäki et al. (2017) for point process testing. Further Mrkvička et al. (2017) extended the notion of this global envelope for general multivariate test vectors. This extension applies, e.g., to the case where the multivariate vector consists of values of two or more functions at once. We first recall the measures and associated $p$-values introduced in Myllymäki et al. (2017). Second, we define the global extreme rank length envelope as a refinement of the global rank envelope.

Assume the general multivariate vector of the form

$$V = (V_1, \ldots, V_d).$$

Let $V_1, \ldots, V_s$ be the multivariate vectors generated by permutations under the null hypothesis. Let $V_1$ denote the vector obtained by identical permutation.

First we define the extreme rank $R_i$ of the vector $V_i = (V_{i1}, \ldots, V_{id})$ as the minimum of its pointwise ranks, namely

$$R_i = \min_{k=1,\ldots,d} R_{ik}. \quad (5)$$

where the pointwise rank $R_{ik}$ is the rank of the element $V_{ik}$ among the corresponding elements $V_{1k}, V_{2k}, \ldots, V_{sk}$ of the $s$ vectors such that the lowest ranks correspond to the most extreme values of the statistics. How the pointwise ranks are determined, depends on whether a one-sided or a two-sided envelope test is to be performed: Let $r_{1k}, r_{2k}, \ldots, r_{sk}$ be the raw ranks of $V_{1k}, V_{2k}, \ldots, V_{sk}$, such
that the smallest $V_{ik}$ has rank 1. In the case of ties, the raw ranks are averaged. The pointwise ranks are then calculated as

$$R_{ik} = \begin{cases} r_{ik}, & \text{for one-sided test, small } V \text{ is considered extreme} \\ s + 1 - r_{ik}, & \text{for one-sided test, large } V \text{ is considered extreme} \\ \min(r_{ik}, s + 1 - r_{ik}), & \text{for two-sided test.} \end{cases}$$

The extreme ranks can contain many ties, e.g. in a one-sided test with $d$-variate vectors, up to $d$ out of the $s$ vectors can take the rank 1. Therefore we need to break these ties in an efficient way. Ordering of the vectors by the extreme rank length (Myllymäki et al., 2017) refines the extreme rank ordering in order to minimize the possibility of ties.

Consider the vectors of pointwise ordered ranks $R_i = (R_i[1], R_i[2], \ldots, R_i[d])$, where $\{R_i[1], \ldots, R_i[d]\} = \{R_1[1], \ldots, R_id\}$ and $R_i[k] \leq R_1'[k]$ whenever $k \leq k'$. The extreme rank given in (5) corresponds to $R_i = R_i[1]$. The extreme rank length measure of the vectors $R_i$ is equal to

$$R_{i}^{\text{erl}} = \frac{1}{s} \sum_{i'=1}^{s} 1(R_i \prec R_i')$$

(7)

where

$$R_i \prec R_i' \iff \exists n \leq d : R_i[k] = R_i'[k] \forall k < n, R_i[n] < R_i'[n].$$

We remark here that Narisetty and Nair (2016) independently defined the two-sided extreme rank length measure as a functional depth and called it extremal depth.

2.4.1. $p$-values

We distinguish three different $p$-values attached to the rank envelope test. All the $p$-values are based on Monte Carlo testing principles. The conservative and liberal $p$-values are given as

$$p_+ = \frac{1}{s} \sum_{i=1}^{s} 1(R_i \leq R_1)/s, \quad p_- = \frac{1}{s} \sum_{i=1}^{s} 1(R_i < R_1)/s. \quad \text{(8)}$$

The $p$-value based on the extreme rank length ordering is given as

$$p^{\text{erl}} = 1 - \frac{1}{s} \sum_{i=1}^{s} 1(R_1 \prec R_i)/s. \quad \text{(9)}$$

According to Myllymäki et al. (2017, Proposition 6.1), it holds that $p_- < p^{\text{erl}} \leq p_+$. Note here, that there still can appear some ties in between $R_1$ and $R_i$, $i = 2, \ldots, s$. But since these ties are unlikely to happen, we define $p^{\text{erl}}$ as the conservative $p$-value. Alternatively the ties can be broken by randomization.
2.4.2. The new graphical envelope

Myllymäki et al. (2017) defined the graphical global envelope with respect to the ordering of the ranks \( R_i, i = 1, \ldots, s \). This ordering can have a lot of ties and consequently the graphical envelope based on this ordering requires a lot of permutations in order to be precise. We eliminate this problem in this paper by defining the graphical envelope with respect to the extreme rank length ordering (7).

Assuming that all the \( V_i \) follow the same joint distribution, we construct rank envelopes with level \( (1 - \alpha) \) as sets \( \{ V^{(\alpha)}_{\text{low}}, V^{(\alpha)}_{\text{upp}} \} \) such that, the probability that \( V = (V_1, \ldots, V_d) \) falls outside this envelope in any of the \( d \) points is less or equal to \( \alpha \),

\[
Pr(\exists k \in \{1, \ldots, d\} : V_k \notin [V^{(\alpha)}_{\text{low}}, V^{(\alpha)}_{\text{upp}}]) \leq \alpha.
\]

Let \( I_{\alpha} = \{ i \in 1, \ldots, s : R_{\text{erl}}^{(\alpha)} \geq R_{\text{erl}}^{(i)} \} \) be the index set of vectors, and define

\[
V^{(\alpha)}_{\text{low}} = \min_{i \in I_{\alpha}} V_{ik}, \quad V^{(\alpha)}_{\text{upp}} = \max_{i \in I_{\alpha}} V_{ik}
\]

for the two-sided test, following the idea of Narisetty and Nair (2016) to define an envelope as a convex hull. For one-sided tests, let \( V^{(\alpha)}_{\text{low}} = -\infty \) or \( V^{(\alpha)}_{\text{upp}} = \infty \), respectively, for all \( k = 1, \ldots, d \). By choosing \( R_{\text{erl}}^{(\alpha)} \) to be the largest value in \( \{ R_{\text{erl}}^{1}, \ldots, R_{\text{erl}}^{s} \} \) for which

\[
\sum_{i=1}^{s} \mathbb{1}(R_{i}^{\text{erl}} < R_{(\alpha)}^{\text{erl}}) \leq \alpha s, \quad (10)
\]

we get the graphical interpretation described in the next subsection.

The following theorem states that inference based on the \( p_{\text{erl}} \) and the global envelope specified by \( V^{(\alpha)}_{\text{low}} \) and \( V^{(\alpha)}_{\text{upp}} \) are equivalent. Therefore, we can refer to this envelope as the 100 \( \cdot (1 - \alpha) \)% global extreme rank length envelope.

**Theorem 2.2.** Let \( p_{\text{erl}} \) be as given in (9), and \( V^{(\alpha)}_{\text{low}}, V^{(\alpha)}_{\text{upp}} \) define the 100 \( \cdot (1 - \alpha) \)% global extreme rank length envelope. Then, assuming that there are no pointwise ties with probability 1, it holds that:

1. \( V_{ik} < V^{(\alpha)}_{\text{low}} \) or \( V_{ik} > V^{(\alpha)}_{\text{upp}} \) for some \( k = 1, \ldots, d \) iff \( p_{\text{erl}} \leq \alpha \), in which case the null hypothesis is rejected;
2. \( V^{(\alpha)}_{\text{low}} \leq V_{ik} \leq V^{(\alpha)}_{\text{upp}} \) for all \( k = 1, \ldots, d \) iff \( p_{\text{erl}} > \alpha \), and thus the null hypothesis is not rejected;

**Proof.** According to the definition of \( p_{\text{erl}} \) is \( p_{\text{erl}} \leq \alpha \) iff number of \( R_i \) smaller or equal to \( R_1 \) is smaller or equal to \( \alpha \). That is equivalent, according to the definition of \( R_{(\alpha)}^{\text{erl}} \), to the \( R_{1}^{\text{erl}} < R_{(\alpha)}^{\text{erl}} \). This holds iff \( 1 \notin I_{\alpha} \), which is equivalent to \( V_{ik} < V^{(\alpha)}_{\text{low}} \) or \( V_{ik} > V^{(\alpha)}_{\text{upp}} \) for some \( k = 1, \ldots, d \) according to the definition of the extreme rank length envelope.

The second part of the proof can be proven equivalently.
In the following theorem, we will prove that the global extreme rank length envelope is contained in the global rank envelope. The \( l \)-th rank envelope was defined in Myllymäki et al. (2017) as
\[
V^{(l)}_{\text{low}k} = \min_{i=1,...,s} V_{ik} \quad \text{and} \quad V^{(l)}_{\text{upp}k} = \max_{i=1,...,s} V_{ik} \quad \text{for} \quad k = 1,\ldots,d, \tag{11}
\]
where \( \min \) and \( \max \) denote the \( l \)-th smallest and largest values, respectively, and \( l = 1, 2, \ldots, \lfloor (s + 1)/2 \rfloor \).

**Lemma 2.1.** Let
\[
W^{(l)}_{\text{low}k} = \min_{i \in I_l} V_{ik}, \quad W^{(l)}_{\text{upp}k} = \max_{i \in I_l} V_{ik}
\]
for \( I_l = \{ i \in 1, \ldots, s : R_i \geq l \} \). Then \( W^{(l)}_{\text{low}k} \geq V^{(l)}_{\text{low}k} \) and \( W^{(l)}_{\text{upp}k} \leq V^{(l)}_{\text{upp}k} \).

**Proof.** Since all vectors in \( I_l \) have the extreme rank greater or equal to \( l \) and thus \( R_{ik} \geq l \) or \( R_{ik} \leq s - l + 1 \) (for two-sided test), the envelope defined by \( W \) is contained in the envelope defined by \( V \). \( \square \)

**Theorem 2.3.** The \( 100 \cdot (1 - \alpha) \)% global extreme rank length envelope is contained in the \( 100 \cdot (1 - \alpha) \)% global rank envelope.

**Proof.** This is a direct consequence of Lemma 2.1 and the fact that \( I_\alpha \) contains a smaller number of functions than \( I_l \) for \( l = l_\alpha = \max \{ l : \sum_{i=1}^s 1( R_i < l ) \leq \alpha s \} \), which is the critical rank for the \( 100 \cdot (1 - \alpha) \)% global rank envelope. \( \square \)

Remark here that the \((l + 1)\)-th rank envelope given by \( V^{(l+1)}_{\text{low}k} \) and \( V^{(l+1)}_{\text{upp}k} \) is not necessarily contained in the \( l \)-th extreme rank length envelope given by \( W^{(l)}_{\text{low}k} \) and \( W^{(l)}_{\text{upp}k} \).

### 2.5. One-way graphical functional ANOVA test

The proposed methods are performed in three steps. First the test vector is chosen. Second \( s \) permutations are applied to the raw functions (or on the rescaled functions in the case of unequal variance) and the chosen test vector is computed for each permutation. Third the combined rank envelope test is applied to the set of \( s \) test vectors. The following theorem specifies the graphical interpretation of our proposed tests and claims the exactness of the graphical method which is given up to the level of ties between the extreme rank lengths.

**Theorem 2.4.** Consider one-way graphical functional ANOVA test with \( T \), \( T' \) or \( T_2 \) chosen as the test vector and equal variances. Let \( p^{\text{crit}} \) be as given in (9), and \( T^{\alpha}_{\text{low}k} \) and \( T^{\alpha}_{\text{upp}k} \) define the \( 100 \cdot (1 - \alpha) \)% global extreme rank length envelope. Then, assuming that there are no pointwise ties with probability 1 in the stochastic process \( SP(\mu, \gamma) \), it holds that:

1. \( T_{1k} < T^{\alpha}_{\text{low}k} \) or \( T_{1k} > T^{\alpha}_{\text{upp}k} \) for some \( k \) iff \( p^{\text{crit}} \leq \alpha \), in which case the null hypothesis is rejected;
2. \( T_{\text{low}}^k \leq T_{1k} \leq T_{\text{upp}}^k \) for all \( k \) iff \( \text{per} \geq \alpha \), and thus the null hypothesis is not rejected.

Theorem 2.4 is a direct consequence of Proposition 2.1 and Theorem 2.2. For the case of unequal variances the above interpretation is due to the Proposition 2.1 achieved only asymptotically.

Since we apply here the global extreme rank length envelope and not the global rank envelope as it was the case in our previous works, a lower number of permutations can be used. Anyway we recommend to use some thousands of permutations at minimum. In case of many groups the number of permutations has to increase in order to not lose the power of the test, as it is demonstrated in the simulation study.

It is important to mention here that the graphical interpretation automatically identifies which groups are responsible for the potential rejection and also it identifies which parts of the functions are responsible for the rejection. This is very important for the interpretation of the result of the test.

Note that for the test vector \( T_2 \) the one-sided rank test has to be used, whereas for the other test vectors the two-sided rank test is used.

### 2.6. Comparison to other permutation methods

The nonparametric permutation methods often used in the brain image statistics are similar to our proposed methods, therefore we would like to stress the differences. The single threshold test (Nichols and Holmes, 2001) of a certain statistic whose maximum is permuted is limited to the statistics that are homogeneous across the functional domain, in order to be sensitive in the whole functional domain and not only in the part of the domain where the functions are the most varying. The \( p \)-min permutation procedure used e.g. in Pantazis et al. (2005) solves this problem. This method can be viewed as our rank envelope \( F \)-test. However, the \( p \)-min permutation procedure uses the conservative \( p \)-value of our rank envelope \( F \)-test, i.e. the upper bound of the \( p \)-interval, \( p_+ \) in (9). On the other hand, our rank envelope \( F \)-test is equipped also with the extreme rank length \( p \)-value which solves the problem of ties in the \( p \)-min distribution and therefore it significantly reduces the conservativeness of the test.

Further, our graphical functional ANOVA test gives the graphical interpretation in the original space of functions and for each group of functions, whereas the \( p \)-min test gives it only in the transformed space of \( p \)-values and for all groups simultaneously. Therefore the \( p \)-min test is only able to identify the regions of rejection. Our graphical functional ANOVA test is also equipped with the global extreme rank length envelope which informs the user about the variability of the curves in the study. Finally, the graphical functional ANOVA test is defined here also for combining several post-hoc tests together in one test and therefore it indicates which two groups are different and where they are different.
3. Simulation study

Our simulation study has two parts. First we compared our methods with some existing methods on a design taken from the study of Cuevas et al. (2004) in order to check if our methods are comparable in power and significance level to the existing methods. Second we changed the design from comparing three groups into comparing ten groups in order to check how much of the power is lost by having a long test vector $\mathbf{T}$ or $\mathbf{T}'$ with respect to other procedures.

In our study we included the following procedures from the literature:

- asymptotic version of the $F$-test (AsF) introduced in Cuevas et al. (2004),
- random univariate projection method (RPM) introduced in Cuesta-Albertos and Febrero-Bande (2010),
- bootstrapped version of $F$-type statistic test (Fb) introduced in Zhang (2014),
- globalizing pointwise $F$-test (GPF) introduced in Zhang (2014),
- a method based on a basis function representation (FP) introduced in Górecki and Smaga (2015),
- one-dimensional integral permutation test (IPT) introduced in Hahn (2012),
- $F$-max permutation procedure described in Nichols and Holmes (2001) and
- $p$-min permutation procedure described in Pantazis et al. (2005).

Our methods included in the study were

- the graphical functional ANOVA applied on group means (GFAM) where test vector $\mathbf{T}$ is used,
- the graphical functional ANOVA applied on group means contrasts (GFAC) where test vector $\mathbf{T}'$ is used and
- the rank envelope $F$-test (REF).

First we compared all the above methods using an artificial example of $J = 3$ groups and $n = 10$ functions in each group observed in the interval $[0, 1]$ through 100 evenly spread discrete points. Four different models with two different autocorrelation error structures were considered. The models were

- M1: $T_{ij}(r) = r(1 - r) + e_{ij}(r), \ i = 1, 2, 3, j = 1, \ldots, 10,$
- M2: $T_{ij}(r) = r^i(1 - r)^{6-i} + e_{ij}(r), \ i = 1, 2, 3, j = 1, \ldots, 10,$
- M3: $T_{ij}(r) = r^{i/5}(1 - r)^{6-i/5} + e_{ij}(r), \ i = 1, 2, 3, j = 1, \ldots, 10,$
- M4: $T_{ij}(r) = 1 + i/50 + e_{ij}(r), \ i = 1, 2, 3, j = 1, \ldots, 10.$

The first autocorrelation structure of errors was the white noise, i.e. $e_{ij}(r_k), k = 1, \ldots, 100,$ were iid random variables with $N(0, \sigma)$. In the second structure, the errors $e_{ij}(r)$ were modelled by standard Brownian process with dispersion parameter $\sigma$. For each combination of the four models and two autocorrelation structures, we considered six different contaminations for the deterministic part of the model given by six different standard deviations $\sigma_1 = 0.05, \sigma_2 = 0.1, \sigma_3 = 0.15, \sigma_4 = 0.2, \sigma_5 = 0.4, \sigma_6 = 0.8$. Every standard deviation was twice the
previous one, except $\sigma_3$ which was added in the middle of $\sigma_2$ and $\sigma_4$ in order to increase the sensitivity of the simulation study.

The model M1 corresponds to the situation where $H_0$ is true. Thus, in this case, the predetermined significance level was estimated. It was set to $\alpha = 0.05$ in all cases. The other models represent different situations where $H_0$ is false. The mean functions in model M2 and M3 have different shape, whereas the mean functions in model M4 are constant.

All the tested procedures were run using 2000 Monte Carlo replications or permutations. The RPM was run with 30 random projections and the false discovery rate $p$-value computed out of these projections was used as a final output of this procedure. The extreme rank length $p$-value was used as the output of all our new tests (GFAM, GFAC, REF). We performed 1000 simulations for each combination of model, autocorrelation structure and standard deviation, and we computed the proportion of rejections to obtain the estimates of significance levels and powers. All the results are summarized in Table 1 for the white noise cases and in Table 2 for the Brownian error cases.

The empirical significance level should be in the interval $(0.037, 0.064)$ with the probability 0.95 (given by the 2.5% and 97.5% quantiles of the binomial distribution with parameters 1000 and 0.05). This is satisfied for most of the studied cases. The only exception are the AsF, Fb and GPF procedures in the case of the independent errors. This exception is caused by the high number of discretized points. For 20 discretized points these methods did not show this feature.

The estimated powers of our three procedures are significantly greater than the powers of AsF, RPM, Fb and GPF methods for all three studied models especially for the case of the Brownian errors. They are comparable to the powers of $p$-min and $F$-max procedures, because they have similar nature. FP and IPT tests are more powerful for iid errors and less powerful for Brownian errors with respect to our proposed tests, this is caused by completely different nature of the corresponding test statistics. Surprisingly the powers of the REF method are not greater than the powers of the RECM and RECMD methods which are fully nonparametric (the errors are normally distributed in the considered models).

In the second part of our simulation study we took the model M3 and extended this model for ten groups, considering

$$M: T_{ij}(r) = r^{i/5}(1 - r)^{6-i/5} + e_{ij}(r), \ i = 2, \ldots, 11, \ j = 1, \ldots, 10.$$ 

We used again the two correlation structures in the model and six levels of contamination as in the first part of the study. The results are summarized in Table 3 for the white noise cases in the upper part and for the Brownian error in the lower part of the table.

The relations between powers of different methods were the same as in the case of the three groups. Also we do not see any decrease in the power for the GFAM and GFAC methods with respect to the REF and $p$-min methods in the iid case. On the other hand there was such decrease in the Brownian error case.
This loss of power in the GFAM and GFAC tests can be prevented by increasing number of permutations: We performed the experiment also with 10000 permutations and obtained very similar powers as with 2000 permutations for all the other methods except GFAM and GFAC: the power of GFAM increased from 0.360 to 0.884 and the power of GFAC from 0.360 to 0.688 in the case of $\sigma_6$ and Brownian errors. Thus the power of the graphical tests was comparable to the power of REF and $p$-min tests with 10000 permutations.

4. Real case study - comparison of groups of temperature curves

As the first real case example we chose classical year temperature curves. We used the water temperature data sampled at the water level of Rimov reservoir in Czech republic every day from 1979 to 2014. The water temperature data are naturally smoothed. We artificially constructed three groups of year temperature curves in order to see if there are any changes in the year temperature curves. The first group was taken to consist of years 1979–1990, the second group of years 1991–2002 and the third group group of years 2003–2014. Figure 1 shows the temperature curves in the three groups.

![Figure 1. The temperature curves in the three groups.](image)

Figure 2 shows the comparison of the three groups of functions by means of $T$, i.e. the GFAM method. Each subplot shows the comparison of a group with respect to variability of all groups. The result shows that only the third group deviates from the overall mean significantly and that this deviation is around 120th day of the year as can be seen on the bottom right panel in details. This result corresponds to the finding that the summer comes earlier in the last decade than earlier. The 95% global envelope demonstrates here the variability of the average curve in the original space of curves under the $H_0$ hypothesis. The test was performed with 2500 permutations.
Rank envelope test (using extreme rank length): 0.004

Figure 2. Rank envelope test for comparison of three groups of temperature curves. The subplots correspond to the comparison of the certain group average with respect to the variability in all groups. The right bottom subplot is a zoom to the only significant region seen in the third group. The grey areas represent the 95% global envelope.
Figure 3 shows the result of the comparison of three groups of curves via difference of group averages when $T'$ was used as test vector, i.e. the GFAC method. Each subplot shows the comparison of two groups. The test statistic being negative corresponds to the situation that the second group temperature is higher than the first group temperature in the comparison. This test contains implicitly the post hoc comparison of groups.

As an illustration of the rank envelope $F$-test, we computed the one-sided rank test for the $F$-statistics. The resulted 95% global upper envelope (using 2500 permutations) can be seen in Figure 4 showing the days responsible for the rejection, but this representation is not in the space of temperature curves as the two previous tests.
Figure 4. Rank envelope test (using extreme rank length) for comparison of three groups of temperature curves via the $F$-statistic of the ANOVA. Solid line corresponds to the data $F$-values, the dashed line is the 95% global upper envelope, and the dotted line corresponds to the average of $F$-values from the permutations.
5. Real case study - comparison of groups of point patterns

As second real case example we chose an example from point patterns statistics because the summary functions of point processes have often different variances. The usual practice in point pattern literature is to represent a structure of the point pattern by a summary function which is a function of the distance $r$, usually the distance between two points in the pattern. Thus in fact each point pattern is replaced by one summary function and thus the comparison of groups of point patterns leads to the functional ANOVA problem.

To describe our approach we reanalyse the data of Diggle et al. (1991) containing three groups of pyramidal neurons in the cingulate cortex of humans, the normal (control) group, schizoaffective group and schizophrenic group. One representative pattern from each group can be seen in Figure 5. Each point pattern is rescaled to the unit observing window. (The original window was 1421 $\mu$m $\times$ 1421 $\mu$m.) We discarded the point patterns with less than 20 points prior to the analysis, in order to concentrate at meaningful point patterns, which led to 12 point patterns in the normal group, 7 patterns in the schizoaffective group and 7 patterns in the schizophrenic group.

As the summary function we chose the estimator of the centred $L$-function with the isotropic correction, since it is the most commonly used summary function in point pattern analysis. Shortly, this function represents the number of neighboring points within the given distance from a typical point of the pattern compared with the expected number of neighbors in the case of complete spatial randomness. That is, if the centred $L$-function is positive for a given distance, it indicates attraction of points up to the distance, while negative values of the function indicate repulsion of points. The details can be found e.g. in Illian et al. (2008). Figure 6 shows these estimated $L$-functions in the three groups.
5.1. Scaled L-functions

Since the point patterns have different numbers of points, the L-functions are estimated with different precisions. To take into account these differences, prior to the functional ANOVA test, we scale the estimated centred L-functions, \( \hat{L}_{ij}^c(r) \). Because the variance of the estimated L-function behaves approximately as \( 1/m_{ij} \), where \( m_{ij} \) is the number of points in the \( ij \)-th point pattern, we first stabilize the variances in the same manner as for unequal group variances:

\[
S_{ij}(r) = \frac{\hat{L}_{ij}^c(r) - \bar{L}^c(r)}{\sqrt{1/m_{ij}}} \tau + \bar{L}^c(r),
\]

where \( \bar{L}^c(r) \) is the sample mean of all functions and \( \tau = \frac{\sum_{ij} \sqrt{1/m_{ij}}}{\sum_{ij} 1} \) is the estimator of the overall standard deviation of the functions. Then all \( S_{ij}(r) \) have (approximately) equal variances and we can take (3) or (4) as our test vector.

Figure 7 shows the scaled L-functions.

5.2. Test results

Figure 8 shows the result of the GFAC method when the above scaling (12) is applied to the original L-functions in order to take into account their unequal variances. The number of \( r \)-values was set to 500 for each L-function. Each subplot shows the comparison of 2 groups. The test statistic being positive corresponds to the situation that the first group is more clustered than the second group in the comparison. Our result shows no differences at the significance level 0.05 between groups similarly as in the originally study.
Figure 7. The scaled estimated centred L-functions (12) in the three groups.

Figure 8. Rank envelope test for comparison of the three groups of L-functions via difference of group weighted averages using 2500 permutations. The left subplot corresponds to the difference between the first and second group, the middle subplot corresponds to the difference between the first and third group and the right subplot corresponds to the difference between the second and third group. The grey area represents the 95% global envelope.
Figure 9 further shows the result of the GFAM method applied to the scaled $L$-functions (12). Each subplot shows the comparison of a group with respect to the rest of the groups. The test statistic being positive corresponds to the situation that the group is more clustered than the rest of groups. Our result again shows no differences between groups. We also applied the rank envelope $F$-test. Also this test showed no significant differences between the groups ($p = 0.10$). (The figure is omitted.)

![Graph](image)

**Figure 9.** Rank envelope test for comparison of the three groups of $L$-functions via difference of the averages of a group and the rest of the groups using 2500 permutations. The left, middle and right subplots correspond to the difference of the first, second and third groups with respect to the remaining groups, respectively. The grey area represents the 95% global envelope.

From the first subplot of Figure 8, we can observe that the first and second group are rather similar. Therefore we joined the first and second group as was done in the original study. The result of the comparison of first and second group with third group is shown in Figure 10. Thus, we detect a significant difference between the first and second group with respect to the third group ($p = 0.035$), whereas in the original study by Diggle et al. (2000) differences were not detected even though Hahn (2012) proved that the original method was liberal due to the permutation of functional residuals. From the graphical envelope we see that differences occur for small distances around 100 $\mu m$ and indicate that the patterns in the third group are more regular than those in the first and the second group.

6. Discussion

A new one-way graphical functional ANOVA test was introduced in this paper. The test has exact type I error and it provides a graphical interpretation which is essential for the interpretation of the results. Also corrections for unequal variances were presented. Two different test vectors can be used leading to two
**Figure 10.** Rank envelope test for comparison of the two groups of scaled $L$-functions (12) via difference of group averages using 2500 permutations. The plot corresponds to the difference between the joined first and second group, and the third group. The grey area represents the 95% global envelope.
different graphical interpretations. The first option compares every group with the rest of the groups. The second option compares differences between every pair of groups similarly like a Tukey post-hoc test in the univariate ANOVA. A positive side effect of our methods is that the post-hoc test is provided together with the main ANOVA procedure at the given significance level. Further the rank envelope $F$-test was proposed which can be used when the graphical interpretation is not required in the space of the functions.

Since the proposed test works in a highly dimensional multivariate settings, no smoothness of the functions is required. On the other hand, the same discretization of functions is required for every function. If this is not the case, a smoothing technique has to be applied followed by further identical discretization of functions.

Our new methods were compared to the other functional ANOVA procedures, available through the software R, with respect to their power. The permutation procedures were uniformly more powerful than asymptotic $F$-test and random projection methods, bootstrapped $F$-type statistic test and globalizing point-wise $F$-test in our study. Considering the permutation methods, none of the procedures was uniformly the most powerful one. The $F$-max and $p$-min procedures were very similar in power to our REF method (they are the most similar methods). The integral based procedure IPT and procedure based on the basis representation were very different in power than the other tests. For the iid case they were more powerful whereas for the Brownian case they were less powerful than our proposed methods. The opposite can be said about comparison of $F$-max and $p$-min with our proposed procedures. An important feature of all studied permutation methods is that these procedures do not loose their power when the functions are more densely discretized.

Importantly, our simulation study shows that there is no procedure which would be uniformly more powerful than our proposed procedures. Therefore, we believe that our procedures are useful in practice due to their graphical interpretation and post-hoc nature.

Our simulation study also shows that our methods can lose some power (with respect to the powers of other permutation methods, but not with respect to the methods of other nature) when a lot of groups are analyzed. In such a case where the test vector is very long, the number of permutations has to be increased in order to eliminate this problem.

Our new procedures were designed for the one-way functional ANOVA design. A question of our future research is how these procedures can be extended into multi-way design. Since the permutation of the functional residuals leads to a liberal method, the problem has to be solved in a more complex way.

Acknowledgements

Mrkvička has been financially supported by the Grant Agency of Czech Republic (Project No. 16-03708S) and Mari Myllymäki by the Academy of Finland (project numbers 295100, 306875). Hahn’s research has been supported by the
Centre for Stochastic Geometry and Advanced Bioimaging, funded by the Vil-lum Foundation. We are also greatful to Petr Znachor, the solver of the Grant Agency of Czech Republic (Project No. 15-13750S), for providing the temperature data.

References

Abramovich, F. and C. Angelini (2006). Testing in mixed-effects \{FANOVA\} models. *Journal of Statistical Planning and Inference* 136(12), 4326 – 4348.

Aliy, M., J. Seguin, A. Fischer, N. Mignet, L. Wendling, and T. Hurtut (2013). Comparison of the spatial organization in colorectal tumors using second-order statistics and functional anova. *Image and Signal Processing and Analysis (ISPA), 2013 8th International Symposium on*, 257–261.

Cuesta-Albertos, J. and M. Febrero-Bande (2010). A simple multiway anova for functional data. *Test* 19(3), 537–557.

Cuevas, A., M. Febrero, and R. Fraiman (2004). An anova test for functional data. *Computational Statistics & Data Analysis* 47(1), 111 – 122.

Diggle, P. J., N. Lange, and F. M. Beneš (1991). Analysis of variance for replicated spatial point patterns in clinical neuroanatomy. *Journal of the American Statistical Association* 86, 618–625.

Diggle, P. J., J. Mateu, and H. E. Clough (2000). A comparison between parametric and non-parametric approaches to the analysis of replicated spatial point patterns. *Advances in Applied Probability* 32, 331–343.

Ferraty, F., P. Vieu, and S. Viguier-Pla (2007). Factor-based comparison of groups of curves. *Computational Statistics & Data Analysis* 51(10), 4903 – 4910.

Górecki, T. and L. Smaga (2015). A comparison of tests for the one-way anova problem for functional data. *Computational Statistics* 30, 987 – 1010.

Hahn, U. (2012). A studentized permutation test for the comparison of spatial point patterns. *Journal of the American Statistical Association* 107(498), 754–764.

Illian, J., A. Penttinen, H. Stoyan, and D. Stoyan (2008). *Statistical Analysis and Modelling of Spatial Point Patterns* (1 ed.). Statistics in Practice. John Wiley & Sons, Ltd.

Mrkvička, T., M. Myllymäki, and U. Hahn (2017). Multiple monte carlo testing, with applications in spatial point processes. *Statistics and Computing* 27(5), 1239 – 1255.

Myllymäki, M., T. Mrkvička, P. Grabarnik, H. Seijo, and U. Hahn (2017). Global envelope tests for spatial processes. *J. R. Statist. Soc. B* 79, 381–404.

Narisetty, N. N. and V. J. Nair (2016). Extremal depth for functional data and applications. *Journal of the American Statistical Association* 111(516), 1705–1714.

Nichols, T. E. and E. Holmes (2001). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain M* 15, 1–25.
Pantazis, D., T. E. Nichols, S. Baillet, and R. M. Leahy (2005). A comparison of random field theory and permutation methods for the statistical analysis of meg data. *Neuroimaging* 25, 383–394.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ramsay, J. and B. Silverman (2006). *Functional Data Analysis* (2 ed.). Springer Series in Statistics. Springer.

Schladitz, K., A. Särkkä, I. Pavenstädt, O. Haferkamp, and T. Mattfeldt (2003). Statistical analysis of intramembranous particles using freeze fracture specimens. *Journal of Microscopy* 211, 137–153.

Zhang, J.-T. (2014). *Analysis of Variance for the functional data*. Chapman & Hall.
Table 1
The proportions of rejections (at level 0.05) over 1000 runs in the case of independent errors for models M1, M2, M3, M4. See text for the model specifications and descriptions of different test abbreviations.

| Model | σ1 = 0.05 | σ2 = 0.1 | σ3 = 0.15 | σ4 = 0.2 | σ5 = 0.4 | σ6 = 0.8 |
|-------|-----------|-----------|-----------|-----------|-----------|-----------|
| M1    | AsF 0.000 | 0.000     | 0.000     | 0.000     | 0.000     | 0.000     |
|       | RPM 0.060 | 0.071     | 0.071     | 0.052     | 0.044     | 0.054     |
|       | Fb 0.000  | 0.000     | 0.000     | 0.000     | 0.000     | 0.000     |
|       | GPF 0.001 | 0.001     | 0.002     | 0.001     | 0.001     | 0.002     |
|       | FP 0.055  | 0.059     | 0.038     | 0.041     | 0.055     | 0.045     |
|       | IPT 0.044 | 0.052     | 0.044     | 0.041     | 0.048     | 0.056     |
|       | F-max 0.060| 0.051     | 0.048     | 0.041     | 0.061     | 0.043     |
|       | p-min 0.052 | 0.039   | 0.049     | 0.046     | 0.051     | 0.051     |
|       | GFAM 0.049 | 0.048    | 0.054     | 0.048     | 0.048     | 0.049     |
|       | GFAC 0.046 | 0.051    | 0.050     | 0.050     | 0.045     | 0.054     |
|       | REF 0.043 | 0.058     | 0.048     | 0.039     | 0.055     | 0.048     |
| M2    | AsF 0.998 | 0.041     | 0.004     | 0.000     | 0.000     | 0.000     |
|       | RPM 0.758 | 0.178     | 0.104     | 0.076     | 0.060     | 0.060     |
|       | Fb 0.954 | 0.000     | 0.000     | 0.000     | 0.000     | 0.000     |
|       | GPF 1.000 | 0.254     | 0.047     | 0.010     | 0.007     | 0.001     |
|       | FP 1.000 | 0.929     | 0.767     | 0.562     | 0.161     | 0.076     |
|       | IPT 1.000 | 0.667     | 0.282     | 0.138     | 0.063     | 0.051     |
|       | F-max 0.932 | 0.197   | 0.095     | 0.077     | 0.059     | 0.048     |
|       | p-min 0.869 | 0.189   | 0.086     | 0.079     | 0.063     | 0.048     |
|       | GFAM 0.993 | 0.298    | 0.131     | 0.088     | 0.062     | 0.052     |
|       | GFAC 0.995 | 0.291    | 0.123     | 0.082     | 0.058     | 0.043     |
|       | REF 0.936 | 0.190     | 0.101     | 0.077     | 0.069     | 0.052     |
| M3    | AsF 1.000 | 0.931     | 0.102     | 0.008     | 0.000     | 0.000     |
|       | RPM 0.998 | 0.618     | 0.249     | 0.160     | 0.098     | 0.072     |
|       | Fb 1.000 | 0.615     | 0.003     | 0.000     | 0.000     | 0.000     |
|       | GPF 1.000 | 0.997     | 0.403     | 0.137     | 0.007     | 0.000     |
|       | FP 1.000 | 1.000     | 0.960     | 0.845     | 0.351     | 0.094     |
|       | IPT 1.000 | 1.000     | 0.895     | 0.509     | 0.122     | 0.052     |
|       | F-max 1.000 | 0.920    | 0.388     | 0.193     | 0.071     | 0.050     |
|       | p-min 1.000 | 0.877   | 0.346     | 0.164     | 0.067     | 0.050     |
|       | GFAM 1.000 | 0.979    | 0.536     | 0.253     | 0.071     | 0.050     |
|       | GFAC 1.000 | 0.993    | 0.554     | 0.248     | 0.083     | 0.038     |
|       | REF 1.000 | 0.932     | 0.384     | 0.192     | 0.077     | 0.054     |
| M4    | AsF 1.000 | 0.181     | 0.005     | 0.000     | 0.000     | 0.000     |
|       | RPM 0.923 | 0.273     | 0.141     | 0.101     | 0.086     | 0.077     |
|       | Fb 1.000 | 0.010     | 0.000     | 0.000     | 0.000     | 0.000     |
|       | GPF 1.000 | 0.598     | 0.079     | 0.026     | 0.005     | 0.004     |
|       | FP 1.000 | 0.983     | 0.859     | 0.737     | 0.277     | 0.087     |
|       | IPT 1.000 | 0.933     | 0.451     | 0.218     | 0.097     | 0.055     |
|       | F-max 0.809 | 0.196    | 0.123     | 0.077     | 0.050     | 0.053     |
|       | p-min 0.750 | 0.183    | 0.104     | 0.066     | 0.057     | 0.054     |
|       | GFAM 0.978 | 0.338    | 0.161     | 0.096     | 0.068     | 0.057     |
|       | GFAC 0.988 | 0.342    | 0.153     | 0.107     | 0.061     | 0.055     |
|       | REF 0.863 | 0.193     | 0.103     | 0.072     | 0.056     | 0.059     |
Table 2
The proportions of rejections (at level 0.05) over 1000 runs in the case of brownian errors for models M1, M2, M3, M4. See text for the model specifications and descriptions of different test abbreviations.

|   | σ₁ = 0.05 | σ₂ = 0.1 | σ₃ = 0.15 | σ₄ = 0.2 | σ₅ = 0.4 | σ₆ = 0.8 |
|---|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| **M1** | AsF | RPM | Fb | GPF | FP | IPT | F-max | p-min | GFAM | GFAC | REF |
| AsF | 0.066 | 0.063 | 0.060 | 0.060 | 0.060 | 0.060 | 0.060 | 0.060 | 0.060 | 0.060 | 0.060 |
| RPM | 0.055 | 0.039 | 0.042 | 0.042 | 0.038 | 0.045 | 0.038 | 0.045 | 0.038 | 0.045 | 0.038 |
| Fb | 0.033 | 0.032 | 0.038 | 0.032 | 0.030 | 0.039 | 0.030 | 0.039 | 0.030 | 0.039 | 0.030 |
| GPF | 0.068 | 0.071 | 0.068 | 0.071 | 0.071 | 0.077 | 0.071 | 0.077 | 0.071 | 0.077 | 0.077 |
| FP | 0.047 | 0.056 | 0.050 | 0.048 | 0.048 | 0.054 | 0.048 | 0.054 | 0.048 | 0.054 | 0.048 |
| IPT | 0.050 | 0.044 | 0.053 | 0.056 | 0.049 | 0.062 | 0.056 | 0.049 | 0.062 | 0.056 | 0.062 |
| F-max | 0.058 | 0.041 | 0.059 | 0.050 | 0.047 | 0.058 | 0.050 | 0.047 | 0.058 | 0.050 | 0.058 |
| p-min | 0.046 | 0.042 | 0.050 | 0.047 | 0.045 | 0.057 | 0.047 | 0.045 | 0.057 | 0.047 | 0.057 |
| GFAM | 0.047 | 0.047 | 0.056 | 0.050 | 0.052 | 0.054 | 0.050 | 0.052 | 0.054 | 0.050 | 0.054 |
| GFAC | 0.053 | 0.050 | 0.066 | 0.044 | 0.049 | 0.053 | 0.044 | 0.049 | 0.053 | 0.044 | 0.053 |
| REF | 0.053 | 0.037 | 0.054 | 0.048 | 0.048 | 0.057 | 0.048 | 0.048 | 0.057 | 0.048 | 0.057 |
| **M2** | σ₁ = 0.05 | σ₂ = 0.1 | σ₃ = 0.15 | σ₄ = 0.2 | σ₅ = 0.4 | σ₆ = 0.8 |
| AsF | 0.660 | 0.134 | 0.091 | 0.080 | 0.077 | 0.062 |
| RPM | 0.993 | 0.623 | 0.255 | 0.134 | 0.065 | 0.049 |
| Fb | 0.393 | 0.065 | 0.048 | 0.037 | 0.042 | 0.026 |
| GPF | 1.000 | 0.663 | 0.279 | 0.171 | 0.102 | 0.071 |
| FP | 0.645 | 0.106 | 0.066 | 0.064 | 0.065 | 0.049 |
| IPT | 1.000 | 0.600 | 0.227 | 0.129 | 0.074 | 0.049 |
| F-max | 1.000 | 0.958 | 0.599 | 0.337 | 0.110 | 0.059 |
| p-min | 1.000 | 0.954 | 0.584 | 0.330 | 0.105 | 0.050 |
| GFAM | 1.000 | 0.949 | 0.548 | 0.321 | 0.112 | 0.053 |
| GFAC | 1.000 | 0.930 | 0.540 | 0.308 | 0.116 | 0.048 |
| REF | 1.000 | 0.955 | 0.598 | 0.338 | 0.108 | 0.053 |
| **M3** | σ₁ = 0.05 | σ₂ = 0.1 | σ₃ = 0.15 | σ₄ = 0.2 | σ₅ = 0.4 | σ₆ = 0.8 |
| AsF | 1.000 | 0.506 | 0.194 | 0.122 | 0.073 | 0.055 |
| RPM | 1.000 | 0.997 | 0.894 | 0.652 | 0.159 | 0.064 |
| Fb | 0.997 | 0.230 | 0.098 | 0.047 | 0.040 | 0.031 |
| GPF | 1.000 | 1.000 | 1.000 | 0.996 | 0.329 | 0.099 |
| FP | 1.000 | 0.455 | 0.158 | 0.092 | 0.057 | 0.042 |
| IPT | 1.000 | 1.000 | 1.000 | 0.994 | 0.259 | 0.073 |
| F-max | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.608 |
| p-min | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.592 |
| GFAM | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.436 |
| GFAC | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.476 |
| REF | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.598 |
| **M4** | σ₁ = 0.05 | σ₂ = 0.1 | σ₃ = 0.15 | σ₄ = 0.2 | σ₅ = 0.4 | σ₆ = 0.8 |
| AsF | 0.746 | 0.226 | 0.132 | 0.096 | 0.105 | 0.063 |
| RPM | 0.920 | 0.288 | 0.110 | 0.072 | 0.060 | 0.040 |
| Fb | 0.550 | 0.144 | 0.077 | 0.060 | 0.053 | 0.038 |
| GPF | 1.000 | 0.677 | 0.295 | 0.190 | 0.115 | 0.074 |
| FP | 0.691 | 0.193 | 0.116 | 0.077 | 0.079 | 0.049 |
| IPT | 1.000 | 0.596 | 0.240 | 0.143 | 0.091 | 0.052 |
| F-max | 1.000 | 1.000 | 0.981 | 0.808 | 0.188 | 0.075 |
| p-min | 1.000 | 1.000 | 0.978 | 0.791 | 0.180 | 0.068 |
| GFAM | 1.000 | 0.999 | 0.903 | 0.628 | 0.155 | 0.070 |
| GFAC | 1.000 | 0.997 | 0.893 | 0.648 | 0.146 | 0.069 |
| REF | 1.000 | 1.000 | 0.981 | 0.798 | 0.187 | 0.074 |
Table 3

The proportions of rejections (at level 0.05) over 1000 runs for model M. The white noise cases are shown in the upper part and the Brownian error cases are shown in the lower part of the table. See text for the model specification and descriptions of different test abbreviations.

| iid   | $\sigma_1 = 0.05$ | $\sigma_2 = 0.1$ | $\sigma_3 = 0.15$ | $\sigma_4 = 0.2$ | $\sigma_5 = 0.4$ | $\sigma_6 = 0.8$ |
|-------|------------------|------------------|------------------|------------------|------------------|------------------|
| AsF   | 1.000            | 1.000            | 0.988            | 0.360            | 0.004            | 0.000            |
| RPM   | 1.000            | 0.992            | 0.812            | 0.440            | 0.116            | 0.056            |
| Fb    | 1.000            | 1.000            | 0.412            | 0.000            | 0.000            | 0.000            |
| GPF   | 1.000            | 1.000            | 1.000            | 0.988            | 0.156            | 0.016            |
| FP    | 1.000            | 1.000            | 1.000            | 1.000            | 0.872            | 0.244            |
| IPT   | 1.000            | 1.000            | 1.000            | 0.996            | 0.328            | 0.072            |
| F-max | 1.000            | 1.000            | 0.996            | 0.740            | 0.120            | 0.076            |
| p-min | 1.000            | 1.000            | 0.988            | 0.696            | 0.136            | 0.068            |
| GFAM  | 1.000            | 1.000            | 0.996            | 0.828            | 0.132            | 0.060            |
| GFAC  | 1.000            | 1.000            | 1.000            | 0.972            | 0.196            | 0.040            |
| REF   | 1.000            | 1.000            | 0.996            | 0.764            | 0.112            | 0.076            |

| Brown | $\sigma_1 = 0.05$ | $\sigma_2 = 0.1$ | $\sigma_3 = 0.15$ | $\sigma_4 = 0.2$ | $\sigma_5 = 0.4$ | $\sigma_6 = 0.8$ |
|-------|------------------|------------------|------------------|------------------|------------------|------------------|
| AsF   | 1.000            | 0.996            | 0.656            | 0.364            | 0.092            | 0.064            |
| RPM   | 1.000            | 1.000            | 1.000            | 0.996            | 0.544            | 0.104            |
| Fb    | 1.000            | 0.944            | 0.208            | 0.148            | 0.036            | 0.012            |
| GPF   | 1.000            | 1.000            | 1.000            | 1.000            | 0.724            | 0.148            |
| FP    | 1.000            | 1.000            | 0.564            | 0.268            | 0.048            | 0.052            |
| IPT   | 1.000            | 1.000            | 1.000            | 1.000            | 0.656            | 0.124            |
| F-max | 1.000            | 1.000            | 1.000            | 1.000            | 1.000            | 0.912            |
| p-min | 1.000            | 1.000            | 1.000            | 1.000            | 1.000            | 0.892            |
| GFAM  | 1.000            | 1.000            | 1.000            | 1.000            | 0.976            | 0.360            |
| GFAC  | 1.000            | 1.000            | 1.000            | 1.000            | 1.000            | 0.360            |
| REF   | 1.000            | 1.000            | 1.000            | 1.000            | 1.000            | 0.904            |