MOX-Report No. 10/2019

Domain selection and family-wise error rate for functional data: a unified framework

Abramowicz, K.; Pini, A.; Schelin, L.; Sjostedt de Luna, S.;

Stamm, A.; Vantini, S.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it http://mox.polimi.it

Domain selection and family-wise error rate for functional data: a unified framework

Konrad Abramowicz[*], Alessia Pini[†], Lina Schelin[‡], Sara Sjöstedt de Luna[*], Aymeric Stamm[§], Simone Vantini[¶]

Abstract

Functional data are smooth, often continuous, random curves, which can be seen as an extreme case of multivariate data with infinite dimensionality. Just as component-wise inference for multivariate data naturally performs feature selection, subset-wise inference for functional data performs domain selection. In this paper, we present a unified null-hypothesis testing framework for domain selection on populations of functional data. In detail, p-values of hypothesis tests performed on point-wise evaluations of functional data are suitably adjusted for providing a control of the family-wise error rate (FWER) over a family of subsets of the domain. We show that several state-of-the-art domain selection methods fit within this framework and differ from each other by the choice of the family over which the control of the FWER is provided. In the existing literature, these families are always defined a priori. In this work, we also propose a novel approach, coined threshold-wise testing, in which the family of subsets is instead built in a data-driven fashion. The method seamlessly generalizes to multidimensional domains in contrast to methods based on a-priori defined families. We provide theoretical results with respect to exactness, consistency, and strong and weak control of FWER for the methods within the unified framework.

Keywords: Permutation test, adjusted p-value function, multidimensional domain

1 Introduction

Functional data analysis (FDA) is a field of statistics that pertains to the study of data sets in which the sample unit is a smooth curve. Besides estimation, clustering or prediction, it is also of critical importance to design appropriate statistical methodologies for making inference on populations of functional data, which is the objective of the present work.

For example, suppose that samples of continuous random functions are observed from two populations over some domain, and that we want to test if the mean functions,

[*]Department of Mathematics and Mathematical Statistics, Umeå University, Umeå, Sweden
[†]Department of Statistical Sciences, Universit Cattolica del Sacro Cuore, Milan, Italy
[‡]Unit of Statistics, Umeå School of Business, Economics and Statistics, Umeå University, Umeå, Sweden
[§]UMR CNRS 6629, Laboratoire de Mathmatiques Jean Leray, Nantes, France
[¶]MOX - Department of Mathematics, Politecnico di Milano, Milan, Italy

μ_1 and μ_2, are the same in both populations, testing $H_0 : \mu_1(\cdot) = \mu_2(\cdot)$ over the whole domain versus $H_1 : \mu_1(\cdot) \neq \mu_2(\cdot)$ over at least some part of the domain. Many methods have been devised in the literature to form so called *global* (overall) tests for this situation as well as for more general scenarios, both parametrically (e.g. Spitzner et al. 2003; Cuevas et al. 2004; Abramovich and Angelini 2006; Horváth and Kokoszka 2012; Staicu et al. 2014) and non-parametrically (e.g. Hall and Tajvidi 2002; Cardot et al. 2004; Corain et al. 2014). If H_0 is rejected, it is typically also of interest to identify the parts of the domain where significant differences occur, performing so called *local* inference, with control of the type I error.

In this paper, we focus on local inference for functional data, which we refer to as *domain selection*. Few attempts have been made in this direction in the literature. A natural approach pertains to discretizing the functional domain, performing pointwise inference and using the resulting pointwise p-values to select regions responsible for rejecting a null hypothesis. For instance, Fan and Zhang (2000); Reiss et al. (2010); Ramsay and Silverman (2005) derive pointwise confidence bands for functional data. This however only provides a pointwise control of the probability of type I errors. If such a control is desirable on the entire domain, it is necessary to adjust the pointwise p-values in order to account for the multiplicity of hypothesis tests that are jointly performed when analyzing the whole domain. This issue already arises in multivariate analysis and has given birth to many adjustment procedures (see, e.g., Marcus et al. 1976; Holm 1979; Holmes et al. 1996; Winkler et al. 2014). It naturally translates to functional data. However, functional data differ from multivariate data in that functional data feature unique properties such as natural ordering and continuity, which can be used to improve upon classic methods for performing domain selection.

Vsevolozhskaya et al. (2013, 2014) propose a method for local inference of functional data that relies on the availability of a partition of the domain that allows them to perform dimensionality reduction. Specifically, they perform functional hypothesis tests on the elements of the partition and resort to a closed testing procedure (Marcus et al. 1976) to adjust the resulting p-values and achieve strong control of the family-wise error rate (FWER) between the elements of the partition. Hence, the resulting inference might heavily depend on the partition itself. In addition, the coarseness of the partition defines the depth into which local inference is performed, and the approach is actually of practical relevance only for relatively small partitions. In the remainder of the paper, we refer to this method as *partition closed testing* (PCT). Another approach – introduced for functional t-tests in Pini and Vantini (2017) and extended to functional-on-scalar linear models in Abramowicz et al. (2018) – is the *interval-wise testing* (IWT). The procedure simultaneously tests a family of hypotheses generated by restricting the functional data to any interval of the domain. The resulting unadjusted p-values are turned into a continuous function of adjusted p-values by borrowing ideas from closed testing procedures in order to control the FWER on every interval of the domain. The adjusted p-value function is then used for domain selection. This is however of practical use only for functional data defined on one-dimensional domains as it is unclear how to define "multidimensional intervals" and, in any event, that would be computationally over-demanding. Furthermore, IWT only provides strong control of the FWER on intervals: if the subset of the domain on which the null hypothesis is true is more complex (e.g. a union of intervals), IWT fails at controlling the FWER.

In this paper, following the recent literature on domain selection, we focus on local inferential techniques that aim at providing control of the FWER. We start by formalizing the concepts of weak and strong control of the FWER in the context of functional data. Then, we introduce a general framework for performing local inference that leads to domain selection. The ground principles of the methodology are based on standard pointwise inferential procedures and their set-wise counterparts for a chosen family of subsets. The framework is related to a wide class of inferential problems (e.g., properties and comparisons of populations, significance tests for coefficients in models), as we utilize general concepts of null and alternative hypotheses. Furthermore, it can be applied either to a parametric or a nonparametric analysis. Using the properties of the underlying pointwise tests, we formulate and prove finite sample and asymptotic properties for methods within this framework. We show how well-established methods from the literature of inference for functional data can be described in the light of our proposed unifying framework. In addition, we present a novel method with desirable theoretical properties, where the computational complexity is independent of the dimension of the domain of the functional data.

The paper is outlined as follows. In Section 2, we formulate the inferential problem and define different types of FWER control. In Section 3, we present the unified framework for local inference and the theoretical results describing the properties of the resulting methods. In Section 4, we present how some of the existing methods are special cases of this framework and introduce some other methods that feature interesting theoretical properties. We also present an illustrative example to elucidate the construction of families and the domain selection mechanisms that can be obtained. A simulation study designed to exemplify the theoretical properties of the introduced method is presented in Section 5. In Section 6, we discuss a generalization to multidimensional domains and show the full potential of this approach in a neuroimaging application. Finally, Section 7 contains conclusions.

2 Definitions and the inferential problem

Consider a space of continuous random functions defined on domain D, where D is a compact subset of $\mathbb{R}^d$, $d \geq 1$. Let us consider a general inferential problem based on a sample of n independent functional observations. Without loss of generality, assume that we aim at testing a functional null hypothesis H_0 against an alternative hypothesis H_1 not only globally but also from a local perspective. For instance, it could be the functional two-sample t-test where $H_0 : \mu_1(\cdot) = \mu_2(\cdot)$ is tested against $H_1 : \mu_1(\cdot) \neq \mu_2(\cdot)$, cf. Section 4.3. Let $\mathcal{D}_0$ and $\mathcal{D}_1$ denote the regions of D where the null hypothesis is true and false, respectively. Our goal is to construct an inferential method that correctly identifies regions $\mathcal{D}_0$ and $\mathcal{D}_1$ and controls the type I error along the domain.

Formally, assume that we observe a random sample of n random continuous functions $y_1(t), y_2(t), \ldots, y_n(t), t \in D$, possibly with attached functional or scalar covariates. For all $t \in D$, we denote by H_0^t and H_1^t the restrictions of H_0 and H_1 to the point t, respectively. Assume that, based on the data, we can obtain a test statistic $T_n(t)$ for testing H_0^t against H_1^t at point t, where H_0^t is rejected for large values of $T_n(t)$. We denote by $p_n(t)$ the p-value of the test at point t based on the statistic $T_n(t)$. In the

notation, we stress the dependence of the test statistic and the p-value functions on the sample size n, as some properties investigated in our work are asymptotic. Depending on the assumptions on the generative process of the functional data, on the sample size, and on the test, $p_n(t)$ can be computed with parametric, asymptotic, or nonparametric tests. Below, we define some of the properties that are typically required for pointwise tests.

Definition 2.1. *We say that the pointwise test of H_0^t against H_1^t based on the statistic $T_n(t)$ with p-value $p_n(t)$ is*

- ***exact*** *if for any $\alpha \in (0,1)$ and any $n \in \mathbb{N}_+$ the probability of rejecting H_0^t at level α is equal to α when H_0^t is true:*

$$t \in \mathcal{D}_0 \quad \Rightarrow \quad \mathbb{P}[p_n(t) \leq \alpha] = \alpha,$$

- ***asymptotically exact*** *if for any $\alpha \in (0,1)$ the probability of rejecting H_0^t at level α is asymptotically equal to α when H_0^t is true:*

$$t \in \mathcal{D}_0 \quad \Rightarrow \quad \lim_{n\to\infty} \mathbb{P}[p_n(t) \leq \alpha] = \alpha,$$

- ***consistent*** *if for any $\alpha \in (0,1)$ the probability of rejecting H_0^t when H_0^t is false is asymptotically one:*

$$t \in \mathcal{D}_1 \quad \Rightarrow \quad \lim_{n\to\infty} \mathbb{P}[p_n(t) \leq \alpha] = 1.$$

Remark 2.1. *In Definition 2.1, we specify in general terms $n \to \infty$. However, depending on the test that is performed, some more specific assumptions about the sample size may be required. For example, when performing a test comparing two populations, both sample sizes are required to go to infinity and not only the total sample size n.*

Since our goal is to provide testing methods resulting in domain selection we need to introduce concepts of tests on sets of the domain. Let us introduce the following hypotheses defined on any set $A \subset D$,

$$H_0^A = \{H_0^t \text{ is true } \forall t \in A\} \text{ and } H_1^A = \{\exists t \in A : H_1^t \text{ is true}\}.$$

We assume that tests of H_0^A against H_1^A are performed using the following statistic

$$T_n^A = \int_A T_n(t)\mathrm{d}t, \tag{1}$$

where the integral is defined in a Lebesgue sense. Let p_n^A be the corresponding p-value. Similarly to pointwise inference, we provide the definitions of exactness and consistency for the test on set A. Let $|A|$ denote the Lebesgue measure of a set A.

Definition 2.2. *For any $A \subseteq D$ s.t. $|A| > 0$, we say that the test of H_0^A against H_1^A based on the statistic $T_n(t)$ with p-value p_n^A is*

- ***exact*** *if for any $\alpha \in (0,1)$ and for any $n \in \mathbb{N}_+$*

$$|A \cap \mathcal{D}_1| = 0 \quad \Rightarrow \quad \mathbb{P}[p_n^A \leq \alpha] = \alpha;$$

- ***asymptotically exact*** *if for any* $\alpha \in (0,1)$

$$|A \cap \mathcal{D}_1| = 0 \quad \Rightarrow \quad \lim_{n\to\infty} \mathbb{P}[p_n^A \leq \alpha] = \alpha;$$

- ***consistent*** *if for any* $\alpha \in (0,1)$

$$|A \cap \mathcal{D}_1| > 0 \quad \Rightarrow \quad \lim_{n\to\infty} \mathbb{P}[p_n^A \leq \alpha] = 1.$$

In the nonparametric permutation test framework, it is straightforward to directly build exact and consistent tests on sets from the corresponding pointwise tests under some mild assumptions. Specifically, following Pesarin and Salmaso (2010), pp. 122–124, we know that if permutation tests are used and we use the same permutations for all points of the set, the (asymptotic) exactness of the pointwise tests implies (asymptotic) exactness of the tests on sets. Further, if $\forall t \in D$, $T_n(t)$ is non-negative and stochastically greater under H_1^t than under H_0^t, we have that consistency of the pointwise tests implies consistency of the tests on sets.

Let us now turn our attention to domain selection. Say that we use the pointwise p-values for selecting the parts of the domain imputable for the rejection of H_0 by performing a thresholding at level $\alpha \in (0,1)$. Even though in this way we can select a region, the probability that this region - or part of it - has been wrongly selected is not controlled. In detail, since the p-values, $p_n(t)$, are only computed pointwise, we can not guarantee any control of the probability of committing at least one type I error over the domain.

In multivariate statistical analysis - when several tests are performed - the single p-values are adjusted to provide a global control of the type I error rate. The selection of the variables for the rejection of the null hypothesis is performed by thresholding of properly adjusted p-values instead of the original unadjusted ones. One of the families of adjustement strategies are those controlling the family wise error rate (FWER, i.e., the probability of rejecting at least one true null hypothesis). There are two classical types of control of the FWER that have been defined in the multivariate literature; weak control of the FWER holds if the FWER is controlled when $\mathcal{D}_0 = D$, while strong control of the FWER holds if the FWER is controlled when $\mathcal{D}_0$ is any subset of D.

Our aim is to define an adjusted p-value function, $\tilde{p}_n(t)$, $t \in D$, that can be thresholded to select the regions of D imputable for the rejection of H_0 providing control of FWER over the domain. The different FWER control types of a procedure for local testing based on an adjusted p-value function $\tilde{p}_n(t)$ can be formally defined as follows.

Definition 2.3. *We say that a procedure is provided with a* ***weak control of the FWER*** *if for any* $n \in \mathbb{N}_+$ *its adjusted p-value function* $\tilde{p}_n(t)$, $t \in D$ *is such that,* $\forall \alpha \in (0,1)$*:*

$$\mathcal{D}_0 = D \quad \Rightarrow \quad \mathbb{P}(\exists t \in \mathcal{D}_0 : \tilde{p}_n(t) \leq \alpha) \leq \alpha.$$

Definition 2.4. *We say that a procedure is provided with a* ***strong control of the FWER*** *if for any* $n \in \mathbb{N}_+$ *its adjusted p-value function* $\tilde{p}_n(t)$, $t \in D$ *is such that,* $\forall \alpha \in (0,1)$*:*

$$\mathbb{P}(\exists t \in \mathrm{cl}(\mathcal{D}_0) : \tilde{p}_n(t) \leq \alpha) \leq \alpha,$$

where $\mathrm{cl}(\mathcal{D}_0)$ *denotes the closure of the set* $\mathcal{D}_0$*.*

The reason why we introduce the strong control on the closure of $\mathcal{D}_0$ is detailed in Remark 2.2. Some situations require an asymptotic strong control of the FWER, i.e., the control of the FWER when the sample size n goes to infinity.

Definition 2.5. *We say that the procedure is provided with an* ***asymptotic strong control of the FWER*** *if its adjusted p-value function $\tilde{p}_n(t)$, $t \in D$ is such that, $\forall \alpha \in (0,1)$:*

$$\limsup_{n\to\infty} \mathbb{P}(\exists t \in \mathrm{cl}(\mathcal{D}_0) : \tilde{p}_n(t) \leq \alpha) \leq \alpha.$$

Asymptotic weak control of the FWER can be defined similarly. Finally, similarly to pointwise inference we define consistency of an inferential procedure, assuring that it asymptotically detects the parts of the domain where H_1 holds, i.e., $\mathcal{D}_1$.

Definition 2.6. *We say that the procedure is* ***consistent*** *if its adjusted p-value function $\tilde{p}_n(t)$, $t \in D$ is such that, $\forall \alpha \in (0,1)$:*

$$\lim_{n\to\infty} \mathbb{P}(\forall t \in \mathrm{Int}(\mathcal{D}_1) : \tilde{p}_n(t) \leq \alpha) = 1,$$

where $\mathrm{Int}(\mathcal{D}_1)$ denotes the interior of set $\mathcal{D}_1$.

Remark 2.2. *Since tests on subsets are performed using an integrated pointwise test statistic, deviations from the null hypothesis at only one point or at a set of null Lebesgue measure can not be detected. In particular, the boundary of the set $\mathcal{D}_1$ cannot be detected, since it has null measure. Hence, strong control of the FWER is extended to the closure of the set $\mathcal{D}_0$, while consistency can be reached only for the interior of $\mathcal{D}_1$.*

3 A unified framework

In this section we describe a unified framework for testing local functional hypotheses over the domain D, given a set of n random functions. We present a class of methods that can be used to adjust the pointwise p-values $p_n(t)$ in order to provide a control of the FWER over the domain D. Consider a non-empty (possibly infinite) family $\mathcal{F}$ of Lebesgue-measurable subsets of the domain of non-null measure, such that: $\cup_{S\in\mathcal{F}} S = D$. The testing procedure that we propose is based on performing tests on the restrictions of H_0 and H_1 to all subsets of the family, and then adjusting the p-value according to the results of such tests. First, we formally describe the testing procedure for a general $\mathcal{F}$, and provide a characterization of the inferential properties of the methods depending on the choice of $\mathcal{F}$. Further, we describe in detail several methods that can be obtained for some particular choices of $\mathcal{F}$. The unified framework consists in the following steps (presented graphically in Section B, supplementary file).

1. **Computation of p-values for all subsets.** For all $S \in \mathcal{F}$, compute the p-value p_n^S of the test of H_0^S against H_1^S, based on the test statistic T_n^S in (1).

2. **Computation of the adjusted p-value function.** For all $t \in D$, compute the adjusted p-value function defined as:

$$\tilde{p}_n(t;\mathcal{F}) = \sup_{S\in\mathcal{F}:t\in S} p_n^S. \tag{2}$$

3. Domain selection. Select the subsets of the domain D where H_0 is rejected at level $\alpha \in (0,1)$ as

$$\{t \in D: \quad \tilde{p}_n(t;\mathcal{F}) \leq \alpha\}.$$

We consider two types of families $\mathcal{F}$: a predefined type, where all subsets belonging to $\mathcal{F}$ are defined a priori, and a data-driven type, where the subsets belonging to the family depend on the data at hand. For sake of clarity, we denote the predefined families by $\mathcal{F}_-$ and the data driven ones by $\mathcal{F}_n$. The inferential properties of the described procedure are characterized by the following results.

Theorem 3.1. *Let $\mathcal{F}_-$ be a predefined non-empty family of Lebesgue-measurable subsets of the domain D. Let $\tilde{p}_n(t;\mathcal{F}_-)$, $t \in D$ be the adjusted p-value function defined in (2). If the tests of H_0^S against H_1^S are exact $\forall S \in \mathcal{F}_-$, the procedure based on the adjusted p-value function $\tilde{p}_n(t;\mathcal{F}_-)$, $t \in D$ is provided with the following control for all $n \in \mathbb{N}_+$:*

$$\forall S \in \mathcal{F}_- : H_0^S \text{ is true} \quad \Rightarrow \quad \mathbb{P}[\exists t \in S : \tilde{p}_n(t;\mathcal{F}_-) \leq \alpha] \leq \alpha.$$

The above theorem states that if the family $\mathcal{F}$ is fixed, the probability of wrongly detecting as significant a set (or part of it) where the null hypothesis is actually true is upperly bounded to α for every set included in the family $\mathcal{F}$. For instance, if the family $\mathcal{F}$ includes every subset of the domain D, we clearly have strong control of the FWER, cf. Definition 2.4. It is not straightforward to extend Theorem 3.1 to data driven families, since the sets belonging to the family are random. Still, under some assumptions on the structure of $\mathcal{F}$, an asymptotic control of the FWER can be provided.

Theorem 3.2. *Let $\mathcal{F}_n$ be a data driven non-empty family of Lebesgue-measurable subsets of the domain D. Let $\tilde{p}_n(t;\mathcal{F}_n)$, $t \in D$ be the adjusted p-value function defined in (2). Assume that all tests of H_0^S against H_1^S are asymptotically exact and that all tests of H_0^t against H_1^t are asymptotically exact and consistent. In addition, assume that there exists a sequence $\{\varepsilon_n\}_{n\geq 1}$ such that $\forall n \in \mathbb{N}_+$ $\varepsilon_n > 0$, $\lim_{n\to\infty} \varepsilon_n = 0$, and $\forall n$, $\mathcal{F}_n$ almost surely contains the set*

$$\{t \in D \text{ s.t. } p_n(t) \geq \varepsilon_n\}.$$

Then, the procedure based on the adjusted p-value $\tilde{p}_n(t;\mathcal{F}_n)$ is provided with an asymptotic strong control of the FWER.

Theorem 3.2 states the conditions under which a procedure based on a data-driven family $\mathcal{F}_n$ is provided with an asymptotic strong control of the FWER. In addition to assuming that tests on subsets are asymptotically exact, also consistency of the point-wise tests based on the unadjusted p-value function is needed. Asymptotically, as $n \to \infty$, the family $\mathcal{F}_n$ must contain the set $P_n = \{t \in D \text{ s.t. } p_n(t) > 0\}$ (that separates the sets on D on which the point-wise p-value function is zero). Indeed, if point-wise tests based on the unadjusted p-value function are consistent, as $n \to \infty$, $\forall t \in \mathcal{D}_1$, $p_n(t)$ converges to zero almost surely, while $\forall t \in \mathcal{D}_0$, $p_n(t)$ is almost surely positive. Hence, asymptotically, the set $\{t \in D \text{ s.t. } p_n(t) > 0\}$ a.s. corresponds to $\mathcal{D}_0$, that is asymptotically included in $\mathcal{F}_n$. Such inclusion, together with the fact that the test of $H_0^{\mathcal{D}_0}$ against $H_1^{\mathcal{D}_0}$ is asymptotically exact, implies asymptotic strong control.

Theorem 3.2 makes a connection between the family $\mathcal{F}_n$ and the unadjusted p-value function. In principle, $\mathcal{F}_n$ can be obtained in many different ways. However, if its construction is related to $p_n(t)$, it is provided with sound theoretical properties.

A consequence of Theorem 3.2, is that if all tests of H_0^S against H_1^S are exact and if all tests of H_0^t against H_1^t are consistent, we still obtain asymptotic strong control of the FWER, given that the family $\mathcal{F}_n$ satisfies the necessary assumptions. Finally, the following theorems provide the conditions on the family $\mathcal{F}$ under which the testing procedure is consistent. We start by the pre-defined families.

Theorem 3.3. *Let $\mathcal{F}_-$ be a predefined non-empty family of Lebesgue-measurable subsets of the domain D. Let $\tilde{p}_n(t;\mathcal{F}_-)$, $t \in D$ be the adjusted p-value function defined in* (2)*. Assume that the Lebesgue measure of $\mathcal{D}_1$ is strictly positive and that tests of H_0^S against H_1^S are consistent $\forall S \in \mathcal{F}_-$. Further assume that $\mathcal{F}_-$ is such that:*

$$\forall S \in \mathcal{F}_- : S \cap \mathrm{Int}(\mathcal{D}_1) \neq \emptyset, \quad |S \cap \mathcal{D}_1| > 0. \tag{3}$$

Then the procedure based on the adjusted p-value function $\tilde{p}_n(t;\mathcal{F}_-)$ is consistent.

Consistency is assured for a predefined family if the Lebesgue measure of $\mathcal{D}_1$ is positive, and if all sets of the family intersecting $\mathcal{D}_1$ in one of its interior points also intersects with $\mathcal{D}_1$ on a set of positive measure. Such requirement is important due to the integral nature of the test statistics used on the elements of the family $\mathcal{F}_-$ and its invariance with respect to values on zero measure sets. For data driven families the content of $\mathcal{F}_n$ may be sample size dependent. Since consistency is a limiting property, without loss of generality, we weaken the formulation of the assumptions in Theorem 3.3.

Theorem 3.4. *Let $\mathcal{F}_n$ be a data driven non-empty family of Lebesgue-measurable subset of the domain D. Let $\tilde{p}_n(t;\mathcal{F}_n)$, $t \in D$ be the adjusted p-value defined in* (2)*. Assume that the Lebesgue measure of $\mathcal{D}_1$ is strictly positive and that all tests of H_0^S against H_1^S are consistent. Further assume that $\mathcal{F}_n$ is such that as $n \to \infty$:*

$$\forall S \in \mathcal{F}_n : S \cap \mathrm{Int}(\mathcal{D}_1) \neq \emptyset, \quad |S \cap \mathcal{D}_1| > 0 \ a.s.$$

Then the procedure based on the adjusted p-value function $\tilde{p}_n(t;\mathcal{F}_n)$ is consistent.

4 Examples of methods within the unified framework

This section discusses test procedures for particular choices of $\mathcal{F}$, and their corresponding theoretical properties. We focus on the case when D is one-dimensional, leaving the discussion about higher dimensions to Section 6. The computational aspects of each procedure are presented in Section C in the supplementary file. In what follows, when using the results from Section 3 all underlying assumptions are assumed valid.

4.1 Pre-defined families $\mathcal{F}_-$

We start by describing how some inferential methods already presented in the literature can be embedded within the unified framework with the pre-defined families $\mathcal{F}_-$.

Global Testing

Consider the method based on the family consisting of one single set, being the whole domain, i.e.,

$$\mathcal{F}_{Glob} := \{D\}.$$

The corresponding test procedure would then perform just one global test over the whole domain D and assign its p-value to all points of D. The adjusted p-value function is $\tilde{p}_n(t; \mathcal{F}_{Glob}) \equiv p_n^D, \forall t \in D$. From Theorem 3.1 it is straightforward to see that this method is provided with a weak control of the FWER for exact tests. In addition, consistency of the procedure follows directly by the consistency of the single test. However, a global test can not provide strong control of the FWER. In addition, since the adjusted p-value function is constant, it can not be used to select specific parts of the domain responsible for the rejection of the null hypothesis.

Borel-Wise Testing

The Borel-wise testing procedure (BWT), being on the other end of the spectrum compared to the Global test, is based on the choice

$$\mathcal{F}_{BWT} := \mathcal{B}(D), \tag{4}$$

where $\mathcal{B}(D)$ denotes all Borel sets of the domain D of non-null measure. We exclude the Borel subsets of zero measure from the procedure since the test statistic (1) is not definite on such sets. For simplicity, we still denote the procedure Borel-wise testing. The resulting procedure is the continuous extension of the closed testing procedure (see, e.g., Marcus et al. 1976), that has been proposed in multivariate analysis to adjust p-values.

If all tests are exact, Theorem 3.1 implies that BWT is provided with a strong control of the FWER. However, Theorem 3.3 does not apply to the family $\mathcal{F}_{BWT}$ since even if $|\mathcal{D}_1| > 0$, there exist Borel sets that intersect interior points of $\mathcal{D}_1$ in zero-measure sets. Analogously to the global test, the adjusted p-value function for this method is constant $\tilde{p}_n(t; \mathcal{F}_{BWT}) \geq \max_{t \in D} p_n(t)$ (see Proposition A.1 in Section A, supplementary file). It is therefore clear that BWT is not consistent, and can not be used for domain selection.

Partition Closed Testing

Assume that interest lies in performing tests on an a priori selected partition of the original domain. Let $\{S_j\}_{j=1}^J$ for some finite $J \in \mathbb{N}_+$ define the sets of the partition, satisfying $S_j \subseteq D$, $S_j \cap S_{j'} = \emptyset \ \forall j \neq j'$, and $\bigcup_{j=1}^J S_j = D$. Assume that S_j is Lebesgue-measurable for all j. Then, the Partition Closed Testing procedure (PCT, Vsevolozhskaya et al. 2013) is the inferential procedure based on a family containing all possible unions between sets S_j, i.e.,

$$\mathcal{F}_{PCT} = \{\cup_{j \in \mathbb{I}} S_j\}_{\mathbb{I} \subseteq \{1,\ldots,J\}}.$$

From Theorem 3.1 it follows that the PCT procedure is provided with a control that is in between the weak and the strong control of the FWER: strong control holds

between the S_j sets in the sense that the probability of selecting at least one set S_j, where the null hypothesis $H_0^{S_j}$ is true, is controlled. *Within* the sets S_j, however, the control is only weak, since if the adjusted p-value $\tilde{p}_n(t; \mathcal{F}_{PCT})$ for $t \in S_j$ is below α, we only know that S_j presents a statistically significant deviation from the null hypothesis in at least one of its points. With this method, it is not possible to decide which set of points that are responsible for the rejection of H_0. Furthermore, for every finite J, the PCT method is consistent as a consequence of Theorem 3.3. In the general case, since the method is based on performing tests on unions of sets S_j, the adjusted p-value $\tilde{p}_n(t; \mathcal{F}_{PCT})$ is a stepwise constant function assuming the same value for all points belonging to the same element of the partition.

It is straightforward to see that for $J = 1$ the PCT method coincides with the global testing. Further, consider two equisized partitions of the domain D, the first of size J_0, $J_0 \in \mathbb{N}_+$ and the second of size $J_1 = kJ_0$, for an arbitrary $k \in \mathbb{N}_+, k > 1$. By definition, the adjusted p-value function for the PCT method based on the partition of size J_1, cannot be smaller than the one corresponding to size J_0. Moreover, if at any $t_0 \in D$ the unadjusted p-value function is above the significance level, the corresponding adjusted p-value function increases with k, and at some point exceeds the significance level on the whole domain, resulting in no domain selection. Finally if the measure of all elements of the partition goes to zero, the PCT and BWT methods coincide.

Interval-Wise Testing

The Interval-Wise Testing (IWT, Pini and Vantini 2017) is based on performing a test on every interval of the domain. The method fits under the unified framework with the family

$$\mathcal{F}_{IWT} = \{[t_1, t_2] : t_2 > t_1\}_{t_1, t_2 \in D}\,.$$

The procedure is provided with a weak (but not strong) control of the FWER but not with a strong control of the FWER. From Theorem 3.1, it follows that the IWT is provided with a control of the FWER over intervals. The attained interval-wise control of the FWER (for a formal definition, see, Pini and Vantini 2017), is also in between the weak and the strong control. Moreover the IWT is consistent as a consequence of Theorem 3.3. Further, the pointwise test statistic is a continuous function, and the test statistic (1) is also continuous with respect to the limits of integration. This implies that the IWT-adjusted p-value function $\tilde{p}_n(t; \mathcal{F}_{IWT})$ is continuous on D, providing us with a viable tool for domain selection.

Other methods similar to IWT can be defined by replacing intervals with more complex subsets of the domain. For instance, an apparently straightforward extension of IWT would be the extension of the underlying family so that it also includes families of countable unions of intervals. However, such a generalization does not lead to a method with desired properties. Indeed, for a fixed integer K, consider the testing procedure based on the family

$$\mathcal{F}_K = \left\{\cup_{j=1}^{K}[t_{1j}, t_{2j}] \ : t_{2j} > t_{1j}\right\}_{t_{1j}, t_{2j} \in D,\, j=1,\dots,K}$$

that is, the family of all possible unions of at most K disjoint intervals. It can be shown (see Proposition A.2 in Section A, supplementary file) that the adjusted p-value

function $\tilde{p}_n(t;\mathcal{F}_K)$ is such that on the one hand, $\forall K \geq 2$, $\tilde{p}_n(t;\mathcal{F}_K)$ is constant on D and such that $\tilde{p}_n(t;\mathcal{F}_K) \geq \max_{t\in D} p_n(t)$. On the other hand, $\forall K < \infty$, $\tilde{p}_n(t;\mathcal{F}_K)$ is not provided with a finite-sample strong control of the FWER. Hence the adjusted p-value function is constant, making the method unsuitable for domain selection. Further, the finite sample strong control of the FWER is not possible to obtain for any such family smaller than $\mathcal{F}_{BTW}$.

4.2 Data-driven families $\mathcal{F}_n$

Section 4.1 shows that in the case of pre-defined families it is not possible to guarantee both the possibility of performing domain selection and strong control of the FWER. This limitation introduces the urgency of focusing on a wider class of families which could possibly overcome this limitation, i.e., data-driven families. In the following, we show that in this enlarged setting and with large sample sizes, it is possible to identify families that could both provide an (asymptotic) strong control of the FWER and allow for domain selection at the same time. As an example of this novel apporach, we introduce a novel Treshold-Wise Testing procedure and briefly comment on this.

Threshold-Wise Testing

The idea behind the Threshold-Wise Testing (TWT) is to costruct a family $\mathcal{F}_{TWT_n}$ derived from the unadjusted p-value function, thus being data dependent. We define the family as made of the sublevel and superlevel sets of the the unadjusted p-value function. Formally, the TWT is based on the data driven family

$$\mathcal{F}_{TWT_n} = \{\{t \in D : p_n(t) \leq y\}, \{t \in D : p_n(t) \geq y\}\}_{y\in[0,1]}. \tag{5}$$

For finite n, TWT is only provided with a weak control of the FWER. However, it meets the conditions of Theorem 3.2, and is hence provided with an asymptotic strong control of the FWER. In addition, due to the limiting behaviour of the unadjusted p-value function, as $n \to \infty$ the sets of $\mathcal{F}_{TWT_n}$ intersect $\mathcal{D}_1$ on sets of positive measure. Hence, Theorem 3.4 implies that the method is consistent and can be used for domain selection.

Other data driven families can also be constructed using preimages of the unadjusted p-value function, corresponding to a suitable family of subsets of the codomain $[0, 1]$. Such families, that also satisfy the assumptions of Theorems 3.2 and 3.4 will share the same asymptotic properties as the TWT method. For example, $\mathcal{F}_{TWT_n}$ can be enlarged so it contains the counter images of every closed interval in the codomain of the unadjusted p-value function. A reduced family of $\mathcal{F}_{TWT_n}$ is exemplified by the two level sets of the unadjusted p-value function:

$$\mathcal{F}_{0_n} = \{\{t \in D : p_n(t) > 1/n\}, \{t \in D : p_n(t) \leq 1/n\}\}.$$

It is straightforward to assess the inferential impacts of reducing or enlarging a given family. Removing some elements from the family means removing the corresponding tests in the computation of the adjusted p-value function (eq. 2) thus weakening the control of the FWER (being the control guaranteed just on the performed tests) but possibly increasing the statistical power of the procedure (being the new adjusted

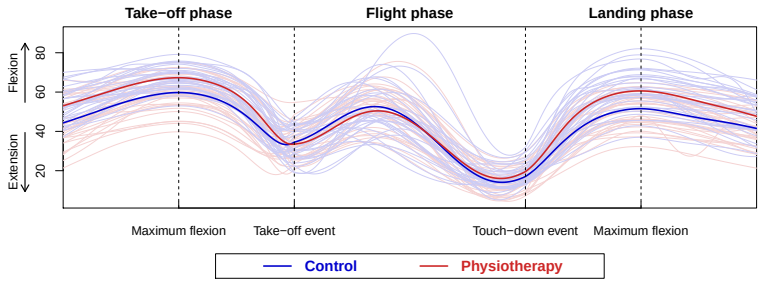

Figure 1: Pre-processed kinematics data. The bold lines correspond to the average curves within each group. The landmarks used for registration are marked on the x-axis.

p-value function uniformly equal to or smaller than the original one). If instead new elements (and thus new tests) are introduced in the family, the control of the FWER is strengthened and thus the convergence towards a strong control of the FWER possibly fastened. The price for this is of course a natural loss of statistical power.

In the following sections, we use the TWT in (5) as a representative of the class of data-driven family methods. It illustrates how data-driven family methods successfully may overcome the pre-defined family methods both in terms of FWER and power. However, it is out of the scope of this work to show that the TWT is, in ways yet to be defined, the "best" method based on the segmentation of the codomain of the unadjusted p-value function.

Finally, the computational costs of TWT and of all other possible methods based on the segmentation of the codomain of the unadjusted p-value function are not affected by the dimensionality of the domain. This makes them naturally suited to deal with functional data defined on multidimensional domains or even on smooth manifolds, cf. Section 6.

4.3 Illustrative 1D example

We present the results of the p-value adjustment methods on a knee kinematics data example. The data are obtained from a long-term follow up after injury to the anterior cruciate ligament (ACL) also including knee-healthy controls (KACL20-study) (see, e.g., Tengman et al. 2013, for details). We consider the knee flexion during a one-leg hop for distance, comparing a first group ($i = 1$, $n_1 = 33$) of ACL-injured individuals treated with physiotherapy only with a second group ($i = 2$, $n_2 = 34$) of knee-healthy controls.

Movements were recorded using a motion capture system with eight cameras (Oqus ®, Qualisys Medical AB, Gothenborg, Sweden). After initial pre-processing we obtain the dynamics of the knee flexion over time. To make the functional objects comparable, we align the data using as landmarks take-off and touch-down events, together with the events of maximal knee flexion during take-off and landing-phases. For details, we refer to Hébert-Losier et al. (2015). The obtained functional data together with landmarks is displayed in Figure 1.

We assume that the observed functional data follow the model $y_{ij}(t) = \mu_i(t) + \varepsilon_{ij}(t)$,

$i = 1, 2$, $j = 1, \ldots, n_i$, $t \in D$. The functions $\mu_i(\cdot)$ are group means and $\varepsilon_{ij}(\cdot)$ are zero-mean *i.i.d.* error functions. Note that we make no assumptions about the dependency of the error term in-between two time points, neither we make any further distributional assumptions about it. We want to test the equality of means in both groups,:

$$H_0 : \mu_1(\cdot) = \mu_2(\cdot) \qquad vs. \qquad H_1 : \mu_1(\cdot) \neq \mu_2(\cdot). \tag{6}$$

To carry out local inference, we make use of the pointwise test statistics:

$$T_n(t) = (\bar{y}_1(t) - \bar{y}_2(t))^2, \tag{7}$$

where $\bar{y}_1(t)$ and $\bar{y}_2(t)$, $t \in D$, are the means of the functional observations at point t in the ACL and control groups, respectively. We utilise permutations to perform the pointwise test and the tests on subsets (see, Pesarin and Salmaso 2010) which guarantees that all tests are exact and consistent and, as such, satisfy the assumptions of the theorems.

Implementation details The domain is discretised into 1000 equisized subintervals. The unadjusted p-values are computed on the right endpoint of each subinterval. For each subset $S \in \mathcal{F}$ the integrated test statistic is computed with a rectangle rule. Permutation tests on subsets are based on 5000 randomly chosen permutations. To avoid exhaustive calculations, for approximating the BWT-adjusted p-value function $\tilde{p}_n(t)$ we use its lower bound $\max_{t \in D} p_n(t)$. For the TWT procedure the codomain of the unadjusted p-value function is discretized into 5000 subintervals. Finally, the PCT method is applied on partitions of $J = 3$ and $J = 18$ equisized intervals.

Results In Figure 2 we present the results of the introduced inferential methods. The figure consists of six panels, each corresponding to a different family $\mathcal{F}$. In each plot, the black and colored lines corresponds to the unadjusted and adjusted p-value function, respectively. For each family we also present some of its members below the corresponding x-axis. The unadjusted p-value ranges from very low values up to 1 (as the mean functions for both groups are crossing twice). For the global test, the effect of the significantly different regions is sufficiently strong to result in an adjusted p-value that is low. On the contrary, the existence of an unadjusted p-values equal to 1, implies that the BWT-adjusted p-value is constantly equal to 1. When looking at the PCT method, we see the difference between the two resolutions used. When $J = 3$, the adjusted p-value function is a 3-step function identifying two regions where H_0 is rejected. On the other hand, when $J = 18$ the adjusted p-value function is an 18-step function, and the potential resolution of regions where the two populations are significantly different is higher. Still, the adjustment for a higher number of elements of the partition makes the procedure more conservative. Indeed, in this case no significant difference are detected. The IWT results in a continuous adjusted p-value function. However, a union of two intervals is detected, so the strong control of the FWER on the selected region is not guaranteed. Finally, the TWT method results in an adjusted p-value function that is relatively similar to the IWT one. Even though no strong control of the FWER is provided for finite samples, asymptotically such property is attained. A figure showing the resulting domains for all methods are found in Section D in the supplementary file.

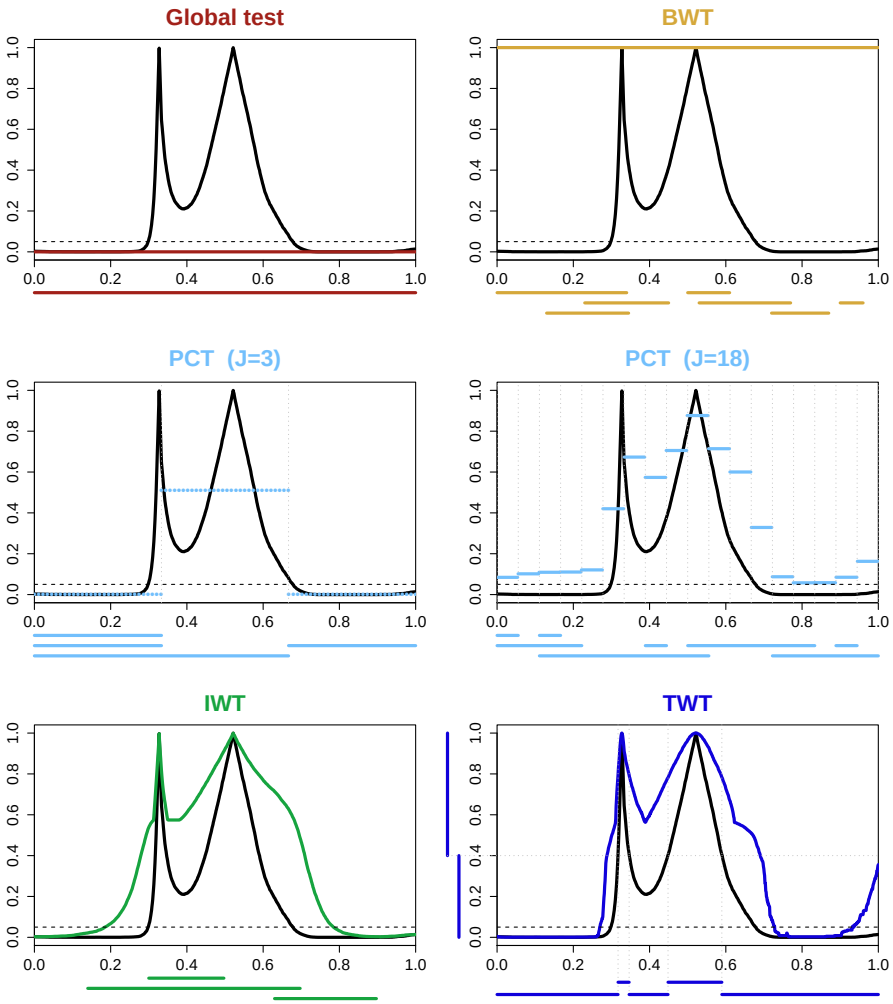

Figure 2: Plots representing resulting adjusted p-value functions for introduced methods (colored). In each plot, the solid (black) line corresponds to the unadjusted p-value function. Below each plot, a sample of members of the family $\mathcal{F}$ is presented. A horizontal line at level 5% is added to help the reader visualise the results of the adjustment on domain selection.

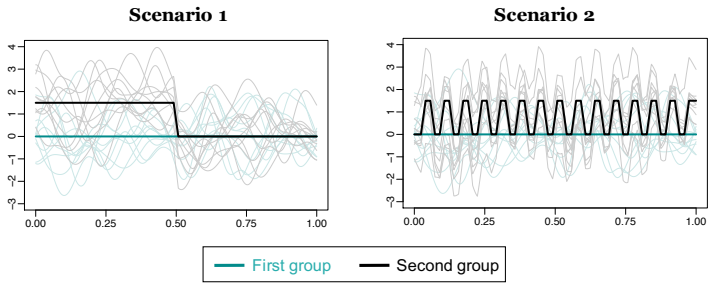

Figure 3: Mean functions (bold lines) together with the sample functions (light lines) for both groups. Left and right panels correspond to Scenario 1 and 2, respectively.

5 Simulation study

To compare the performance of the methods in a finite sample setting and further exemplify the properties, we designed the following simulation study. The inferential problem at hand is the same as in our illustrative data example, i.e., the comparison of means of two functional populations. We consider equal size samples of two groups: $y_{ij}(t) = \mu_i(t) + \varepsilon_{ij}(t)$ $i = 1, \ldots, n$, $j = 1, 2$, $t \in D = [0, 1]$. The error functions $\varepsilon_{ij}(t)$ are zero mean, Gaussian and independent between individuals and populations. We assume a squared exponential covariance structure on time domain, i.e., $\mathrm{Cov}(\varepsilon_{ij}(t_1), \varepsilon_{ij}(t_2)) = \exp(-(t_1 - t_2)^2), t_1, t_2 \in D$. We consider two scenarios for the mean function, depicted in Figure 3. In both cases the mean value functions are equal on about half of the domain and different on the remaining part. The amplitude difference, when present, is at the same level in both cases. However, the difference between the two scenarios is the distribution of the equality region on the domain. In Scenario 1, the difference is present on one unique interval, in the beginning of the domain, while in Scenario 2, there are 15 alternating equality and inequality regions. Similarly to the illustrative example we want to test the two sample mean equality hypothesis (6) using permutation tests with the pointwise test statistics defined in (7).

Implementation details The domain of the functional data is discretised into 60 equisized subintervals, and the unadjusted p-values are computed on the right endpoints of the 60 subintervals. For each set $s \in \mathcal{F}$ the integrated test statistic is computed with a rectangle rule. Each permutation test is based on 5000 randomly chosen permutations. The lower bound $\max_{t \in D}$ is used for approximating the BWT-adjusted p-value function $\tilde{p}_n(t)$. For the TWT method the codomain of the p-value function is discretized into 100 subintervals. The PCT method is applied on partitions of 4, 5, and 10 equisized intervals.

Performance measures The performance of the methods is measured by estimating the following measures based on 1000 simulated realizations.

- The FWER, defined as $P(\exists t \in \mathcal{D}_0 : \tilde{p}_n(t) \leq \alpha)$, by the proportion of the simulated realizations where we wrongly rejected H_0 in at least one point $t \in \mathcal{D}_0$.
- The sensitivity, defined as $\mathrm{E}\left[|t \in \mathcal{D}_1 : \tilde{p}_n(t) \leq \alpha| / |\mathcal{D}_1|\right]$, by the average proportion of the domain $\mathcal{D}_1$ where a difference is correctly discovered.

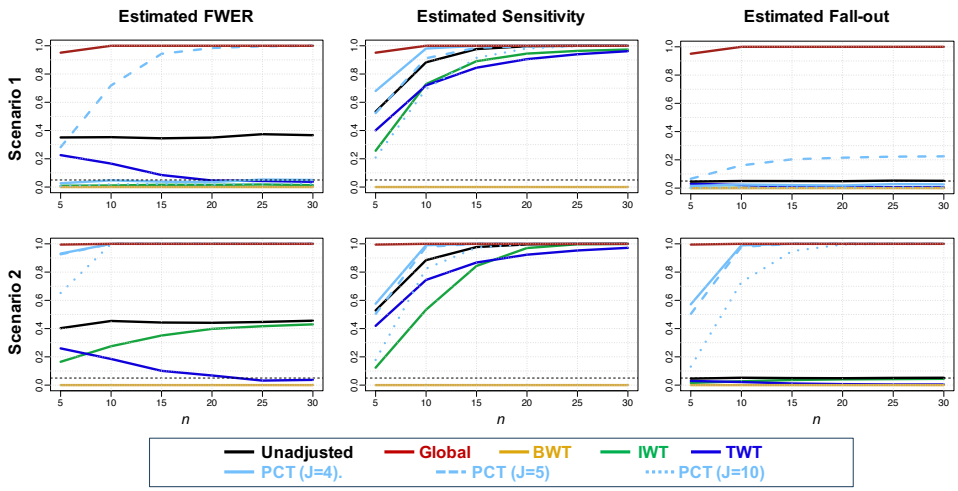

Figure 4: Effect of increased sample size n on the estimated FWER (first column), sensitivity (second column) and fall-out (third column) for the introduced methods in the first (upper row) and the second (lower row) scenarios. Line colors correspond to different methods, while line types correspond to different sizes of the partition for the PCT method.

- The fall-out, defined as $\mathrm{E}\left[|t \in \mathcal{D}_0 : \tilde{p}_n(t) \leq \alpha|/|\mathcal{D}_0|\right]$, by the average proportion of the domain $\mathcal{D}_0$ where a difference is wrongly discovered.

In Figure 4, we present the dynamics of the estimated quantities, as a function of n. As expected, for both scenarios, the sensitivity of all the methods, except BWT, increases as n increases. On the contrary, the BWT is the only method always controlling the FWER in both scenarios. Observe thought that in practice the method does not detect any significant differences and hence is not of any practical use. The IWT and PCT methods, control the FWER only if the underlying partition into $\mathcal{D}_0$ and $\mathcal{D}_1$ can be captured by the corresponding family of subsets. In Scenario 1, since the null hypothesis is true on an interval, the IWT results in a finite sample control of the FWER. Further, the interval can also be constructed using a partition defined by the PCT method with $J = 4$ and 10 intervals, but not with J=5. The visualisation of PCT partitions can be found in Section E, in the supplementary file. In Scenario 2, none of the PCT partitions result in a separation of $\mathcal{D}_0$ and $\mathcal{D}_1$ and therefore no control, neither final sample, nor asymptotic. Finally, the TWT is the only method which possibly allows the selection of portions of the domain and provides strong control of the FWER. In detail, this control is here reached for a reasonable small sample size (i.e, $n \approx 25$) which further supports its possible usefulness in the statistical practice.

6 Analysis of MRI data

This section aims at demonstrating the performance of the proposed TWT framework in the context of statistical testing with domain selection over complex domains. We will emphasize this aspect using data from brain magnetic resonance imaging (MRI).

6.1 Brain structural connectivity from diffusion MRI

A brain image constitutes a complex spatial domain, since it is a subspace of $\mathbb{R}^3$ with a complex shape. In this application, the complex domain over which we will perform p-value correction is defined by the voxels (3D pixels of the imaged brain) that are intersected by the so-called corpus callosum (CC), which is the set of neurons connecting the two hemispheres of our brain. The CC neurons are arranged along a surface, so our domain is a two-dimensional manifold of $\mathbb{R}^3$.

Clinical Context. The major constituents of the brain white matter are (i) the axons that are prolongations of neurons that connect them together, and (ii) the glial cells, which forms the tissue that support the axons. Both types of cells have impermeable membranes, so water trapped within these cells will undergo restricted diffusion. This gave birth to the idea of using diffusion MRI for mapping the white matter microstructure (Moseley et al. 1990). Typically, this is performed by assuming a parametric model of the diffusion in each voxel and by estimating its parameters given the observed MR signals, which depict diffusion in Fourier space. Many models for describing the diffusion voxelwise have been devised in the literature, including the widely used zero-mean 3-dimensional Gaussian distribution (generalization of Brownian motion to anisotropic media), giving its name to the single tensor model (hereafter referred to as STM, Basser et al. 1994). In STM the diffusion tensor is defined as 3-dimensional symmetric definite positive matrix proportional to the covariance matrix of the assumed Gaussian diffusion. The fractional anisotropy (FA, Pierpaoli and Basser 1996) is a scalar index that can be computed from the diffusion tensor, which takes value between 0 (isotropy, all the eigenvalues of the tensor are the same) and 1 (asymptotic anisotropy, degenerated tensor with a single non-zero eigenvalue). FA has been widely adopted as a proxy for quantifying axonal damage (Horsfield and Jones 2002; Assaf and Pasternak 2008) which provokes a drop in FA.

Clinical Question. Diffusion MRI however has a rather low spatial resolution (typically about 2 mm^3 voxel size). For this reason, it has been shown that modeling diffusion in a voxel as a single Gaussian distribution provides an inaccurate description of the microstructure. This is because more than one tissue population compose each voxel of the white matter (different axon orientations, possible presence of glial cells). Hence, when a drop of FA is observed, it might be the sign of lesional damage but it might also simply be a sign of model mis-specification, i.e., a failure to account for the presence of multiple tissue populations. As a response to this criticism, mixture models of the diffusion (a.k.a. multi-compartment models, MCM) have been proposed (Panagiotaki et al. 2012). MCM are based on adding mixture components for modelling additional tissue populations. In this case study, we specifically add components that capture free water diffusion outside the cells and restricted diffusion within glial cells. While it is believed that such models should provide more specific markers for lesional damage, there has not been, to the best of our knowledge, any study that aimed at statistically quantifying this claim. We hereby propose to demonstrate that improving upon the widely adopted STM of the diffusion by using MCM does result in maintaining high FA values where expected.

Design of experiment. To achieve this goal, we processed diffusion MRI data of 30 healthy subjects from the Human Connectome Project (Van Essen et al. 2013) to

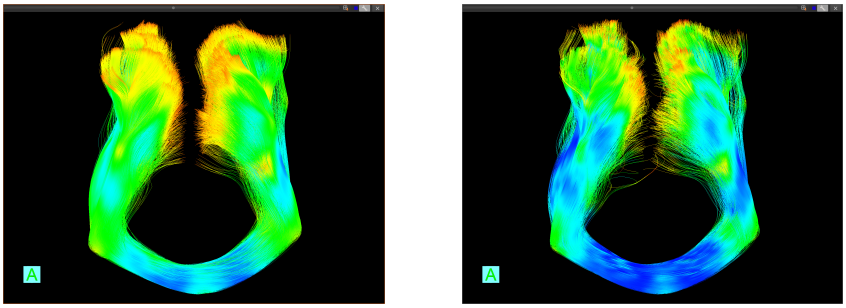

Figure 5: **Corpus Callosum Reconstruction via Diffusion MR Tractography.** Coronal view of the reconstructed corpus callosum of one healthy subject from the Human Connectome Project database, performed using the single tensor model (STM) on the left and the multi-compartment model (MCM) on the right. Colors encode FA (red: 0, blue: 1).

obtain a reconstruction of the CC of each subject. We chose the CC because its reconstruction is relatively easy and it should present a high FA everywhere on healthy subjects. We performed CC reconstructions using both STM and MCM. We fitted the models by maximum likelihood (Stamm et al. 2016) and coregistered the subjects' brain onto the common MNI template (Brett et al. 2002) in which we subsequently performed the CC reconstructions using the fiber assignment continuous tracking (FACT) algorithm (Mori et al. 1999). Figure 5 shows the reconstructed CC for one of the subjects using both models. Finally, we defined the common domain of the CC as the set of all the voxels of size 1.25 mm^3 that were intersected by the CCs of all the 30 healthy subjects, which provided us with a domain of 950 voxels georeferenced in 3 dimensions and lying on a two-dimensional manifold.

6.2 Comparing structural connectivity maps with TWT

We test the null hypothesis that the two distributions that generated the FA maps we observed from the two different diffusion models are the same with a paired one-tailed permutation test. The alternative hypothesis is defined by the voxelwise test statistic that one chooses for performing the test. Using the Kolmogorov-Smirnov (KS) statistic, we look for global differences of the two distributions. Using the difference between the sample means, we instead look for mean differences specifically. Using the variance ratio, we look for variance differences. We used all three test statistics but only report the results for variance comparisons since we obtained very low p-values on the whole domain in the other two cases. In the latter case, we tested the alternative hypothesis that the variance of FA in MCM is lower than the variance of FA in STM.

Domain selection is of paramount importance in brain applications where we need spatial localization of the differences. However, as shown theoretically and by simulations, global testing and BWT are not of practical use because they select either the whole domain or nothing at all. It is also difficult to make a sensible choice of the domain partition that would be required for applying PCT and generalizations of intervals to complex manifolds such as the one defined by the CC surface are not straightforward, which prevents us from using IWT. Hence, domain selection can only be performed through TWT. For completeness, we included also maps of unadjusted p-

values and adjusted p-values using Holm adjustment, a classical procedures for strong control of the FWER (Holm 1979). The three panels of Figure 6 shows the restriction of resulting p-values maps on a two-dimensional section of the three-dimensional domain.

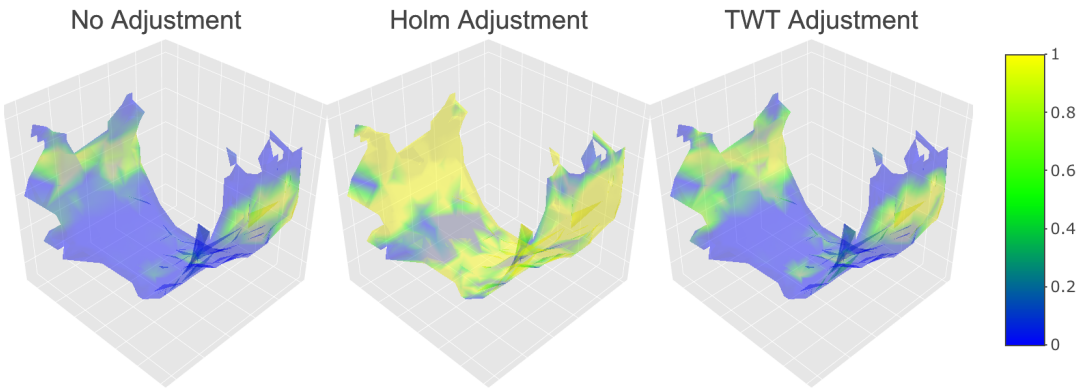

Figure 6: ***P*-Values Maps on the Common Corpus Callosum Surface.** Tested alternative hypothesis is that FA variance is smaller with MCM w.r.t. STM.

We can see that, while the TWT offers strong control of the FWER while maintaining high statistical power, the Holm method loose a significant amount of statistical power. The TWT approach identifies two symmetric areas (one in each brain hemisphere) where the FA variance cannot be claimed to be significantly lower in the MCM with respect to the STM. This is very interesting from a neurological perspective because these two areas are precisely the regions where the CC tract crosses with two other well known tracts, namely the superior longitudinal fasciculus and the pyramidal tract. Hence, while it makes sense that adding extra isotropic component mixtures helps in increasing the FA mean of the CC, we indeed do not expect improvement in the FA variance unless we add extra anisotropic components that would specifically model the additional tracts.

7 Conclusions

In this paper, we introduce a general framework for local inference for functional data, which covers existing and opens up the possibility of introducing new such methods. We provide tools that verify the properties of the methods within the framework. For the methods discussed, we focused on their inferential and computational aspects. A summary of the methods and their properties can be found in Section F, supplementary file. Among the methods resulting from pre-defined choices of $\mathcal{F}$, none om them are resulting in strong control of the FWER, consistency and continuous domain selection properties. Further, the complexity of pre-defined families increases with the dimension and without special care the methods become infeasible for $d > 1$. On the contrary, the complexity of the proposed TWT method, does not depend on the dimension and can be used for complex data.

Supplementary material

The file **Supplementary.pdf** contains proofs related to the results in Section 3, a graphical representation of the unified framework and a comparison between the introduced methods within the framework (including computational aspects), as well as additional figures related to the real and simulated examples. The file **Simulation.R** contains the R-code for reproducing the simulation study in Section 5.

Acknowledgement

This work was supported by the Swedish Research Council (grant numbers 2016-02763 and 340-2013-5203). We are grateful to Charlotte K. Häger for providing data from the KACL20-study, that we here use as an illustrative example.

References

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

Abramowicz, K., Häger, C. K., Pini, A., Schelin, L., Sjöstedt de Luna, S. and Vantini, S. (2018), 'Nonparametric inference for functional-on-scalar linear models applied to knee kinematic hop data after injury of the anterior cruciate ligament', *Scandinavian Journal of Statistics* **45**(4), 1036–1061.

Assaf, Y. and Pasternak, O. (2008), 'Diffusion tensor imaging (dti)-based white matter mapping in brain research: a review.', *Journal of Molecular Neuroscience* **34**(1), 51–61.

Basser, P. J., Mattiello, J. and LeBihan, D. (1994), 'MR diffusion tensor spectroscopy and imaging.', *Biophysical journal* **66**(1), 259–67.

Brett, M., Johnsrude, I. S. and Owen, A. M. (2002), 'The problem of functional localization in the human brain', *Nature Reviews Neuroscience* **3**(3), 243–249.

Cardot, H., Goia, A. and Sarda, P. (2004), 'Testing for no effect in functional linear regression models, some computational approaches', *Communications in Statistics - Simulation and Computation* **33**(1), 179–199.

Corain, L., Melas, V. B., Pepelyshev, A. and Salmaso, L. (2014), 'New insights on permutation approach for hypothesis testing on functional data', *Advances in Data Analysis and Classification* **8**(3), 339–356.

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

Fan, J. and Zhang, J. T. (2000), 'Two-step estimation of functional linear models with applications to longitudinal data', *Journal of the Royal Statistical Society Series B* **62**(2), 303–322.

Hall, P. and Tajvidi, N. (2002), 'Permutation tests for equality of distributions in high-dimensional settings', *Biometrika* **89**(2), 359–374.

Hébert-Losier, K., Pini, A., Vantini, S., Strandberg, J., Abramowicz, K., Schelin, L. and Häger, C. K. (2015), 'One-leg hop kinematics 20 years following anterior cruciate ligament rupture: Data revisited using functional data analysis', *Clinical Biomechanics* **30**(10), 1153–1161.

Holm, S. (1979), 'A simple sequentially rejective multiple test procedure', *Scandinavian Journal of Statistics* **6**, 65–70.

Holmes, A. P., Blair, R. C., Watson, J. D. G. and Ford, I. (1996), 'Nonparametric analysis of statistic images from functional mapping experiments', *Journal of Cerebral Blood Flow & Metabolism* **16**(1), 7–22.

Horsfield, M. and Jones, D. (2002), 'Applications of diffusion-weighted and diffusion tensor mri to white matter diseases - a review.', *NMR in Biomedicine* **15**(7-8), 570–577.

Horváth, L. and Kokoszka, P. (2012), *Inference for functional data with applications*, Vol. 200, Springer.

Marcus, R., Peritz, E. and Gabriel, K. R. (1976), 'On closed testing procedures with special reference to ordered analysis of variance', *Biometrika* **63**(3), 655–660.

Mori, S., Crain, B. J., Chacko, V. P. and van Zijl, P. C. (1999), 'Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging.', *Annals of neurology* **45**(2), 265–9.

Moseley, M. E., Cohen, Y., Kucharczyk, J., Mintorovitch, J., Asgari, H. S., Wendland, M. F., Tsuruda, J. and Norman, D. (1990), 'Diffusion-weighted MR imaging of anisotropic water diffusion in cat central nervous system.', *Radiology* **176**(2), 439–45.

Panagiotaki, E., Schneider, T., Siow, B., Hall, M., Lythgoe, M. and Alexander, D. (2012), 'Compartment models of the diffusion mr signal in brain white matter: a taxonomy and comparison.', *Neuroimage* **59**(3), 2241–2254.

Pesarin, F. and Salmaso, L. (2010), *Permutation tests for complex data: theory, applications and software*, John Wiley & Sons Inc.

Pierpaoli, C. and Basser, P. J. (1996), 'Toward a quantitative assessment of diffusion anisotropy.', *Magnetic resonance in medicine* **36**(6), 893–906.

Pini, A. and Vantini, S. (2017), 'Interval-wise testing for functional data', *Journal of Nonparametric Statistics* **29**(2), 407–424.

Ramsay, J. O. and Silverman, B. W. (2005), *Functional data analysis*, Springer, New York.

Reiss, P. T., Huang, L. and Mennes, M. (2010), 'Fast function-on-scalar regression with penalized basis expansions', *International Journal of Biostatistics* **6**(1).

Spitzner, D. J., Marron, J. S. and Essick, G. K. (2003), 'Mixed-model functional ANOVA for studying human tactile perception', *Journal of the American Statical Association* **98**(462), 263–272.

Staicu, A., Li, Y., Crainiceanu, C. M. and Ruppert, D. (2014), 'Likelihood ratio tests for dependent data with applications to longitudinal and functional data analysis', *Scandinavian Journal of Statistics* .

Stamm, A., Commowick, O., Warfield, S. K. and Vantini, S. (2016), Comprehensive Maximum Likelihood Estimation of Diffusion Compartment Models Towards Reliable Mapping of Brain Microstructure, *in* 'MICCAI 2016, Part III, Lecture Notes in Computer Science', Vol. 9902, Springer, pp. 622–630.

Tengman, E., Grip, H. and Häger, C. (2013), 'One-leg hop-kinematic analysis 20 years after anterior cruciate ligament injury', *Gait & Posture* **38**, S25–S26.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., Ugurbil, K. and WU-Minn HCP Consortium (2013), 'The WU-Minn Human Connectome Project: an overview.', *NeuroImage* **80**, 62–79.

Vsevolozhskaya, O. A., Greenwood, M. C., Bellante, G. J., Powell, S. L., Lawrence, R. L. and Repasky, K. S. (2013), 'Combining functions and the closure principle for performing follow-up tests in functional analysis of variance', *Computational Statistics and Data Analysis* **67**, 175–184.

Vsevolozhskaya, O., Greenwood, M. and Holodov, D. (2014), 'Pairwise comparison of treatment levels in functional analysis of variance with application to erythrocyte hemolysis', *The Annals of Applied Statistics* **8**(2), 905–925.

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M. and Nichols, T. E. (2014), 'Permutation inference for the general linear model', *NeuroImage* **92**, 381–397.

Supplementary material

This file contains supplementary material related to the manuscript "Domain selection and family-wise error rate for functional data: a unified framework"

A Proofs

Proof of Theorem 3.1. Assume that $S \in \mathcal{F}_-$ and H_0^S is true. Since $S \in \mathcal{F}_-$, the test of H_0^S against H_1^S is included in the testing procedure, and its p-value p_n^S is included in the maximization in (2). Hence, by the definition of the adjusted p-value function, we have that $\forall t \in S$, $\forall n \geq 1$, $\tilde{p}_n(t; \mathcal{F}) \geq p_n^S$. By assumption the test of H_0^S is exact and H_0^S is true. Hence we have $\forall \alpha \in (0,1)$ and $\forall n \geq 1$:

$$\mathbb{P}[p_n^S \leq \alpha] = \alpha.$$

This implies

$$\mathbb{P}[\exists t \in S : \tilde{p}_n(t; \mathcal{F}) \leq \alpha] \leq \mathbb{P}[p_n^S \leq \alpha] = \alpha,$$

which ends the proof.

Proof of Theorem 3.2. The FWER can only be defined if $\mathcal{D}_0$ is not empty. Hence, assume that $\mathcal{D}_0 \neq \emptyset$. Depending on the nature of $\mathcal{D}_0$ and $\mathcal{D}_1$, we distinguish between two cases.

Case 1. If $|\mathcal{D}_1| = 0$ (and consequently $\mathrm{cl}(\mathcal{D}_0) = D$), we have that H_0^t is true almost everywhere on D. Since by the definition of $\mathcal{D}_0$, $H_0^{\mathcal{D}_0}$ is true and the test of $H_0^{\mathcal{D}_0}$ against $H_1^{\mathcal{D}_0}$ is asymptotically exact, its p-value $p_n^{\mathcal{D}_0}$ is such that $\forall \alpha \in (0,1)$:

$$\lim_{n\to\infty} \mathbb{P}[p_n^{\mathcal{D}_0} \leq \alpha] = \alpha.$$

Note that, due to the integral nature of the test statistic, the p-value $p_n^{\mathcal{D}_0}$ of the test on $\mathcal{D}_0$ coincides with the p-value of the test on the closure of $\mathcal{D}_0$:

$$p_n^{\mathcal{D}_0} = p_n^{\mathrm{cl}(\mathcal{D}_0)} = p_n^D.$$

Furthermore, since the global test is included in the family $\mathcal{F}_n$, we have that $\forall n \geq 1$ and $\forall t \in D$, $\tilde{p}_n(t; \mathcal{F}_n) \geq p_n^D = p_n^{\mathcal{D}_0}$. Hence, we have:

$$\limsup_{n\to\infty} \mathbb{P}[\exists t \in D,\ \tilde{p}_n(t; \mathcal{F}_n) \leq \alpha] \leq \limsup_{n\to\infty} \mathbb{P}[\exists t \in D,\ p_n^D \leq \alpha] = \alpha.$$

and we have the thesis.

Case 2. Assume now $|\mathcal{D}_1| > 0$, i.e., there exists at least one subset of D of non-null measure where the null hypothesis is false. Since all tests of H_0^t against H_1^t are consistent we have that $\forall t \in \mathcal{D}_1$, $p_n(t) \xrightarrow{a.s.} 0$ as $n \to \infty$. Since all tests of H_0^t against H_1^t are asymptotically exact, we also have that $\forall t \in \mathcal{D}_0$, $p_n(t)$ converges in distribution to a uniform distribution on $[0,1]$, such that $\forall n \geq 1\ \mathbb{P}[p_n(t) > 0] = 1$, and by consequence we have also $\lim_{n\to\infty} \mathbb{P}[p_n(t) > 0] = 1$. So with probability one:

$$\forall t_0 \in \mathcal{D}_0, t_1 \in \mathcal{D}_1 : \lim_{n\to\infty} p_n(t_1) = 0 < \lim_{n\to\infty} p_n(t_0),$$

i.e., the images of the sets $\mathcal{D}_0$ and $\mathcal{D}_1$ through the function $p_n(t)$ separate asymptotically on the interval $[0,1]$. Now, we know that for every n the family $\mathcal{F}_n$ includes almost surely the set $\{t \in D \text{ s.t. } p_n(t) \geq \varepsilon_n\}$, $\tilde{\varepsilon}_n > 0$. Since $\lim_{n\to\infty} \varepsilon_n = 0$, for $n \to \infty$ and with probability one, $\mathcal{F}_n$ will asymptotically include the set $\mathcal{D}_0$. Consequently, $\forall t \in \mathcal{D}_0$, $\limsup_{n\to\infty} \tilde{p}_n(t;\mathcal{F}_n) \geq \limsup_{n\to\infty} p_n^{\mathcal{D}_0}$ almost surely. Due to the integral nature of the test statistic, the p-value $p_n^{\mathcal{D}_0}$ of the test on $\mathcal{D}_0$ coincides with the p-value of the test on the closure of $\mathcal{D}_0$:

$$p_n^{\mathcal{D}_0} = p_n^{\mathrm{cl}(\mathcal{D}_0)}.$$

Finally, since the test of $H_0^{\mathcal{D}_0}$ against $H_1^{\mathcal{D}_0}$ is asymptotically exact and $H_0^{\mathcal{D}_0}$ is true, we have

$$\mathbb{P}[p_n^{\mathrm{cl}(\mathcal{D}_0)} \leq \alpha] = \mathbb{P}[p_n^{\mathcal{D}_0} \leq \alpha] \to \alpha.$$

Hence:

$$\limsup_{n\to\infty} \mathbb{P}[\forall t \in \mathrm{cl}(\mathcal{D}_0),\ \tilde{p}_n(t;\mathcal{F}_n) \leq \alpha] \leq \limsup_{n\to\infty} \mathbb{P}[p_n^{\mathrm{cl}(\mathcal{D}_0)} \leq \alpha] = \alpha.$$

Proof of Theorem 3.3. Consider a point $t \in \mathrm{Int}(\mathcal{D}_1)$ (H_0^t is false). Then, for every set S such that $t \in S$, we have that H_0^S is false, since on $S_1 = (S \cap \mathcal{D}_1)$ the null hypothesis is false. In addition, since $t \in \mathrm{Int}(\mathcal{D}_1)$, condition (3) implies that $|S_1| > 0$. Since all tests are consistent, we have

$$\forall S : t \in S \quad p_n^S \xrightarrow[n\to\infty]{a.s.} 0$$

Since the adjusted p-value at point t is

$$\tilde{p}_n(t;\mathcal{F}_-) = \sup_{S\in\mathcal{F}_- : t\in S} p_n^S,$$

we also have $\tilde{p}_n(t;\mathcal{F}_-) \to 0$ almost surely as $n \to \infty$. Hence, we have

$$\forall t \in \mathrm{Int}(\mathcal{D}_1) \quad \Rightarrow \lim_{n\to\infty} \mathbb{P}(\tilde{p}_n(t;\mathcal{F}_-) \leq \alpha) = 1.$$

Further, note that $\mathbb{P}\left[\forall t \in \mathrm{Int}(\mathcal{D}_1),\ \tilde{p}_n(t;\mathcal{F}_-) \leq \alpha\right] = \mathbb{P}[\sup_{t\in\mathrm{Int}(\mathcal{D}_1)} \tilde{p}_n(t;\mathcal{F}_-) \leq \alpha]$. Then, by the definition of $\tilde{p}_n(t;\mathcal{F}_-)$:

$$\sup_{t\in\mathrm{Int}(\mathcal{D}_1)} \tilde{p}_n(t;\mathcal{F}_-) = \sup_{t\in\mathrm{Int}(\mathcal{D}_1)} \sup_{S:t\in S} p_n^S = \sup_{S:|S\cap\mathrm{Int}(\mathcal{D}_1)|>0} p_n^S$$

where the fact that $|S \cap \mathrm{Int}(\mathcal{D}_1)| > 0$ is a direct consequence of condition (3). Being $H_0^{S\cap\mathrm{Int}(\mathcal{D}_1)}$ false when $|S \cap \mathrm{Int}(\mathcal{D}_1)| > 0$, we have that $\sup_{S:|S\cap\mathrm{Int}(\mathcal{D}_1)|>0} p_n^S \to 0$, almost surely. This proves the thesis, i.e., $\forall \alpha \in (0,1)$:

$$\lim_{n\to\infty} \mathbb{P}(\forall t \in \mathrm{Int}(\mathcal{D}_1) : \tilde{p}_n(t) \leq \alpha) = 1.$$

Proof of Theorem 3.4. Let $A \subseteq \mathrm{Int}(\mathcal{D}_1)$ and consider a point $t \in A$ (H_0^t is false). As $n \to \infty$, we know that for every set S such that $t \in S$, the measure of $S \cap \mathcal{D}_1$ is almost surely positive. Since the null hypothesis is false in $S \cap \mathcal{D}_1$, it is also false on S. Since all tests are consistent, we have

$$\forall S : t \in S \quad p_n^S \xrightarrow[n\to\infty]{a.s.} 0$$

And the thesis follows with the same argument used for Theorem 3.3.

Proposition A.1. *Let $\mathcal{F}_{BWT}$ be the BWT family defined in equation (4). For all $t \in D$, let $\tilde{p}_n(t; \mathcal{F}_{BWT})$ be the corresponding adjusted p-value function. Then the adjusted p-value function is constant, and such that:*

$$\tilde{p}_n(t; \mathcal{F}_{BWT}) \equiv p^* \geq \max_{t \in D} p_n(t).$$

Proof. Denote with $S^* \in \mathcal{F}_{BWT}$ the Borel set such that $p^{S^*} \geq p^S \quad \forall S \in \mathcal{F}_{BWT}$. We have straightforwardly that $\forall t \in S^*$, $\tilde{p}_n(t; \mathcal{F}_{BWT}) = p_n^{S^*}$. Consider a point $t \notin S^*$, and denote with S_ε^* the set $S_\varepsilon^* = S^* \cup (B_\varepsilon(t) \cap D)$, where $B_\varepsilon(t)$ is the closed ball of t with radius $\varepsilon \geq 0$, i.e.: $B_\varepsilon(t) = \{t' \in \mathbb{R}^d : \ d(t', t) \leq \epsilon\}$, where $d(\cdot, \cdot)$ is the Euclidean distance in $\mathbb{R}^d$. Clearly, $S_\varepsilon^* \in \mathcal{F}_{BWT}$, so by definition of adjusted p-value, we have

$$\forall \varepsilon \geq 0 : \quad \tilde{p}_n(t; \mathcal{F}_{BWT}) \geq p_n^{S_\varepsilon^*}. \tag{8}$$

The test statistic (1) is such that

$$\lim_{\varepsilon \to 0} T_n^{S_\varepsilon^*} = T_n^{S^*} \Rightarrow \lim_{\varepsilon \to 0} p_n^{S_\varepsilon^*} = p_n^{S^*}.$$

So, by taking the limit in inequality (8), we have $\tilde{p}_n(t; \mathcal{F}_{BWT}) \geq p_n^{S^*}$. Finally, observing that by definition of S^*, we have $\tilde{p}_n(t; \mathcal{F}_{BWT}) \leq p_n^{S^*}$, it must be $\tilde{p}_n(t; \mathcal{F}_{BWT}) = p_n^{S^*}$. In conclusion, we have that $\tilde{p}_n(t; \mathcal{F}_{BWT}) \equiv p_n^{S^*} \geq \max_{t \in D} p_n(t)$, $\forall t \in D$.

Proposition A.2. *For integer K, consider the testing procedure based on the family*

$$\mathcal{F}_K = \left\{ \cup_{j=1}^K [t_{1j}, t_{2j}] \ : t_{2j} > t_{1j}, \right\}_{t_{1j}, t_{2j} \in D, j = 1, \ldots, K}$$

The adjusted p-value of such family $\tilde{p}_n(t; \mathcal{F}_K)$ is such that:

1. *$\forall K \geq 2$, $p_n(t; \mathcal{F}_K)$ is constant on D and such that $p_n(t; \mathcal{F}_K) \geq \max_D p_n(t)$;*
2. *$\forall K < \infty$, $p_n(t; \mathcal{F}_K)$ is not provided with a finite-sample control of the FWER.*

Proof. The first point of the statement follows from the same argument that was used for the BWT procedure, by replacing S^* with S_K^* that is the set in $\mathcal{F}_K$ such that $p_n^{S_L^*} \geq p^S \quad \forall S \in \mathcal{F}_K$.
For the second point of the statement, let $A = \cup_{j=1}^{K+1} [t_{1_j}, t_{2_j}]$ with $[t_{1_j}, t_{2_j}]$ being $K+1$ disjoint intervals of non null measure. Clearly $A \notin \mathcal{F}_K$. Furthermore, A can not be closely approximated with any set of the family $\mathcal{F}_K$. Assume now that H_0^A is true. Since A can not be closely approximated by any set in $\mathcal{F}_K$, the adjusted p-value function $p_n(t; \mathcal{F}_K)$ is not necessarily lower than p_n^A for all $t \in A$. Hence, the probability

$$\mathbb{P}[\exists t \in A : p_n(t; \mathcal{F}_K) \leq \alpha]$$

is not controlled and can in general be higher than α. Hence, the FWER is not strongly controlled.

B Graphical representation of the unified framework

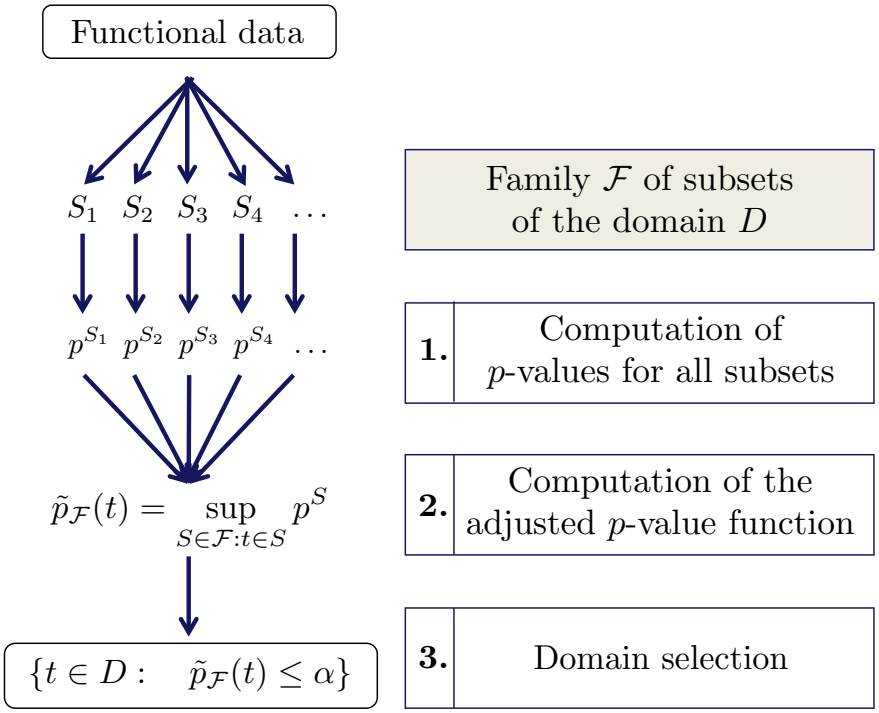

Figure 7: Graphical representation of the unified framework.

C Computational aspects of introduced methods

Borel-Wise Testing From the practical point of view, the implementation of the BWT requires a discretization of the domain of the functional data on a fine grid in order to be able to perform tests. In detail, we can partition the domain D into k equally-spaced subsets $S_1, \ldots, S_k$ and perform a test for every possible union of the subsets. The BWT-adjusted p-value can be estimated by applying formula (2) to these discretized subsets. As $k \to \infty$, the obtained approximation of the BWT-adjusted p-value function converges to its theoretical expression. This procedure poses another problem: the number of tests to be performed for such a procedure is $\mathcal{O}(2^k)$, with k large enough to reduce the approximation error. Consequently, the BWT is computationally unfeasible.

Partition Closed Testing The implementation of the method, for a given partition of size J is straightforward, requiring 2^J tests to be performed, potentially causing computational difficulties for large values of J. On the other hand, since the family consists of a finite number of well defined sets, there is no information loss related to a discretisation of the domain.

Interval-Wise Testing The implementation of the IWT requires a discretization of the domain into a fine grid. Then, each element of the grid is tested, together with each interval composed by elements of the grid. The number of tests to be performed to apply IWT is $\mathcal{O}(k^2/2)$, where k is the number of elements of the grid. For further details on the implementation, see Pini and Vantini (2017).

Threshold-Wise Testing The numerical implementation of TWT first requires the computation of the unadjusted p-value function on a fine grid of the domain. The codomain of the unadjusted p-value function, $[0, 1]$, is discretized into a fine grid (of size M) and each anchored discrete interval is mapped on a set of the domain through the unadjusted p-value function. Finally, the adjusted p-value is computed by identifying the largest p-value of tests on sets $\mathcal{F}_{TWT_n}$. Since anchored intervals only depend on one parameter (one of the two extremes of the interval is always fixed), the total number of tests that is required here is $\mathcal{O}(M)$.

D Selected domain and adjusted p-value functions comparison in Illustrative 1D example (Subsection 4.3)

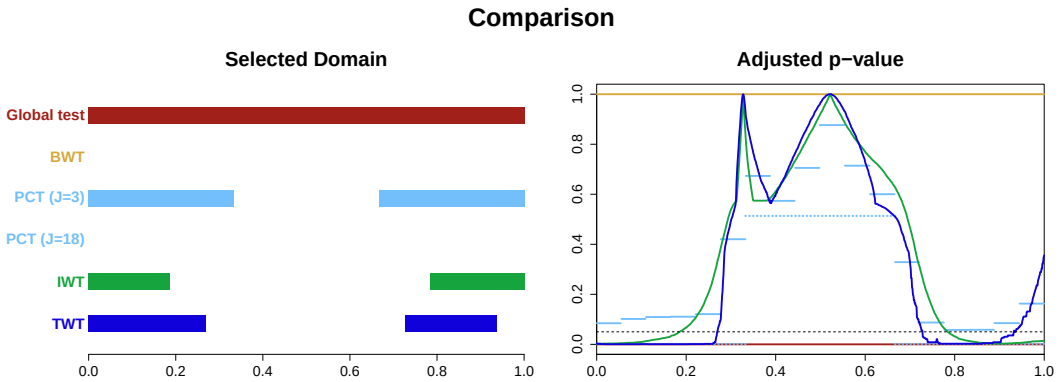

Figure 8: Comparison between the regions with significant difference detected by the six considered procedures (left) and their adjusted p-value functions (right).

E PCT partition visualization for Simulation study in Section 5.

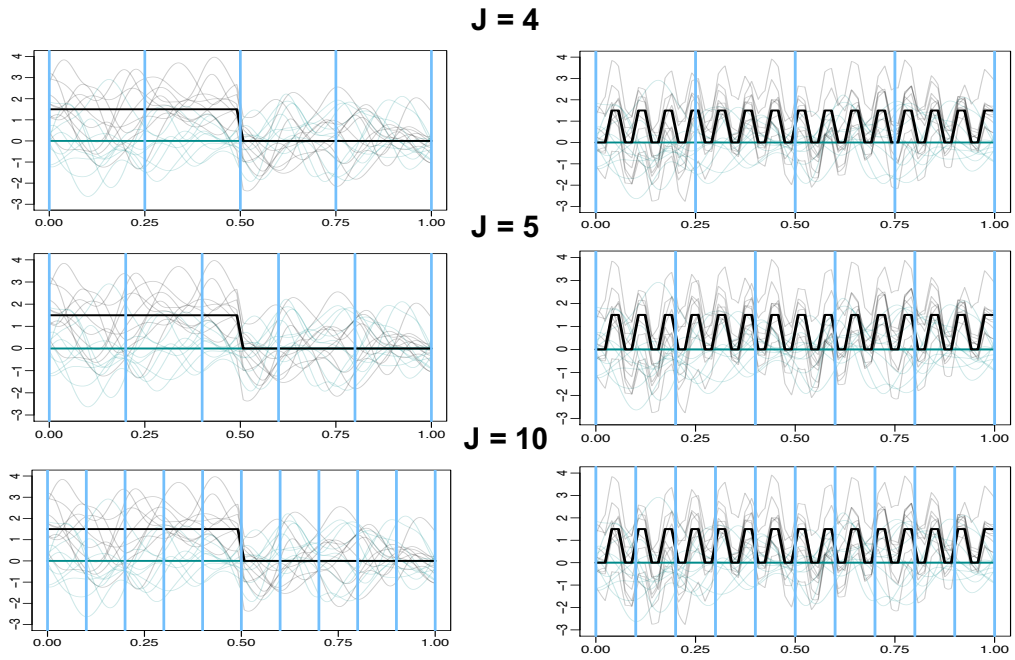

Figure 9: Visualization of the equisized partitions of size J used in the CTP procedure in simulation study (vertical lines) in both scenarios: Scenario 1 (left column) and Scenario 2 (right column). Similarly to Figure 3 we present the mean functions (bold lines) together with the sample functions (light lines) for both groups (with groups color coded)

F Tabular summary of the introduced methods

Procedure	**Family $\mathcal{F}$**	**Weak control of FWER**	**Strong control of FWER**	**Consistent**	**Type of $\tilde{p}_n(\cdot;\mathcal{F})$**	**Complexity**
Global test	$\{D\}$	Yes	No	Yes	Constant	1
BWT	$\mathcal{B}(D)$	Yes	Yes	No	Constant	$\mathcal{O}(2^k)$
PCT	$\{\cup_{j\in\mathbb{I}} S_j\}_{\mathbb{I}\subseteq\{1,\ldots,J\}}$	Yes	No	Yes	Stepwise	$\mathcal{O}(2^J)$
IWT	$\{[t_1,t_2] : t_2 > t_1\}_{t_1,t_2\in D}$	Yes	No	Yes	Continuous	$\mathcal{O}(k^2/2)$
TWT	$\Big\{ \{t\in D : p_n(t)\leq y\}, \{t\in D : p_n(t)\geq y\} \Big\}_{y\in[0,1]}$	Yes	Asymptotically	Yes	Continuous	$\mathcal{O}(M)$

Table 1: Tabular summary of all introduced methods for local testing functional hypotheses on domain D. k: number of discretization points on the domain D. J: size of the partition used in PCT. M: number of discretization points on the codomain of the unadjusted p-value function $[0,1]$.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

09/2019 Antonietti, P.F.; Facciola', C.; Verani, M.
Unified analysis of Discontinuous Galerkin approximations of flows in fractured porous media on polygonal and polyhedral grids

08/2019 Prouse, G.; Stella, S.; Vergara, C.; Engelberger, S.; Trunfio, R.; Canevascini, R.; Quarteroni, A
Computational analysis of turbulent haemodynamics in radiocephalic arteriovenous fistulas with different anastomotic angles

07/2019 Dal Santo, N.; Manzoni, A.
Hyper-reduced order models for parametrized unsteady Navier-Stokes equations on domains with variable shape

04/2019 Delpopolo Carciopolo, L.; Formaggia, L.; Scotti, A.; Hajibeygi, H.
Conservative multirate multiscale simulation of multiphase flow in heterogeneous porous media

05/2019 Gasperoni, F.; Ieva, F.; Paganoni, A.M.; Jackson, C.; Sharples, L.
Evaluating the effect of healthcare providers on the clinical path of Heart Failure patients through a novel semi-Markov multi-state model

06/2019 Pagani, S.; Manzoni, A.; Carlberg, K.
Statistical closure modeling for reduced-order models of stationary systems by the ROMES method

03/2019 Ratti, L.; Verani, M.
A posteriori error estimates for the monodomain model in cardiac electrophysiology

02/2019 Micheletti, S.; Perotto, S.; Soli, L.
Topology optimization driven by anisotropic mesh adaptation: towards a free-form design

01/2019 Regazzoni, F.; Dedè, L.; Quarteroni, A.
Machine learning for fast and reliable solution of time-dependent differential equations

63/2018 Brugiapaglia, S.; Micheletti, S.; Nobile, F.; Perotto, S.
Wavelet-Fourier CORSING techniques for multi-dimensional advection-diffusion-reaction equations