Interval-Wise Testing for Functional Data

A. Pini[a], S. Vantini[a]

[a] MOX– Department of Mathematics, Politecnico di Milano, Milan Italy
alessia.pini@polimi.it
simone.vantini@polimi.it

Abstract

We propose an inferential procedure for functional data, able to select the intervals of the domain imputable of rejecting a functional null hypothesis. The procedure is based on three different steps: (*i*) a functional test is performed on any interval of the domain; (*ii*) an unadjusted and an adjusted p-value function are defined from the results of the previous tests; (*iii*) the significant intervals of the domain are selected by thresholding the unadjusted or the adjusted p-value functions, depending on the desired type of control of the family-wise error rate (i.e., point-wise or interval-wise, respectively). In detail, we prove that the newly defined unadjusted p-value function provides a control of the point-wise error rate (i.e., given any point of the domain where the null hypothesis is not violated - in an L^2 sense to be suitably defined - the probability of wrongly selecting it as significant is controlled) and that it is point-wise consistent (i.e., given any point of the domain where the null hypothesis is violated - in an L^2 sense to be suitably defined - the probability of selecting it as significant goes to one as the sample size goes to infinity). Similarly, we prove that the newly defined adjusted p-value function provides instead a control of the interval-wise error rate (i.e., given any interval of the domain where the null hypothesis is almost-everywhere not violated the probability of wrongly selecting it as significant is controlled) and that it is interval-wise consistent (i.e., given any interval of the domain where the null hypothesis is almost-everywhere violated the probability of selecting it as significant goes to one as the sample size goes to infinity). The procedure is also applied - together to other two state-of.the-art procedures - to the analysis of of the Canadian daily temperatures, to test for pairwise differences between four climatic regions. In detail, we show how the new procedure hereby proposed is able to give a new deeper and useful insight on the possible rejection of the null hypothesis that consists in the selection of the periods of the years presenting significant differences between each couple of regions.

Keywords: Functional Data Analysis, Inference, Domain Selection.

1 Introduction

Functional Data Analysis (FDA), that is the analysis of data sets characterized by the observations of one or more functions for each statistical unit, has recently achieved a considerable importance in modern statistical scientific investigation (Ramsay and Silverman 2002, 2005; Ferraty and Vieu 2006; Horváth and Kokoszka 2012).

A lively area in the field of FDA is the one of statistical inference, which is currently approached from two different perspectives: parametric and non-parametric inference. The former approach commonly relies on distributional assumptions on functional data and/or on asymptotic results (Cuevas et al. 2004; Abramovich and Angelini 2006; Antoniadis and Sapatinas 2007; Horváth and Kokoszka 2012; Staicu et al. 2014; Zhang and Liang 2014). The latter approach relies instead on permutation or bootstrap computational intensive techniques (Hall and Tajvidi 2002; Cardot et al. 2007; Hall and Van Keilegom 2007).

Although being based on different models and assumptions, all these works share a common aim, that is, the one of performing a global test on the whole domain of the functional data. This aim is reached by looking at functions as the data points. Even though mathematically very appealing and providing powerful inferential tools, this type of approach is in many applicative cases not fully satisfactory. Indeed, for instance, whenever the null hypothesis of no difference between the two populations is rejected, practitioners are not just satisfied to know that the two functional populations are significantly different, but they are often interested in understanding which features of the data significantly differ between the two populations. In a finite-dimensional multivariate case, this issue is referred as *feature selection*. In the FDA framework, since data are functions, the feature selection might translate into a selection of the portions of the domain presenting significant differences between the two populations. We will refer to it as *domain selection*. In the very recent literature some examples of domain selection can be found in cluster analysis (Chen et al. 2014), in functional regression with LASSO penalization (Zhao et al. 2012), and thresholded wavelet and sparse representations (Donoho and Johnstone 1994; Dong et al. 2011).

Focussing on inferentially driven feature/domain selection techniques, in the finite-dimensional multivariate case a feature selection can be performed by exploiting multiple testing techniques (Dudoit and Van Der Laan 2007). The commonly used approach is to test all marginal univariate hypotheses pertaining the single features of the data, applying a Bonferroni-Holm (Holm 1979), or Benjamini-Hochberg (Benjamini and Hochberg 1995) multiplicity correction, and then thresholding the resulting adjusted p-values.

Extending to the FDA framework the latter technique poses at least two major challenges: (i) point-wise p-values are in general not trivially defined in functional spaces. If, for instance, data are embedded in the L^2 space (that is the natural extension of the Euclidean geometry to the FDA framework), the point-wise evaluation of data is meaningless; (*ii*) the multiplicity correction would involve a family of infinite tests, and both the Bonferroni Holm and the Benjamini Hochberg procedures are not able to deal with such a case.

In the recent literature, two works have been proposed on this line of research, trying to solve the issue of multiplicity in the FDA framework in different ways: the discretization of functional data, and the discretization of the domain. In Pini and Vantini (2013), data are projected on a finite dimensional functional basis, and a family of tests is performed on the coefficients of the basis expansion, by controlling the probability of falsely rejecting any set of consecutive components of the basis expansion. In Vsevolozhskaya et al. (2014) a finite partition of the domain in sub-intervals is a priori defined, and the significant sub-intervals are selected, with a control of the probability of falsely selecting any set of sub-intervals. Since the first approach relies on a basis expansion of data, conclusions could change depending on the particular basis chosen to project data. In the second case the choice of the initial partition in sub-intervals can affect the conclusions.

In this paper we consider these two works as a starting point, exploiting the sound properties of both procedures. In detail we propose a procedure which neither rely on a basis expansion, nor on an a priori partition in sub-intervals, to obtain a purely non-parametric inferential procedure able to detect the portions of the domain presenting a rejection of the null hypothesis for functional data embedded in L^2. In detail, we propose an inferential procedure providing: (*i*) an unadjusted p-value function providing for each point of the domain a p-value controlling the point-wise error rate (i.e., given any point where the null hypothesis is not violated in a L^2 sense that will be defined, the probability of wrongly selecting it as significant is controlled); and (*ii*) an adjusted p-value function controlling the interval-wise error rate (i.e., given any interval of the domain where the null hypothesis is not violated, the probability of wrongly selecting it is controlled). We show how these p-value functions can be thresholded to select the statistically significant parts of the domain.

The paper is structured as follows: in Section 2 we present the procedure in a general inferential framework in FDA, giving the details of its implementation in the particular case of testing the difference between two functional populations. Finally, Section 3 reports the application of the procedure to a benchmark functional dataset, to test the differences between the mean daily temperatures of different Canadian regions.

2 Methodology and theoretical properties

2.1 Functional interval-wise testing procedure

Suppose that, based on the observation of a set of L^2 random functions over the domain $T = (a, b) \subset \mathbb{R}$, we aim at testing a functional null hypothesis H_0 against an alternative H_1. For instance, in a two-sample framework, if we denote as μ_1

and μ_2 the means of the two functional populations ($\mu_j \in L^2(T)$, $j = 1, 2$), the null hypothesis H_0 can be written as $\mu_1 = \mu_2$ and the alternative one as $\mu_1 \neq \mu_2$. Let $\mathcal{I} \subseteq T$ be an interval or a complementary interval of the form $\mathcal{I} = (t_1, t_2)$ or $\mathcal{I} = T \backslash (t_1, t_2)$, with $a \leq t_1 < t_2 \leq b$, respectively. We further define as $H_0^{\mathcal{I}}$ and $H_1^{\mathcal{I}}$ the restriction of the null and alternative hypotheses on $\mathcal{I}$, respectively (e.g., in a two-sample framework, $H_0^{\mathcal{I}} : \mu_1^{\mathcal{I}} = \mu_2^{\mathcal{I}}$ and $H_1^{\mathcal{I}} : \mu_1^{\mathcal{I}} \neq \mu_2^{\mathcal{I}}$, with $\mu_j^{\mathcal{I}}$ $j = 1, 2$, the restrictions of μ_j over $\mathcal{I}$). Note that $H_0 = \bigcap_{\mathcal{I} \subseteq T} H_0^{\mathcal{I}}$.

In the general case, given an $\mathbb{R}$-valued functional test statistic that could be used to test H_0 globally, we propose a procedure based on the following steps.

- **Interval-wise testing**: for any $\mathcal{I} \subseteq T$, $p^{\mathcal{I}}$ is defined as the p-value of the functional test of $H_0^{\mathcal{I}}$ vs. $H_1^{\mathcal{I}}$, based on the restriction of the chosen test statistic on $\mathcal{I}$.

- **Definition of the p-value functions**: the unadjusted p-value at point t (i.e., $p(t)$) and the adjusted p-value at point t (i.e., $\tilde{p}(t)$) are defined as:

$$p(t) = \limsup_{\mathcal{I} \to t} p^{\mathcal{I}}; \quad \tilde{p}(t) = \sup_{\mathcal{I} \ni t} p^{\mathcal{I}}(t),$$

  where with the notation $\mathcal{I} \to t$, we indicate that both the extremes of the interval $\mathcal{I}$ converge to t.

- **Domain selection**: the significant intervals of the domain obtained by controlling point-wise error rate or the inteval-wise error rate are selected by thresholding the p-value functions $p(t)$ or $\tilde{p}(t)$, respectively.

2.2 Test of mean difference between two populations

We here further detail the procedure in the two-population framework, for testing mean differences between two populations. Everything described here could be possibly extended to more complex testing problems.

Suppose to observe two independent samples of random functions on the $L^2(T)$ space: ξ_{ji}, where $j = 1, 2$ is the population index, $i = 1, \ldots, n_j$ is the unit index, and $T = (a, b) \subset \mathbb{R}$. Let μ_j with $j = 1, 2$ denote the functional means of the two populations. We want to test the following hypotheses:

$$H_0 : \mu_1 = \mu_2 \text{ against } H_1 : \mu_1 \neq \mu_2, \tag{1}$$

where the equality is defined in the L^2 sense (i.e., $\mu_1 = \mu_2 \Leftrightarrow \int_T (\mu_1(t) - \mu_2(t))^2 \mathrm{d}t = 0$). In the case of rejection of the null hypothesis, we aim at selecting the portions of the domain presenting a significant mean difference between the two groups. To achieve this target we apply the procedure sketched in Subection 2.1, that is detailed in the following.

Interval-wise testing. In this phase a functional test is performed for every open interval, and complementary interval of the domain. In particular, for any interval $\mathcal{I}$ of the form (t_1, t_2), or $T \backslash (t_1, t_2)$, with $a \leq t_1 < t_2 \leq b$, we perform the functional test:

$$H_0^{\mathcal{I}} : \mu_1^{\mathcal{I}} = \mu_2^{\mathcal{I}} \text{ against } H_1^{\mathcal{I}} : \mu_1^{\mathcal{I}} \neq \mu_2^{\mathcal{I}}, \tag{2}$$

where $\mu_j^{\mathcal{I}}$ is the restriction of μ_j over $\mathcal{I}$.

For testing (2), different strategies are possible. In the two-population case, when assuming the functional normality of data, it is possible to use parametric functional tests over the corresponding domains, such as the ones proposed by Horváth and Kokoszka (2012).

If the assumption of functional normality seems not realistic, one can rely on non-parametric permutation methods to perform the tests. In detail, we here propose to use for each interval the permutation test described in Hall and Tajvidi (2002), using a test statistic based on the L^2 distance between the two sample means over $\mathcal{I}$:

$$T^{\mathcal{I}}(\xi_{1,1}, \dots, \xi_{1,n_1}, \xi_{2,1}, \dots, \xi_{2,n_2}) = \frac{\|\bar{\xi}_1 - \bar{\xi}_2\|^2_{L^2(\mathcal{I})}}{|\mathcal{I}|} = \frac{1}{|\mathcal{I}|} \int_{\mathcal{I}} \left(\bar{\xi}_1(t) - \bar{\xi}_2(t)\right)^2 \mathrm{d}t, \tag{3}$$

where $\bar{\xi}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \xi_{ji}$, $j = 1, 2$.

An exact permutation test for (2) can then be achieved by evaluating the test statistic $T^{\mathcal{I}}$ over all possible permutations of the data over the sample units. In this framework the p-value of the corresponding test, denoted by $p^{\mathcal{I}}$, is the proportion of the corresponding test statistics evaluated on permuted data exceeding the statistics evaluated on the original data set. As usual, high values of $p^{\mathcal{I}}$ indicates no differences between the two functional means in $\mathcal{I}$, while low values of $p^{\mathcal{I}}$ indicate significant differences in $\mathcal{I}$ between the two functional means.

Definition of the p-value functions. In the second phase, an unadjusted and an adjusted p-value function are evaluated over the domain T for testing (1). In detail, the family of interval-wise tests defined in the first phase, are used to define the unadjusted and adjusted functional p-values.

Definition 2.1. *Consider the set of curves $\xi_{ji} \in L^2(T)$, with $j = 1, 2$, $i = 1, \dots, n_j$, and $T = (a, b) \subset \mathbb{R}$. Let ξ_{ji} be a random sample from a functional population of mean $\mu_j \in L^2(T)$, $j = 1, 2$. Let H_0 and H_1 denote, respectively, the null and alternative functional hypotheses, and $H_0^{\mathcal{I}}$, $H_1^{\mathcal{I}}$ their restriction on $\mathcal{I}$. Finally, let $p^{\mathcal{I}}$ denote the p-value of the functional test of $H_0^{\mathcal{I}}$ against $H_1^{\mathcal{I}}$. Then, the unadjusted p-value function $p(t)$ and the adjusted p-value function $\tilde{p}(t)$ are defined $\forall t \in T$ as:*

$$p(t) = \limsup_{\mathcal{I} \to t} p^{\mathcal{I}}; \quad \tilde{p}(t) = \sup_{\mathcal{I} \ni t} p^{\mathcal{I}}.$$

Note that, even though in the L^2 framework it is meaningless to talk about point-wise evaluations of the functions ξ_{ji}, the boundedness of the p-values $p^{\mathcal{I}}$ guarantees that both the unadjusted and the adjusted p-values are instead well defined $\forall t \in T$.

In addition, being T an open limited interval of $\mathbb{R}$, we have that $L^2(T) \subset L^1(T)$, and thus the integral mean value theorem guarantees that, almost everywhere:

$$p(t) = \lim_{\mathcal{I} \to t} p^{\mathcal{I}},$$

and $p(t)$ coincides almost everywhere with the p-value of the permutation test based on the test statistic $(\bar{\xi}_1(t) - \bar{\xi}_2(t))^2$, computed from the point-wise evaluations of the almost everywhere continuous representatives of the L^2-equivalence classes of data ξ_{ji}. Moreover, in the special case of data embedded in $L^2(T) \cap C^0(T)$, the identity above holds $\forall t \in T$.

We here report the theoretical properties of the unadjusted and adjusted p-value functions $p(t)$ and $\tilde{p}(t)$.

- The unadjusted p-value $p(t)$ is provided with a control of the **point-wise error rate** (Theorem 1 of the Appendix), that is, $\forall \alpha \in (0, 1)$:

$$\forall t \in T \text{ s.t. } \exists \mathcal{I} \ni t : \, H_0^{\mathcal{I}} \text{ is true} \Rightarrow \mathbb{P}[p(t) \leq \alpha] \leq \alpha. \tag{4}$$

- The adjusted p-value $\tilde{p}(t)$ is provided with a control of the **interval-wise error rate** (Theorem 2 of the Appendix), that is, $\forall \alpha \in (0, 1)$:

$$\forall \mathcal{I} \subseteq T : \, H_0^{\mathcal{I}} \text{ is true} \Rightarrow \mathbb{P}\left[\forall t \in \mathcal{I}, \tilde{p}(t) \leq \alpha\right] \leq \alpha. \tag{5}$$

In detail, property (4) implies that the probability of wrongly rejecting a point belonging to a "true" interval is controlled. Property (5) implies instead that the probability of wrongly rejecting a "true" interval is controlled.

Note that, since for almost every $t \in T$, $p(t)$ coincides with the p-value of the permutation test based on the test statistic $(\bar{\xi}_1 - \bar{\xi}_2)^2$, we also have that, $\forall \alpha \in (0, 1)$:

$$\text{for almost every } t \in T \text{ s.t. } \exists \mathcal{I} \ni t : \, H_0^{\mathcal{I}} \text{ is true} \Rightarrow \mathbb{P}[p(t) \leq \alpha] = \alpha.$$

In the special case of data embedded in $L^2(T) \cap C^0(T)$, point-wise error rate and interval-wise error rate can be more easily defined starting from the point-wise evaluations of data. Indeed, in this case, for each point t, $p(t)$ allows the control of the probability that the two means are wrongly detected as different in the point t (i.e., the probability of detecting a difference in t whe $\mu_1(t) = \mu_2(t)$). On the other hand, $\tilde{p}(t)$ allows to control the probability that the two means are wrongly detected as different over an interval $\mathcal{I}$ such that $\forall t \in \mathcal{I}$, $\mu_1(t) = \mu_2(t)$.

Finally, it is worth mentioning that both the unadjusted and the adjusted p-value functions are consistent (Theorem 3 of the Appendix). In detail, $\forall \alpha \in (0, 1)$:

$$\forall t \in T \text{ s.t. } \nexists \mathcal{I} \ni t : \, H_0^{\mathcal{I}} \text{ is true} \Rightarrow \mathbb{P}\left[p(t) \leq \alpha\right] \xrightarrow[n \to \infty]{} 1;$$

$$\forall \mathcal{I} \subseteq T \text{ s.t. } \nexists \mathcal{J} \subseteq \mathcal{I} : H_0^{\mathcal{J}} \text{ is true} \Rightarrow \mathbb{P}\left[\forall t \in \mathcal{I}, \tilde{p}(t) \leq \alpha\right] \xrightarrow[n\to\infty]{} 1.$$

The special case of data embedded in $L^2(T) \cap C^0(T)$ provides a clear interpretation of such consistency property. Indeed, for the unadjusted p-value, we have that the probability of truly detecting any point t for which $\mu_1(t) \neq \mu_2(t)$ converge to one as the sample size increases. In addition, for the adjusted p-value, we have that the probability of truly detecting any interval $\mathcal{I}$ for which $\mu_1(t) \neq \mu_2(t)$ $\forall t \in \mathcal{I}$ also converge to one as the sample size increases.

Domain selection. The intervals of the domain presenting a significant mean difference between the two populations are selected by thresholding the p-value functions evaluated in the previous step. In detail, if we are only interested in controlling the point-wise error rate at level $\alpha \in (0, 1)$, we select the points $t \in T$ such that $p(t) \leq \alpha$. If instead we are interested in controlling the interval-wise error rate at level α, we select the points $t \in T$ such that $\tilde{p}(t) \leq \alpha$.

3 Case study: analysis of Canadian daily temperatures

To illustrate the potential of the functional testing procedure illustrated in this paper, we show here its application to the well known Canadian daily temperature dataset (Ramsay and Silverman 2005). The case study reported in this section is inspired by the pairwise comparisons between daily temperatures of the Australian climatic regions reported in Hall and Tajvidi (2002). In detail, we compare the information provided by the application of the functional test proposed in Hall and Tajvidi (2002) with the result of the functional testing procedure here described. In addition, since in this case one could apply a natural partition of the year into 12 months, we also compare our procedure with the one proposed by Vsevolozhskaya et al. (2014). We first describe the dataset and the results obtained from the analysis. At the end of this section we report some details about the implementation of the test on these data.

The data set contains the average daily temperatures (over 30 years) recorded at 35 weather stations of Canada. The 35 weather stations are divided into four climate zones: Atlantic, Pacific, Continental, and Arctic. The station locations and the functional data are reported in the bottom and top panels of Figure 1, respectively. The four different colors are associated to the different climatic regions. As done by Hall and Van Keilegom (2007), we test for differences between the average temperature of the four regions, in a pairwise perspective.

The Hall Tajvidi test provides for each comparison a global p-value p^T. These p-values are reported in Table 1. This test detects here significant differences between all pairs of climatic zones. Nevertheless, it is not able to identify the periods of the year presenting significant differences.

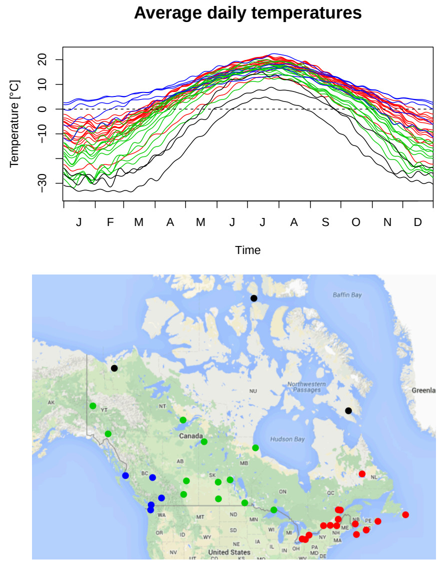

Figure 1: Top: functional data set of average daily temperatures in the 35 weather stations. Bottom: locations of the 35 Canadian weather stations. Each color of both panels is associated with a different region: black-Arctic; green-Continental; red-Atlantic; blue-Pacific.

Pair difference	p^T
Arctic-Atlantic	0.000
Arctic-Continental	0.001
Arctic-Pacific	0.014
Atlantic-Continental	0.000
Atlantic-Pacific	0.000
Continental-Pacific	0.000

Table 1: Global p-values p^T of the test of $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$ over the whole interval T associated to each pairwise difference.

The unadjusted and adjusted p-value functions evaluated according to the procedure described in Section 2.2 are instead displayed in Figure 2, where each different panel of the left column corresponds to a different couple of regions. In detail, the unadjusted p-value function is reported in dashed black line, and the adjusted p-value function is reported in full black line.

Focusing on the adjusted p-value $\tilde{p}(t)$, it is possible to give a very clear interpretation of these results. We notice that both Atlantic and Pacific areas (red and blue curves, respectively) differ from the Arctic climate (black curves) over the entire year (i.e., $\tilde{p}(t)$ is strongly significant along the whole time domain). The temperatures of these two areas also differ from the Continental ones (green curves) during winter only (i.e., $\tilde{p}(t)$ is strongly significant during the winter period). The Continental and Arctic climates are significantly different during the whole year but the winter months (i.e., $\tilde{p}(t)$ is strongly significant from February to November). Finally, the Atlantic and Pacific climates are pointed out as significantly different during the first months of the year, from January to March.

Let us now identify the periods of the year in which the differences occur by limiting to 5% the probability that any period is wrongly selecting as significant. This task is simply obtained by thresholding the adjusted p-value function $\tilde{p}(t)$ at the 5% level. The detected periods are reported in gray in the right panels of Figure 2, together with the curves of the two gropus. The starting and ending days of the significanlty different year periods are summarized in Table 2.

All results are interpretable in terms of Canadian climatic conditions (Stanley 2002). Indeed, due to the mitigating effect of the sea during winter, the temperatures in both the Atlantic and the Pacific regions are higher with respect to the ones in both the Continental and the Arctic regions. During the rest of the year, on the other hand, the temperature in the Continental region rises, reaching the same level as the ones of the Atlantic and Pacific regions, whereas the Arctic area temperatures stay lower. Finally, the difference between the Atlantic and Pacific regions is exclusively due to the warmer period from January to March of the latter, due to the influence of the warmer maritime air of Pacific Ocean.

Pair difference	Period of the year
Arctic-Atlantic	01 Jan. – 31 Dec.
Arctic-Continental	02 Feb. – 15 Dec.
Arctic-Pacific	01 Jan. – 31 Dec.
Atlantic-Continental	17 Sep. – 24 Mar.
Atlantic-Pacific	13 Jan. – 16 Mar.
Continental-Pacific	13 Oct. – 08 Apr.

Table 2: Periods of the year presenting a significant difference between each pair of regions controlling the interval-wise error rate at 5% level.

Finally, we report here a comparison between the conclusions that can be drawn by applying the procedure described in this work and by applying the

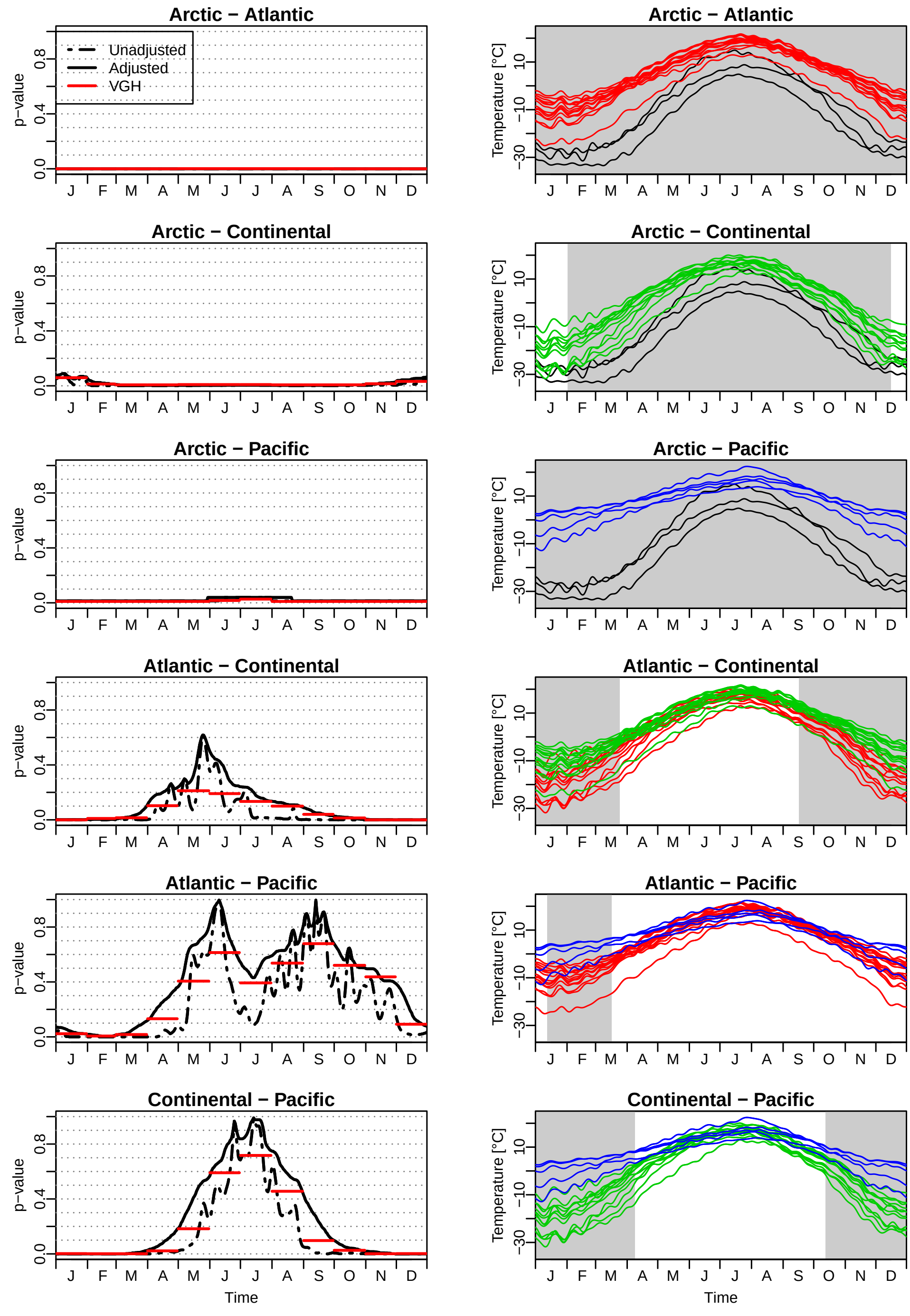

Figure 2: Left: unadjusted (black dashed line) and adjusted (black full line) p-value functions associated to the pairwise differences between the average daily temperatures of four Canadian climatic zones. The red lines represents the adjusted p-values associated to each month according to the VGH test. Right: functional data of each pairwise comparison, and year periods presenting a significant difference between each pair of regions controlling the interval-wise error rate at 5% (gray areas).

testing procedure proposed in Vsevolozhskaya et al. (2014) (we will refer to it as VGH test). The latter one is based on partitioning the domain T into pre-defined sub-intervals and then associating to each sub-interval a unique adjusted p-value obtained by applying the close testing procedure over the sub-intervals. In terms of control of the family-wise error rate, the adjusted p-values provided by VGH test provide a strong control of the family-wise error rate over the family of sub-intervals (i.e., for any set of sub-intervals onto which $\mu_1 = \mu_2$ the probability that at least one of the sub-intervals is wrongly detected as significant is controlled) but only a weak control of the family-wise error rate within each sub-interval (i.e., the procedure does not distinguish between points of the same sub-interval and thus it either rejects or does not reject the null hypothesis $\mu_1 = \mu_2$ on the entire the sub-interval. Consequently the control of the family-wise error rate on portions of the sub-interval is lost). Two are the consequences of this kind of control of the family-wise error rate: (i) differently from the procedure here proposed, the VGH test does not guarantee any control of the family-wise error rate on generic intervals of the domain unless the pre-defined ones or their unions; (ii) as the sample size goes to infinity, an interval (a, b) of the domain included in one of the pre-defined sub-intervals has a probability one of being entirely pointed out as significant also in the undesired case where the null hypothesis $\mu_1 = \mu_2$ is not violated on (a, b) but on an other part of the sub-interval which (a, b) belongs to.

In the case of Canadian daily temperatures, a natural partition of the year domain can be constituted by the 12 months of the year. The adjusted p-values associated to each month evaluated with the VGH test are reported with a solid red line in the left panels of Figure 2. The results of the procedure we here proposed and of the monthly-based VGH-procedure are generally coherent. The main difference is here related to the temporal resolution in the selection of the periods of the year presenting a significant difference. For instance, by means of the VGH test, we can conclude that there is a significant difference at 5% level between the temperatures of the Arctic and the Continental regions during December (the associated adjusted p-value is 0.032). By means of the adjusted p-value function $\tilde{p}(t)$ we are able to provide a more precise information, that is, that this difference is observed in the first 15 days of the month.

Of course, the temporal resolution of the VGH test could be in principle improved by selecting a larger number of pre-defined sub-intervals. However in the practice, when the number k of sub-intervals increases, the power of the VGH test decreases (as in any procedure based on close testing), and moreover its computational cost (which grows exponentially as 2^k) becomes quickly unaffordable.

Details on the implementation. As all FDA techniques, the procedure to evaluate the unadjusted and adjusted p-value functions and select the significant intervals of the domain described in Section 2 has to be numerically approximated to deal with the analysis of real data.

Coherently with previous analyses of the Canadian daily temperature data set presented in the literature (Ramsay and Silverman 2005), the functional data have been obtained by means of a Fourier smoothing on 65 harmonics. The number of harmonics was selected through a cross-validation procedure.

The test statistic (3) is evaluated by means of a numerical integration method, through a trapezoidal rule based on a fine uniform grid of 365 knots. The same evaluation is used to perform the test proposed by Vsevolozhskaya et al. (2014).

Since the number of permutations of the data to be explored to evaluate the p-values is extremely high in at least some of the studied cases, a Conditional Monte Carlo (CMC) algorithm is applied to estimate the p-values of the tests of $H_0^{\mathcal{I}}$ against $H_1^{\mathcal{I}}$ for any interval $\mathcal{I}$ (Pesarin and Salmaso 2010). In the case of the analysis reported in this section a CMC algorithm based on 1000 randomly chosen permutations was employed. Note that, to decrease the computational cost of the permutation method applied to any interval $\mathcal{I}$, it is possible to perform all tests based on the same set of permutations.

Finally, the p-values of the tests $H_0^{\mathcal{I}}$ against $H_1^{\mathcal{I}}$ for any interval $\mathcal{I}$ are discretized on a sufficiently fine grid. Relying on the continuity of the test statistic (3) with respect to the extremes of the integration interval, the max and $\limsup$ of Definition 2.1 have been approximated with their discrete counterparts.

4 Discussion

In this paper we presented a non-parametric domain-selective inferential procedure for functional data embedded in the $L^2(T)$ space (where T is any limited open interval of $\mathbb{R}$). We defined the unadjusted and the adjusted p-value functions, respectively $p(t)$ and $\tilde{p}(t)$, at each point t of the domain. We showed how the unadjusted p-value function is provided with a control of the point-wise error rate, while the adjusted p-value function is provided with a control of the interval-wise error rate. Finally, based on the unadjusted and adjusted p-value functions, we provided a strategy to select the intervals of the domain leading to a rejection of the null hypothesis.

The main novelty of the proposed procedure consists in the extension of the current works on inference for functional data, by direclty focusing on the domain of the curves, instead of providing an overall result on the whole domain, and introducing the concept of unadjusted and adjusted p-value functions. Since the procedure is based on non-parametric permutation tests, it is fully non-parametric. In detail, itsapplication neither requires to specify the distribution of the functional data, nor to expand the data on a functional basis.

To show the potential of the proposed procedure in applications, we performed in Section 3 an analysis of the Canadian daily temperatures data set (Ramsay and Silverman 2005). We applied the procedure to test for pairwise differences between the daily temperatures of four different climatic regions in

Canada. The application of the proposed procedure to this data set show how the latter is able to select the periods of the year presenting significant differences between each couple of regions, providing a clear interpretation of the test results.

Appendix

In this section we provide the formal proofs of the theoretical results characterizing the control provided by the unadjusted and adjusted p-value functions. In the following, suppose then that we are addressing a testing problem in a functional framework, over the domain T. Let $p(t)$, and $\tilde{p}(t)$ the p-value functions defined according to the procedure described in Section 2. Note that all results hold for the unadjusted and adjusted p-values associated to any type of functional test of a null hypothesis H_0 against an alternative H_1 for which an $\mathbb{R}$ valued test statistic is identified. The results only rely on the fact that, for any interval $\mathcal{I}$, the tests of $H_0^{\mathcal{I}}$ against $H_1^{\mathcal{I}}$ is exact and consistent, i.e.,

$$\forall \mathcal{I} \subseteq T : H_0^{\mathcal{I}} \text{ true} \Rightarrow \ \mathbb{P}[p^{\mathcal{I}} \leq \alpha] = \alpha \text{ (exactness);}$$

$$\forall \mathcal{I} \subseteq T : H_0^{\mathcal{I}} \text{ false} \Rightarrow \ \mathbb{P}[p^{\mathcal{I}} \leq \alpha] \xrightarrow[n\to\infty]{} 1 \text{ (consistency).}$$

Our first result characterizes the type of control provided by the unadjusted p-value $p(t)$, that is a control of the point-wise error rate.

Theorem 1. *The unadjusted p-value function $p(t)$ is provided with a control of the point-wise error rate. In detail, $\forall \alpha \in (0, 1)$:*

$$\forall t \in T \textit{ s.t. } \exists \mathcal{I} \ni t : \ H_0^{\mathcal{I}} \textit{ is true} \Rightarrow \mathbb{P}[p(t) \leq \alpha] \leq \alpha.$$

Proof. Let $t \in T$, and $\mathcal{I} \ni t$ s.t. $H_0^{\mathcal{I}}$ is true. This implies that, $\forall \mathcal{J} \subseteq \mathcal{I}$, $H_0^{\mathcal{J}}$ is also true. Since every test is exact, $\mathbb{P}[p^{\mathcal{J}} \leq \alpha] = \alpha$, $\forall \mathcal{J} \subseteq \mathcal{I}$, $\forall \alpha \in (0, 1)$. The unadjusted p-value $p(t)$ is then $p(t) = \limsup_{\mathcal{J} \to t} p^{\mathcal{J}}$. Note that the distribution of $p^{\mathcal{J}}$ is stochastically dominated by the one of $p(t)$. Indeed, if the latter was not true, it would imply that $\exists \alpha \in (0, 1)$ s.t. $\mathbb{P}[p(t) \leq \alpha] > \mathbb{P}[p^{\mathcal{J}} \leq \alpha]$. Hence, since $p(t) = \inf_{\substack{|\mathcal{I}|<\varepsilon \\ \mathcal{I} \ni t}} \sup_{\substack{\mathcal{J} \subseteq \mathcal{I} \\ \mathcal{J} \ni t}} p^{\mathcal{J}}$, the latter would imply that $\exists \mathcal{J} \subseteq \mathcal{I}$ s.t. $\mathbb{P}[p^{\mathcal{J}} \leq \alpha] > \alpha$, that is in contradiction with the exactness of all tests. Hence, we have $\forall \alpha \in (0, 1)$: $\mathbb{P}[p(t) \leq \alpha] \leq \alpha$.

In addition, the unadjusted p-value function $p(t)$ coincides almost everywhere with the p-value of the permutation test based on the test statistic $(\bar{\xi}_1(t) - \bar{\xi}_2(t))^2$, computed from the point-wise evaluations of the almost everywhere continuous representatives of the L^2-equivalence classes of data ξ_{ji}. Such permutation test is an exact test for the restriction of the null hypothesis on the point t H_0^t against H_1^t. Since $\forall \mathcal{I} \ni t$, $H_0^t \subset H_0^{\mathcal{I}}$, we also have:

$$\text{for almost every } t \in T \text{ s.t. } \exists \mathcal{I} \ni t : \ H_0^{\mathcal{I}} \text{ is true} \Rightarrow \mathbb{P}[p(t) \leq \alpha] = \alpha.$$

The following result characterizes instead the type of control provided by the adjusted p-value function $\tilde{p}(t)$, that is, a control of the interval-wise error rate.

Theorem 2. *The adjusted p-value function $\tilde{p}(t)$ is provided with a control of the interval-wise error rate. In detail, $\forall \alpha \in (0,1)$:*

$$\forall \mathcal{I} \subseteq T: \; H_0^{\mathcal{I}} \text{ is true} \Rightarrow \mathbb{P}\left[\forall t \in \mathcal{I}, \tilde{p}(t) \leq \alpha\right] \leq \alpha.$$

Proof. Let $H_0^{\mathcal{I}}$ hold. Since the test of $H_0^{\mathcal{I}}$ is exact, $\mathbb{P}_{H_0^{\mathcal{I}}}[p^{\mathcal{I}} \leq \alpha] = \alpha$. For any $t \in \mathcal{I}$, we have $\tilde{p}(t) \geq p^{\mathcal{I}}$, so $\mathbb{P}_{H_0^{\mathcal{I}}}[\forall t \in T, \, \tilde{p}(t) \leq \alpha] \leq \mathbb{P}_{H_0^{\mathcal{I}}}[p^{\mathcal{I}} \leq \alpha] = \alpha$. $\square$

Finally, the following Theorem prove the consistency of the tests based on both the adjusted and the adjusted p-value functions.

Theorem 3. *The unadjusted p-value function $p(t)$ and the adjusted p-value function $\tilde{p}(t)$ are consistent. In detail, $\forall \alpha \in (0,1)$:*

$$\forall t \in T \text{ s.t. } \nexists \mathcal{I} \ni t: \; H_0^{\mathcal{I}} \text{ is true} \Rightarrow \mathbb{P}\left[p(t) \leq \alpha\right] \xrightarrow[n \to \infty]{} 1;$$

$$\forall \mathcal{I} \subseteq T \text{ s.t. } \nexists \mathcal{J} \subseteq \mathcal{I}: \; H_0^{\mathcal{J}} \text{ is true} \Rightarrow \mathbb{P}\left[\forall t \in \mathcal{I}, \, \tilde{p}(t) \leq \alpha\right] \xrightarrow[n \to \infty]{} 1.$$

Proof. First of all note that, by the definition of unadjusted and adjusted p-vale functions, we have, $\forall t \in T$, $p(t) \leq \tilde{p}(t)$. Hence, it is sufficient to prove that $\tilde{p}(t)$ is consistent. Suppose that the interval $\mathcal{I} \subseteq T$ is such that, $\nexists \mathcal{J} \subseteq \mathcal{I}$ s.t. $H_0^{\mathcal{J}}$ is true. This means that all intervals contained in $\mathcal{I}$ are "false" intervals. Let $t \in \mathcal{I}$. Then, for any interval $\mathcal{K}$ containing the point t, we have that $H_0^{\mathcal{K}}$ is false, since on $(\mathcal{K} \cap \mathcal{I}) \subseteq \mathcal{I}$ the null hypothesis is false. Since each test is consistent, we have that, for $n \to \infty$, the p-value $p^{\mathcal{K}} \to 0$ almost surely, $\forall \mathcal{K} \ni t$. Since $\tilde{p}(t) = \sup_{\mathcal{K} \ni t} p^{\mathcal{K}}$, we also have $\tilde{p}(t) \to 0$ almost surely. Hence, $\forall \alpha \in (0,1)$, $\forall t \in \mathcal{I}$, and for $n \to \infty$, $\mathbb{P}[\tilde{p}(t) \leq \alpha] \to 1$. The latter holds for any $t \in \mathcal{I}$, hence:

$$\forall \mathcal{I} \subseteq T \text{ s.t. } \nexists \mathcal{J} \subseteq \mathcal{I}: \; H_0^{\mathcal{J}} \text{ is true} \Rightarrow \mathbb{P}\left[\forall t \in \mathcal{I}, \, \tilde{p}(t) \leq \alpha\right] \xrightarrow[n \to \infty]{} 1.$$

References

Abramovich, F. and Angelini, C. (2006), Testing in mixed-effects FANOVA models, *J. Statist. Plann. Inference* **136**(12), 4326–4348.

Antoniadis, A. and Sapatinas, T. (2007), Estimation and inference in functional mixed-effects models, *Comput. Statist. Data Anal.* **51**(10), 4793–4813.

Benjamini, Y. and Hochberg, Y. (1995), Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Stat. Soc. Ser. B Stat. Methodol.* pp. 289–300.

Cardot, H., Prchal, L. and Sarda, P. (2007), No effect and lack-of-fit permutation tests for functional regression, *Comput. Statist.* **22**(3), 371–390.

Chen, H., Reiss, P. T. and Tarpey, T. (2014), Optimally weighted l2 distance for functional data, *Biometrics* **70**(3), 516–525.

Cuevas, A., Febrero, M. and Fraiman, R. (2004), An ANOVA test for functional data, *Comput. Statist. Data Anal.* **47**(1), 111–122.

Dong, W., Zhang, D., Shi, G. and Wu, X. (2011), Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization, *Image Processing, IEEE Transactions on* **20**(7), 1838–1857.

Donoho, D. L. and Johnstone, J. M. (1994), Ideal spatial adaptation by wavelet shrinkage, *Biometrika* **81**(3), 425–455.

Dudoit, S. and Van Der Laan, M. J. (2007), *Multiple testing procedures with applications to genomics*, Springer, New York.

Ferraty, F. and Vieu, P. (2006), *Nonparametric functional data analysis: theory and practice*, Springer.

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

Hall, P. and Van Keilegom, I. (2007), Two-sample tests in functional data analysis starting from discrete data, *Stat. Sinica* **17**(4), 1511.

Holm, S. (1979), A simple sequentially rejective multiple test procedure, *Scand. J. Stat.* **6**, 65–70.

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

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

Pini, A. and Vantini, S. (2013), The interval testing procedure: Inference for functional data controlling the family wise error rate on intervals., Technical Report 13/2013, MOX, Politecnico di Milano.

Ramsay, J. O. and Silverman, B. W. (2002), *Applied functional data analysis: methods and case studies*, Vol. 77, Springer.

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

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, *Scand. J. Stat.* .

Stanley, D. (2002), *Canada's maritime provinces*, Vol. 1, Lonely Planet.

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

Zhang, J. and Liang, X. (2014), One-way ANOVA for functional data via globalizing the pointwise F-test, *Scand. J. Stat.* **41**(1), 51–71.

Zhao, Y., Ogden, R. T. and Reiss, P. T. (2012), Wavelet-based lasso in functional linear regression, *J. Comput. Graph. Stat.* **21**(3), 600–617.