Politecnico di Milano
Department of Mathematics
Doctoral Program In Mathematical Models
and Methods in Engineering

Global and Component-Wise Distribution-Free Inference for Functional Data: Methods and Applications

Doctoral Dissertation of:
Alessia Pini

Supervisor:
Dr. Simone Vantini

The Chair of the Doctoral Program:
Prof. Roberto Lucchetti

Year 2014 – XXVII Cycle

"... you never learn anything unless you are willing to take a risk and tolerate a little randomness in your life."

Heinz Rudolf Pagels

Acknowledgments

First of all, I would like to thank my supervisor Dr. Simone Vantini for guiding me during these three years with patience, enthusiasm and motivation. He started believing in me before I did, and all his insightful suggestions and detailed, precise, and patient reviews made me improve more and more.

I would like to thank all the persons I worked with during my PhD, for their precious help and the fruitful collaboration. In particular, I would like to thank Aymeric Stamm from Harvard Medical School, Elisa Ferrario from École Centrale Paris, Konrad Abramowicz, Sara de Luna, Charlotte Häger, Lina Schelin, and Johan Strandberg from Umeå University, Kim Hébert Losier from National Sports Institute of Malaysia, Bianca Maria Colosimo, Daniele Colombo, and Barbara Previtali from Politecnico di Milano.

I am very grateful to the three families that supported me during this PhD. Thanks to my family, for supporting and helping me in every decision. Many thanks to the MOXstat family, for all the interesting discussions, and for all the crazy ones, too. A special mention to Eil, for the mutual support during the PhD, and for the "cappuccino pact". Thanks to the Dysfunctional family, for making me feel at home during the cold Swedish winter.

Dulcis in fundo, thanks to Luca. Because it is even unnecessary to say why.

Contents

Introduction

Functional Data Analysis, i.e., the statistical analysis of sets of curves, is a lively area of statistics (Ramsay and Silverman, 2002, 2005; Ferraty and Vieu, 2006). New possibilities introduced by recent technologies of recording and storing high-resolution data representable as functions (e.g., time or space-varying data, frequency spectra), leave statisticians with the problem of developing novel methodologies to analyze this type of data. In this work we address one of the debated issues in this field, that is how to perform inference on functional data.

Functional data are typically modeled as random elements of an infinite dimensional separable Hilbert space, usually constituted by the L^2 space (Tarpey, 2003). One of the main issues of this framework is the impossibility of defining a probability density function for random functions (Delaigle and Hall, 2010). However, the problem of developing suitable inferential tools for functional data is of high importance for practitioners. For instance, a test for deciding whether several groups of curves have the same (functional) mean could be applied to a wide range of situations. For example, we may think of the case in which the curves express the movement through time of groups of subjects affected by different pathologies. In this case, a test of mean comparison between groups could provide a quantitative way to understand whether they express the same behavior.

Statistical inference for functional data is currently approached from two different perspectives: parametric and non-parametric inference. Non-parametric inference commonly relies on computational intensive permutation or bootstrap techniques (Hall and Tajvidi, 2002; Cardot et al., 2007; Cuesta-Albertos and Febrero-Bande, 2010; Hall and Van Keilegom, 2007*a*). Parametric inference relies instead on distributional assumptions on functional data (e.g., normality) and on asymptotic results (Horváth and Kokoszka, 2012; Spitzner et al., 2003; Cuevas et al., 2004; Fan and Lin, 1998; Schott, 2007).

In the case of functional data, normality is a very demanding assumption. Indeed, it implies that the projection of the random function on each element of the functional space (e.g., L^2) is a normal random variable (Tarpey, 2003). Such an assumption is practically impossible to verify. For this reason we choose in this work to base inference on non-parametric permutation tests (Good, 2005; Pesarin and Salmaso, 2010). For instance, suppose that we want to test for differences between two groups of curves. A permutation test can be constructed by evaluating the distribution of a test statistic measuring the distance between groups over the space of all possible permutations of data across groups. The p-value of such test can be evaluated as the proportion of permutations leading to a test statistic higher than the one evaluated with the non-permuted data.

The resulting test is exact, and it does not require any parametric distributional assumption.

In this thesis, we present and apply different inferential methods for functional data based on permutation tests. In detail, the thesis includes three parts: Part I describes the different methodological approaches developed to make inference on functional data; Part II presents various applications of the methodologies described in Part I; Part III presents details on the `fdatest` `R` package that contains implementation of inferential procedures for functional data.

Methods

In Chapter 1, we introduce an inferential method for functional data based on a generalization of Hotelling's T^2 statistic (Hotelling, 1931) in functional Hilbert spaces. The methodology introduced is a global inferential method, in the sense that it provides a unique result over the whole domain of the curves. Specifically, we introduce a novel test statistic, the functional Hotelling's T^2, to perform inference on functional data in the L^2 geometry. The test statistics corresponds to a semi-distance between the sample mean and the actual mean, induced by the sample covariance operator. Functional Hotelling's T^2 is the natural extension in the infinite-dimensional framework of the statistical tools for testing the mean with unknown variance, from the works of Gosset and Fisher at the beginning of the twentieth century (Gosset, 1908; Fisher, 1925*a*), up to the earlier extensions of Hotelling's T^2 to high dimensional data (Secchi et al., 2013). Furthermore, even if functional Hotelling's T^2 is defined in the L^2 geometry, as it is the natural extension of Euclidean geometry exploited in the classical multivariate framework, we show how the statistic can be extended on any functional Hilbert space. Finally, we suggest permutation tests based on functional Hotelling's T^2 to test for the mean of one functional population, and to compare the means of two functional populations.

As the majority of works proposing inferential methods for functional data analysis, the inferential method described in Chapter 1 is a global testing procedure, i.e., a procedure that provides a unique result over the whole domain of the curves. For instance, when comparing the means of two different groups, the result of the test is whether the mean functions are statistically different from one another.

In some cases, knowing that two groups are statistically different from one another is not fully satisfactory for practitioners. Indeed, one this is discovered, one of the main interests is to discover how groups differ. Differences might be expressed only in a particular part of the domain, or on a particular frequency band. In these cases, we would want to identify the differences -when present- with a control on false positives.

For this reason, in Chapter 2, we approach the problem of inference in a component-wise perspective, in order to provide a way to answer the question of how data differ. In detail, we suggest a novel technique: Interval Testing Procedure (ITP) that enables inference for functional data whenever data are described via a basis expansion. As its final result, ITP provides a family of adjusted p-values associated to each component of the basis expansion. Hence, significantly different basis components can be selected by comparing corresponding

adjusted p-values with the desired significance level α. When applied to a local basis, such as B-splines, it provides a selection of the statistically significant intervals of the domain. When applied to a Fourier basis, it provides instead a selection of the statistically significant frequency bands. The procedure is provided with an interval-wise control of the Family Wise Error Rate, which is a novel type of error control particularly suited for functional data that guarantees that the probability of falsely rejecting any interval of true null hypotheses is controlled.

In Chapter 2, we introduce ITP in the context of testing differences between two functional populations, as well as for testing the mean of one functional population. Then, in Chapter 3, the procedure is extended to the case of testing a functional-on-scalar linear model, i.e., a model in which the functional responses are expressed as a sum of a functional common mean and functional fixed regression coefficients multiplied by scalar covariates. In this case, we provide: (i) a functional F-test (that we prove to be exact and consistent) for testing the regression model; and (ii) functional t-tests (that we prove to be asymptotically exact and consistent) for testing the effects of single covariates. All tests are performed in a non-parametric way by applying suitable permutation tests. A major advantage over previous works dealing with this type of model is that our inferential approach requires neither normality assumptions on functional residuals nor specific covariance structure.

Applications

In Part II the ITP is employed to analyze and test functional data sets coming from different applications.

In Chapter 4 we present a functional one-way Analysis of Variance applied to knee movement data, on a follow-up study on Anterior Cruciate Ligament (ACL) ruptures. The knee movements of individuals suffering from an ACL injury, treated with surgery (first group) or physiotherapy (second group) and uninjured controls (third group) are investigated, using data from a long term follow-up study (about 20 years after the injury). The aim of the analysis is to compare the kinematics during a one-leg hop, i.e., a task during which all individuals jumped horizontally as far as possible with one leg in a controlled fashion. The main objective of the analysis is to test for significant differences in knee motion between the three groups, and thereafter identify in which part of the hopping task the groups differ. For this purpose, a B-spline-based ITP is applied to the one-way ANOVA model, and the time intervals presenting significant differences between the three groups are detected. Furthermore, since pairwise group comparisons are performed, the specific groups showing dissimilar patterns are identified, showing how individuals that were treated with physiotherapy present significantly different movement patterns with respect to the other two groups both during take-off and during landing.

Chapter 5 presents the analysis of the same data using a more sophisticated model, in which we consider additional individual-specific covariates to describe movements. In the previous chapter, we focused on differences between groups, showing that the individuals treated with physiotherapy present different movements with respect to the other two groups. In this chapter, we want

to investigate whether this result is solely due to the treatment method (i.e., physiotherapy), or can it be explained by mans of other variable characterizing the individuals (e.g., gender, BMI, age). An ANCOVA model is considered, in which functional data (the flexion of the knee during the jump) are modeled by means of a linear model with fixed scalar covariates and time-varying fixed regression coefficients, in the framework described in Chapter 3. ITP-based tests on the regression coefficients of the model are performed, and intervals presenting significant effects selected. The results of the analysis show that, even after having discounted for the jump length (the only covariate that turns out to be significant) the movement patter of physiotherapy group is significantly different with respect to the other two groups during take-off and landing.

Chapter 6 presents an application of ITP for the remote monitoring of laser welding. A functional two-way ANOVA is performed to investigate the effects of the gap between the welded plates and the location of the laser beam on the laser emission spectra. The effects of both the gap and the location are estimated throughout all recorded wavelength range, and wavelength bands presenting significant effects are inferentially selected. The result of such technique is the selection of the bands of wavelengths that can be used to remotely monitor the gap between the plates during the welding process. Indeed, we select a band in the thermal emission domain in which the gap effect is significant and the location one is not. This suggests the use of emission data on this band to monitor the gap between the plates during the welding process at any possible location, as in this band the emission is significantly influenced by the gap and not by the location.

In Chapter 7, ITP is applied on daily temperatures and irradiation data, for a renewable energy application. Mean and variance functions of daily irradiation and temperature in a close area in southern Spain are tested by means of a Fourier-based ITP. As a result, significant frequencies in the description of mean and variance functions of data are selected, and time-varying estimates of the mean and variance are calculated as a Fourier expansion of the sample mean coefficients, restricted to the frequencies selected by the test. Finally, a time-varying beta model is defined from the mean and variance estimates. This model is then used to calculate some relevant quantities for the installation of a residential photovoltaic plant. The model is used to evaluate the Expected Energy Not Supplied, being a relevant parameter in photovoltaic applications. A comparison with the approach mostly used in literature (based on a constant distribution in time) shows that the use of a time-varying distribution gives a more precise result, which allows avoiding overestimation of the size of the panels.

Software

Part III reports some details about the implementation of the ITP. Implementations of the ITP with B-spline and Fourier basis, for both the two-population and the one-population cases, have been made freely available in the R-package `fdatest` (Pini and Vantini, 2014), downloadable from CRAN. The package is briefly described in Chapter 8, and its `R` documentation, giving a detailed description of each function together with some examples, is reported.

Part I

Methods

Chapter 1

Global Inference on Functional Data: Hotelling's T^2 in Functional Hilbert Spaces

Abstract

The field of statistics is at the cusp of a revolution in the way data is collected by measuring instruments. Massive information is retrieved in real-time and/or spatially-referenced, hence producing new kind of data: functional data. Statistical inference for functional data is particularly challenging as it is an extreme case of high-dimensional data for which, no matter how large the sample is, information will always be insufficient to fully characterize the underlying model. In detail, after a historical excursus over the test statistics introduced for approaching the problem of testing the mean, we provide a generalization of Hotelling's T^2 on any functional Hilbert space, naturally dubbed functional Hotelling's T^2. We discuss a nonparametric permutational framework that enables statistical testing for the mean function of a population as well as for the difference between the mean functions of two populations. Within this framework, we show how a number of state-of-the-art test statistics can be seen as approximations of functional T^2 statistic hereby proposed.

Keywords: Hotelling's T square, Functional Data, Inference, Permutation Test

1.1 State of the art

The tremendously fast technological developments pertaining to measuring instruments have brought the field of Statistics at the cusp of a revolution, with real-time and/or spatially-referenced continuous information as the elementary datum to be analyzed. Various constraints (time, economical or ethical issues) on the other hand often prevent data analysts from collecting large samples. This brings the statistician out of his comforting zone where enough information is available to fully characterize all the variables under study and urges the demand for new inferential procedures that make the most out of the available

information to provide the best possible inference. Traditionally, the number of variables under study is referred to as the *dimension* of the problem and often denoted p, while the number of observations of these variables is referred to as the *sample size* and often denoted n. Hence, traditional samples with more observations than variables are termed small p large n data while modern samples with more variables than observations are termed large p small n data. Functional data is an extreme case of large p small n data with $p \to \infty$. In this paper, we propose a chronological overview and evolution of the statistical approach to the inference for the mean from the early works of De Moivre and Gauss back at the beginning of the XX century to the most recent advances. We will show how this evolution is tightly related to the sample characteristics and we will address this specific problem for functional data (extreme case of high-dimensional setting) by introducing a new test statistic.

z-test. In the XIX century, the German mathematician and astronomer Carl Friedrich Gauss, while trying to measure distances between stars, realized that he could not obtain perfectly reproducible measurements (Gauss, 1809). Rather, his measurements were clustered around a central value, with more frequently close to this value and less frequently further away. He named this distribution of measurements the Normal distribution, also named after his name nowadays. As a matter of fact, this distribution was introduced 60 years before by the French mathematician Abraham de Moivre in the privately circulated pamphlet "Approximatio ad summam terminorum binomii $(a + b)^n$ in seriem expansi" (De Moivre, 1733) in response to the Bernoulli brothers' paper 23 years earlier where he derived a simple approximation to the Bernoulli distribution. In this work, de Moivre unveils the mathematical expression of the Normal distribution curve, well known as the "Bell curve". French mathematician and astronomer Pierre-Simon Laplace further formalized the introduction of the Normal distribution in the "Théorie analytique des probabilités" (Laplace, 1820).

Almost a century later, the English statistician and geneticist Sir Ronald Aylmer Fisher publishes "Statistical methods for research workers" (Fisher, 1925*b*), in which he formalizes the use of the Normal distribution for statistical inference using elementary one-dimensional data. Let $(x_1, \ldots, x_n)$ be a sample of n independent measurements following the Normal distribution with mean μ and standard deviation (SD) σ. Fisher interprets the area under de Moivre's curve as a measure of probability. Hence, if σ is known and the hypothesis $\mu = \mu_0$ is formulated, he defines the so-called z-score $z_0 = \sigma^{-1}(\overline{x} - \mu_0)/\sqrt{n}$, where $\overline{x}$ is the sample mean and shows that z_0 follows a centered Normal distribution with unit standard deviation under the null hypothesis. Subsequently, he argues that the farther away from 0 the z-score z_0, the more evidence there is against the hypothesis $\mu = \mu_0$ since it implies that the occurrence of such a z-score was very unlikely under this assumption. This is the basis of the z-test, which enables for the first time to make inference for the mean of one-dimensional data.

However, most of one-dimensional data are not normally distributed and the above theory relies on the cornerstone that z_0 follows a centered Normal distribution with unit SD. The validity of this assumption is somehow guaranteed

by the Central Limit Theorem (CLT). Hence, most of the inferential procedures proposed in the early 1900s pertain to large samples. We refer to this period as the $1 = p < n = +\infty$ age of Statistics (see Figure 1.1).

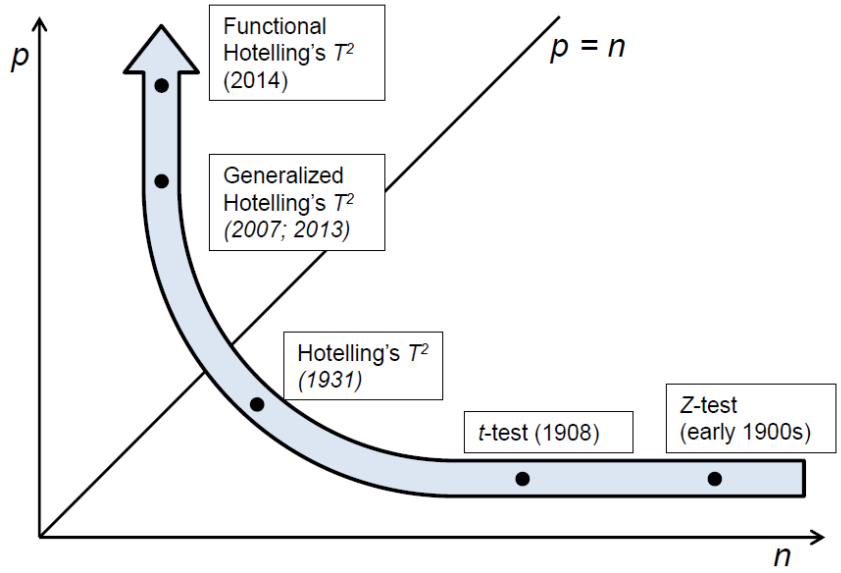


Figure 1.1: Timeline of the principal results on statistical inference for the mean.

t-test. Eventually, in many fields of applied statistics, it turned out that large samples were not feasible, mainly for time, economical or ethical reasons. This became problematic for applying the z-test for two reasons: (i) the SD of the measurement distribution is never known in practical situations but large sample theory provides an unbiased estimator of it, which is not accurate for small finite sample sizes and (ii) the measurement distribution is rarely Normal but large sample theory ensures that the sample mean is Normal (CLT), which is not guaranteed anymore in the small finite sample size setting.

The English statistician William Sealy Gosset was the first scientist to acknowledge this fact. At that time, as reported in Box (1981, 1987), he was working on a study about breweries and had only a few observations for making inference. Hence, he circumvented this issue by introducing the t-distribution under the pen-name *Student* in his work "The Probable Error of a Mean" (Gosset, 1908). He accounted for the variability of the sample standard deviation in the z-statistic, which becomes non-negligible at low sample sizes. To avoid confusion, he labels it as the t-statistic and characterizes its distribution under the assumption of normality of the data. Fisher further studies the t-distribution in "Applications of 'Student's' distribution" (Fisher, 1925*a*).

Jointly working together, Gosset and Fisher thus introduced the Student's t-distribution and formulated the corresponding t-test, which enables inference for the mean of one-dimensional Normal data using small samples. We refer to this period as the $1 = p < n < +\infty$ age of Statistics (see Figure 1.1).

Hotelling's T^2 test. A few years later, a growing interest arose in study-

ing multiple features (variables) associated to the same underlying statistical unit (observation). A simple example of this can be formulated as the following question: what are the averaged height and weight of the US population? One can obviously treat the two questions separately but would not account for the obvious correlation between the two variables by doing so. In other words, the scientific community was in need of inferential procedures for jointly distributed multi-dimensional data. Building on Indian statistician Prasanta Chandra Mahalanobis's work "Analysis of Race Mixture in Bengal" (Mahalanobis, 1927) where the distance named after him is introduced, American statistician Harold Hotelling introduces the T^2-statistic as a multivariate generalization of the t-statistic in "The generalization of Student's ratio" (Hotelling, 1931). In essence, the T^2-statistic is the Mahalanobis distance between the multivariate sample mean and a multivariate hypothesized mean. Hotelling derives the statistical distribution of the T^2-statistic under the assumption of multivariate normality with dimension p smaller than the sample size n, which provided the scientific community with adequate inferential procedures for simultaneously testing for the mean of multiple features.

Hotelling thus introduced the T^2-statistic, which follows a Fisher distribution under the assumption of multivariate normality with $p < n$. We refer to this period as the $1 < p < n < +\infty$ age of Statistics (see Figure 1.1).

High-dimensional tests. At the end of the XX Century, probably one of the most dramatic changes of paradigm in the history of modern statistics occurred. So far, due to technological limitations, it was a luxury to be able to measure multiple features at the same time (and so p easily remained smaller than n). The major breakthroughs that measuring instruments underwent during the second half of the XX century yielded data with more features than observations (and thus it became usual that p exceeds n at least by an order of magnitude). In other words, statistical research translated from a world with enough information to fully characterize the features of interest ($p < n$) to a world with insufficient information to do so ($p \geq n$). DNA micro-arrays for gene expression are one of the most famous examples of such data. They are characterized by thousands of variables being evaluated on only a few replicates.

Due to the increasing number of such large p small n data, many efforts have been made to extend Hotelling's result to the $p > n$ case for enabling inference for the mean of multi-dimensional data which dimension exceeds the sample size. The work of Srivastava (2007) is pioneering in this direction. He proposes a generalized T^2-statistic and shows that it follows a Fisher distribution for each n and p, with $n < p < +\infty$, under the assumption of multivariate normality and of proportionality of the variance-covariance matrix to the identity (which implies the independence among components). In Secchi et al. (2013), a generalized T^2-statistic is presented in a less stringent framework, i.e., without relying on the assumption of independence among components (even though still requiring multivariate normality). Under some conditions on the trace of the variance-covariance matrix, they show that it follows a χ^2 distribution with $n-1$ degrees of freedom in the $p \to \infty$ regime. We refer to this period as the $1 < n < p < +\infty$ age of Statistics (see Figure 1.1).

Functional tests. Some areas of applied statistics are interested in a particular kind of data: they aim at making inference for a single variable acquired in a continuous fashion by cutting-edge measuring instruments. The occurrence of such functional data is growing rapidly in these areas and raises the demand for appropriate inferential tools. Functional data analysis (FDA) has been one of the focuses of statisticians in the XXI century (Ramsay and Silverman, 2002, 2005; Ferraty and Vieu, 2006). The curve describing the continuous variable can be viewed as an infinity of points, or variables, and is thus the obvious extreme case of large p small n data. In addition, each "variable" describing a given point on the observed curves cannot be assumed independent from the other points on the same curves. We are thus entering in a new age of Statistics at the antipodes with respect to the beginning of the XX century that we shall refer to as the $1 < n < p = +\infty$ age of Statistics.

A commonality between the different inferential procedures provided during the last two centuries is the normality assumption of the data. This yielded parametric tests that are particularly appealing because (i) they generally achieve great statistical power and (ii) they only require the computation of a single test statistic, which is computationally easy and the comparison with tabulated critical values. In contrast, nonparametric approaches to the problem of inference, such as permutation tests, also introduced during the XX century (Fisher, 1936), were not widely used because available technologies back in these days could not cope with the high computational burden that these procedures generated.

This was not really a concern during the $1 = p < n = +\infty$ age. Indeed, after Russian mathematician Aleksandr Mikhailovich Lyapunov proved the CLT under very wide assumptions in the "Nouvelle forme du théoreme sur la limite de probabilité" (Lyapunov, 1901), the z-test could be easily applied to non-normal data. During the $1 = p < n < +\infty$ age, even though asymptotic normality of the sample mean was not sufficient anymore to ensure that the t-statistic follows the t-distribution, inferential procedure for testing the assumption of normality of one-dimensional data already existed and was thus not a debated point.

Debates really started with the $1 < p < n < +\infty$ age. Indeed, Hotelling's T^2 test strongly relies on the assumption of multivariate normality, which can be assessed in the bivariate case but becomes more and more challenging to assess as the dimension p increases. This is known as the "curse of dimensionality" (Hastie et al., 2009). These concerns grew even more during the $1 < n < p < +\infty$ age as most statistical procedures proposed for their analysis, such as the tests proposed in Srivastava (2007); Secchi et al. (2013), also strongly rely on the assumption of multivariate normality and, in addition, have been shown not to be robust with respect to violation of this assumption (Secchi et al., 2013). Similar concerns remain now that we enter the $1 < n < p = +\infty$ age with FDA.

Consequently, in this work, following the approach pioneered by Fisher (Fisher, 1936), we propose a nonparametric permutational framework for the inference on the mean of functional data. This framework does not rely on either multivariate normality or pre-specified variance-covariance structures. As such, it offers an appealing alternative to parametric procedures, the validity of which remains unclear in the new settings we find ourselves into.

In detail, in the present work, we propose a L^2 generalization of the Hotelling's T^2 statistic. We refer to it as functional Hotelling's T^2. We define the statistic and discuss its properties in Section 1.2. In Subsection 1.2.3 we discuss how to compute the functional Hotelling's T^2, and show how its finite-dimensional approximation is related with the multivariate large p small n generalization of T^2 provided in Secchi et al. (2013). In Section 1.3 we discuss a possible application of the functional Hotelling's T^2 statistic to the problem of inference for the mean in FDA, by means of nonparametric permutation tests. In Section 1.4 we compare it with other L^2-based test statistics presented in literature to test for functional data. Finally, in Section 1.5 we extend functional Hotelling's T^2 to any functional Hilbert space. All proofs are reported in the Appendix.

1.2 Hotelling's T^2 in L^2

1.2.1 Theoretical framework

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on the space $L^2(T)$ of all real-valued squared-integrable functions on the domain T (where T is an interval of $\mathbb{R}$ of the form (a, b)). The space $L^2(T)$, endowed with its natural inner product $(\xi_1, \xi_2) = \int_T \xi_1(t)\xi_2(t)dt$ for any $\xi_1, \xi_2 \in L^2(T)$, and associated norm $\|\xi\|_{L^2} = \sqrt{\int_T \xi^2(t)dt}$ (for any $\xi \in L^2(T)$), is a Hilbert space. Let $\mathbb{E}$ denote the integration with respect to the probability measure $\mathbb{P}$. The elementary datum in functional data analysis (FDA) is a random function of which we shall give a proper mathematical definition. Following (Tarpey, 2003), we state:

Definition 1.1. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a random L^2-function, or L^2-valued random variable, is a measurable surjective map from the sample space Ω to $L^2(T)$.

Note that, unlike Tarpey (2003), we here require the random function to be a surjective map from Ω to $L^2(T)$. This assumption means that the random function is nondegenerate, in the sense that it covers the whole $L^2(T)$ space.

Similarly to real-valued one-dimensional (resp., multi-dimensional) variables, for a given random L^2-function ξ, we can define the concepts of mean and dispersion around the mean. In traditional discrete cases, the former is a real one-dimensional (resp., multi-dimensional) vector and the latter is summarized by the variance (resp., variance-covariance matrix). The mean of a random L^2 function on the other hand is a function and its dispersion is characterized by a covariance operator. They are given by the following.

Definition 1.2. Let ξ be a random L^2-function. The mean function $\mu : Tto\mathbb{R}$ and covariance operator $V : L^2(T) \to L^2(T)$ of the random L^2-function ξ are: respectively given by:

$$\mu(t) = \mathbb{E}[\xi(t)], \text{ and } (Vf)(t) = \int_T \sigma(t, s)f(s)ds,$$

where $\sigma : T \times T \to \mathbb{R}$ is the covariance function of ξ:

$$\sigma(t, s) = \mathbb{E}\left[(\xi(t) - \mu(t))(\xi(s) - \mu(s))\right], \text{ for any } (t, s) \in T \times T\,.$$

In the current work, similarly to (Horváth and Kokoszka, 2012; Bosq, 2000), we restrict ourselves to random L^2-functions with finite total variance, i.e., such that:

$$\mathbb{E}\left[\|\xi\|_{L^2}^2\right] < +\infty \text{ (finite total variance).}$$

This covers a huge number of practical situations and confers convenient properties to the covariance operator such as the *spectral theorem decomposition*. Indeed, the covariance function σ can then be shown to belong to $L^2(T \times T)$. As a result, the covariance operator V is an Hilbert-Schmidt operator, i.e., it belongs to the subspace $\mathrm{HS}(L^2(T))$ of the space of limited linear operators $\mathcal{L}(L^2(T))$ on $L^2(T)$ (Arveson, 2002).

At this point and for the rest of the paper, we will assume that we collected a random sample of n independent and identically distributed (iid) random L^2-functions $\xi_1, \ldots, \xi_n$, with common mean function μ and covariance operator V, satisfying the finite total variance assumption.

Unbiased estimators for μ and V are given by the following

Definition 1.3. The sample mean function $\overline{\xi} : T \to \mathbb{R}$ is an unbiased estimator of the mean function μ and is given by:

$$\overline{\xi}(t) = \frac{1}{n}\sum_{i=1}^{n} \xi_i(t)\,.$$

The sample covariance operator $\widehat{V} : L^2(T) \to L^2(T)$ is an unbiased estimator of the covariance operator V and is given by:

$$\left(\widehat{V} f\right)(t) = \int_T \mathcal{S}(t,s) f(s) \mathrm{d}s,$$

where $\mathcal{S}$ is the sample covariance function defined as:

$$\mathcal{S}(t,s) = \frac{1}{n-1}\sum_{i=1}^{n} (\xi_i(t) - \overline{\xi}(t))(\xi_i(s) - \overline{\xi}(s)) \text{ for any } (t,s) \in T \times T\,.$$

The proof of unbiasness of these random variables as estimators of the mean function and covariance operator respectively is straightforwardly obtained by replicating the proof of unbiasness of their multivariate counterparts. Note that $\overline{\xi}$ is a random L^2-function and $\widehat{V}$ is a random $\mathrm{HS}(L^2)$-operator.

1.2.2 Definition of Hotelling's T^2 in L^2

Similarly to the multivariate case, it is possible to break down the total variance in the original functional dataset into two components, one of which only depends on the data. The following theorem states such a decomposition of variance and introduces some useful operators.

Theorem 1.1. *Consider a sample of n iid random functions with mean μ, covariance operator V s.t. $\mathbb{E}[\|\xi_i\|_{L^2}^2] < +\infty$. Then, the following variance decomposition holds:*

$$(n-1)\hat{V} + n\bar{V} = \tilde{V},$$

or, equivalently, $\forall g \in L^2(T)$:

$$(n-1)(g, \hat{V}g) + n(g, \bar{V}g) = (g, \tilde{V}g),$$

where:

- *$\hat{V}$ is the sample covariance operator, with kernel $\mathcal{S}$, that describes the dispersion of data ξ_i around the sample mean $\bar{\xi}$;*
- *$\bar{V}$ is the random operator with kernel $\big(\bar{\xi}(t) - \mu(t)\big)\big(\bar{\xi}(s) - \mu(s)\big)$, that describes the distance between the sample mean $\bar{\xi}$ and the mean μ;*
- *$\tilde{V}$ is the random operator with kernel $\sum_{i=1}^{n} \big(\xi_i(t) - \mu(t)\big)\big(\xi_i(s) - \mu(s)\big)$, that describes the dispersion of data ξ_i around the mean μ.*

The random operators $\overline{V}$ and $\widehat{V}$ introduced in Theorem 1.1 are the key concepts for generalizing Hotelling's T^2 statistic to the functional case. The following definition formally introduces this statistic:

Definition 1.4. The functional Hotelling's T^2-statistic is defined as the L^2 distance between the sample mean function and the true mean function "standardized" to the sample covariance operator. Similarly to the multivariate case, it reads:

$$T^2 = n \max_{g \in \mathrm{Im}(\widehat{V})} \frac{(g, \overline{V}g)}{(g, \widehat{V}g)}. \tag{1.1}$$

The functional T^2-statistic has a number of desirable properties that makes it particularly appealing for inferential purposes:

T^2 is a semi-distance between μ and $\overline{\xi}$. It is important to keep in mind that, although the formulation of the functional T^2-statistic proposed in Definition 1.4 is closely related to the multivariate T^2-statistic that one can find in many textbooks on introduction to multivariate analysis, the two statistics fundamentally differs in their mathematical implications. The multivariate T^2-statistic is defined as the maximum of the squared t-statistics associated to all possible one-dimensional projections of the multi-dimensional data. Differently, the functional T^2-statistic is defined as the maximum over the space $\mathrm{Im}(\widehat{V})$ spanned by the sample covariance operator $\widehat{V}$, which is an $(n-1)$-dimensional random subspace of $L^2(T)$. As a result, T^2 is a distance between μ and $\overline{\xi}$ in the random space $\mathrm{Im}(\widehat{V})$ but is only a semi-distance in $L^2(T)$, for which the identity of indiscernibles does not hold.

T^2 is invariant under similarity transformations. Functional Hotelling's T^2-statistic is invariant under similarity transformations of the data, i.e., under affine transformations $\xi \mapsto aO\xi + \mathbf{f}$, where $a \in \mathbb{R}^+$, $\mathbf{f} \in L^2(T)$ and O is an orthogonal linear limited operator on $L^2(T)$, i.e., O satisfies $(Og_1, Og_2)_{L^2} = (g_1, g_2)_{L^2}$ for any $g_1, g_2 \in L^2(T)$. Lehmann and Romano (2006) have shown that this type of invariance is the largest family of invariance transformations that one can achieve in the framework $p \geq n$. In this sense, the functional T^2-statistic is invariant-optimal.

T^2 "marginal" distributions under functional normality are known. The notion of *functional normality* has been introduced in Tarpey (2003) and stipulates that a random L^2-function is normally distributed if and only if, for all $u \in L^2(T)$, the real-valued one-dimensional random variable (ξ, u) is normally distributed. If we further assume functional normality of our dataset, Theorem

1.1 combined with Cochran's Theorem yields the following (see Proposition 1.1 in the Appendix):

$$n\frac{(g,\overline{V}g)}{(g,\widehat{V}g)} \sim F(1,n-1)\,. \tag{1.2}$$

Equation 1.2 provides the distribution of the ratios involved in the T^2 statistic. However the distribution of its maximum over all functions of $\mathrm{Im}(\widehat{V})$ is not easy to elicit without introducing very strong assumptions on the covariance operator V. In addition, functional normality may be too stringent for many applications and hard to defend and/or prove. For all these reasons, we will tackle the problem of inference for the mean function within a nonparametric permutational framework, based on minimal distributional assumption.

1.2.3 A finite-dimensional approximation of Hotelling's T^2 in L^2

With Definition 1.4, we gave a formal definition of functional Hotelling's T^2 statistic. However, expressed as a maximization problem, T^2 is of little practical interest here. Indeed, permutation tests rely on the evaluation of a sufficient statistic over an enormous number of permuted datasets, which might become computationally too heavy if, for each evalution, a maximization problem has to be solved. Furthermore, in practical scenarios, analytic expressions of the observed functions ξ_i's are often not provided. Rather, finite high-dimensional approximations are available.

Let us consider a countable set of basis functions $\{\phi_k\}_{k\geq 1}$ of $L^2(T)$. It is possible to project the original n observed functions onto the first p elements of such a basis. Let $\boldsymbol{\xi}_i = ((\phi_1,\xi_i),\dots,(\phi_p,\xi_i))$ be the vector of the scores of the *i-th* observed function ξ_i projected onto the first p elements of the basis. Then, we can define the p-dimensional random vector $\overline{\boldsymbol{\xi}}$ as the sample mean of the individual scores and the $p\times p$ matrix S as their sample variance-covariance matrix. Similarly, the mean function μ can be projected into a p-dimensional vector $\boldsymbol{\mu}$ of mean scores. At this point, the finite-dimensional approximation of functional Hotelling's T^2 can be computed directly without solving any maximization problem, as shown by the following.

Theorem 1.2. *Consider a sample of n iid random functions with mean μ, covariance operator V s.t. $\mathbb{E}[\|\xi_i\|^2_{L^2}] < +\infty$. Let $\{\phi_k\}_{k\geq 1}$ be a countable set of basis functions of $L^2(T)$. Then, for any $p \geq 1$, the following identity holds:*

$$T^2_p = n \max_{g\in \mathrm{Im}(\widehat{V})\cap\{\phi_1,\dots,\phi_p\}} \frac{(g,\overline{V}g)}{(g,\widehat{V}g)} = n(\overline{\boldsymbol{\xi}}-\boldsymbol{\mu})^\top W^{1/2} S^{+} W^{1/2}(\overline{\boldsymbol{\xi}}-\boldsymbol{\mu})\,, \tag{1.3}$$

where $W \in \mathbb{R}^{p\times p}$ is the matrix of inner products between the basis functions $[W]_{i,j} = (\phi_i,\phi_j)$ and S^+ is the Moore-Penrose generalized inverse (Rao and Mitra, 1971) of the sample variance-covariance matrix S. In addition:

$$T^2_p \xrightarrow[p\to\infty]{a.s.} T^2\,.$$

Theorem 1.2 states that, if the basis used to project the data is orthonormal (i.e., $W = I$), if we limit the search for the maximum in the functional T^2 definition to those functions in $\mathrm{Im}(\widehat{V})$ that are spanned by the first p elements

of any basis of $L^2(T)$, then the resulting maximum can be formulated as a high-dimensional T^2 statistic as introduced in (Secchi et al., 2013). In the case of non-orthonormal basis, this finite-dimensional approximation is still related to the high-dimensional generalization provided in (Secchi et al., 2013), but the generalized inverse of the covariance matrix is rescaled, by considering the inner products between the basis functions. In addition, as $p \to \infty$, the sequence of such statistics converges almost surely to the functional T^2 statistic.

Note that, with the basis of principal components of $\widehat{V}$, we have the equality $T^2 = T^2_{n-1}$, i.e., the functional Hotelling's T^2 can be exactly evaluated by means of the first $n-1$ sample principal components.

1.3 Permutation test L^2 based on Hotelling's T^2

The problem of inference for functional data has been addressed in the literature from both a parametric and a nonparametric perspective. The former approach commonly relies on distributional assumptions on functional data and on asymptotic results (Horváth and Kokoszka, 2012; Spitzner et al., 2003; Cuevas et al., 2004; Fan and Lin, 1998; Schott, 2007). The latter approach relies instead on permutation or bootstrap techniques, which are computationally intensive (Hall and Tajvidi, 2002; Cardot et al., 2007; Cuesta-Albertos and Febrero-Bande, 2010; Pini and Vantini, 2013; Hall and Van Keilegom, 2007*a*). The method that we propose for testing functional data relies on this latter approach.

in detail, we now show how functional Hotelling's T^2 can be used in nonparametric permutation procedures for making inference on the mean of a random L^2 function (Section 1.3.1) and on the difference between the means of two random L^2 functions (Section 1.3.2).

1.3.1 One-population test

Let $(\xi_1, \dots, \xi_n)$ be n i.i.d. random L^2-functions with mean function μ and covariance operator V that satisfy the finite total variance assumption ($\mathbb{E}[\|\xi_i\|^2_{L^2}] < +\infty$, for all $i \in \{1, n\}$).

Assuming that we want to test the following null hypothesis on the mean function:

$$H_0 : \mu = \mu_0, \text{ vs. } H_1 : \mu \neq \mu_0\,, \text{ with } \mu_0 \in L^2(T)\,, \tag{1.4}$$

one can compute, under the null hypothesis H_0, the functional T^2 statistic (Definition 1.4):

$$T^2_0 = n \max_{g \in \mathrm{Im}(\widehat{V})} \frac{(g, \overline{V_0} g)}{(g, \widehat{V} g)}\,, \tag{1.5}$$

where $\overline{V_0}$ is the random operator with kernel $\overline{\sigma_0}(t,s) = (\overline{\xi}(t) - \mu_0(t))(\overline{\xi}(s) - \mu_0(s))$ for any $t, s \in T$ and $\widehat{V}$ is the sample covariance operator with kernel $\mathcal{S}$.

One can use the T^2_0 statistic in a permutational framework for testing the null hypothesis H_0. Instead of the normality assumption often required in this framework (see for instance Horváth and Kokoszka 2012), we make in a permutation framework the much weaker assumption of symmetry of the distribution of the data around the mean. Then, a permutation test can be constructed by evaluating the test statistic (1.5) over all possible reflections of

data with respect to the center of symmetry under H_0, i.e., the transformations $\xi_i(t) \mapsto \xi_i^*(t) = \mu_0(t) + (-1)^{c_i}(\xi_i(t) - \mu_0(t))$, with $i = 1, ..., n$, and $c_i \in \{0, 1\}$. The p-value of test (1.4) is the proportion of permuted $T_0^2(\xi_1^*, \xi_2^*, ..., \xi_n^*)$ exceeding the value $T_0^2(\xi_1, \xi_2, ..., \xi_n)$ evaluated on the original data set.

1.3.2 Two-population test

Let $(\xi_{11}, \ldots, \xi_{n_1 1})$ and $(\xi_{12}, \ldots, \xi_{n_2 2})$ be two independent samples of size n_1 and n_2 respectively. Let $(\xi_{11}, \ldots, \xi_{n_1 1})$ be i.i.d. random L^2-functions with mean function μ_1 and covariance operator V and let $(\xi_{12}, \ldots, \xi_{n_2 2})$ be i.i.d. random L^2-functions with mean function μ_2 and covariance operator V. In addition, we assume that the assumption of finite total variance is met in the two samples.

Assuming that we want to test the following null hypothesis:

$$H_0 : \mu_1 = \mu_2, \text{ vs. } H_1 : \mu_1 \neq \mu_2 \,, \tag{1.6}$$

one can compute, under H_0, the functional T^2 statistic (Definition 1.4):

$$T_0^2 = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1} \max_{g \in \mathrm{Im}(\widehat{V_{\text{pooled}}})} \frac{(g, \overline{V_0} g)}{(g, \widehat{V_{\text{pooled}}} g)} \,, \tag{1.7}$$

where $\overline{V_0}$ is the random operator with kernel

$$\overline{\sigma_0}(t, s) = \left[\overline{\xi_1}(t) - \overline{\xi_2}(t)\right]\left[\left[\overline{\xi_1}(s) - \overline{\xi_2}(s)\right]\right], \forall t, s \in T$$

with $\overline{\xi_1}$ and $\overline{\xi_2}$ being the sample mean functions of the first and the second populations respectively and $\widehat{V_{\text{pooled}}}$ is the pooled sample covariance operator with pooled covariance function $\mathcal{S}_{\text{pooled}}$ defined as:

$$\begin{aligned}\mathcal{S}_{\text{pooled}}(t, s) = \frac{1}{n_1 + n_2 - 2} \Bigg[&\sum_{i=1}^{n_1} \left(\xi_{i1}(t) - \overline{\xi_1}(t)\right)\left(\xi_{i1}(s) - \overline{\xi_1}(s)\right) \\ &+ \sum_{i=1}^{n_2} \left(\xi_{i2}(t) - \overline{\xi_2}(t)\right)\left(\xi_{i2}(s) - \overline{\xi_2}(s)\right) \Bigg], \forall t, s \in T \,.\end{aligned}$$

A permutation test can be constructed by evaluating the test statistic (1.7) over all permutations of data over the sample units $(\xi_{11}, ..., \xi_{n_1 1}, \xi_{12}, ..., \xi_{n_2 2}) \mapsto (\xi_{11}^*, ..., \xi_{n_1 1}^*, \xi_{12}^*, ..., \xi_{n_2 2}^*)$. The p-value of the corresponding test is then the proportion of $T_0^2(\xi_{11}^*, ..., \xi_{n_1 1}^*, \xi_{12}^*, ..., \xi_{n_2 2}^*)$ exceeding $T_{f,0}^2(\xi_{11}, ..., \xi_{n_1 1}, \xi_{12}, ..., \xi_{n_2 2})$ evaluated on the original data set.

1.4 Other L^2-based test statistics

To perform a permutation test on the mean of one functional population (or two functional populations), we only need to define a distance or semi-distance between the sample mean function (or difference between the sample mean functions) and the mean function under the null hypothesis H_0 (or difference between the two sample mean functions). In the literature of permutation testing, the following distances have been proposed for random L^2 functions:

The L^2 distance.

$$\Delta_{L^2}^2 = \int_T \left(\overline{\xi}(t) - \mu(t)\right)^2 dt\,.$$

This test statistic and associated permutation test have been proposed in Hall and Tajvidi (2002); Hall and Van Keilegom (2007*a*). It is also possible to derive parametric or asymptotic tests based on the same statistic under the assumption of functional normality (see for instance Horváth and Kokoszka 2012). The statistic Δ_{L^2} can be expressed as the norm of an appropriate operator in L^2 as:

$$\Delta_{L^2}^2 = n \max_{g \in L^2(T)} \frac{(g, \overline{V}g)}{(g, g)}\,.$$

Hence, $\Delta_{L^2}^2$ can be seen as an approximation of the functional Hotelling's T^2, where the sample covariance operator $\widehat{V}$ is assumed to be the identity operator. Note that this statistic neither accounts for the point-wise variance of the data nor its covariance structure. It instead gives equal weight to equally-long intervals of the domain T.

The standardized L^2 distance. (i.e., the L^2 distance between standardized data)

$$\Delta_{L_t^2}^2 = \int_T \frac{\left(\overline{\xi}(t) - \mu(t)\right)^2}{\mathcal{S}(t,t)} dt\,,$$

where $\mathcal{S}(t,t)$ is the point-wise sample variance. This test statistic has been introduced in Hall and Tajvidi (2002) and can be seen as a weighted version of the L^2 statistic. Similarly to the L^2 statistic, the statistic $\Delta_{L_t^2}^2$ can be expressed as the norm of an appropriate operator in L^2 as:

$$\Delta_{L_t^2}^2 = n \max_{g \in L^2(T)} \frac{(g, \overline{V}g)}{(g, D_\sigma g)}\,.$$

Hence, $\Delta_{L_t^2}^2$ can be seen as a more sophisticated approximation of the functional Hotelling's T^2 statistic. The sample covariance operator is indeed assumed to be "diagonal" and reads $(D_\sigma g)(t) = \mathcal{S}(t,t)g(t)$. The $\Delta_{L_t^2}^2$ statistic thus makes use of the point-wise estimates $\mathcal{S}(t,t)$ of the variance of the data but does not account for its auto-correlation structure $\mathcal{S}(t,s)$.

Note that, unlike the functional T^2 statistic, the $\Delta_{L^2}^2$ and $\Delta_{L_t^2}^2$ statistics are distances in $L^2(T)$ (and not semi-distances). On the other hand, they share no commonality with traditional test statistics used for null hypothesis statistical testing in multivariate analysis and they are not invariant under similarity transformations.

1.5 Hotelling's T^2 in functional Hilbert spaces

In the previous sections we presented the functional Hotelling's T^2 in the L^2 geometry as the natural extension of finite-dimensional Euclidean geometry to the space $L^2(T)$. Nevertheless, functional Hotelling's T^2 can be extended to every functional Hilbert space. Indeed, its definition only requires the evaluation

of mean function and covariance operator, which directly derive from the notion of inner product.

In particular, let H be a functional Hilbert space, endowed with the inner product $(\cdot,\cdot)_H$ and associated norm $\|\cdot\|_H$. Let $(\xi_1,\dots,\xi_n)$ be n i.i.d. H-valued random variables with mean $\mu \in H$ and covariance operator $V \in \mathcal{L}(H)$. A sample estimate of the mean in H is the Fréchet mean: $\bar{\xi} = \text{argmin}_{\xi\in H} \sum_{i=1}^n \|\xi_i - \xi\|_H^2$. Hence, functional Hotelling's T^2 can be defined in the space H as:

$$T^2 = n \max_{g\in \text{Im}(\hat{V})} \frac{(g,\bar{V}g)_H}{(g,\hat{V}g)_H}. \tag{1.8}$$

where:

- $\hat{V} \in \mathcal{L}(H)$ is the sample covariance operator in the space H, (defined according to the scalar product in H), describing the dispersion of data ξ_i around the Fréchet mean $\bar{\xi}$. Indeed, $\hat{V}$ is such that $(g,\hat{V}g)_H$ is the sample variance of the scores of the orthogonal projections of ξ_i on g, with respect to the inner product in H.
- $\bar{V} \in \mathcal{L}(H)$ is a random operator associated to the distance between the Fréchet mean $\bar{\xi}$ and the mean μ. Indeed, $\bar{V}$ is such that $(g,\bar{V}g)_H$ is the square distance between the scores of the orthogonal projections of $\bar{\xi}$ and μ over g, with respect to the inner product in H.

In the following, we report two concrete examples of geometry where we explicit the definition of these operators: (i) the Sobolev space $H^k(T)$ of k-differentiable squared-integrable real functions with squared-integrable derivatives (Section 1.5.1) and (ii) the Bayes linear space $B^2(T)$ of non-negative real functions on T with squared-integrable logarithm (Boogaart et al., 2014) (Section 1.5.2).

1.5.1 Example: Hotelling's T^2 in Sobolev spaces

Consider the Sobolev space $H^k(T)$, that is, the space of k-differentiable functions $g \in L^2(T)$ such that, for $j \le k$, $D^j g \in L^2(T)$ (where $D^j g$ denotes the j-*th* derivative of g). The space $H^k(T)$ is a Hilbert space, endowed with the following inner product:

$$(f,g)_{H^k} = \sum_{j=0}^k (D^j f, D^j g)_{L^2} = \sum_{j=0}^k \int_T (D^j f)(t)\cdot(D^j g)(t)dt\,. \tag{1.9}$$

Let $(\xi_1,\dots,\xi_n)$ be n i.i.d. $H^k(T)$-valued random variables with mean $\mu \in H^k(T)$ defined as $\mu = \text{argmin}_{m\in H^k(T)} \mathbb{E}\left[\|\xi_i - m\|_{H^k}^2\right]$. The functional Hotelling's T^2 in $H^k(T)$ then reads:

$$T^2 = n \max_{g\in \text{Im}(\widehat{V})} \frac{(g,\overline{V}g)_{H^k}}{(g,\widehat{V}g)_{H^k}}\,, \tag{1.10}$$

where the operators $\widehat{V}$ and $\overline{V}$ can be explicitly defined using the inner product in $H^k(T)$ given by Eq.(1.9). In details,

- the operator $\hat{V} \in \mathcal{L}(H^k)$ is defined as:

$$\left(\hat{V}f\right)(t) = \int_T \sum_{j=0}^{k} \mathcal{S}_{0j}(t,s) D^j f(s) \mathrm{d}s,$$

where $\mathcal{S}_{lj}(t,s)$ is the sample covariance between lth and jth derivatives of data ξ_i: $\mathcal{S}_{lj}(t,s) = \frac{1}{n-1}\sum_{i=1}^{n}(D^l\xi_i(t) - D^l\bar{\xi}(t))(D^j\xi_i(s) - D^j\bar{\xi}(s))$, and $\bar{\xi}(t)$ is the Fréchet mean of the ξ_i:

$$\bar{\xi} = \underset{m \in H^k}{\operatorname{argmin}} \sum_{i=1}^{n} \parallel \xi_i - m \parallel^2_{H^k};$$

- the operator $\bar{V} \in \mathcal{L}(H^k)$ is defined as:

$$\left(\bar{V}f\right)(t) = \int_T \sum_{j=0}^{k} (\bar{\xi}(t) - \mu(t))(D^j\bar{\xi}(s) - D^j\mu(s)) D^j f(s) \mathrm{d}s.$$

To have a better insight into the interpretation of Hotelling's T^2 in the Sobolev space $H^k(T)$, we can rely on the following identities (Lemma 1.1 of the Appendix):

$$\begin{aligned}(g, \hat{V}g)_{H^k} &= \widehat{Var}\left[(g,\xi_1)_{H^k}, \ldots, (g,\xi_n)_{H^k}\right];\\ (g, \bar{V}g)_{H^k} &= ((g, \bar{\xi} - \mu)_{H^k})^2.\end{aligned}$$

These identities show that Hotelling's T^2 in $H^k(T)$ can be interpreted as the maximum over all elements in the image space of $\widehat{V}$ of the ratio between: (i) the squared distance between the scores of the orthogonal projections of $\bar{\xi}$ and μ on g, with respect to the inner product in $H^k(T)$; and (ii) the sample variance of the scores of the orthogonal projections of the ξ_i's on g, with respect to the inner product in $H^k(T)$.

1.5.2 Example: Hotelling's T^2 in the Bayes linear space

Another example of functional Hilbert space recently introduced in the Functional Data Analysis literature is the Bayes linear space $B^2(T)$, that is, the space of absolutely continuous density functions on the compact set T with squared-integrable logarithm. The interested reader can find detailed descriptions of Bayes spaces in Egozcue et al. (2006); Egozcue and Pawlowsky-Glahn (2006); Menafoglio et al. (2013); Boogaart et al. (2014); Hron et al. (2014). As shown by Egozcue et al. (2006), $B^2(T)$ is a functional Hilbert space when proper addition $\oplus$, scalar multiplication $\odot$ and inner product $(\cdot,\cdot)_{B^2}$ operations are defined. In details, for any $f, g \in B^2(T)$ and $\alpha \in \mathbb{R}$:

$$\begin{aligned}(f \oplus g)(t) = \frac{f(t)g(t)}{\int_T f(s)g(s)ds}, \qquad (\alpha \odot f)(t) = \frac{f(t)^\alpha}{\int_T f(s)^\alpha ds},\\ (f,g)_{B^2} = \frac{1}{2|T|}\iint_{T\times T} \ln\frac{f(t)}{f(s)} \ln\frac{g(t)}{g(s)} dtds,\end{aligned} \tag{1.11}$$

where $|T|$ is the measure of the compact set T.

An isometric isomorphism between $B^2(T)$ and $L^2(T)$ is defined by the centred log-ratio (clr) transformation (Boogaart et al., 2014; Menafoglio et al., 2013):

$$\mathrm{clr}(f)(t) = \ln f(t) - \frac{1}{|T|}\ln f(s)\mathrm{d}s. \tag{1.12}$$

Using both the Hilbert geometry conferred from the addition, scalar multiplication and inner product proposed by Egozcue et al. (2006) and the isomorphism in Eq.(1.12), we can provide a functional Hotelling's T^2 statistic in $B^2(T)$ useful for making inference on the mean of populations of density functions on a compact support.

Let $(\xi_1, \ldots, \xi_n)$ be n i.i.d. $B^2(T)$-valued random variables with mean $\mu \in B^2(T)$ defined as $\mu = \mathrm{argmin}_{m\in B^2(T)} \mathbb{E}\left[\|\xi_i - m\|^2_{B^2}\right]$. The functional Hotelling's T^2 in $H^k(T)$ then reads:

$$T^2 = n \max_{g\in \mathrm{Im}(\widehat{V})} \frac{(g, \overline{V}g)_{B^2}}{(g, \widehat{V}g)_{B^2}}, \tag{1.13}$$

where the operators $\widehat{V}$ and $\overline{V}$ can be explicitly defined using the inner product in $B^2(T)$ given by Eq.(1.11) and the isomorphism given by Eq.(1.12). In details,

- the sample covariance operator $\hat{V} \in \mathcal{L}(B^2)$ is defined as:

$$\left(\hat{V}f\right)(t) = \mathrm{clr}^{-1}\left(\int_T \mathcal{S}_c(t,s)\mathrm{clr}(f)(s)\mathrm{d}s\right),$$

  where clr^{-1} is the inverse centered log-ratio transformation, and $\mathcal{S}_c(t,s)$ is the sample covariance between clr-transformed data:
  $\mathcal{S}_c(t,s) = \frac{1}{n-1}\sum_{i=1}^n \left(\mathrm{clr}(\xi_i)(t) - \mathrm{clr}(\bar{\xi})(t)\right)\left(\mathrm{clr}(\xi_i)(s) - \mathrm{clr}(\bar{\xi})(s)\right)$.

- the operator $\bar{V} \in \mathcal{L}(B^2)$ is defined as:

$$\left(\bar{V}f\right)(t) = \mathrm{clr}^{-1}\Bigg(\int_T \left(\mathrm{clr}(\bar{\xi})(t) - \mathrm{clr}(\mu)(t)\right) \left(\mathrm{clr}(\bar{\xi})(s) - \mathrm{clr}(\mu)(s)\right)\mathrm{clr}(f)(s)\mathrm{d}s\Bigg),$$

  where $\bar{\xi}(t)$ is the Fréchet mean of the ξ_i:

$$\bar{\xi} = \underset{m\in B^2}{\mathrm{argmin}} \sum_{i=1}^n \| \xi_i - m \|^2_{B^2}.$$

Similarly to we did in Sobolev spaces, to have a better insight into the interpretation of Hotelling's T^2 in the Bayes space $B^2(T)$, we can rely on the following identities (Lemma 1.2 of the Appendix):

$$\begin{aligned}(g, \hat{V}g)_{B^2} &= \widehat{Var}[(g,\xi_1)_{B^2}, \ldots, (g,\xi_n)_{B^2}] \\ (g, \bar{V}g)_{B^2} &= ((g, \bar{\xi} - \mu)_{B^2})^2\end{aligned}$$

Hence, Hotelling's T^2 in $B^2(T)$ is the the maximum over all elements in the image space of $\widehat{V}$ of the ratio between: (i) the squared distance between the scores of the orthogonal projections of $\bar{\xi}$ and μ on g, with respect to the inner product in $B^2(T)$; and (ii) the sample variance of the scores of the orthogonal projections of the ξ_i's on g, with respect to the inner product in $B^2(T)$.

1.6 Conclusions

After a historical excursus on how the problem of inference for the mean evolved in the statistical research, from the early works of De Moivre and Gauss back at the beginning of the XX century to the most recent advances, we presented a generalization of Hotelling's T^2 (functional Hotelling's T^2) in functional Hilbert spaces, and demonstrated how it can be used for hypothesis testing for the mean of functional data within a permutational framework.

The functional Hotelling's T^2 is presented as a natural extension of Euclidean geometry to the functional L^2 space. It is a semi-distance based on a semi-metric in L^2. In essence, the functional T^2 statistic maximizes the ratio of an operator that assesses the distance between the sample mean of an i.i.d. functional dataset and its actual mean to another operator that assesses the variability of such a functional dataset around its sample mean. We presented a practical way of computing this statistic without resorting to optimization algorithms by projecting the dataset onto any basis of the image space of the sample covariance operator.

For inferential purposes, we set up a permutational framework for making inference on the mean (or difference between means) of functional data. We discussed the advantage of our proposed functional T^2 statistic, which, unlike all other statistics proposed in the literature, fully accounts for the covariance structure of the input data. Moreover, we showed that already existing test statistics recently presented in the literature are in fact approximations of functional Hotelling's T^2, where the variance and/or correlation of the data is ignored.

Finally, even though we presented functional Hotelling's T^2 in the L^2 geometry, as the natural functional extension of Euclidean geometry, we also showed how our functional T^2 statistic can be defined and used in virtually any Hilbert space. Examples included in this work are the Sobolev spaces $H^k(T)$ and the Bayes linear space $B^2(T)$.

An interesting and challenging future development of this work would be the extension of T^2 to the larger family of functional metric spaces (e.g., Banach spaces), following the direction of some lively and very recent areas of statistical research, such as object-oriented data analysis and shape analysis (see for instance Marron and Alonso 2014). This extension requires a definition of T^2 exclusively based on a metric that relies neither on the notion of inner product nor on the one of vector space.

Appendix

1.A Proofs

Proof. [Theorem 1.1] Note that, by their definition, these 3 operators have respectively 1, $n-1$ and n degrees of freedom. Moreover, we have that, $\forall \omega \in \Omega$, $\forall t, s \in T$:

$$\begin{aligned}
(n-1)\mathcal{S}(\omega)(t,s) &= \sum_{i=1}^{n} \big(\xi_i(\omega)(t) - \bar{\xi}(\omega)(t)\big)\big(\xi_i(\omega)(s) - \bar{\xi}(\omega)(s)\big) \\
&= \sum_{i=1}^{n} \Big[\big(\xi_i(\omega)(t) - \mu(t) + \mu(t) - \bar{\xi}(\omega)(t)\big) \\
&\times \big(\xi_i(\omega)(s) - \mu(s) + \mu(s) - \bar{\xi}(\omega)(s)\big)\Big] \\
&= \sum_{i=1}^{n} \big(\xi_i(\omega)(t) - \mu(t)\big)\big(\xi_i(\omega)(s) - \mu(s)\big) \\
&+ \sum_{i=1}^{n} \big(\bar{\xi}(\omega)(t) - \mu(t)\big)\big(\bar{\xi}(\omega)(s) - \mu(s)\big) \\
&- \sum_{i=1}^{n} \big(\xi_i(\omega)(t) - \mu(t)\big)\big(\bar{\xi}(\omega)(s) - \mu(s)\big) \\
&- \sum_{i=1}^{n} \big(\bar{\xi}(\omega)(t) - \mu(t)\big)\big(\xi_i(\omega)(s) - \mu(s)\big) \\
&= \sum_{i=1}^{n} \big(\xi_i(\omega)(t) - \mu(t)\big)\big(\xi_i(\omega)(s) - \mu(s)\big) \\
&- n\big(\bar{\xi}(\omega)(t) - \mu(t)\big)\big(\bar{\xi}(\omega)(s) - \mu(s)\big).
\end{aligned}$$

Hence, we have:

$$\begin{aligned}
(n-1)\mathcal{S}(\omega)(t,s) + n\big(\bar{\xi}(\omega)(t) - \mu(t)\big)\big(\bar{\xi}(\omega)(s) - \mu(s)\big) \\
= \sum_{i=1}^{n} \big(\xi_i(\omega)(t) - \mu(t)\big)\big(\xi_i(\omega)(s) - \mu(s)\big),
\end{aligned} \tag{1.14}$$

and the thesis follows. □

Proposition 1.1. *Consider a sample of n iid random functions $\xi_1, ..., \xi_n$ with mean μ, covariance operator V s.t. $\mathbb{E}[\|\xi_i\|_{L^2}^2] < +\infty$, and $\mathrm{Im}(V) = L^2(T)$. Let the random functions be normally distributed, i.e., $\forall u \in L^2(T)$, (ξ, u) is a real univariate gaussian random variable. Then, we have:*

$$n\frac{(g, \bar{V}g)}{(g, \hat{V}g)} \sim F(1, n-1).$$

Proof. Let $g \in \mathrm{Im}(\hat{V})$. Under the normality assumption we have:

- $(g, \tilde{V}g) \sim (g, Vg)\chi^2(n)$;
- $n(g, \bar{V}g) \sim (g, Vg)\chi^2(1)$.

Indeed, for the first one, we have:

$$(g, \tilde{V}g) = \sum_{i=1}^{n} \Big(\int_T \big(\xi_i(t) - \mu(t)\big) g(t) dt \Big)^2$$

We know that the random functions $\xi_i - \mu$, $i = 1, \ldots, n$, are independent and identically distributed as $\mathcal{N}_\infty(0, V)$ (Gaussian random function with mean 0 and covariance operator V). Thus, the random variables

$$\int_T \big(\xi_i(t) - \mu(t)\big) g(t) dt, \ i = 1, \ldots, n$$

are independent and identically distributed as $\mathcal{N}_1\big(0, (g, Vg)\big)$, thanks to the definition of gaussian random function. The thesis follows immediately by definition of the χ^2 distribution.

The second statistic can be written:

$$n(g, \bar{V}g) = \Big(\int_T \sqrt{n}\big(\bar{\xi}(t) - \mu(t)\big) g(t) dt \Big)^2$$

Similar arguments give the distribution of $N(g, \bar{V}g)$.

This result put us in the conditions to use Cochran's theorem (Johnson and Wichern, 2007). It leads then to

- $(n-1)(g, \hat{V}g) \sim (g, Vg)\chi^2(n-1)$;
- $n(g, \bar{V}g)$ and $(n-1)(g, \hat{V}g)$ are independent.

These 2 points carry with them the following consequence: given $\hat{V}$, $\forall g \in \mathrm{Im(V)} \cap \mathrm{Im}(\hat{\mathrm{V}})$, i.e., $\forall g \in \mathrm{Im}(\hat{\mathrm{V}})$,

$$n \frac{(g, \bar{V}g)}{(g, \hat{V}g)} \sim F(1, n-1) \tag{1.15}$$

Finally, we know that $\mathrm{Ker}(\hat{V})$ has null measure in $L^2(T)$. Hence, $\mathbb{P}[g \notin \mathrm{Im}(\hat{V})] = 0$. This last condition leads to the thesis. □

Proof. [Theorem 1.2] For the first part of the statement it is sufficient to note that T_p^2 is a monotonic increasing sequence which converges to the functional statistic T_f^2 defined in (1.1), as the basis $\{\phi_k\}_{k \geq 1}$ is dense in L^2.

Now, at p fixed, we aim at finding the expression of T_p^2. It requires first to write the decomposition of each function involved on the basis $\{e_k\}_{k \geq 1}$, and project them on the space generated by the first p basis components. We have:

$$g = \sum_{k=1}^{p} g_k \phi_k \qquad \xi_{i,p} = \sum_{k=1}^{p} \xi_{ik} \phi_k \qquad \bar{\xi}_p = \sum_{k=1}^{p} \bar{\xi}_k \phi_k \qquad \mu_p = \sum_{k=1}^{p} \mu_k \phi_k.$$

Note that we are now working with finite-dimensional approximations $\xi_{i,p}, \bar{\xi}_p, \mu_p$ of the functions $\xi_i, \bar{\xi}, \mu$, and that all approximations converge to the respective function for $p \to \infty$.

Now, the projection of the quantity $(g, \bar{V}g)$ in the p-dimensional space generated by the first p ϕ_k can be written as:

$$(g, \bar{V}g)_p = \Big(\int_T \big(\bar{\xi}_p(t) - \mu_p(t)\big) g(t) dt \Big)^2 = \Big(\int_T \sum_{k=1}^{p} \sum_{l=1}^{p} (\bar{\xi}_k - \mu_k) g_l \phi_k(t) \phi_l(t) dt \Big)^2$$

At this point, note that, by definition:

$$\int_T \phi_k(t) \phi_l(t) dt = W_{kl}$$

Thus, we obtain:

$$(g, \bar{V}g)_p = \Big(\sum_{k=1}^{p} \sum_{l=1}^{p} (\bar{\xi}_k - \mu_k) W_{kl} g_l \Big)^2 = \big((\bar{\boldsymbol{\xi}} - \boldsymbol{\mu})' W \mathbf{g} \big)^2 = \big((\bar{\boldsymbol{\xi}} - \boldsymbol{\mu})' W^{1/2} W^{1/2} \mathbf{g} \big)^2,$$

where

$$\mathbf{g} = (g_1, \ldots, g_p)'$$
$$\bar{\boldsymbol{\xi}} - \boldsymbol{\mu} = (\bar{\xi}_1 - \mu_1, \ldots, \bar{\xi}_p - \mu_p)'$$

Similarly, we have:

$$\begin{aligned}(n-1)(g,\hat{V}g)_p &= \sum_{i=1}^{n}\Big(\int_T \big(\xi_{i,p}(t)-\bar{\xi}_p(t)\big)g(t)dt\Big)^2 \\ &= \sum_{i=1}^{n}\Big(\int_T \sum_{k=1}^{p}\sum_{l=1}^{p}(\xi_{ik}-\bar{\xi}_k)g_l\phi_k(t)\phi_l(t)dt\Big)^2 \\ &= \sum_{i=1}^{n}\Big(\sum_{k=1}^{p}\sum_{l=1}^{p}(\xi_{ik}-\bar{\xi}_k)W_{lk}g_l\Big)^2 = (n-1)\mathbf{g}'W^{1/2}SW^{1/2}\mathbf{g},\end{aligned}$$

where

$$\begin{aligned}S &= \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{\xi}_i-\bar{\boldsymbol{\xi}})(\boldsymbol{\xi}_i-\bar{\boldsymbol{\xi}})' \\ \boldsymbol{\xi}_i-\bar{\boldsymbol{\xi}} &= (\xi_{i1}-\bar{\xi}_1,\dots,\xi_{ip}-\bar{\xi}_p)'\end{aligned}$$

Thus, we obtain the following:

$$T_p^2 = n\max_{\mathbf{g}\in\mathrm{Im(S)}}\frac{\big((\bar{\boldsymbol{\xi}}-\boldsymbol{\mu})'W^{1/2}W^{1/2}\mathbf{g}\big)^2}{\mathbf{g}'W^{1/2}SW^{1/2}\mathbf{g}}.$$

It can be written in another interesting way thanks to the Maximization Lemma in Johnson and Wichern (2007). We get the final representation:

$$T_p^2 = n(\bar{\boldsymbol{\xi}}-\boldsymbol{\mu})'W^{1/2}S^{+}W^{1/2}(\bar{\boldsymbol{\xi}}-\boldsymbol{\mu}),$$

where S^+ is the Moore-Penrose inverse of S.

□

Lemma 1.1. *Let $\{\xi_i\}_{i=1,\dots,n}$ a set of random elements of $H^k(T)$, with common mean μ, and let $\hat{V}$ and $\bar{V}$ be the two H^k operators defined in Subsection 1.5.1. The two operators $\hat{V}$ and $\bar{V}$ are such that, for any $g \in H^k$:*

- $(g,\hat{V}g)_{H^k} = \widehat{Var}[(g,\xi_i)_{H^k}]$;
- $(g,\bar{V}g)_{H^k} = ((g,\bar{\xi}-\mu)_{H^k})^2$.

Proof. For any $g \in \mathrm{Im}(\hat{V})$, we have:

$$\begin{aligned}(g,\hat{V}g)_{H^k} &= \sum_{l=0}^{k}(D^l g, D^l(\hat{V}g))_{L^2} \\ &= \sum_{l=0}^{k}\int_T D^l g(t)D^l\left[\int_T\sum_{j=0}^{k}\mathcal{S}_{0j}(t,s)D^j g(s)\mathrm{d}s\right]\mathrm{d}t \\ &= \sum_{l=0}^{k}\int_T D^l g(t)\int_T\sum_{j=0}^{k}\partial_t^l\mathcal{S}_{0j}(t,s)D^j g(s)\mathrm{d}s\mathrm{d}t \\ &= \sum_{l=0}^{k}\sum_{j=0}^{k}\iint_{T\times T}D^l g(t)\mathcal{S}_{lj}(t,s)D^j g(s)\mathrm{d}s\mathrm{d}t,\end{aligned}$$

where in the last equality, we used the fact that:

$$\begin{aligned}\partial_t^l\mathcal{S}_{0j}(t,s) &= \partial_t^l\frac{1}{n-1}\sum_{i=1}^{n}(\xi_i(t)-\bar{\xi}(t))(D^j\xi_i(s)-D^j\bar{\xi}(s)) \\ &= \frac{1}{n-1}\sum_{i=1}^{n}(D^l\xi_i(t)-D^l\bar{\xi}(t))(D^j\xi_i(s)-D^j\bar{\xi}(s)) = \mathcal{S}_{lj}(t,s).\end{aligned}$$

Furthermore, we have:

$$\begin{aligned}
&\sum_{l=0}^{k}\sum_{j=0}^{k}\iint_{T\times T} D^l g(t)\mathcal{S}_{lj}(t,s)D^j g(s)\mathrm{d}s\mathrm{d}t\\
=&\sum_{l=0}^{k}\sum_{j=0}^{k}\iint_{T\times T} D^l g(t)\frac{1}{n-1}\sum_{i=1}^{n}(D^l\xi_i(t)-D^l\bar{\xi}(t))(D^j\xi_i(s)-D^j\bar{\xi}(s))D^j g(s)\mathrm{d}s\mathrm{d}t\\
=&\frac{1}{n-1}\sum_{i=1}^{n}\sum_{l=0}^{k}\sum_{j=0}^{k}\left(\int_T D^l g(t)(D^l\xi_i(t)-D^l\bar{\xi}(t))\mathrm{d}t\right)\left(\int_T D^j g(t)(D^j\xi_i(t)-D^j\bar{\xi}(t))\mathrm{d}t\right)\\
=&\frac{1}{n-1}\sum_{i=1}^{n}\left(\sum_{l=0}^{k}\int_T D^l g(t)(D^l\xi_i(t)-D^l\bar{\xi}(t))\mathrm{d}t\right)^2\\
=&\widehat{Var}\left[\sum_{l=0}^{k}\int_T D^j g(t)D^j\xi_i(t)\mathrm{d}t\right]\\
=&\widehat{Var}[(g,\xi_i)_{H^k}],
\end{aligned}$$

i.e., $(g,\bar{V}g)_{H^k}$ is the sample variance of the scores of the orthogonal projections of ξ_i on g, $\widehat{Var}[(g,\xi_i)_{H^k}]$.

In the same way, for any $g\in H^k$, we have:

$$\begin{aligned}
(g,\bar{V}g)_{H^k} &= \sum_{l=0}^{k}(D^l g, D^l(\bar{V}g))_{L^2}\\
&= \sum_{l=0}^{k}\int_T D^l g(t)D^l\left[\int_T\sum_{j=0}^{k}(\bar{\xi}(t)-\mu(t))(D^j\bar{\xi}(s)-D^j\mu(s))D^j g(s)\mathrm{d}s\right]\mathrm{d}t\\
&= \sum_{l=0}^{k}\int_T D^l g(t)\int_T\sum_{j=0}^{k}\partial_t^l(\bar{\xi}(t)-\mu(t))(D^j\bar{\xi}(s)-D^j\mu(s))D^j g(s)\mathrm{d}s\mathrm{d}t\\
&= \sum_{l=0}^{k}\sum_{j=0}^{k}\iint_{T\times T} D^l g(t)(D^l\bar{\xi}(t)-D^l\mu(t))(D^j\bar{\xi}(s)-D^j\mu(s))D^j g(s)\mathrm{d}s\mathrm{d}t\\
&= \sum_{l=0}^{k}\sum_{j=0}^{k}\left(\int_T D^l g(t)(D^l\bar{\xi}(t)-D^l\mu(t))\mathrm{d}t\right)\left(\int_T D^j g(t)(D^j\bar{\xi}(t)-D^j\mu(t))\mathrm{d}t\right)\\
&= \left(\sum_{l=0}^{k}\int_T D^l g(t)(D^l\bar{\xi}(t)-D^l\mu(t))\mathrm{d}t\right)^2\\
&= \left((g,\bar{\xi}-\mu)_{H^k}\right)^2,
\end{aligned}$$

that is, $(g,\bar{V}g)_H$ is the square distance between the scores of the orthogonal projections of $\bar{\xi}$ and μ over g. □

Lemma 1.2. *Let $\{\xi_i\}_{i=1,\dots,n}$ a set of random elements of $B^2(T)$, with common mean μ, and let $\hat{V}$ and $\bar{V}$ be the two H^k operators defined in Subsection 1.5.2. The two operators $\hat{V}$ and $\bar{V}$ are such that, for any $g\in B^2(T)$:*

- $(g,\hat{V}g)_{B^2} = \widehat{Var}[(g,\xi_i)_{B^2}]$;
- $(g,\bar{V}g)_{B^2} = ((g,\bar{\xi}-\mu)_{B^2})^2$.

Proof. For any $g\in\mathrm{Im}(\hat{V})$, exploiting the isomorphism (1.12), we have:

$$\begin{aligned}
(g,\hat{V}g)_{B^2} &= (\mathrm{clr}(g),\mathrm{clr}(\hat{V}g))_{L^2}\\
&= \left(\mathrm{clr}(g),\int_T\mathcal{S}_c(t,s)\mathrm{clr}(g)(s)\mathrm{d}s\right)_{L^2}\\
&= \iint_{T\times T}\mathrm{clr}(g)(t)\mathcal{S}_c(t,s)\mathrm{clr}(g)(s)\mathrm{d}s\mathrm{d}t\\
&= (\mathrm{clr}(g),\hat{V}_c\mathrm{clr}(g))_{L^2},
\end{aligned}$$

where $\hat{V}_c \in \mathcal{L}(L^2)$ is the integral operator of kernel $\mathcal{S}_c(t,s)$. Finally, we have:

$$\begin{aligned}(\mathrm{clr}(g), \hat{V}_c \mathrm{clr}(g))_{L^2} &= \widehat{Var}((\mathrm{clr}(\xi_i), \mathrm{clr}(g))_{L^2}) \\ &= \widehat{Var}((\xi_i, g))_{B^2}).\end{aligned}$$

In the same way, for any $g \in B^2(T)$:

$$\begin{aligned}(g, \bar{V}g)_{B^2} &= (\mathrm{clr}(g), \mathrm{clr}(\bar{V}g))_{L^2} \\ &= \left(\mathrm{clr}(g), \int_T (\mathrm{clr}(\bar{\xi})(t) - \mathrm{clr}(\mu)(t))(\mathrm{clr}(\bar{\xi})(s) - \mathrm{clr}(\mu)(s))\mathrm{clr}(g)(s)\mathrm{d}s\right)_{L^2} \\ &= \iint_{T\times T} \mathrm{clr}(g)(t)(\mathrm{clr}(\bar{\xi})(t) - \mathrm{clr}(\mu)(t))(\mathrm{clr}(\bar{\xi})(s) - \mathrm{clr}(\mu)(s))\mathrm{clr}(g)(s)\mathrm{d}s\mathrm{d}t \\ &= (\mathrm{clr}(g), \bar{V}_c \mathrm{clr}(g))_{L^2},\end{aligned}$$

where $\bar{V}_c \in \mathcal{L}(L^2)$ is the integral operator of kernel $(\mathrm{clr}(\bar{\xi})(t) - \mathrm{clr}(\mu)(t))(\mathrm{clr}(\bar{\xi})(s) - \mathrm{clr}(\mu)(s))$. Finally, we have:

$$\begin{aligned}(\mathrm{clr}(g), \bar{V}_c \mathrm{clr}(g))_{L^2} &= \left((\mathrm{clr}(\bar{\xi}) - \mathrm{clr}(\mu), \mathrm{clr}(g))_{L^2}\right)^2 \\ &= \left((\bar{\xi} - \mu, g)_{B^2}\right)^2.\end{aligned}$$

□

Chapter 2

Component-Wise Inference on Functional Data: the Interval Testing Procedure

Abstract

We propose a novel technique (Interval Testing Procedure, or ITP) that enables inference for functional data, whenever the use of a basis expansion to describe data is advisable. The procedure is very general and indeed it can be used to perform different hypothesis testing (e.g., equality in distribution between two or more functional populations, mean function of a functional population). The procedure involves three steps: (*i*) representing data on a (possibly high-dimensional) functional basis; (*ii*) testing each possible set of consecutive basis coefficients; (*iii*) computing the adjusted p-values associated to each basis component, using a novel strategy here proposed. The procedure is provided with an interval-wise control of the Family Wise Error Rate: a new type of error control, which we hereby define, that is particularly suited for functional data. A simulation study comparing the ITP with other testing procedure is reported. The ITP is then applied to the analysis of hemodynamical features involved with the cerebral aneurysm pathology. The procedure is implemented in the `fdatest` R package.

Keywords: Family Wise Error Rate, Functional Data, Inference, Permutation Test

2.1 Introduction

"Are these two groups of curves statistically different?". "Well, since you say that they differ, could you tell me which the differences are?". "What is the probability that these differences just popped up by chance?". Such kinds of questions are becoming more and more urging in many research areas, due to the fast development of more and more precise acquisition devices. Despite the recent breakthrough of functional data analysis (FDA) as a method for analyzing data sets made of curves, and the development of many statistical

tools to answer questions similar to the former one, very few are instead currently available to answer questions similar to the latter ones. We here develop a new inferential procedure able not only to assess the equality in distribution between two functional populations, but also to point out the differences by controlling the probability of false discoveries.

In FDA, data are functions defined in a continuous domain and lying in an infinite dimensional separable Hilbert space (Ramsay and Silverman, 2002, 2005). The major issue for making inference in FDA is that classical multivariate inferential tools (e.g., Hotelling's theorem) become pretty useless in this framework, since they require the number of sample units to be greater than the dimension of the space in which data are observed. In addition, it is not easy to define a suitable model for objects lying in infinite-dimensional spaces. Indeed, the normality and the dependency structure of data in the functional setting are difficult to assess. Hence, the growing interest for the analysis of this type of data is urging the development of inferential techniques suited for this kind of data.

Many methods dealing with inferential problems for functional data are currently object of statistical investigation. The most common approach is the one of "global inference" on the curves. Such techniques are made by a global test that provides a unique p-value. Examples of such techniques are usually based on asymptotic results or on strong modeling assumptions on data distribution (e.g., Abramovich and Angelini 2006; Antoniadis and Sapatinas 2007; Cuevas et al. 2004; Fan and Lin 1998; Horváth and Kokoszka 2012; Schott 2007; Spitzner et al. 2003; Staicu et al. 2014; Zhang and Liang 2014). Many methods have been proposed to test functional data also in a permutational framework, thus not relying on strong distributional assumptions (e.g., Cardot et al. 2007; Hall and Tajvidi 2002). All these procedures are meant to state if there is evidence to reject the assumption of equality in distribution, but not for imputing the rejection to specific features of the data. This could make them not satisfactory in some applications because they do not give any insight to the reasons that has led to rejection.

Although functional data are theoretically infinite-dimensional objects, in the practice, statisticians deal with them by projecting them on a finite dimensional space spanned by a suitable truncated basis, which may be fixed (e.g., Fourier basis, B-splines, wavelets, polynomials, Ramsay and Silverman 2005), data driven (e.g., functional principal components Hall and Van Keilegom (2007*b*)), or even random (Cuesta-Albertos and Febrero-Bande, 2010), and whose dimension can be low or high, depending on the application and on the computational cost. Thus, similarly to these latter works, whenever functional data are described through a basis expansion, we here propose to base inference directly on the set of coefficients representing the data. In this perspective, the functional test can be replaced by a family of tests pertaining the components of the basis expansion.

A natural approach could be to test each component, and then to correct the test results in order to provide the control of the level of the test for each possible set of true null hypotheses (i.e., strong control of the Family Wise Error Rate). The correction can be made, for instance, using the Bonferroni or the

Bonferroni-Holm procedures (Holm, 1979). An example of application of such technique to FDA can be found, for instance, in Spitzner et al. (2003). The resulting procedures provide a strong control of the Family Wise Error Rate (FWER) and enable the selection of a smaller subset of significant components. Nevertheless, they are generally not suited for cases in which the number of basis components is large (namely, computational cost of might explode and/or their power can become very low), that is often the case in FDA.

The Interval Testing Procedure (ITP), which we here propose, deals with the previous issue, and is meant for dealing with functional data. Similarly to Bonferroni-like component-wise inferential techniques, it is able, in case of rejection, to highlight which components have led to the rejection. Differently from them, even when the number of components is very large, its power remains comparable with the one provided by global inference techniques. Since "there is no such thing as a free lunch" the ITP lacks the strong control of the FWER. Indeed it just provides an "interval-wise" control of the FWER (which is stronger than the weak control provided by global tests but weaker than the strong control provided by component-wise procedures). As we will show, in the FDA framework this is a minor drawback since this kind of control might be sufficient in the practice.

For example, as we will show in the following Sections, when testing for the difference between two functional populations relying on the B-spline representation, the "interval-wise" control implies the control of the FWER on intervals of the domain. This means that, for any sub-interval of the domain, if there is no difference between the two populations, the probability that they are detected as significantly different on this interval is controlled to the desired level.

The paper is outlined as follows: in Section 2.2 the ITP is described for the two population test (i.e., testing for differences between two functional populations). The Section describes the algorithm, and presents a discussion on the theoretical properties of the ITP, both in terms of control of the FWER and of its power. In Section 2.3, we indicate how to extend the ITP in other frameworks (e.g., test for the mean of one population, test for ANOVA, and tests for linear models) is provided. In Section 2.4 we present the results of a simulation study comparing the performances of the ITP with other component-wise techniques based on the Bonferroni-Holm and Benjamini-Hockberg (Benjamini and Hochberg, 1995) corrections. In Section 2.5 the ITP is applied to a case study devoted to the analysis the Aneurisk data set (Sangalli et al., 2009), and concerning the comparison between geometric and hemodynamic features of the internal carotid artery in two groups of patients associated to different levels of severity of the cerebral aneurysm pathology. The proofs of all Theorems are reported in Appendix 2.A. In the Appendix we also report some details of the implementation of permutation tests for univariate and multivariate data (Section 2.B), and a second case study on the Fourier-based analysis of Milan daily temperatures data (Section 2.C).

The R-package `fdatest` implementing the ITP for one or two populations of functional data evaluated on a uniform grid is available on CRAN (Pini and Vantini, 2014). All computations and images have been created using R (R Core Team, 2012).

2.2 The ITP in the two-population framework

2.2.1 ITP algorithm

Suppose to have two independent samples of sizes n_1 and n_2 of independent random functions on a separable Hilbert space. We aim at testing the null hypothesis of no differences in distribution between the two populations of curves which the two samples have been drawn.

The testing procedure we propose is composed by the following steps:

1. **Basis Expansion**: functional data are represented through the coefficients of a basis expansion;
2. **Interval-Wise Testing**: statistical tests are performed on each interval of basis coefficients;
3. **Multiple Correction**: for each component of the basis expansion, an adjusted p-value is computed from the p-values of the tests performed in the previous step.

First step: basis expansion
Theoretically, each function can be uniquely represented through a countable sequence of coefficients associated to a basis of the functional space (i.e., Fourier harmonics, B-splines, wavelets, ...). In practice, very rarely functional data come with an analytic expression. More often, just some point-wise evaluations of a function (possibly with some noise) are available, and thus just a reduced number of components can be estimated. It is thus necessary to represent data by means of a finite-dimensional representation $y_{ij}(t)$ obtained through an expansion on a reduced basis $\{\phi^{(k)}\}_{k=1,...,p}$:

$$y_{ij}(t) = \sum_{k=1}^{p} c_{ij}^{(k)} \phi^{(k)}(t), \tag{2.1}$$

where i is the unit index, j the population index, and k the basis component index. This projection constitutes the first step in most FDA procedures. The integer p represents the finite dimension of the functional space in which data are represented. It is important to point out that the choice of the basis and of the dimension p of the truncated basis expansion is not a critical issue for the ITP, that can deal with any functional basis, and with any dimension p, independently on the sample size. The choice of the basis used to represent data, of the truncation p, and of the method used to estimate the coefficients is deeply discussed in the FDA literature. We refer to Ramsay and Silverman (2005) for an overview on this issue.

In the end, we can represent each of the $n = n_1 + n_2$ units by means of the corresponding p coefficients $\{c_{ij}^{(k)}\}_{k=1,...,p}$, $i = 1, ..., n_j$, $j = 1, 2$ associated to the expansion (2.1). The assumptions made for the functional populations can be re-stated in terms of the expansion coefficients: we thus have for each k, that $c_{11}^{(k)}, ..., c_{n_1 1}^{(k)}, c_{12}^{(k)}, ..., c_{n_2 2}^{(k)}$ are independent, and $c_{11}^{(k)}, ..., c_{n_1 1}^{(k)} \sim C_1^{(k)}$, $c_{12}^{(k)}, ..., c_{n_2 2}^{(k)} \sim C_2^{(k)}$, where $C_1^{(k)}$ and $C_2^{(k)}$ denote the unknown distributions of the kth basis coefficient in the two populations. Note that we do not assume

here independence between basis coefficients pertaining to different components (for example, when using a B-spline basis expansion, we have typically that successive basis coefficients associated to the same sample unit have positive dependence), nor the joint or marginal normality of basis coefficients (that can be unrealistic or anyhow difficult to assess), nor the orthogonality of the basis. We just assume independence among units.

Second step: interval-wise testing

The second step of the ITP consists in obtaining tests on each interval of components. In particular, each basis component k is marginally tested ($H_0^{(k)} : C_1^{(k)} \stackrel{d}{=} C_2^{(k)}$); then, a bivariate test is performed on each couple of successive basis components ($H_0^{(k,k+1)} : H_0^{(k)} \cap H_0^{(k+1)}$); then, a three-variate test is performed on each triple of successive basis components ($H_0^{(k,k+1,k+2)} : H_0^{(k)} \cap H_0^{(k+1)} \cap H_0^{(k+2)}$), and so on, up to the global p-variate test, jointly on all components ($H_0^{(1,\dots,p)} : H_0^{(k)} \bigcap_{k=1}^{p} H_0^{(k)}$). Finally, we obtain a family of tests with their associated p-values (e.g., Figure 2.1(a)). We denote with $\lambda^{(\mathbf{k})}$ the p-value of the multivariate test for $H_0^{(\mathbf{k})} = \bigcap_{k \in \mathbf{k}} H_0^{(k)}$ (where $\mathbf{k}$ is a vector of successive indexes in $\{1, ..., p\}$). In addition to all possible tests on intervals, we add the multivariate tests on the complementary sets of each interval, i.e., we also test each hypothesis $H_0^{(\mathbf{k}^c)} = \bigcap_{k \notin \mathbf{k}} H_0^{(k)}$ (as shown in Figure 2.1(a)). The advantage of considering also the complementary sets of the intervals is pertaining to the multiple correction phase, and will be detailed in the next paragraph.

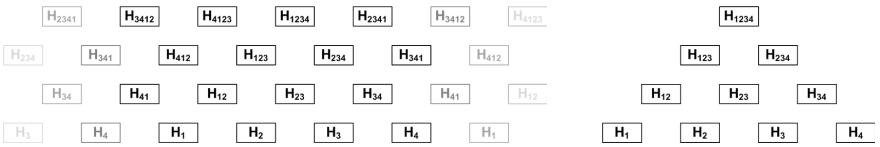


Figure 2.1: Example (with $p = 4$) of the family of multivariate tests explored by the ITP (on the left). Version focusing only on intervals, excluding the complementary sets (on the right).

The interval-wise tests can be performed in different ways, depending on the sample size and assumptions on the distributions of $C_j^{(k)}$. The best-case scenario is the one in which coefficients are jointly normal and $n > p$. In this case, we can base the ITP on Hotelling's T-square tests on each interval.

On the contrary, in a more realistic scenario, we cannot assess the normality of coefficients and $n \leq p$. A possible approach to deal with this issue is to exploit the Non Parametric Combination Procedure (NPC), presented in Pesarin and Salmaso (2010). The NPC is a procedure that enables to build multivariate permutation tests by means of combining univariate synchronized permutation tests. The resulting tests are correct for any n and p even in the presence of dependence among components.

The use of permutation tests to build the family of interval-wise tests has another advantage, that is the fact that the hypothesis of independence between sample units can be relaxed. Indeed, permutation tests only require exchangeability between units under the null hypothesis. This allows, for instance, to base the ITP on the coefficients of a data driven basis expansion.

For a more detailed discussion about the permutation tests, and the NPC procedure, see Appendix 2.B.

Third step: multiple correction
The final step of the ITP consists in the computation of the adjusted p-values associated to each basis component, in order to provide an interval-wise control of the FWER.

We obtain the adjusted p-value for the kth component $\lambda^{(k)}_{ITP}$ by associating to the kth component the maximum p-value observed over the p-values of all interval-wise tests of the previous family whose null hypothesis implies $H^{(k)}_0$:

$$\lambda^{(k)}_{ITP} = \max\left(\max_{\mathbf{k} \text{ s.t. } \mathbf{k}\ni k} \lambda^{(\mathbf{k})}, \max_{\mathbf{k}^c \text{ s.t. } \mathbf{k}^c\ni k} \lambda^{(\mathbf{k}^c)}\right).$$

In the next section we will prove that (if we reject $H^{(k)}_0$ when the kth adjusted p-value $\lambda^{(k)}_{ITP} \leq \alpha$), for any interval $\mathbf{k}$ the probability of rejecting all $H^{(k)}_0$ pertaining to the interval is lower or equal to α, if all the $H^{(k)}_0$ pertaining to that interval are true. This property reads interval-wise control of the FWER.

Note that in a general multivariate setting, where the order of the variables is arbitrary, this latter control is of poor interest. On the contrary, in this setting where the component k has a natural order (e.g., in B-splines k is related to the domain, in the Fourier basis k is related to frequency), this control becomes of immediate interest from the practitioners' point of view. For instance when using B-splines the control holds on any sub-interval of the domain with edges in correspondence of the knots, or when using Fourier expansion the control holds on any frequency band. We can thus control the probability of wrongly detecting sub-intervals of the domain or frequency bands.

As a final remark, note that we could have based the procedure only on tests on intervals of components, excluding the ones on complementary sets, as shown in Figure 2.1(b). Nevertheless, according to this combination strategy the hypotheses in the "middle" are tested more times than the ones at the "edges" (in the example of Figure 2.1(b) with $p = 4$, $H^{(2)}_0$ and $H^{(3)}_0$ are included in 6 tests, whereas the hypotheses $H^{(1)}_0$ and $H^{(4)}_0$ are only tested 4 times). This asymmetry may favor the rejection of the hypotheses at the edges, since tested less times. This is why we introduced the tests on the complementary intervals. The resulting procedure has two major advantages: (i) each components is tested the same number of times (i.e., $p(p+1)/2$), and (ii) the FWER is controlled not only on intervals but also on their respective complementary sets.

2.2.2 Theoretical properties of the ITP

In this subsection, we present some theoretical results regarding the control of the FWER and the power of the ITP. All the proofs of the theorems are reported in Appendix 2.A. Not to overload the notation, through this theoretical section we will indicate with "intervals" both proper intervals and complementary sets of intervals.

We start by formally defining the novel type of error control that we introduced, i.e., the interval-wise control of the FWER.

Definition 2.1. Given a p-dimensional expansion of a functional data set, an inferential procedure is provided with an *interval-wise control of the FWER* if, for any interval $\mathbf{k}$ of components and for all level $\alpha \in [0, 1]$, the probability of rejecting at least one of the null hypotheses pertaining the components of the interval is less than α, when all these hypotheses are true:

$$\forall \text{ interval } \mathbf{k} \subseteq \{1, ..., p\} : \quad \mathbb{P}_{H_0^{(\mathbf{k})}\text{true}} \left[\exists k \in \mathbf{k} : H_0^{(k)} \text{ is rejected} \right] \leq \alpha.$$

The following result characterizes the control of the FWER provided by the ITP.

Theorem 2.1. *The ITP based on the p components of any basis expansion is provided with an interval-wise control of the FWER:*

$$\forall \textit{ interval } \mathbf{k} \subseteq \{1, \dots, p\} : \quad \mathbb{P}_{H_0^{(\mathbf{k})}true} \left[\exists k \in \mathbf{k} : \lambda_{ITP}^{(k)} < \alpha \right] \leq \alpha.$$

In simple words, interval-wise control of the FWER means that, given any interval of components associated to true null hypotheses, the probability that at least one of the null hypotheses associated to the interval is wrongly detected as false is always controlled at the desired level. This kind of control guarantees, among the others, the control on the entire set of components and on single components, as extreme cases of intervals. From these, we have indeed the following.

Corollary 2.1. *The ITP based on the p components of any basis expansion is provided with a weak control of the FWER, i.e., the probability of rejecting at least one null hypothesis when all null hypotheses are true is controlled:*

$$\mathbb{P}_{H_0^{(\{1,\dots,p\})}true} \left[\exists k \in \{1, \dots, p\} : \lambda_{ITP}^{(k)} < \alpha \right] \leq \alpha.$$

Corollary 2.2. *The ITP based on the p components of any basis expansion is provided with a control of the Comparison-Wise Error Rate, i.e., for each component the probability that the null hypothesis pertaining the component is rejected when true is controlled:*

$$\forall k \in \{1, \dots, p\} : \quad \mathbb{P}_{H_0^{(k)}true} \left[\lambda_{ITP}^{(k)} < \alpha \right] \leq \alpha.$$

To further characterize the inferential properties of the ITP, it is useful to introduce two other testing procedures, derived from the use of different families of multivariate tests in the second step of the procedure described in Subsection 2.2.1. The Global Testing Procedure (GTP), which is associated to a degenerative family made by the global test only, and the Closed Testing Procedure (CTP), which is associated to the family made by all $2^p - 1$ possible multivariate tests. Note that the CTP, even though theoretically sound, becomes quickly unfeasible in practice when the dimension p is high, due to the high number of tests it is based on. The number of tests indeed grows exponentially in the number of components p for the CTP, quadratically for the ITP and is constant equal to one for the GTP.

The ITP is intermediate between the GTP and CTP. Indeed, the GTP provides a weak control of the FWER (the probability of wrongly rejecting at least

one null hypothesis is controlled only if all null hypotheses are true) while the CTP provides a strong control of the FWER (the probability of wrongly rejecting at least one null hypothesis is controlled over any set made of true null hypotheses).

The following results make a comparison between the ITP, GTP, and CTP in terms of power and error rate. In particular, Theorems 2.2 and 2.3 prove (globally and component-wise, respectively) that the power of the GTP is always higher than the power of the ITP which is always higher than the power of the CTP. The same theorems prove also (globally and component-wise, respectively) that the CTP is always more conservative than the ITP which is always more conservative than the GTP, which is indeed exact.

Theorem 2.2 (Global properties)**.** *Let us consider a CTP, an ITP, and a GTP based on the p components of a basis expansion. The actual global levels of the CTP, of the ITP, and of the GTP (i.e., the probability of rejecting at least one $H_0^{(k)}$ when all null hypotheses are true) satisfy:*

$$\alpha_{CTP} \leq \alpha_{ITP} \leq \alpha_{GTP} = \alpha \ .$$

The powers of the CTP, of the ITP, and of the GTP (i.e., the probability of rejecting at least one $H_0^{(k)}$ when at least one of the null hypotheses is false) satisfy:

$$\pi_{CTP} \leq \pi_{ITP} \leq \pi_{GTP} \ .$$

Theorem 2.3 (Component-wise properties)**.** *Let us consider a CTP, an ITP, and a GTP based on the p components of a basis expansion. The Comparison-Wise Error Rates of the CTP and of the ITP on each component (i.e., the probability of rejecting $H_0^{(k)}$ when true) satisfy:*

$$CWER_{CTP}^{(k)} \leq CWER_{ITP}^{(k)} \leq \alpha \ .$$

The component-wise powers of the CTP and of the ITP on each component (i.e., the probability of rejecting $H_0^{(k)}$ when false) and the power of the GTP satisfy:

$$\pi_{CTP}^{(k)} \leq \pi_{ITP}^{(k)} \leq \pi_{GTP} \ ,$$

with π_{GTP} the power of the global test.

Previous theorems explicit the tradeoff between the control of the FWER and the power both globally (Theorem 2.2) and component-wise (Theorem 2.3). Indeed the weaker control of the FWER of the ITP with respect to the CTP is counterbalanced by the fact that the ITP is less conservative and more powerful (globally and component-wise) than the CTP. On the contrary, the stronger control of the FWER of the ITP with respect to the GTP is counterbalanced by the fact that the ITP is more conservative and less powerful than the GTP. This latter power loss is anyway countered by a big gain in interpretability of the test results with respect to the GTP. Indeed, differently from the GTP, the ITP is able to highlight the basis elements which the rejection is due to.

Note that all results hold for any implementation of the ITP. Indeed, the corresponding proofs exclusively rely on the exactness of the family of interval-wise tests described in Subection 2.2.1 and not on the nature of the latter ones.

Thus, the theoretical results depend neither on the type and dimension of the basis used nor on the type of tests. In practice, the choice of the basis matters instead. Indeed, for instance, if one desires a control on intervals of the domain, a B-spline basis expansion should be used, while if the control is desired in the frequency domain a Fourier expansion should be used.

In conclusion, when dealing with functional data, the ITP provides a good compromise between the CTP and GTP gathering the best of both procedures. Indeed, like the CTP, the ITP performs a selection of the significant components; and, like the GTP, its computational costs remain affordable even for large values of p; moreover, its control of the FWER and its power are intermediate between the ones provided by the CTP and GTP.

2.3 Extending the ITP to different frameworks

The idea of performing a family of multivariate tests on each interval of consecutive basis components of a functional basis, along the line described in Section 2.2 is very general, and may be applied more or less straightforwardly to situations more complex than the comparison between two functional populations. To apply the ITP to other tests in FDA, we just need to define a way to build exact multivariate tests on the intervals of components of the basis expansion.

For instance, we can think about a situation in which we aim at testing for differences between two paired populations. This case can be treated by applying suitable paired tests in the second step of the algorithm. Furthermore, we can apply the ITP in the one-population framework, where the objective is testing for the mean of a functional population. In this case, we just have to provide exact multivariate tests for the mean vectors of the coefficient intervals.

We can also provide ITP-based tests for a functional ANOVA framework, where the objective is to detect differences among $g > 2$ independent functional populations. In this case, the tests can be based on MANOVA-like tests on each interval of components. By changing the type of tests performed in the second phase of the ITP, we can also deal with more complicated situations, such as tests for variance, two-way ANOVA, ANCOVA, and linear models. In detail, when all tests performed in the second phase of the algorithm are exact, the resulting ITP will provide an interval-wise control of the FWER.

2.4 Simulation study

The aim of the simulation study here reported is to investigate the performances of the ITP. In particular, we want to compare the performances of the ITP and the Benjamini-Hochberg procedure (BH, Benjamini and Hochberg 1995), as the number of false hypotheses increases, being the latter one currently one of the most used approaches to multiple testing. The BH is indeed a multiple testing procedure widely spread in many research fields, who deals with the problem of extremely large families of tests. If applied to a basis expansion, it controls the False Discovery Rate (FDR) over components, i.e., the expected proportion of falsely rejected components among those being rejected. The FDR control is weaker than the strong FWER control (i.e., the BH is only provided with weak

control of the FWER), but it leads to procedures that are more powerful than the ones provided with strong FWER control (e.g., Bonferroni-Holm).

We also compared the ITP with the Bonferroni-Holm procedure (which is provided with a strong control of the FWER, Holm 1979). Due to the large number of components (i.e., 50), the Bonferroni-Holm never rejected any hypothesis. Hence, the results for the performances of the Bonferroni-Holm are not reported here, being not informative.

2.4.1 Simulation setting

The simulation study is divided into two parts: in the first part for each component we compare the probability that the component is rejected; in the second part we compare the probability of rejecting at least one of the components which are known to be identically distributed in the two populations (i.e., false discoveries) and the probability of rejecting at least one of the components which are known to be differently distributed in the two populations (i.e., true discoveries).

In the entire study, we consider a standard scenario for data generation. We perform a test for the differences between two independent populations of functional data on $L^2[0,1]$ generated via the cubic B-spline coefficients in a 50-dimensional space. Let $c_1^{(1)}, c_1^{(2)}, .., c_1^{(50)}$ be the random B-spline coefficients associated to units of the first population, and $c_2^{(1)}, c_2^{(2)}, .., c_2^{(50)}$ the random B-spline coefficients associated to units of the second population, and let $\mu_1^{(k)}$ and $\mu_2^{(k)}$ indicate the means of the coefficients of the first and second populations, respectively. We generate the coefficients $c_j^{(k)}$ from a normal distribution, with mean $\mu_1^{(k)} = 0$, $\mu_2^{(k)} \in [0,1]$. We suppose that the functional means of the two populations differ on a closed interval of length $h \in [0,1]$. In this interval, the mean of the differences between coefficients will be equal to a constant $v \in [0,1]$.

Two different scenarios of alternative hypotheses are explored, by varying the parameters h and v. In the first scenario, we take a constant $v = 0.5$, and vary h (which is the length of the interval where the two means differ), between the extremes values of 0 (i.e., no difference along the entire domain) and 1 (i.e., difference along the entire domain). In the second scenario, we take a constant $h = 0.5$, and vary the parameter v (which is the maximal difference between the two coefficients' means) between 0 and 1.

The different components are generated independently, i.e., the variance covariance matrix of the 50-dimensional vector of differences is $\Sigma = \sigma^2 I$, with $\sigma^2 = 0.25$. Other simulations have been performed with different choices for Σ, showing that the described results do not change considering a more complicate covariance structure. Finally, we suppose to observe $n_1 = n_2 = 10$ different realizations from the two populations. An instance of the simulated data for $h = 0.5, v = 1$ is reported in Figure 2.2. In all scenarios, 5000 different functional data sets are simulated.

2.4.2 Component-wise probability of rejection

On Figure 2.3, we report the component-wise probability of rejection at a 5% level (i.e., the probability that a component is rejected at 5% level). In particular, on each panel we report a different scenario for h and v, and for each

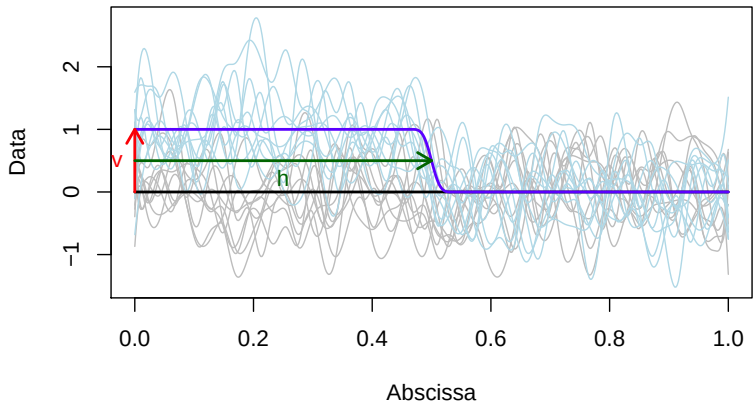


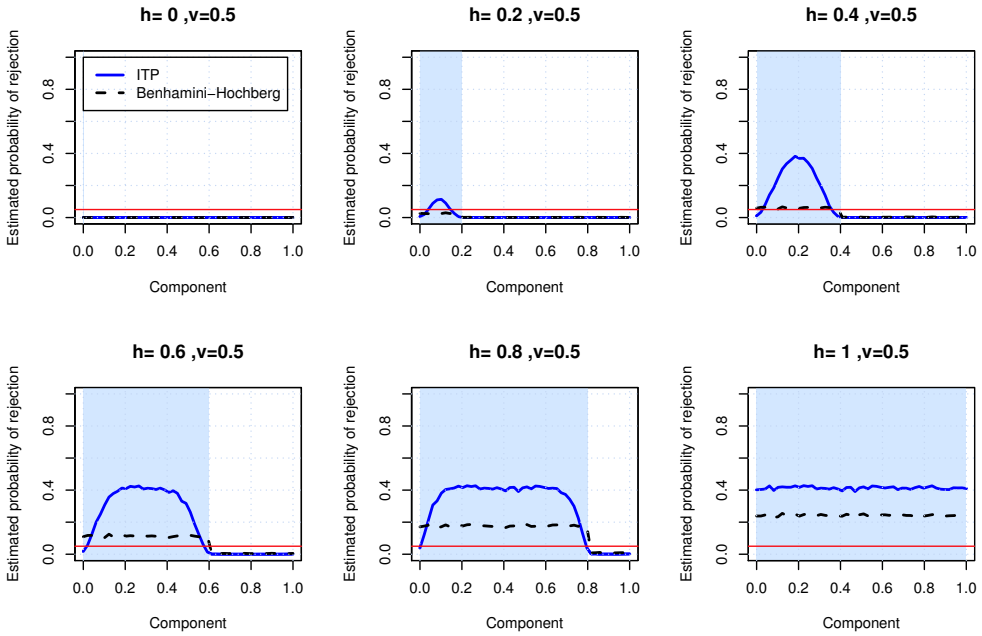
Figure 2.2: Functional data of the first and second group simulated for the study (gray and lightblue lines, respectively), functional means (black and blue lines, respectively), in the example $h = 0.5, v = 1$.

component along the abscissa we report on the vertical-axis the probability that the component is rejected. In addition, the shaded gray part of each panel indicates the interval in which the two means are actually different, according to the simulation setting. Thus, in the white part of the plot the graph represents the probability of rejecting a component which should not be rejected (i.e., component-wise error rate), while in the gray part the probability of rejecting a component which should be rejected (i.e., component-wise powers). The top panels of Figure 2.3 show the results obtained varying the parameter h with $v = 0.5$, whereas the lower panels of the same figure show the results obtained varying the parameter v with $h = 0.5$.

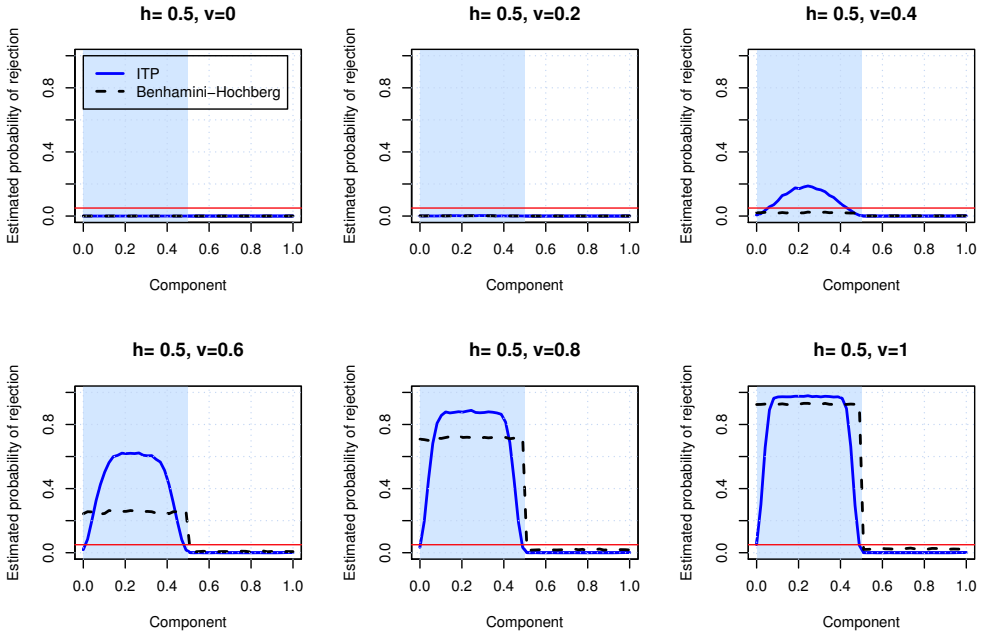
The simulation shows that both procedures assure the control of the CWER in each scenario, as expected by theory. We notice that the ITP maximizes the component-wise power at the center of the intervals where the difference occurs, while the BH component-wise power has a flat shape over all the components. This is due to the fact that the ITP exploits the ordered structure of the components, gaining power in the center of the false hypotheses interval, thanks to the structure of the family of multivariate tests that are performed. On the other hand, the BH procedure adjusts the p-values in a step-wise way, without considering the fact that coefficients are ordered. Hence, the ITP appears to be more powerful in detecting the presence of a significant interval, but more conservative with respect to the amplitude of the interval (i.e., its power at the boundaries is lower). The BH procedure is, on the other hand, less powerful in detecting the presence of a significant interval, but once detected, it targets its actual amplitude.

2.4.3 Family-wise probability of rejection

In the top panels of Figure 2.4 we report, for each testing procedure, the estimated probability of having at least one false discovery, that is the FWER (i.e.,



(a) Estimated component-wise probability of rejection as a function of h



(b) Estimated component-wise probability of rejection as a function of v

Figure 2.3: Estimated component-wise probability of rejection at a 5% level for the considered multiple testing procedures on each scenario.

the probability of rejecting at least one of the components which are known to be identically distributed in the two populations). The estimated probability is obtained by varying the parameters h (left) and v (right). We notice that, coherently with the theory, the ITP controls this probability for any scenario, while this control is in general not guaranteed by the BH procedure. Note that, if the difference between the two populations is on a fixed domain (top-right panel), as the difference increases, the BH procedure starts rejecting components that should not be rejected, up to a 40% probability.

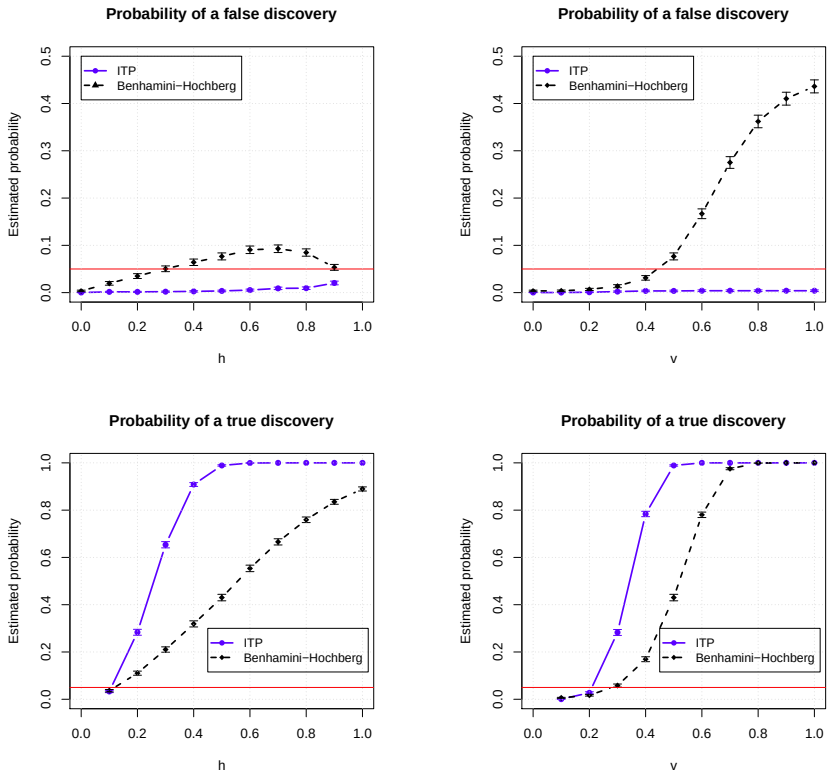


Figure 2.4: Estimated probability of having at least one false discovery (top) and estimated probability of having at least one true discovery (bottom) as functions of h (left) and v (right), for $h, v \in \{0, 0.1, 0.2, ..., 1\}$. The error bands indicate the 95% confidence intervals for the real probability.

In the lower panels of Figure 2.4 we report for both procedures the estimated probability of having at least one true discovery, that is the power on false components (i.e., the probability of rejecting at least one of the components which are known to be differently distributed in the two populations). In all explored cases, the ITP outperforms the BH procedure, showing the ITP to be more powerful in detecting differences. This confirms what noticed with the component-wise results, i.e., that the ITP is more powerful in detecting the presence of an interval presenting differences between the two populations.

Note that the probability of having at least one false discovery is not defined in the case $h = 1$ (only false null hypotheses), while the probability of having at least one true discovery is not defined in the cases $h = 0$ or $v = 0$ (only true null hypotheses).

2.5 Analysis of the Aneurisk data set

Aim of this section is to show the usefulness of the ITP in practical situations. In detail, we present the analysis of the Aneurisk data set (Passerini et al., 2012), which deals with the geometrical and hemodynamical features of the internal carotid arteries (ICA) of patients affected by a cerebral aneurysm.

The aim of this analysis is to assess whether the geometry and/or the hemodynamics of the internal carotid artery can be related to the type and severity of the pathology. In particular, we look for possible differences in the distributions of vessel-radius, centerline-curvature, and wall-shear-stress (WSS) - as functions of the arch-length along the carotid centerline - between subjects affected by a severe form of the pathology (i.e., upper group, 25 subjects with an aneurysm in the upper part of the brain within the skull) and subjects affected by a minor form of the pathology or healthy (i.e., lower group, 25 subjects with an aneurysm in the lower part of the head outside the skull or without any aneurysm). A detailed description of data gathering and processing can be found in Passerini et al. (2012). The projection of data on $p = 128$ uniformly spaced B-splines of order $m = 3$ for radius, curvature, and WSS are reported in the bottom panels of Figure 2.5. Upper group functions are reported in blue while the lower group ones in red.

In detail, we perform three separated analyses for the radius, curvature, and WSS functions, respectively, and we implement the ITP for the differences between two independent functional populations (Section 2.2). The interval-wise multivariate tests of the second step are obtained using the NPC procedure based on the Fisher combination function (Pesarin and Salmaso, 2010). Note that the original data are very smooth, and thus the procedure is robust as the order m of the basis and/or its dimension p vary.

The adjusted p-values of the three ITPs are reported in the central panels of Figure 2.5. At level $\alpha = 5\%$, we do not detect any statistical difference between upper and lower groups pertaining neither the radius nor the curvature functions. A difference in terms of WSS is instead detected. In detail, being here the support of the B-splines localized with respect to the arc-length (i.e., x-axis), we can impute the rejection to the segment of the carotid associated to the arc-length interval $(-2.783\text{cm}, -1.632\text{cm})$ (gray region in bottom panels of Figure 2.5). In particular, we found lower WSS for very severe subjects (i.e., upper group) while higher WSS for less severe subjects (i.e., lower group).

Hemodynamics could explain this finding: the latter region corresponds to the second bend of the ICA (i.e., the segment where a second peak of curvature is present and where the ICA becomes getting narrower). The bends of the ICA are indeed "guardians" of the arteries of upper part of the brain, which are among the weakest in the entire body (being not surrounded by any muscular tissue). Thanks to the passage through the bends the unsteady blood flow from the heart is made steadier before entering the brain. This "stabilizing" effect is related to the loss of energy which is in turn related to the magnitude of the wall-shear-stress within the bends.

Hence, the ITP provides the statistician (S) with a tool to answer the questions related to this application pointed out by the practitioner (P).

P: "Are the curves of the upper and lower groups statistically different?"

S: "They are not with respect to radius and curvature, but they are with respect to WSS."

P: "Could you tell me which the differences in the WSS curves are?"

S: "There is a significant differences between the two groups in the segment of the carotid in the arc-length interval $(-2.783\text{cm}, -1.632\text{cm})$."

P: "Are you sure that these differences did not just pop up by chance?"

S: "We can state that the probability that this result popped up by chance is lower than 5%. Indeed, if there is no difference in distribution between the two population in the segment $(-2.783\text{cm}, -1.632\text{cm})$, the probability of pointing out that segment would be less than 5%."

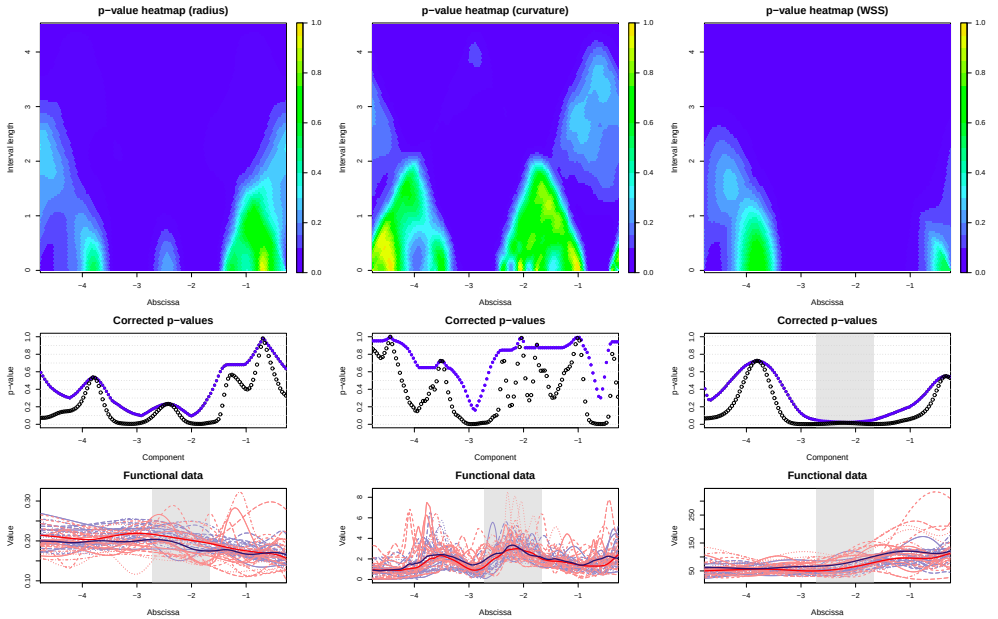
Finally, just for theoretical interest, on the top panels of Figure 2.5, we added the heat-map of all p-values of interval-wise tests performed in the second step of the procedure and used to compute the adjusted p-values, and in the middle panels both the adjusted and the unadjusted component p-values (full dots and empty dots, respectively).

The CTP is in this application unfeasible (i.e., more than 10^{38} multivariate tests would be needed) and the GTP on WSS rejects the null hypothesis of no difference in the WSS between the groups, but it cannot detect in which segment of the carotid this difference is. The Bonferroni-Holm correction is not able to detect any difference between the two groups, while the Benjamini-Hochberg correction detects, a larger interval than the one detected by the ITP (i.e., the interval $(-3.239\text{cm}, -1.210\text{cm})$). This is consistent with the weaker control of the FWER provided by the Benjamini-Hochberg correction, and comes with the property that, on average, up to the 5% of this interval is expected to be composed by false discoveries.

2.6 Discussion

We presented a novel inferential procedure suited for functional data analysis (FDA). The procedure, named Interval Testing Procedure (ITP), involves three steps: (i) representing functional data on a functional basis; (ii) performing a family of multivariate tests on each interval of components; (iii) computing adjusted p-values associated to each basis component. The procedure is very general and it can be easily declined to deal with several inferential problems occurring in FDA: for example, the comparison of two or more functional populations, or testing for the mean function of a functional population. The inference carried out by the ITP is semi-parametric in the sense that we make use of a parametric basis expansion to represent data, but we do not introduce any strong distributional assumption on the coefficients of the expansion (e.g., we do not assume gaussianity).

We introduced the definition of interval-wise control of the Family Wise Error Rate (FWER) which is particularly meaningful in the framework of FDA and which the ITP is provided with. In detail, interval-wise control of the FWER



(a) Radius (b) Curvature (c) WSS

Figure 2.5: Aneurisk case study analysis of radius (left), curvature (center) and WSS (right). Top: p-value heat-maps; center: uncorrected (empty dots) and corrected (full dots) p-values; bottom: curves of the upper and lower groups (blue and red, respectively), and sample means associated to the two groups (bold blue and red curves). The shaded part indicates the interval where significant differences area found in terms of WSS.

refers to the property of controlling the FWER over all intervals of components of the basis expansion, meaning that, for any interval, if there is no difference in distribution between the investigated populations the probability of incorrectly detecting as significant at least one of the components of the interval is controlled. As an example this control, which lies in between the weak and the strong control of the FWER, if associated to a B-spline expansion implies that, given any interval of the domain in which there is no difference between the two functional populations, the probability that at least a part of the domain is wrongly detected as significant is controlled to the desired level.

In addition to having proved the interval-wise control property of the ITP, we also proved that the component-wise and global statistical power of the ITP is always higher than the one provided by the Closed Testing Procedure (which provides a strong control of the FWER but it is computationally unfeasible in the functional framework). On the contrary, we proved that the component-wise and global power of the ITP is always lower than the Global Testing Procedure one (which however provides only a weak control of the FWER and does not provide any guide to the interpretation in terms of components of the test result).

Even though all theoretical properties shown in this work hold for any dimension p of the basis, it is of course of major interest for future research to explore the trade-off, in terms of power of the procedure, as p increases. Indeed, being usually the data available as noisy evaluations observed at some points, an extreme large value of p may lead to overfitting, and thus it results in a loss of power, due to the presence of noise in data representation. On the contrary, a too low value of p may lead to oversmoothing, and it results in a loss of power,

due to excessive flattening of data differences.

A comparison with Bonferroni-Holm and Benjamini-Hochberg procedures has been carried out through a simulation study. The major finding that can be drawn from simulations is that the ITP is more powerful than the Bonferroni-Holm procedure. Moreover, it appears to be more powerful than the Benjamini-Hochberg procedure in detecting the presence of a significant interval, but more conservative with respect to the amplitude of the interval. The Benjamini-Hochberg procedure is, on the other hand, less powerful in detecting the presence of a significant interval, but once detected, it targets its actual amplitude.

We reported the application of the ITP to a case study to show its potential in practice. In detail, we performed a B-spline-based inference for the difference between radius, curvature and wall shear stress curves along the Internal Carotid Artery of two pathologically-different groups of subjects. We compared the findings highlighted by the ITP with the ones pointed out by the Bonferroni-Holm correction procedure (which provides a strong control of the FWER) and the Benjamini-Hochberg correction procedure (which provides a control of the False Discovery Rate and thus just a weak control of the FWER). The ITP turned out to be more powerful than the Bonferroni-Holm procedure and comparable with the Benjamini-Hochberg correction procedure.

An R-package (`fdatest`) implementing the ITP is available on CRAN (Pini and Vantini, 2014). The current version of the package requires functional data evaluated on a uniform grid; it performs a projection of each function on a chosen functional basis using Ordinary Least Squares; it performs the entire family of interval-wise multivariate tests; and, finally, it provides the vector of the adjusted p-values, which can be used to select the statistically significant basis components at level α. The package provides also a plotting function creating a graphical output like the ones presented in Figure 2.5: the p-value heat-map, the plot of the adjusted p-values, and the plot of functional data.

We conclude by mentioning Vsevolozhskaya et al. (2014), a very recent work appeared on *The Annals of Applied Statistics* that we have been recently aware of, and that proposes an inferential procedure for functional data along the same line of research of the present manuscript. In very few words, for selecting segments of the domain, Vsevolozhskaya et al. (2014) propose to perform a family of global tests on pre-selected segments of the domain, and then apply a closed testing procedure over this family. Despite the similarity in the aims, the two procedures present major differences with respect to the control of the FWER. Indeed, while the ITP guarantees the strong control of the FWER on any segment of the domain, the procedure proposed in Vsevolozhskaya et al. (2014) provides only a weak control within each pre-selected segment and lacks any control on segments different from the pre-selected ones. On the other hand, the latter procedure has a strong control over any couplet, triplet, ... of pre-selected segments, which the ITP is instead missing. Anyhow, the two approaches propose two ways to reach a shared target, that is, searching for a good compromise between the advantages and disadvantages of global testing procedure and closed testing procedure.

Appendix

2.A Proofs

Proof. **(Theorem 2.1)**
Let $\mathbf{k} = \{k_1, k_2, ...k_d\}$ be a set of indices defining an interval in $\{1, 2, ..., p\}$. Let $\mathcal{R}^{(k_i)}_{\alpha,ITP}$ be the event "$H_0^{(k_i)}$ is rejected by the ITP at level α" and $\mathcal{R}^{(\mathbf{k})}_{\alpha,ITP} = \bigcup_{k_i \in \mathbf{k}} \mathcal{R}^{(k_i)}_{\alpha,ITP}$ the event "at least one of the $H_0^{(k_i)}$ is rejected by the ITP at level α". Proving the interval-wise control of the FWER of the ITP means proving that, for any $\mathbf{k}$ and for any α, $\mathbb{P}[\mathcal{R}^{(\mathbf{k})}_{\alpha,ITP}] \leq \alpha$ when $\mathbf{H}_0^{(\mathbf{k})} = \bigcap_{k_i \in \mathbf{k}} H_0^{(k_i)}$ is true (i.e., when all $H_0^{(k_i)}$ are true).

Let us indicate with $\mathcal{R}^{(\mathbf{k})}_{\alpha}$ the event "$\mathbf{H}_0^{(\mathbf{k})} = \bigcap_{k_i \in \mathbf{k}} H^{(k_i)}$ is rejected at level α by the corresponding multivariate test". This is the conclusion of the test included in the family of tests explored within the ITP that is derived from the aggregation of the univariate tests for $H_0^{(k_i)}$ with $k_i \in \mathbf{k}$. Thanks to the structure of the ITP, for any $k_i \in \mathbf{k}$, the latter test is among the ones used to correct the k_ith p-value. Thus, the ITP cannot reject $H_0^{(k_i)}$ if the latter test does not reject $\mathbf{H}_0^{(\mathbf{k})}$ (i.e., $\mathcal{R}^{(k_i)}_{\alpha,ITP} \subseteq \mathcal{R}^{(\mathbf{k})}_{\alpha}$). This inclusion holds for all $k_i \in \mathbf{k}$ and thus we have that $\mathcal{R}^{(\mathbf{k})}_{\alpha,ITP} = \bigcup_{k_i \in \mathbf{k}} \mathcal{R}^{(k_i)}_{\alpha,ITP} \subseteq \mathcal{R}^{(\mathbf{k})}_{\alpha}$; and consequently that $\mathbb{P}[\mathcal{R}^{(\mathbf{k})}_{\alpha,ITP}] \leq \mathbb{P}[\mathcal{R}^{(\mathbf{k})}_{\alpha}]$. Finally, due to the exactness of all tests included in the family explored by the ITP, we have that, when $\mathbf{H}_0^{(\mathbf{k})} = \bigcap_{k_i \in \mathbf{k}} H_0^{(k_i)}$ is true, the second term of the latter inequality is equal to α and thus, under the same assumption, that $\mathbb{P}[\mathcal{R}^{(\mathbf{k})}_{\alpha,ITP}] \leq \alpha$. □

Proof. **(Corollaries 2.1 and 2.2)**
The proofs are straightforward, and come directly from the fact that the global test on all components (Corollary 2.1) and the marginal tests on each component (Corollary 2.2) are extreme types of intervals included in the ITP. □

Proof. **(Theorem 2.2)**
Let $\mathcal{R}^{(k)}_{\alpha,ITP}$ be the event "$H_0^{(k)}$ is rejected by the ITP at level α", $\mathcal{R}^{(k)}_{\alpha,CTP}$ be the event "$H_0^{(k)}$ is rejected by the CTP at level α", and $\mathcal{R}_{\alpha,GTP}$ be the event "$\mathbf{H}_0 = \bigcap_{k=1,...,p} H_0^{(k)}$ is rejected by the GTP at level α". Thanks to the structure of the ITP and of the CTP, all multivariate tests used to correct the kth p-value in the ITP are used to correct the CTP but not vice versa. Moreover, the global test is among the test used in both the CTP and the ITP to correct the kth p-value. Thus, every time the CTP rejects $H_0^{(k)}$ also the ITP rejects it and every time the ITP rejects $H_0^{(k)}$ also the GTP rejects it. Thus we have that $\mathcal{R}^{(k)}_{\alpha,CTP} \subseteq \mathcal{R}^{(k)}_{\alpha,ITP} \subseteq \mathcal{R}_{\alpha,GTP}$, and consequently that $\mathbb{P}[\mathcal{R}^{(k)}_{\alpha,CTP}] \leq \mathbb{P}[\mathcal{R}^{(k)}_{\alpha,ITP}] \leq \mathbb{P}[\mathcal{R}_{\alpha,GTP}]$. Let us now consider the event "at least one of the $H_0^{(k)}$ is rejected by the ITP at level α" (i.e., $\bigcup_{k=1,...,p} \mathcal{R}^{(k)}_{\alpha,ITP}$) and the event "at least one of the $H_0^{(k)}$ is rejected by the CTP at level α" (i.e., $\bigcup_{k=1,...,p} \mathcal{R}^{(k)}_{\alpha,CTP}$). We have that $\bigcup_{k=1,...,p} \mathcal{R}^{(k)}_{\alpha,CTP} \subseteq \bigcup_{k=1,...,p} \mathcal{R}^{(k)}_{\alpha,ITP} \subseteq \mathcal{R}_{\alpha,GTP}$ and thus $\mathbb{P}[\bigcup_{k=1,...,p} \mathcal{R}^{(k)}_{\alpha,CTP}] \leq \mathbb{P}[\bigcup_{k=1,...,p} \mathcal{R}^{(k)}_{\alpha,ITP}] \leq \mathbb{P}[\mathcal{R}_{\alpha,GTP}]$. Now, if the state of nature implies that $\mathbf{H}_0 = \bigcap_{k=1,...,p} H_0^{(k)}$ is true, the left term defines the actual global level of CTP, the second term the actual global level of ITP, and the third one the actual global level of the GTP which is equal to α. Thus, the first thesis is proven.

On the contrary, if the state of nature implies that $\mathbf{H}_0 = \bigcap_{k=1,...,p} H_0^{(k)}$ is false, the left

term defines the power of CTP, the second term the power of ITP, and the third one the power of the GTP. Thus, also the second thesis is proven. □

Proof. **(Theorem 2.3)**
Let $k \in \{1, 2, ..., p\}$ be an index referring to the kth component of the basis representation. Let $\mathcal{R}^{(k)}_{\alpha,ITP}$ be the event "$H^{(k)}_0$ is rejected by the ITP at level α", $\mathcal{R}^{(k)}_{\alpha,CTP}$ be the event "$H^{(k)}_0$ is rejected by the CTP at level α", and $\mathcal{R}_{\alpha,GTP}$ be the event "$\mathbf{H}_0 = \bigcap_{k=1,...,p} H^{(k)}_0$ is rejected by the GTP at level α". Thanks to the structure of the ITP and of the CTP, all multivariate tests used to correct the kth p-value in the ITP are used to correct the CTP but not vice versa. Moreover, the GTP is among the test used in both the CTP and the ITP to correct the kth p-value. Thus, every time the CTP rejects $H^{(k)}_0$ also the ITP rejects it and every time the ITP rejects $H^{(k)}_0$ also the GTP rejects it. Thus we have that $\mathcal{R}^{(k)}_{\alpha,CTP} \subseteq \mathcal{R}^{(k)}_{\alpha,ITP} \subseteq \mathcal{R}_{\alpha,GTP}$, and consequently that $\mathbb{P}[\mathcal{R}^{(k)}_{\alpha,CTP}] \leq \mathbb{P}[\mathcal{R}^{(k)}_{\alpha,ITP}] \leq \mathbb{P}[\mathcal{R}_{\alpha,GTP}]$. Now, if the state of nature implies that $H^{(k)}_0$ is true, the left term defines the CWER of the CTP and the second term the CWTR of the ITP. Moreover, being single components special kind of intervals, Theorem 2.1 proves that also the CWTR is controlled by the ITP. Thus, the first thesis $CWER^{(k)}_{CTP} \leq CWER^{(k)}_{ITP} \leq \alpha$ is proven.

On the contrary, if the state of nature implies that $H^{(k)}_0$ is false, the left term defines the component-wise power of the CTP, the second term the component-wise power of the ITP, and the third one the power of the GTP. Thus, also the second thesis $\pi^{(k)}_{CTP} \leq \pi^{(k)}_{ITP} \leq \pi_{GTP}$ is proven. □

2.B Permutation tests and non parametric combination

The second step of the ITP consists in performing univariate and multivariate tests on the coefficients of the basis expansion (2.1).

We discuss here in detail a possible approach that can be used to perform the tests. This approach is based on non-parametric permutation (NPC) tests and can be used for any n and p (in particular, also when $p \geq n$). It is based on the combination of joint univariate permutation tests, based on different test statistics.

2.B.1 Univariate tests on basis components

We start describing the method that we propose for the univariate test on basis components. We aim at testing the differences between the two populations for each $k = 1, ..., p$ by means of a univariate test on the kth coefficient, defined by:

$$H^{(k)}_0 : C^{(k)}_1 \overset{\text{d}}{=} C^{(k)}_2 \quad \text{vs} \quad H^{(k)}_1 : C^{(k)}_1 \overset{\text{d}}{\neq} C^{(k)}_2 \tag{2.2}$$

In order to perform a marginal test for each k, we introduce a suitable permutation test, based on a family of data transformations preserving the likelihood under $H^{(k)}_0$, and a suitable test statistic, stochastically larger under $H^{(k)}_1$ than under $H^{(k)}_0$.

In particular, fix the basis component k, and let $\mathbf{c}^{(k)} = (\mathbf{c}^{(k)}_1, \mathbf{c}^{(k)}_2)$ the $n_1 + n_2$ dimensional vector of the coefficients associated to units of the two groups, and $\mathbf{c}^{(k)^*} = (\mathbf{c}^{(k)^*}_1, \mathbf{c}^{(k)^*}_2)$ the vector of the permuted coefficients. We have the total exchangeability under $H^{(k)}_0$, thus the family of transformations is composed by any permutation over the sample units of the observed values. On the other hand, in the case of paired coefficients corresponding to paired functional data,

the exchangeability under H_0 is only within and between couples, i.e., couples cannot be split. It is important to note that, being the different components $C_1, C_2, \ldots, C_p$ possibly dependent, the permutations of the coefficients need to be jointly performed, i.e., each permutation is applied simultaneously to the entire set of coefficients. This is the key to build the multivariate interval-wise tests.

The test statistic $T(\mathbf{c}^{(k)^*})$ used for the univariate permutation tests of the expansion coefficients may depend on the functional basis used to describe data. Indeed, in the permutation framework the test statistic has to be properly chosen in order to reflect the characteristics of data which are expected to change the most under the alternative hypothesis.

Once chosen the test statistic, for each k, the p-value of the corresponding test (2.2) is estimated through a conditional MC algorithm (for details, see Pesarin and Salmaso (2010)), as the proportion of $T(\mathbf{c}^{(k)^*})$ exceeding the value $T(\mathbf{c}^{(k)})$ calculated on the original data set.

To better understand how a test statistic may be selected and how the test statistic may depend on the type of test and on the basis used for the analysis, we report some examples that can be used in different applications.

Example 1: B-spline Basis Suppose that a difference between the two functional populations is suspected to occur exclusively on an unknown region of the domain. Then, a quite natural choice to target this problem is the use of the B-spline basis. In particular, we fix a grid of knots along the abscissa, and express each data through the p coefficients associated to the B-spline basis functions $b_m^{(k)}(t)$ of order m: $y_{ij}(t) = \sum_{k=1}^{p} c_{ij}^{(k)} b_m^{(k)}(t)$ (Bosq, 2000).

If we consider the unpaired case, then, a possible test statistic for each test (2.2) can be defined as the difference between the two sample means of the coefficients (which is inferentially equivalent to the classic t-test statistic though computationally less time consuming and thus more commonly used in permutation tests):

$$T(\mathbf{c}^{(k)^*}) = \frac{1}{n_1} \sum_{i=1}^{n_1} c_{i1}^{(k)^*} - \frac{1}{n_2} \sum_{i=1}^{n_2} c_{i2}^{(k)^*}.$$

If, on the contrary, we consider the paired scenario, the same test statistic can be properly rewritten as the sample mean of the differences between the paired coefficients:

$$T(\mathbf{c}^{(k)^*}) = \frac{1}{n_1} \sum_{i=1}^{n_1} \left(c_{i1}^{(k)^*} - c_{i2}^{(k)^*} \right).$$

Example 2: Fourier Basis Suppose now that data are T-periodic curves, and that we expect a difference between the two populations in a frequency band. Thus, it is natural to express data in the frequency domain by means of a Fourier expansion, which can be expressed in both following representations:

$$y_{ij}(t) = m_{ij}^{(0)} + \sum_{k=1}^{p} \left(a_{ij}^{(k)} \cos\left(\frac{2\pi}{T} kt\right) + b_{ij}^{(k)} \sin\left(\frac{2\pi}{T} kt\right) \right); \qquad (2.3)$$

$$y_{ij}(t) = m_{ij}^{(0)} + \sum_{k=1}^{p} \alpha_{ij}^{(k)} \cos\left(\frac{2\pi}{T} kt + \phi_{ij}^{(k)}\right), \tag{2.4}$$

The first expression (2.3) is exactly of the type (2.1), and associates each frequency k to the coefficients $a_{ij}^{(k)}$ and $b_{ij}^{(k)}$. The second expression (2.4) associates instead each frequency to an amplitude and to a phase coefficient (i.e., $\alpha_{ij}^{(k)}$ and $\phi_{ij}^{(k)}$) leading to a more interesting interpretation. Coherently, for the "$0th$" frequency, we can define the amplitude and phase coefficients as $\alpha_{ij}^{(0)} = |m_{ij}^{(0)}|$ and $\phi_{ij}^{(0)} = \pi[1 - \text{sign}(m_{ij}^{(0)})]/2$. Amplitude and phase coefficients have different properties: amplitude coefficients are defined on $[0, +\infty)$, while phase coefficients are angles defined on $[0, 2\pi]$ and invariant by 2π translations. Thus, it is clear that different test statistics need to be used for testing the two quantities.

In particular in the unpaired scenario, for the amplitude coefficients we will rely on the logarithmic distance of geometric sample means:

$$T_{amp}(\boldsymbol{\alpha}^{(k)^*}) = \left| \log \left(\frac{\left(\prod_{i=1}^{n_1} \alpha_{i1}^{(k)^*}\right)^{1/n_1}}{\left(\prod_{i=1}^{n_2} \alpha_{i2}^{(k)^*}\right)^{1/n_2}} \right) \right|.$$

In the paired case, the same test statistic can be more properly rewritten as:

$$T_{amp}(\boldsymbol{\alpha}^{(k)^*}) = \left| \log \left(\prod_{i=1}^{n_1} \frac{\alpha_{i1}^{(k)^*}}{\alpha_{i2}^{(k)^*}} \right)^{1/n_1} \right|.$$

Instead, for testing the phase coefficients in the unpaired scenario, we will rely on the signed geodesic distance (on the circle S^1) between the geodesic sample means:

$$T_{ph}(\boldsymbol{\phi}^{(k)^*}) = \text{sign}(m_{geo}(\boldsymbol{\phi}_2^{(k)^*}) - m_{geo}(\boldsymbol{\phi}_1^{(k)^*}))d_{geo}(m_{geo}(\boldsymbol{\phi}_1^{(k)^*}), m_{geo}(\boldsymbol{\phi}_2^{(k)^*})).$$

In the paired case, we will use the geodesic sample mean of the signed geodesic distances:

$$T_{ph}(\boldsymbol{\phi}^{(k)^*}) = m_{geo}[\{\text{sign}(\phi_{i2}^{(k)^*} - \phi_{i1}^{(k)^*})d_{geo}(\phi_{i1}^{(k)^*}, \phi_{i2}^{(k)^*})\}_{i=1,\dots,n_1}],$$

with the signed geodesic distance and the geodesic sample mean defined according to:

$$d_{geo}(\phi_1, \phi_2) = min\{|\phi_1 - \phi_2|, |2\pi - (\phi_1 - \phi_2)|\}, \quad \phi_1, \phi_2 \in [0, 2\pi);$$

$$m_{geo}(\phi_1, \phi_2, \dots, \phi_q) = \underset{\phi}{\text{argmin}} \sum_{l=1}^{q} [d_{geo}(\phi_l, \phi)]^2 \quad \phi_i \in [0, 2\pi);$$

$$\text{sign}(\phi_2 - \phi_1) = \begin{cases} +1 & \text{if } 0 \le \phi_2 - \phi_1 \le \pi \text{ or } -2\pi \le \phi_2 - \phi_1 < -\pi \\ -1 & \text{if } -\pi < \phi_2 - \phi_1 \le 0 \text{ or } \pi < \phi_2 - \phi_1 \le 2\pi \end{cases}.$$

2.B.2 Multivariate tests on intervals of components

To build the multivariate tests on each interval, we can exploit the NPC methodology, which consists in the construction of suitable combinations of the univariate test statistics in order to obtain multivariate tests on each interval. In the

following, as an example, we will illustrate how to obtain a bivariate test from two univariate tests along the NPC philosophy. The extension to multivariate NPC's is straightforward, and detailed in Pesarin and Salmaso (2010).

Let us indicate with $T_0^{(1)^*}$ and $T_0^{(2)^*}$ the observed values of the two univariate statistics related to variables X_1 and X_2 and with $T_b^{(1)^*}$ and $T_b^{(2)^*}$ the values induced by the permutation b of the bivariate data set containing the realizations of (X_1, X_2). Then, select a combining function, i.e., a continuous non increasing function $\psi : [0,1]^2 \to \mathbb{R}$ which is symmetric on the two arguments and attain its maximal value when at least one argument attains zero. Some possible examples are the Fisher combining function: $\psi_F(x_1, x_2) = -2(\log x_1 + \log x_2)$; the Liptak combining function, based on the inverse of the Normal cdf Φ: $\psi_L(x_1, x_2) = (\Phi^{-1}(1 - x_1) + \Phi^{-1}(1 - x_2))$; the Tippett combining function, based on the maximum test statistic: $\psi_L(x_1, x_2) = \max(1 - x_1; 1 - x_2)$. Finally define $T_b^{(1,2)^*} = \psi(L_b^{(1)}, L_b^{(2)})$ where $L_b^{(1)}$ and $L_b^{(2)}$ are the marginal survival functions of the two test statistics $T^{(1)}$ and $T^{(2)}$ evaluated in $T_b^{(1)^*}$ and $T_b^{(2)^*}$, respectively. Analogously define $T_0^{(1,2)^*} = \psi(L_0^{(1)}, L_0^{(2)})$. The p-value of the joint bivariate test is now simply defined as the proportion of permutations providing $T_b^{(1,2)^*} > T_0^{(1,2)^*}$. Note that $L_0^{(1)}$ and $L_0^{(2)}$ coincide with the p-values of the two original univariate tests. In the practice, the marginal survival functions (and the descending p-values) can be estimated by means of a conditional MC (i.e., just B randomly selected permutations are used). In this case we have: $\hat{L}_b^{(1)} = \frac{\sum_{q=1}^{B} \mathbb{I}(T_q^{(1)^*} \leq T_b^{(1)^*}) + 1/2}{B+1}$ and $\hat{L}_b^{(2)} = \frac{\sum_{q=1}^{B} \mathbb{I}(T_q^{(2)^*} \leq T_b^{(2)^*}) + 1/2}{B+1}$. The p-value of the joint test is of course estimated by $\hat{L}_0^{(1,2)} = \frac{\sum_{q=1}^{B} \mathbb{I}(T_q^{(1,2)^*} \leq T_0^{(1,2)^*}) + 1/2}{B+1}$. For further details about NPCs procedure please refer to Pesarin and Salmaso (2010).

2.C Analysis of the NASA Temperature Data

In this section we report the analysis of daily temperatures registered by NASA satellites in the region ($45^o - 46^o$ North, $8^o - 9^o$ East) including the city of Milan (Italy) from July 1983 to June 2005 and stored in the NASA database *Earth Surface Meteorology for Solar Energy*[1]. The aim of this analysis is to test for the mean function of Milan temperature yearly profiles.

In the application, we identify the 22 years available as sample units ($n = 22$) and the 365 records available for each year as 365 point-wise evaluations of the functional data ($J = 365$) (Figure 2.6), and we aim at testing the mean function of the functional population which data are assumed to be drawn. Because of the periodic nature of these data and because of their daily resolution we perform an ITP starting from the coefficients of a truncated Fourier expansion (2.3) of dimension 365 and period T equal to one year.

In particular we want to see if the functional mean of the data is expressed only on some particular frequencies, by selecting, among the frequencies $k = 0, ..., (J-1)/2 = 182$, the ones whose contribution to the mean function is significantly different from zero. In order to answer this question, we test if the

[1]These data were obtained from the NASA Langley Research Center Atmospheric Science Data Center Surface meteorological and Solar Energy (SSE) web portal supported by the NASA LaRC POWER Project. Data are freely available at: NASA Surface Meteorology and Solar Energy, A Renewable Energy Resource web site (release 6.0): http://eosweb.larc.nasa.gov

mean function is equal to $\mu_0 \equiv 0$ on each frequency. In detail, assuming the functional population to be symmetrically distributed around its mean function, for each frequency $k > 0$ we perform a bivariate test to test the null hypothesis: center$[(A^{(k)}, B^{(k)})] = (0, 0)$, based on the joint changes of the signs of vectors $(a_i^{(k)}, b_i^{(k)})$ and on the Hotelling T^2 statistic:

$$T(\mathbf{a}^{(k)^*}, \mathbf{b}^{(k)^*}) = (\bar{a}^{(k)^*} - a_0^{(k)}, \bar{b}^{(k)^*} - b_0^{(k)})' S_{k,k}^* (\bar{a}^{(k)^*} - a_0^{(k)}, \bar{b}^{(k)^*} - b_0^{(k)}).$$

For the $0 - th$ frequency, we perform a univariate permutation test based on the squared of the univariate Student t statistic and on the change of the signs of the coefficients $m_i^{(0)}$. Finally, we obtain the p-value heat-map (top panel of Figure 2.6) by combining the tests mentioned above as shown in subsection 2.2.1 relying on the Fisher combination function. In the top panel of Figure 2.6, we represent the result of each test included in the family explored by the ITP. In particular, the horizontal axis is associated to the interval central frequencies and the vertical one to the width of the tested band. Each pixel of the image represents a single multivariate test associated to a specific band and its color represents the corresponding p-value (blue corresponds to low p-values and yellow to high p-values). Please remember that p-value heat-map is periodic in the horizontal direction. The adjusted p-value for the kth frequency is the maximal value of the p-values all the tests lying in the upside-down cone with vertex in correspondence with the univariate test for the kth frequency. Thus, the blue triangle of the p-value heat-map resulting from the NASA case study indicates that the only significant frequencies are the two lower ones (i.e., all multivariate tests pertaining to bands including the constant term or the first harmonic present low p-values).

For convenience, the central panel of Figure 2.6 reports for each frequency its unadjusted p-value (empty dots), and its adjusted p-value (full dots), . According to the adjusted p-values, just the first two frequencies (i.e., the constant term and the sinusoids of period one-year) contribute significantly to the mean function.

The ITP thus suggests an easy description of the mean function as a vertically translated sinusoid of period one year. In detail, this sinusoid is characterized by an annual average temperature of 9.023°C and an annual excursion of 21.771°C. Thanks to this reduced representation we can also estimate the 18^{th} January as the coldest day of the year with a mean temperature of -1.891°C and the 20^{th} July as the warmest day of the year with a mean temperature of 19.880°C.

To appreciate the information provided by the ITP, in the lower panel of Figure 2.6, together with the original data (dashed light lines), we report the sample mean (bold solid red line) and the sample mean restricted just to the zero-th and the first frequencies (bold solid blue line), that is the estimate of the mean suggested by the ITP. Note how the high-frequency fluctuations that characterize the sample mean (clearly related to the specific sample at hand) are instead not present in the second estimate, as considered not significant by the ITP.

As a comparison with other inferential procedures that can be applied to the coefficients of the basis expansion, let us mention the fact that: the CTP is not feasible for $p = 365$ (i.e., more than 10^{109} multivariate tests would be needed). The global test of course rejects the null hypothesis that the func-

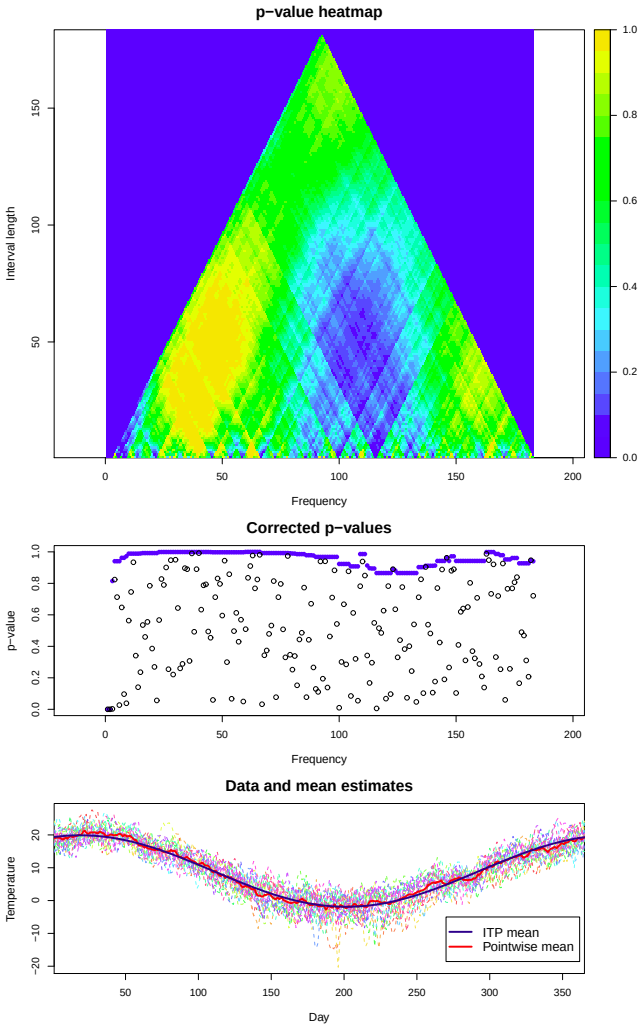


Figure 2.6: NASA case study. Top : p-values heat-map of the ITP. Center: unadjusted p-values (empty dots), and adjusted p-values (full dots) provided by the ITP. Bottom: curves of daily temperatures data (dashed light lines), sample mean (bold solid red line), and mean as estimated according to the ITP results (bold solid blue line).

tion population is centered on zero but it cannot detect which frequencies are not centered on zero. Finally, like the ITP, both the Bonferroni-Holm and the Benjamini-Hochberg corrections of the univariate tests detect just the zero-th and the first frequencies as not centered on zero. Note that the latter correction procedures obtain adjusted p-values by comparing the p-values of univariate tests while the ITP obtains the adjusted p-values by comparing the p-values of a family of multivariate tests (i.e., the ones related to bands) thus exploiting possible dependencies among components.

As a final comment, note that the Fourier expansion of temporal signals is common practice in engineering. Nevertheless, in that field, important frequencies are detected by means of amplitude thresholding and/or frequency filters tuned according to some specific knowledge about the physics (typical amplitude and frequencies of the signal) and/or about the instruments (typical amplitude and frequencies of the noise). The selection criterion derived by the application of the ITP is instead purely statistical and exclusively relies on the observed signals, and it can thus be applied also in context not provided with any quantitative prior knowledge about the problem.

Chapter 3

Component-Wise Inference on Functional-on-Scalar Linear Models

Abstract

We introduce a distribution-free procedure for testing a functional-on-scalar linear model with fixed effects. The procedure does not only test the global hypothesis on all the domain, but also selects the intervals where statistically significant effects are detected. We prove that the proposed tests are provided with an asymptotic interval-wise control of the family-wise error rate, i.e., the probability of falsely rejecting any interval of true null hypotheses. The procedure is then applied to one-leg hop data from a study on anterior cruciate ligament injury. We compare knee kinematics of three groups of individuals, taking individual-specific covariates into account.
Keywords: Functional Data, Inference, Permutation Test, Linear Models

3.1 Introduction

Functional data analysis (FDA) is a relatively new, dynamically developing, research area within the field of statistics. In recent literature, linear models for functional data have been widely studied (see, e.g., Fan and Zhang 2000; Abramovich and Angelini 2006; Cardot et al. 2007; Reiss et al. 2010; Gertheiss et al. 2013; Abramowicz et al. 2014).

In this paper we consider a functional-on-scalar linear model. In detail, we model a functional response with a set of covariates multiplied by functional parameters. Such model finds its application in a wide range of research fields where modern techniques enable collection of high-resolution data. In this context, many of the empirically relevant questions do not only address the effect of covariates on a functional response, but also require identification of significant domain subsets.

We focus on a distribution-free method and therefore propose to use a least squares method for parameter estimation. Parameter estimation of the functional model is handled by first representing the functional response and the

functional regression parameters in terms of a suitable functional basis. The functional estimation problem is thus decomposed into a family of corresponding linear models of univariate response variables, one for each of the coefficients (components) of the basis expansion. Hence, least squares estimation methods for linear models with univariate response variables can be used to estimate the functional linear model (see Section 3.2).

Forming valid tests of various hypotheses about the functional regression parameters, with control of the error rate, is not straightforward. One solution adopted in the literature is to develop global tests for the parameters of the model. Such tests investigate if a covariate has a significant effect on the response, but does not provide any domain selection (Cuevas et al., 2004; Abramovich and Angelini, 2006; Antoniadis and Sapatinas, 2007; Cardot et al., 2007; Schott, 2007; Cuesta-Albertos and Febrero-Bande, 2010; Zhang and Liang, 2014). Another approach, proposed in Fan and Zhang (2000); Reiss et al. (2010); Ramsay and Silverman (2005), is to provide point-wise confidence bands for the functional parameters. The results indicate in which parts of the domain the covariates have an effect, but not at which significance level. As clearly discussed in Ramsay and Silverman (2005, pp. 243–244), point-wise limits are not equivalent to confidence regions for the entire estimated curves. Assuming that data are expressed through a functional basis, inference can be based directly on the expansion coefficients, as proposed by Spitzner et al. (2003). In the latter work, single-component tests are performed, and their p-values adjusted with the Bonferroni-Holm procedure (Holm, 1979). In this way, results are compensated for the many dependent tests performed on the same data set. A drawback with this procedure is that it is typically too conservative, and needs a relevant dimensional reduction of data in order to detect significant functional parameters.

In our work, we follow the same line of research proposed by Spitzner et al. (2003), introducing a less conservative p-values adjustment, which rely on the properties of functional data, and does not require any dimensional reduction of the functional data set. The continuous nature of functional models expressed in terms of a basis expansion such as B-splines, typically implies that neighbouring basis coefficients present a positive dependence. Combinations of neighbouring component-wise (dependent) tests thus have the potential to more easily detect parts of the domain where a functional regression parameter is significantly different from zero. Therefore, in our paper we restrict multiple comparisons of component-wise tests to intervals of neighbouring components and use the Interval Testing Procedure (ITP) introduced by Pini and Vantini (2013), which is based on single- and multiple-component tests. The single-component tests are based on Freedman and Lane permutation schemes (Freedman and Lane, 1983), which do not rely on any distributional assumptions. Further, we use a Non-Parametric Combination (NPC) procedure to obtain simultaneous tests on intervals of components. The NPC procedure is a computationally efficient procedure which preserves the exactness and consistency properties of single-component tests. For further details, we refer to Pesarin and Salmaso (2010). Using the ITP, for each basis component we obtain an adjusted p-value, which is used to select the significant component intervals. Such tests are provided with an interval-wise control of the Family Wise Error Rate (FWER). In detail,

this control implies that the probability of falsely rejecting any interval of basis components associated to true null hypotheses is controlled at the desired significance level. We prove that the proposed tests are exact or asymptotically exact.

The paper is outlined as follows: in Section 3.2, we describe the functional-on-scalar linear model, discussing the methodology proposed for functional parameter estimation and inference. Section 3.3 reports the theoretical properties of the proposed methodology. The proofs of theorems of Section 3.3 are reported in Appendix 3.A. Finally, Appendix 3.B reports some details on the Freedman and Lane permutation scheme, while Appendix 3.C briefly describes the NPC procedure. Two different applications of the methodology described in this chapter are reported in Chapters 5 and 6. In the first case, the methodology is applied to a functional ANCOVA on time-varying knee motion, whereas in the second case, it is applied to a functional two-way ANOVA on laser emission.

3.2 Methodology

3.2.1 The functional-on-scalar linear model

Suppose we have observed a sample of n continuous random functions $\{y_i(t)\}_{i=1,\dots,n}$, over time t: $t \in [a, b]$. We want to study the following functional-on-scalar linear model:

$$y_i(t) = \beta_0(t) + \sum_{l=1}^{L} \beta_l(t) x_{li} + \varepsilon_i(t), \quad i = 1, ..., n, \tag{3.1}$$

where $x_{1i}, ..., x_{li} \in \mathbb{R}$ are known scalar covariates and $\beta_l(t)$, $l = 0, ..., L$, are the fixed functional regression parameters. The errors $\varepsilon_i(t)$, $t \in [a, b]$ are *i.i.d.* (with respect to units) zero-mean random functions (not necessarily Gaussian) with finite total variance, i.e.,

$$\int_a^b \mathbb{E}\left[\varepsilon_i(t)\right]^2 dt < \infty. \tag{3.2}$$

We assume that, for each $i = 1, \dots, n$, $y_i(t)$ can be expressed in terms of basis functions $\{\phi^{(k)}(t)\}_{k=1}^{p}$, i.e.,

$$y_i(t) = \sum_{i=1}^{p} {y_i}^{(k)} \phi^{(k)}(t).$$

Whenever functional data are described through a basis expansion, we can perform inference directly on the set of coefficients representing the data. Therefore, we can project the model (3.1) on the functional space spanned by the basis:

$$\sum_{k=1}^{p} {y_i}^{(k)} \phi^{(k)}(t) = \sum_{k=1}^{p} {\beta_0}^{(k)} \phi^{(k)}(t) + \sum_{l=1}^{L} \sum_{k=1}^{p} {\beta_l}^{(k)} \phi^{(k)}(t) x_{li} + \sum_{k=1}^{p} {\varepsilon_i}^{(k)} \phi^{(k)}(t),$$

for all $t \in [a, b]$, which leads to:

$$\sum_{k=1}^{p} \left[{y_i}^{(k)} - {\beta_0}^{(k)} - \sum_{l=1}^{L} {\beta_l}^{(k)} x_{li} - {\varepsilon_i}^{(k)} \right] \phi^{(k)}(t) = 0 \qquad \forall t \in [a, b]. \tag{3.3}$$

Since $\{\phi^{(k)}(t)\}_{k=1}^{p}$ is a basis, equation (3.3) holds if

$$y_i{}^{(k)} = \beta_0{}^{(k)} + \sum_{l=1}^{L} \beta_l{}^{(k)} x_{li} + \varepsilon_i{}^{(k)}, \quad \forall k = 1, ..., p \tag{3.4}$$

holds. Therefore, we can express model (3.1) as a family of p scalar-on-scalar linear models, with errors pertaining to the same sample unit i possibly dependent. Moreover, we have that:

$$0 = \mathbb{E}[\varepsilon_i(t)] = \mathbb{E}\left[\sum_{k=1}^{p} \varepsilon_i{}^{(k)} \phi^{(k)}(t)\right] = \sum_{k=1}^{p} \mathbb{E}[\varepsilon_i{}^{(k)}] \phi^{(k)}(t) \qquad \forall t \in [a, b],$$

and hence $\mathbb{E}[\varepsilon_i^{(k)}] = 0$ for all $k = 1, ..., p$. From (3.2) and the fact that $\{\phi^{(k)}(t)\}_{k=1}^{p}$ is a basis, we also have that for $k = 1, \ldots, p$, $\mathbb{E}[\varepsilon_i^{(k)^2}] < \infty$. Finally, the independence of the random functions $\varepsilon_i(t)$, $t \in [a, b]$, implies independence across units of the coefficients $\varepsilon_i{}^{(k)}$. Therefore, for fixed k, the error terms $\varepsilon_i{}^{(k)}$, $i = 1, \ldots, n$ are *i.i.d.* zero-mean random variables with finite variance. Note that we are not making assumptions instead on the auto-covariance structure of the $\varepsilon_i(t)$'s. Hence, for fixed i, the errors $\varepsilon_i^{(k)}$, $k = 1, \ldots, p$, are not assumed independent.

In practice, we often can not observe the complete response functions $y_i(t)$, $i = 1, \ldots, n$, and need to estimate them based on the finite number of observations. We refer to, e.g., Ramsay and Silverman (2005) for a discussion about the choice of basis used to represent data and methods used to estimate the coefficients.

3.2.2 Model estimation

The ordinary least squares (OLS) estimators of the functional parameters $\beta_l(t)$, $l = 0, \ldots, L$, can be found by minimizing the sum over units of the L^2 distances between the functional data $y_i(t)$ and the quantity $\beta_0(t) + \sum_{l=1}^{L} \beta_l(t) x_{li}$ with respect to $\beta_l(t)$, $l = 0, \ldots, L$ (Ramsay and Silverman, 2005):

$$\sum_{i=1}^{n} \int_a^b \left(y_i(t) - \beta_0(t) - \sum_{l=1}^{L} \beta_l(t) x_{li} \right)^2 \mathrm{d}t. \tag{3.5}$$

The minimization can be done separately for each coefficient of the basis expansion, even in presence of non-orthonormal basis components. Indeed, when using a basis expansion, (3.5) can be written as:

$$\sum_{i=1}^{n} \int_a^b \left[\sum_{k=1}^{p} \left(y_i^{(k)} - \boldsymbol{\beta}^{(k)'} \mathbf{x}_i \right) \phi^{(k)}(t) \right]^2 \mathrm{d}t, \tag{3.6}$$

where $\boldsymbol{\beta}^{(k)} = (\beta_0^{(k)}, ..., \beta_L^{(k)})'$ and $\mathbf{x}_i$ is the i-th row of the design matrix $X_n \in \mathbb{R}^{(n \times (L+1))}$ ($[X_n]_{i,1} = 1$, $\forall i = 1, ..., n$; $[X_n]_{i,j} = x_{j-1,i}$, $i = 1, ..., n$, $j = 2, ..., L+1$). Equation (3.6) is equivalent to:

$$\sum_{i=1}^{n} \sum_{k_1=1}^{p} \sum_{k_2=1}^{p} \left(y_i^{(k_1)} - \boldsymbol{\beta}^{(k_1)'} \mathbf{x}_i \right) \left(y_i^{(k_2)} - \boldsymbol{\beta}^{(k_2)'} \mathbf{x}_i \right) \int_a^b \phi^{(k_1)}(t) \phi^{(k_2)}(t) \mathrm{d}t,$$

which can be written using matrix notation as

$$\sum_{i=1}^{n} \left(\mathbf{y}_i - \boldsymbol{\beta}' \mathbf{x}_i \right)' W \left(\mathbf{y}_i - \boldsymbol{\beta}' \mathbf{x}_i \right), \tag{3.7}$$

where $\mathbf{y}_i = (y_i^{(1)}, ..., y_i^{(p)})' \in \mathbb{R}^p$, $\boldsymbol{\beta} \in \mathbb{R}^{((L+1)\times p)}$ is the matrix of coefficients, $[\boldsymbol{\beta}]_{l,k} = \beta_l^{(k)}$, and $W \in \mathbb{R}^{p\times p}$ is the matrix of inner products between basis functions $[W]_{k_1k_2} = \int_a^b \phi(t)^{(k_1)}\phi(t)^{(k_2)}\mathrm{d}t$. As shown in Johnson and Wichern (2007), for any positive definite matrix W, we have that:

$$\underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^{n} \left(\mathbf{y}_i - \boldsymbol{\beta}'\mathbf{x}_i\right)' W \left(\mathbf{y}_i - \boldsymbol{\beta}'\mathbf{x}_i\right) = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^{n} \left(\mathbf{y}_i - \boldsymbol{\beta}'\mathbf{x}_i\right)' \left(\mathbf{y}_i - \boldsymbol{\beta}'\mathbf{x}_i\right)$$

that is, in the minimization, W can be replaced with the identity. Note that:

$$\sum_{i=1}^{n} \left(\mathbf{y}_i - \boldsymbol{\beta}'\mathbf{x}_i\right)' \left(\mathbf{y}_i - \boldsymbol{\beta}'\mathbf{x}_i\right) = \sum_{k=1}^{p}\sum_{i=1}^{n} \left(y_i^{(k)} - \boldsymbol{\beta}^{(k)'}\mathbf{x}_i\right)^2$$

and hence the minimization problem on the left hand side with respect to $\boldsymbol{\beta}$ is reduced to the family of p independent minimization problems, one for each component $k = 1, \ldots, p$. For each k, $\sum_{i=1}^{n} \left(y_i^{(k)} - \boldsymbol{\beta}^{(k)'}\mathbf{x}_i\right)^2$ is minimized by the OLS estimate $\hat{\boldsymbol{\beta}}^{(k)} = (\hat{\beta}_0^{(k)}, \ldots, \hat{\beta}_L^{(k)})$ of $\boldsymbol{\beta}^{(k)}$. Therefore $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}^{(1)}, \ldots, \hat{\boldsymbol{\beta}}^{(p)})$ is also the global OLS estimate minimizing (3.6). Hence, for each $l = 0, \ldots, L$, the estimate of the functional regression parameters $\beta_l(t)$ is

$$\hat{\beta}_l(t) = \sum_{k=1}^{p} \hat{\beta}_l^{(k)}\phi^{(k)}(t). \tag{3.8}$$

It is also possible to establish asymptotic properties of the OLS estimates on each basis component k. Consider the following standard conditions:

C1 The matrix $X_m'X_m$ is non-singular for some $m \geq 1$ (implying that it is non-singular for all $n \geq m$), and its inverse $V = X_n'X_n$ is s.t. the elements $[V^{-1}]_{ij} \to 0$ as $n \to \infty$, for all $i, j = 1, ..., L+1$.

C2 For each $k = 1, ..., p$, the regression errors $\varepsilon_i^{(k)}$ satisfy:

$$\sup_{i=1,\ldots,n} \mathbb{E}\left[\varepsilon_i^{(k)^2}\right] < \infty.$$

Under conditions *C1*-*C2*, we have that for each $k = 1, \ldots, p$, the obtained OLS estimates $\hat{\beta}_0^{(k)}, \ldots, \hat{\beta}_L^{(k)}$ are asymptotically strongly consistent estimates of $\beta_0^{(k)}, \ldots, \beta_L^{(k)}$ (see, Lai et al., 1979). Condition *C1* is a sufficient condition for finding an explicit expression of the OLS estimates, and guarantees convergence in probability. Condition *C2* assures almost sure convergence.

3.2.3 Model inference

One of the main challenges with inference for functional linear model (3.1) is performing valid tests of hypotheses on functional regression parameters. Analogously to the classical framework, we are interested in testing the hypotheses on the full model, i.e.,

$$\begin{cases} H_{0,F} : \beta_l(t) = 0 & \forall l \in 1, \ldots, L,\ \forall t \in [a,b] \\ H_{1,F} : \beta_l(t) \neq 0 & \text{for some } l \in \{1, \ldots, L\} \text{ and } t \in [a,b] \end{cases} \tag{3.9}$$

together with tests of significance for specific functional parameter $l \in \{0, \dots, L\}$:

$$\begin{cases} H_{0,l} : \beta_l(t) = 0 & \forall t \in [a, b] \\ H_{1,l} : \beta_l(t) \neq 0 & \text{for some } t \in [a, b]. \end{cases} \tag{3.10}$$

In the most general case we are interested in testing linear hypotheses on the functional parameters of the regression. This is done by specifying a combination matrix C. Let $C \in \mathbb{R}^{(q \times (L+1))}$ be a real-valued full rank matrix, where $q \leq L+1$ denotes the number of hypotheses on the functional regression parameters to be jointly tested. Moreover, let $\mathbf{c}_0(t) = (c_{01}(t), ..., c_{0q}(t))'$ be a vector of fixed functions from the space spanned by the basis functions $\{\phi^{(k)}(t)\}_{k=1,\dots,p}$. Denote by $\boldsymbol{\beta}(t) = (\beta_0(t), ..., \beta_L(t))'$ the vector of functional regression parameters. We are in general interested in testing hypotheses of the form:

$$\begin{cases} H_{0,C} : C\boldsymbol{\beta}(t) = \mathbf{c}_0(t) & \forall t \in [a, b] \\ H_{1,C} : C\boldsymbol{\beta}(t) \neq \mathbf{c}_0(t) & \text{for some } t \in [a, b]. \end{cases} \tag{3.11}$$

where the j-th element of vector $C\boldsymbol{\beta}(t)$ is a function obtained by means of a linear combination of the functional regression parameters $\beta_l(t)$ with weights $[C]_{jl}$: $[C\boldsymbol{\beta}(\boldsymbol{t})]_j = \sum_{l=0}^{L} [C]_{jl}\beta_l(t)$, $j = 1, ..., q$. There are two important special cases of the general functional linear hypotheses:

1. Let $q = L$, $C = (\mathbf{0}|I_L) \in \mathbb{R}^{(L \times (L+1))}$, and $\mathbf{c_0}(\mathbf{t}) = \mathbf{0} \in \mathbb{R}^L$, where I_L is $L \times L$ identity matrix. Then, we obtain a functional F-test on the regression model, reducing the hypotheses in (3.11) to the hypotheses in (3.9);

2. For a fixed l, let $q = 1$, $C \in \mathbb{R}^{1 \times (L+1)}$ with $C_{1r} = 1$ if $r = l$ and 0 otherwise, and $c(t) = 0$. Then we obtain the functional t-test on regression parameter l, reducing the hypotheses in (3.11) to the hypotheses in (3.10).

By using the basis representation of $\boldsymbol{\beta}(t)$ and $\mathbf{c}_0(t)$, functional linear hypotheses are translated into a family of p linear hypotheses pertaining the components of the basis expansion $k = 1, \dots, p$:

$$\begin{cases} H_{0,C}^{(k)} : C\boldsymbol{\beta}^{(k)} = \mathbf{c}_0^{(k)} \\ H_{1,C}^{(k)} : C\boldsymbol{\beta}^{(k)} \neq \mathbf{c}_0^{(k)}, \end{cases} \tag{3.12}$$

where $\mathbf{c}_0^{(k)} \in \mathbb{R}^q$ is a vector composed by the k-th coefficients of the basis expansion of vector $\mathbf{c}_0(t)$ with the basis expansion performed for each element of the vector. Hence, as opposed to model estimation, the problem of inference for multiple components is not straightforward, as it involves a, possibly high-dimensional, family of dependent statistical tests.

In this paper we use interval testing procedure for controlling the probability of falsely rejecting at least one true null hypothesis of the family, i.e., the FWER. The ITP is a three step procedure involving a basis expansion of functional data, the testing of each multiple-component hypothesis pertaining intervals of basis coefficients, and a multiplicity correction providing an interval-wise control of the FWER. In the following paragraphs we first give some details on the starting point of the ITP, that is the single-component testing, and then describe the construction of multiple-component tests and the p-value correction.

Single component testing

For hypotesis testing on a single component k, we use permutation tests based on the Freedman and Lane permutation scheme (Freedman and Lane, 1983), which is briefly described in Appendix 3.B. This permutation strategy is the most commonly used for linear models, and presents many advantages with respect to other techniques (Davison and Hinkley, 1997; Anderson and Legendre, 1999; Anderson and Robinson, 2001; Zeng et al., 2011; Winkler et al., 2014). In particular, it can be shown empirically that its power is typically higher than the one of tests based on other permutation schemes (Anderson and Legendre, 1999; Winkler et al., 2014). Permutation tests based on the Freedman and Lane scheme are based on the permutations of the estimated residuals under the reduced model (i.e., the linear model under the null hypothesis of the test).

For the fixed component k, to perform the F-test with hypotheses

$$\begin{cases} H_{0,F}^{(k)} : \beta_l^{(k)} = 0 & \forall l \in 1, \dots, L \\ H_{1,F}^{(k)} : \beta_l^{(k)} \neq 0 & \text{for some } l \in \{1, \dots, L\} \end{cases} \tag{3.13}$$

we use the F-test statistic:

$$T_F^{(k)} = \frac{(n-L)\sum_{i=1}^n (\hat{y}_i^{(k)} - \bar{y}^{(k)})^2}{L \sum_{i=1}^n (y_i^{(k)} - \hat{y}_i^{(k)})^2}, \tag{3.14}$$

where $\hat{y}_i^{(k)} = \hat{\beta}_0^{(k)} + \sum_{l=1}^L \hat{\beta}_l^{(k)} x_{li}$ are the fitted values of the response coefficients ${y_i}^{(k)}$, and $\bar{y}^{(k)} = \sum_{i=1}^n {y_i}^{(k)}/n$ is the sample mean of the response coefficients.

To perform the t-test for the l-th functional regression parameter, i.e., to test

$$\begin{cases} H_{0,l}^{(k)} : \beta_l^{(k)} = 0 \\ H_{1,l}^{(k)} : \beta_l^{(k)} \neq 0 \end{cases} \tag{3.15}$$

we use the absolute value of the t-test statistic:

$$T_{t,l}^{(k)} = \left| \frac{\hat{\beta}_l^{(k)}}{\text{se}(\hat{\beta}_l^{(k)})} \right|, \tag{3.16}$$

where $\text{se}(\hat{\beta}_l^{(k)})$ is the standard error of $\hat{\beta}_l^{(k)}$. As shown in Pesarin and Salmaso (2010), the test statistic (3.16) is permutationally equivalent to the squared partial correlation coefficient, commonly used in the literature of permutation tests for linear models (see, for instance, Anderson and Robinson 2001).

More in general, to perform each single-component test of linear hypotheses $H_{0,C}^{(k)}$ (3.12), we can use the statistic:

$$T_C^{(k)} = \frac{1}{s^2} \left(C\hat{\boldsymbol{\beta}}^{(k)} - \mathbf{c}_0^{(k)} \right)' \left(C(X_n' X_n)^{-1} C' \right)^{-1} \left(C\hat{\boldsymbol{\beta}}^{(k)} - \mathbf{c}_0^{(k)} \right), \tag{3.17}$$

where $\hat{\boldsymbol{\beta}}^{(k)}$ is the OLS estimate of $\boldsymbol{\beta}^{(k)}$, and $s^2 = (\mathbf{y}^{(k)} - X_n\hat{\boldsymbol{\beta}})'(\mathbf{y}^{(k)} - X_n\hat{\boldsymbol{\beta}})/(n - L + 1)$ is the estimate of the variance of residuals at component k, with $\mathbf{y}^{(k)} = (y_1^{(k)}, \dots, y_n^{(k)})'$.

Multiple tests and p-value correction

The p-values of single-component tests need to be adjusted to provide an interval-wise control of the FWER, according to the interval testing procedure (Pini

and Vantini, 2013). To perform this multiplicity correction, each multivariate hypothesis on intervals of components $\mathcal{I} = \{k_1, k_1 + 1, \ldots, k_2\}$ with $1 \leq k_1 < k_2 \leq p$ need to be tested. In detail, we need to test each hypothesis

$$H_{0,C}^{\mathcal{I}} = \bigcap_{k \in \mathcal{I}} H_{0,C}^{(k)}.$$

Such tests can be approached exploiting the NPC procedure (Pesarin and Salmaso, 2010), which is briefly described in Appendix 3.C. The NPC is a procedure that enables to build multivariate permutation tests by means of combining synchronized univariate permutation tests. The procedure applies in presence of dependence between univariate tests, which is the case in FDA.

Let $\lambda_C^{\mathcal{I}}$ denote the p-value corresponding to the multivariate test on hypothesis $H_{0,C}^{\mathcal{I}}$. The adjusted p-value $\lambda_{ITP,C}^{(k)}$ for the k-th component is then computed as the maximum between all p-values of univariate and multivariate tests containing that component, i.e.:

$$\lambda_{ITP,C}^{(k)} = \max_{\mathcal{I} \ni k} \lambda_C^{\mathcal{I}}.$$

The adjusted p-values can be used to select only the basis components leading to the rejection of the null hypothesis $H_{0,C}^{(k)}$, i.e., the ones with associated adjusted p-value lower than the desired significance level α.

It is important to point out that the ITP takes into account the dependence between the basis coefficients, which in the framework of a functional linear model means that it does not require to specify their covariance structure. Moreover, as tests are based on permutations, the procedure does not require the normality of residuals.

3.3 Theoretical results

In this section, we give theoretical properties of inference on functional-on-scalar linear models performed along the line depicted in Section 3.2. All proofs are reported in Appendix 3.A and the results are valid for the ITP, based on the NPC of tests using the Freedman and Lane scheme.

First, we prove that test of the family of linear hypotheses $\{H_{0,C}^{(k)}\}_{k=1,\ldots,p}$ is provided with an asymptotic interval-wise control of the FWER. Pini and Vantini (2013) proved that, if all univariate and multivariate tests used to build the ITP are exact, the ITP based on the p components of any basis expansion is provided with an interval-wise control of the FWER. This result can be applied directly in the case of the F-test on the regression model, but has to be extended in the more general case of tests on linear hypotheses (including the t-tests on functional regression parameters), as in the latter the exactness of all tests is only asymptotical.

Theorem 3.1. *Under assumptions (C1-C2), the test of the family of linear hypotheses* (3.12) *based on the statistic* $T_C^{(k)}$ (3.17) *is provided with an asymptotic interval-wise control of the FWER. Formally, the ITP-adjusted p-values* $\lambda_{ITP,C}^{(k)}$, $\forall k = 1, ..., p$, *are s.t., for any interval* $\mathcal{I}$ *and any* $\alpha \in (0, 1]$*:*

$$\limsup_{n \to \infty} \mathbb{P}_{H_{0,C}^{\mathcal{I}}} \left[\exists k \in \mathcal{I} \text{ s.t. } \lambda_{ITP,C}^{(k)} \leq \alpha \right] \leq \alpha.$$

Since t-tests are specific cases of linear hypothesis, we obtain directly the following corollary.

Corollary 3.1. *Under assumptions (C1-C2), the test of the family of hypotheses* (3.15) *for the l-th functional regression parameter based on the statistic $T_{t,l}^{(k)}$* (3.16) *is provided with an asymptotic interval-wise control of the FWER.*

Furthermore, the following proposition provides exact results for ITP-based F-test.

Proposition 3.1. *The test of the family of hypotheses* (3.13) *based on the statistic $T_F^{(k)}$* (3.14) *is provided with an exact interval-wise control of the FWER. Formally, the ITP-adjusted p-values $\lambda_{ITP,F}^{(k)}$, $\forall k = 1, ..., p$, are s.t., for any interval $\mathcal{I}$ and any $\alpha \in (0, 1]$:*

$$\mathbb{P}_{H_{0,F}^{\mathcal{I}}}\left[\exists k \in \mathcal{I} \ s.t. \ \lambda_{ITP,F}^{(k)} \leq \alpha\right] \leq \alpha.$$

Next, we focus on the property of consistency of the proposed tests. Let $\mathcal{A}$ denote the set of indexes associated to all components where $H_{0,C}^{(k)}$ is false, i.e., let $H_{1,C}^{\mathcal{A}} = \bigcap_{k\in\mathcal{A}} H_{1,C}^{(k)}$ hold. Then, the following theorem states that the probability of detecting every component of the set $\mathcal{A}$ converges to 1 as the sample size increases.

Theorem 3.2. *The test of the family of linear hypotheses* (3.12) *based on the test statistic $T_C^{(k)}$* (3.17) *is consistent. Formally, for any set $\mathcal{A} \subseteq \{1, ..., p\}$, the ITP-adjusted p-values $\lambda_{ITP,C}^{(k)}$ are s.t.:*

$$\lim_{n\to\infty} \mathbb{P}_{H_{1,C}^{\mathcal{A}}}\left[\forall k \in \mathcal{A}, \lambda_{ITP,C}^{(k)} \leq \alpha\right] = 1.$$

As a consequence, we obtain also consistency results for ITP-based F-test and t-tests.

Corollary 3.2. *The test of the family of hypotheses* (3.13) *based on the F-test statistic $T_F^{(k)}$* (3.14) *is consistent.*

Corollary 3.3. *The test of the family of hypotheses* (3.15) *for the l-th functional regression parameter based on the t-test statistic $T_{t,l}^{(k)}$* (3.16) *is consistent.*

The ITP used in this paper provides the control of FWER on every interval. However, it is possible to consider for the multiplicity correction not only all possible intervals, but also the complementary sets of all intervals. In such way, the control can be extended over the interval complements as well. For details, we refer to Pini and Vantini (2013).

3.4 Discussion

In this work, we introduced a methodology to estimate and test a functional-on-scalar linear model, i.e., a linear model where the response variable is a function and the covariates are fixed scalar variables multiplied by fixed functional parameters. This type of model can be applied whenever functional data are described through a suitable basis expansion. We showed how the initial functional linear

model can be decomposed in a family of dependent linear models, one for each component of the basis expansion.

We provided OLS estimates for the functional regression parameters, as well as tests on the model. Specifically, we provided: (i) a functional F-test for testing the regression model; and (ii) functional t-tests for testing the effects of single covariates. All tests are based on the Interval Testing Procedure (ITP), a non-parametric procedure for testing functional data. We provided theoretical properties for the ITP-based F-test on the regression model and the ITP-based t-tests on the functional regression parameters. In detail, we proved theoretically that the F-test on the regression model is provided with an interval-wise control of the Family Wise Error Rate, implying that the probability of falsely rejecting any interval of true null hypotheses pertaining basis components is controlled. Furthermore, we proved that the F-test on the regression model is consistent, in the sense that the probability of rejecting all false null hypothesis converges to one as the sample size increases. We proved that the t-test on functional regression parameters are provided with an asymptotic interval-wise control of the Family Wise Error Rate, and that they are consistent.

Finally, we mention that the R-package `fdatest` (Pini and Vantini, 2014), available on CRAN, contains the implementation of the inferential procedure on functional-on-scalar linear models presented in this work. The current version of the package requires the evaluation of functional data on a uniform grid. It automatically projects data on a B-spline basis, and performs the F-test on the regression model and all t-tests on regression parameters.

Appendix

3.A Proofs

In this section, we prove the theoretical properties reported in Section 3.3. We first report the theoretical properties of single-components tests based on the Freedman and Lane scheme, i.e., the tests on each component of the basis expansion. Then, we report the theoretical properties of the corresponding multiple-component tests, i.e., the tests on intervals of basis components obtained by means of the NPC of single-components tests. Finally, we prove that the ITP-based tests of linear hypotheses on the functional-on-scalar linear model is provided with an asymptotic interval-wise control of the FWER and that they are consistent. Additionally we show that the ITP-based F-test on the regression model is provided with an exact interval-wise control of the FWER.

3.A.1 Single-component tests

As mentioned above, we first prove the theoretical properties of single-components tests, i.e., the tests on each component of the basis expansion. We start by showing asymptotic exactness of single-component tests on linear hypotheses.

Lemma 3.1. *Under assumptions (C1-C2), and for each component $k = 1, ..., p$, the single-component test of linear hypotheses on the regression parameters* (3.12) *is asymptotically exact.*

Proof. Let $H_{0,C}^{(k)}$ hold, i.e., $C\boldsymbol{\beta}^{(k)} = \mathbf{c}_0^{(k)}$. Under the null hypothesis, the model can be reduced by solving the linear system $C\boldsymbol{\beta}^{(k)} = \mathbf{c}_0^{(k)}$. In detail, since C has full rank, $q \leq L + 1$ regression parameters can be removed from the model. Let $\mathcal{Q}$ denote the set of indexes removed. The reduced model is then $y_i^{(k)} = \sum_{r \notin \mathcal{Q}} \beta_r{}^{(k)} a_r^{(k)} x_{ri} + \varepsilon_i{}^{(k)}$, where $x_{0i} = 1$, $a_r^{(k)}$ are fixed known coefficients (depending only on the solution of linear system $C\boldsymbol{\beta}^{(k)} = \mathbf{c}_0^{(k)}$), and $\varepsilon_i{}^{(k)}$ are *i.i.d.* and zero-mean errors.

The Freedman and Lane permutation scheme is based on the permutations of the residuals $\hat{\varepsilon}_{i,C}^{(k)} = y_i^{(k)} - \sum_{r \notin \mathcal{Q}} \hat{\beta}_{r,C}^{(k)} a_r^{(k)} x_{ri}$, where $\hat{\beta}_{r,C}^{(k)}$, $r \notin \mathcal{Q}$ are the OLS estimate of parameters $\beta_r^{(k)}$ under the reduced model. Under conditions (*C1-C2*), we have strong consistency of the OLS parameter estimates, i.e., in our case: $\hat{\beta}_{r,C}^{(k)} \xrightarrow{a.s.} \beta_r^{(k)}$, $\forall r \notin \mathcal{Q}$. Hence, we also have the strong convergence of the residuals, i.e., $\hat{\varepsilon}_{i,C}^{(k)} \xrightarrow{a.s.} \varepsilon_i^{(k)}$, $\forall i = 1, ..., n$.

The errors $\varepsilon_i{}^{(k)}$ of the reduced linear model are exchangeable. Hence, the likelihood of every permutation is invariant, and equal to $1/n!$. Therefore, the test based on the permutations of the errors $\varepsilon_i{}^{(k)}$ is exact. As $\hat{\varepsilon}_{i,C}^{(k)} \xrightarrow{a.s.} \varepsilon_i^{(k)}$, the residuals are asymptotically exchangeable, i.e., the likelihood of every permutation is asymptotically invariant, and converges to $1/n!$. Hence, the test based on permutations of the residuals is asymptotically exact. □

Asymptotical exactness for the t-test (3.15) is a direct consequence of the above lemma. As an addition to asymptotic results for single component tests on any linear hypothesis, we prove exactness of single-component F-test.

Lemma 3.2. *For each component $k = 1, \dots, p$, the single-component F-test of (3.13) is exact.*

Proof. Under $H_{0,F}^{(k)}$ we have ${y_i}^{(k)} = {\beta_0}^{(k)} + {\varepsilon_i}^{(k)}$. The estimated residuals of this model are $\hat{\varepsilon}_{i,0}^{(k)} = {\beta_0}^{(k)} + {\varepsilon_i}^{(k)} - \hat{\beta}_0^{(k)}$, where $\hat{\beta}_0^{(k)} = \bar{y}^{(k)}$ is the sample mean of the responses ${y_i}^{(k)}$. Note that the quantity ${\beta_0}^{(k)} + {\varepsilon_i}^{(k)} - \hat{\beta}_0^{(k)}$ is permutationally invariant. Hence, the independence between the errors implies the exchangeability of the residuals under $H_{0,F}^{(k)}$. Thus, the test is exact, as it is based on the permutation of exchangeable quantities (Pesarin and Salmaso, 2010). □

In the next step, we verify the consistency of single-component tests on linear hypotheses.

Lemma 3.3. *For each component $k = 1, \dots, p$, the single-component test of linear hypotheses on the regression parameters* (3.12) *based on the test statistic* $T_C^{(k)}$ (3.17) *is consistent.*

Proof. The statement follows directly from the fact that the test statistic $T_C^{(k)}$ is stochastically greater under $H_{1,F}^{(k)}$ than under $H_{0,F}^{(k)}$ (Pesarin and Salmaso, 2010). □

As direct implication of Lemma 3.3, we get the consistency of single-component F-test and t-tests based on test statistics $T_F^{(k)}$ and $T_{t,l}^{(k)}$, respectively.

3.A.2 Multiple-components tests

Next, we investigate the properties of multiple-component tests $H_{0,C}^{\mathcal{I}} = \bigcap_{k \in \mathcal{I}} H_{0,C}^{(k)}$, where $\mathcal{I} = \{k_1, ..., k_2\}$ and $1 \leq k_1 < k_2 \leq p$. To construct these tests from the results of joint single-component tests, we use the NPC methodology. We start by proving the asymptotic exactness of such tests on linear hypotheses.

Lemma 3.4. *Under assumptions (C1-C2), for each interval of components $\mathcal{I}$, the multiple-component test of linear hypotheses on the regression parameters $H_{0,C}^{\mathcal{I}}$ is asymptotically exact.*

Proof. Let $H_{0,C}^{\mathcal{I}}$ hold, i.e., $C\boldsymbol{\beta}^{(k)} = \mathbf{c}_0^{(k)}$, for any $k \in \mathcal{I}$. Under the null hypothesis, and for each $k \in \mathcal{I}$, the model can be reduced by solving the linear system $C\boldsymbol{\beta}^{(k)} = \mathbf{c}_0^{(k)}$. In detail, since C has full rank, $q \leq L+1$ regression parameters can be removed from the model. Let $\mathcal{Q}$ denote the set of indexes removed. The reduced model is then $y_i^{(k)} = \sum_{r \notin \mathcal{Q}} {\beta_r}^{(k)} a_r^{(k)} x_{ri} + {\varepsilon_i}^{(k)}$, where $x_{0i} = 1$, $a_r^{(k)}$ are fixed known coefficients (depending only on the solution of linear systems $C\boldsymbol{\beta}^{(k)} = \mathbf{c}_0^{(k)}$), and ${\varepsilon_i}^{(k)}$ are *i.i.d.* and zero-mean errors.

The NPC applied to the Freedman and Lane permutation scheme is based on the joint permutations (the same for each k) of the residuals $\hat{\varepsilon}_{i,C}^{(k)} = y_i^{(k)} - \sum_{r \notin \mathcal{Q}} \hat{\beta}_{r,C}^{(k)} a_r^{(k)} x_{ri}$, where $\hat{\beta}_{r,C}^{(k)}$, $r \notin \mathcal{Q}$ are the OLS estimate of parameters $\beta_r^{(k)}$ under the reduced model. Under conditions (*C1-C2*), we have strong consistency of the OLS parameters estimates, i.e., in our case: $\hat{\beta}_{r,C}^{(k)} \xrightarrow{a.s.} \beta_r^{(k)}$, $\forall r \notin \mathcal{Q}$, and $\forall k \in \mathcal{I}$. Hence, we also have the strong convergence of the residuals, i.e., $\hat{\varepsilon}_{i,C}^{(k)} \xrightarrow{a.s.} \varepsilon_i^{(k)}$, $\forall i = 1, ..., n$ and $\forall k \in \mathcal{I}$.

The errors ${\varepsilon_i}^{(k)}$ of the linear model are jointly exchangeable. Hence, the likelihood of every joint permutation is invariant, and equal to $1/n!$. So, the test based on the joint permutations of the errors ${\varepsilon_i}^{(k)}$ is exact. As $\hat{\varepsilon}_{i,C}^{(k)} \xrightarrow{a.s.} \varepsilon_i^{(k)}$, $\forall k \in \mathcal{I}$, the residuals are jointly asymptotically exchangeable, i.e., the likelihood of every joint permutation is asymptotically invariant, and converges to $1/n!$. Hence, the test based on joint permutations of the residuals is asymptotically exact. □

Asymptotic exactness of multiple-component t-tests is a direct implication of the above lemma. As in single-component case, we can also show stronger result for multiple-component F-test.

Lemma 3.5. *For each interval of components $\mathcal{I}$, the multiple-component F-test of the regression model is exact.*

Proof. Since all univariate tests are exact and consistent (Corollary 3.2 and Lemma 3.3), the combined test is also exact, due to results of Pesarin and Salmaso (2010). □

We proceed by proving consistency of multiple-component tests on linear hypotheses.

Lemma 3.6. *For each interval of components $\mathcal{I}$, the multiple-component test of linear hypotheses on the regression parameters $H_{0,C}^{\mathcal{I}}$ is consistent.*

Proof. The consistency of the multiple-component test follows directly from the consistency of the corresponding single-component test (Lemma 3.3) and results of Pesarin and Salmaso (2010). □

Once again, since F-test and t-tests are special cases of linear hypothesis tests, the consistency of the multiple-component F-test and t-tests follows immediately from Lemma 3.6.

3.A.3 Properties of IPT-based tests

We start by proving Theorem 3.1, establishing asymptotic interval-wise control of ITP-based tests of linear hypotheses.

Proof of Theorem 3.1. Let $\mathcal{I}$ be an interval of components associated to only true null hypotheses. Consider a component k of the interval, $k \in \mathcal{I}$, and let $\mathcal{K}$ denote the set of all intervals containing the component k. The ITP-adjusted p-value associated to component k is $\lambda_{ITP,C}^{(k)} = \max_{\mathcal{J} \in \mathcal{K}} \lambda_C^{\mathcal{J}}$, where $\lambda_C^{\mathcal{J}}$ is the p-value of the permutation test on the interval $\mathcal{J}$. In particular, as $\mathcal{I} \in \mathcal{K}$, we have that $\lambda_{ITP,C}^{(k)} \geq \lambda_C^{\mathcal{I}}$, and $\mathbb{P}_{H_{0,C}^{\mathcal{I}}}[\lambda_{ITP,C}^{(k)} \leq \alpha] \leq \mathbb{P}_{H_{0,C}^{\mathcal{I}}}[\lambda_C^{\mathcal{I}} \leq \alpha]$. Since all tests are asymptotically exact (Lemmas 3.1 and 3.4), we have:

$$\lim_{n \to \infty} \mathbb{P}_{H_{0,C}^{\mathcal{I}}} \left[\lambda_C^{\mathcal{I}} \leq \alpha \right] = \alpha,$$

and therefore,

$$\limsup_{n \to \infty} \mathbb{P}_{H_{0,C}^{\mathcal{I}}} \left[\exists k \in \mathcal{I} : \lambda_{ITP,C}^{(k)} \leq \alpha \right] \leq \alpha.$$

□

Assertion of Proposition 3.1 follows directly from the results of Pini and Vantini (2013) and the fact that univariate and multivariate tests used to build the procedure are exact (Lemmas 3.2 and 3.5). We now prove Theorem 3.2, which guarantees consistency of ITP-based tests of linear hypothesis.

Proof of Theorem 3.2. Suppose that for a basis component k the alternative hypothesis $H_{1,C}^{(k)}$ is true, i.e. $C\boldsymbol{\beta}^{(k)} \neq \mathbf{c}_0^{(k)}$. Let $\mathcal{K}$ denote the set of every interval containing the component k. If $H_{1,C}^{(k)}$ is true, also each alternative hypothesis pertaining intervals in $\mathcal{K}$ is true. Since each test is consistent (Lemmas 3.3, 3.6), it follows that, for $n \to \infty$, for each $\mathcal{I} \in \mathcal{K}$, the p-value $\lambda_C^{\mathcal{I}}$ converges to zero almost surely. The ITP-adjusted p-value $\lambda_{ITP,C}^{(k)}$ is the maximum among all p-values of the tests containing k, i.e., $\lambda_{ITP,C}^{(k)} = \max_{\mathcal{I} \in \mathcal{K}} \lambda_C^{\mathcal{I}} \to 0$, almost surely. Then, for the ITP-adjusted p-value associated to the k-th component $\lambda_{ITP,C}^{(k)}$, we have:

$$\lim_{n \to \infty} \mathbb{P}_{H_{1,C}^{(k)}} \left[\lambda_{ITP,C}^{(k)} \leq \alpha \right] = 1.$$

The latter holds for any $k \in \mathcal{A}$, where $\mathcal{A}$ denotes the set of all false null hypotheses. Hence, we also have:

$$\lim_{n\to\infty} \mathbb{P}_{H^{\mathcal{A}}_{1,C}} \left[\forall k \in \mathcal{A}, \lambda^{(k)}_{ITP,C} \leq \alpha\right] = 1.$$

□

The Corollary 3.2 and Corollary 3.3 follow immediately from Theorem 3.2, as special cases.

3.B The Freedman and Lane permutation scheme

In this section, we give some details of the implementation of the Freedman and Lane permutation scheme for testing linear hypotheses on regression model for each component k (see eq. (3.4))

$$y_i^{(k)} = \sum_{l=0}^{L} \beta_l{}^{(k)} x_{il} + \varepsilon_i{}^{(k)}, \quad \forall i = 1, \ldots, n,$$

with $x_{i0} = 1$, $\forall i$. Further, we present the the two specific cases: F-test on the regression model; and t-tests on the regression parameters.

The Freedman and Lane permutations are based on the following steps:

i the residuals of the reduced model (that is the linear model under the null hypothesis) are estimated;

ii the residuals of the reduced model are permuted;

iii the permuted responses are computed, through the reduced model and permuted residuals.

For more details about this method, we refer to Freedman and Lane (1983); Anderson and Legendre (1999).

3.B.1 Tests on linear hypotheses

Under the null hypothesis (3.12), the model (3.4) can be reduced by solving the linear system $C\boldsymbol{\beta}^{(k)} = \mathbf{c}_0^{(k)}$. In detail, since C has full rank, $q \leq L+1$ regression parameters can be removed from the model, by expressing them in terms of the others. Let $\mathcal{Q}$ denote the set of indexes of the removed regression parameters. The reduced model is then:

$$y_i^{(k)} = \sum_{r \notin \mathcal{Q}} \beta_r{}^{(k)} \tilde{x}_{ri} + \varepsilon_i{}^{(k)}, \tag{3.18}$$

i.e., the responses can be written in terms of modified covariates $\tilde{x}_{ri} = a_r^{(k)} x_{ri}$, where $a_r^{(k)}$ are fixed known coefficients, depending only on the solution of linear system $C\boldsymbol{\beta}^{(k)} = \mathbf{c}_0^{(k)}$, and $\varepsilon_i{}^{(k)}$ are *i.i.d.* and zero-mean errors.

The residuals of the reduced model can then be estimated as $\hat{\varepsilon}^{(k)}_{i,C} = y_i^{(k)} - \sum_{r\notin\mathcal{Q}} \hat{\beta}^{(k)}_{r,C} \tilde{x}_{ri}$, where $\hat{\beta}^{(k)}_{r,C}$ are the OLS estimates of parameters $\beta_r^{(k)}$, $r \notin \mathcal{Q}$, of model (3.18). Then, the residuals $\hat{\varepsilon}^{(k)}_{i,C}$ are permuted, and the permuted responses are evaluated using the permuted residuals $\hat{\varepsilon}^{(k)^*}_{i,C}$ in the reduced model (3.18):

$$y_i^{(k)^*} = \sum_{r \notin \mathcal{Q}} \hat{\beta}^{(k)}_{r,C} \tilde{x}_{ri} + \hat{\varepsilon}^{(k)^*}_{i,C}. \tag{3.19}$$

3.B.2 F-test for the regression model

In the case of the F-test (3.13), under the null hypothesis all regression parameters except the intercept are null. So, the reduced model is:

$$y_i{}^{(k)} = \beta_0{}^{(k)} + \varepsilon_i{}^{(k)}.$$

The estimated residuals of such model are $\hat{\varepsilon}_{i,F}^{(k)} = y_i{}^{(k)} - \bar{y}^{(k)}$, where $\bar{y}^{(k)}$ is the sample mean of the responses $y_i{}^{(k)}$. Therefore, using the permuted residuals $\hat{\varepsilon}_{i,F}^{(k)^*}$, we get:

$$y_i{}^{(k)^*} = \bar{y}^{(k)} + \hat{\varepsilon}_{i,F}^{(k)^*}.$$

Moreover, in this case permuting the residuals $\hat{\varepsilon}_{i,F}^{(k)}$ is equivalent to permuting the responses $y_i{}^{(k)}$.

3.B.3 t-tests on regression parameters

In the case of t-tests, the model under null hypothesis (3.15) reduces to:

$$y_i^{(k)} = \beta_0{}^{(k)} + \sum_{r \neq l} \beta_r{}^{(k)} x_{ri} + \varepsilon_i{}^{(k)}.$$

The estimated residuals of such model are $\hat{\varepsilon}_{i,l}^{(k)} = y_i^{(k)} - \hat{\beta}_{0,l}^{(k)} + \sum_{r \neq l} \beta_{r,l}{}^{(k)} x_{li}$, where $\hat{\beta}_{r,l}^{(k)}$ are the OLS estimates of the parameters of the reduced model. Then, the permuted responses are:

$$y_i^{(k)^*} = \hat{\beta}_{0,l}^{(k)} + \sum_{r \neq l} \hat{\beta}_{r,l}^{(k)} + \hat{\varepsilon}_{i,l}^{(k)^*}, \tag{3.20}$$

where $\hat{\varepsilon}_{i,l}^{(k)^*}$ are the permuted residuals.

3.C The Non Parametric Combination procedure

The NPC methodology (Pesarin and Salmaso, 2010) allows to build multivariate permutation tests starting from the results of a family of joint univariate permutation tests.

Consider a family of null hypotheses $\{H_0^{(m)}\}_{m \in \mathcal{M}}$, with $\mathcal{M} = \{m_1, \ldots, m_d\}$. Each hypothesis $H_0^{(m)}$ is tested against $H_1^{(m)}$, $m \in \mathcal{M}$, by means of a suitable permutation test, with test statistic $T^{(m)}$. Let $\lambda^{(m)}$ denote the resulting p-value. We want to test the multivariate hypothesis $H_0^{(\mathcal{M})} = \bigcap_{m \in \mathcal{M}} H_0^{(m)}$ against the alternative $H_1^{(\mathcal{M})} = \bigcup_{m \in \mathcal{M}} H_1^{(m)}$, using the results of the univariate tests. The test statistic for such test is the combination of univariate p-values: $\psi(\lambda^{(m_1)}, ..., \lambda^{(m_d)})$, where ψ is any valid combining function, i.e., any function $\psi : [0,1]^d \mapsto \mathbb{R}$ satisfying:

(P1) ψ is non-increasing in each argument;

(P2) ψ is invariant with respect to rearrangements of its arguments:

$$\psi(\lambda^{(m_1)}, ..., \lambda^{(m_d)}) = \psi(\lambda^{(m_1^*)}, ..., \lambda^{(m_d^*)}),$$

where $(\lambda^{(m_1^*)}, ..., \lambda^{(m_d^*)})$ is any rearrangement of $(\lambda^{(m_1)}, ..., \lambda^{(m_d)})$;

(P3) ψ attains its supremum value $\bar{\psi}$ (possibly not finite) even when only one argument attains zero;

(P4) for $\alpha \in (0, 1]$, let ψ_α denote the critical value of the test statistic, i.e., $\psi_\alpha = F_\psi^{-1}(\alpha)$, where F_ψ is the cdf of the test statistic ψ. Then, for a valid combining function, ψ_α is finite and strictly smaller than $\bar{\psi}$.

The following theorem, reported in Pesarin and Salmaso (2010) shows the properties of combined tests:

Theorem 3.3. *If permutation tests for respectively $H_0^{(m)}$ against $H_1^{(m)}$, $m \in \mathcal{M}$ are exact and consistent, then the NPC test based on a combining function ψ satisfying (P1) to (P4) is an exact and consistent test for $H_0^{\mathcal{M}}$ against $H_1^{\mathcal{M}}$.*

Part II

Applications

Chapter 4

One-Way Functional ANOVA: Analysis of Human Movement

Abstract

Regardless of conservative or surgical intervention, anterior cruciate ligament (ACL) ruptures can cause long-term functional deficits. To assist identify and treat deficiencies, three-dimensional motion analysis systems can be used to provide objective data that describes functional performance. In most circumstances, kinematic data are analyzed using traditional methods, which reduce continuous data series into discrete variables. Conversely, functional data analysis (FDA) considers magnitude, shape, and timing of continuous data, thereby providing a more comprehensive overview on motion. Here, we employ a novel ANOVA-based FDA technique to examine the entire time domain of knee-kinematic curves from one-leg hops performed by 95 subjects. All subjects were involved in a long-term follow-up ACL study (KACL20) that involved ACL-ruptured subjects treated ∼20 years ago conservatively with physiotherapy only (ACL_{PT}) or combined with reconstructive surgery (ACL_{R}), and knee-healthy controls (CTRL).
Keywords: Biomechanics, Curve Analysis, Functional Outcomes, Interval Testing Procedure, Lower Extremity, Rehabilitation

4.1 Introduction

Anterior cruciate ligament (ACL) injuries are common worldwide, affecting each year approximately 1/4 million Americans (Silvers and Mandelbaum, 2011) and up to 0.05% of several national populations (Røtterud et al., 2011). In the short-term, acute interventions of such injuries can instigate high socio-economic costs (Brophy et al., 2009), with the injuries themselves causing deficits in knee function, control, and strength still apparent 6 to 9 months post-injury (Xergia et al., 2013). In the middle-to-long term, individuals post-ACL injuries show changes in cartilage morphology up to two-years post-therapeutic interventions (Frobell, 2011); low rates of return to competitive sports despite promising functional outcome scores (Ardern et al., 2011); persisting knee-joint laxity 17-years post-reconstruction on radiographic evaluation in $\sim$ 70% of cases (Ait Si Selmi et al., 2006); and precursors of osteoarthritis or osteoarthritis itself about 25-years

post-reconstruction in $\sim$ 50% (Pernin et al., 2010; Yamaguchi et al., 2006) to 90% of cases (Tengman et al., 2014).

Conservative treatment of ACL ruptures (ACL_{PT}) traditionally involves a physiotherapeutic approach inclusive of analgesic modalities, rehabilitative exercises, and physical activity modification. Alternatively, surgical treatment of ACL ruptures (ACL_R) involves reconstruction combined with physiotherapy. There is a longstanding controversy regarding whether all ACL-ruptured individuals should undergo surgery and the extent to which the long-term functional outcomes differ between conservatively and surgically treated patients (Delincé and Ghafil, 2012). Although some studies report greater knee instability and degeneration and poorer subjective outcomes with conservative treatment (Mihelic et al., 2011), others suggest relatively similar outcomes between treatment approaches (Ageberg et al., 2008; Frobell et al., 2010; von Porat et al., 2006).

In any circumstance, evidence suggests adverse effects persisting more than 20-years post-ACL ruptures (Mihelic et al., 2011; Yamaguchi et al., 2006), with knee function and kinematics never fully returning to normative values (Delincé and Ghafil, 2012). Still, only a few studies have investigated the long-term effects (i.e., $>$ 15 years) of both types of ACL-rupture interventions on knee-joint kinematics during functionally demanding tasks, particularly in relation to matched knee-healthy controls. The existing research in this area provides equivocal findings, with no consistent conclusions regarding the presence or extent of long-term biomechanical functional deficits in conservatively- and surgically-treated groups (Roos et al., 2014; Stensdotter et al., 2013; von Porat et al., 2006). For instance, a recent investigation reported inferior one-leg balancing ability in ACL_{PT} individuals 20-years post-intervention when compared to matched knee-healthy controls (CTRL), but comparable abilities between ACL_{PT} and ACL_R (Stensdotter et al., 2013). On the other hand, von Porat et al. (2006) observed no marked differences in the kinematic and kinetic characteristic of gait, stair ascent and descent, and cross-over hops between such groups (i.e., ACL_{PT}, ACL_R, and CTRL) when ACL-ruptured subjects were 16-years post-treatment.

In biomechanics, the use of traditional statistical analyses may explain such divergent findings given that continuous kinematic and kinetic data series are generally reduced to discrete variables (e.g., maximum knee flexion angles and time-to-peak forces), thereby removing information pertinent to the timing and shape of the continuous data series. Alternatively, functional data analysis (FDA) considers the magnitude, shape, and timing of continuous data series, and may therefore discern differences overlooked by traditional analyses. As such, employing FDA to treat biomechanical data has emerged over the last decade (Godwin et al., 2010; Pataky et al., 2013; Ryan et al., 2006) with the benefit of conserving the time-dependent structure of the data. In particular, the Interval Testing Procedure (ITP) (Pini and Vantini, 2013) identifies time intervals in which populations of interest differ, without dimensional reduction since the latter intervals are inferentially identified.

Here, we propose using this novel FDA method named ITP on one-leg hop (OLH) kinematic data taken from the KACL20 long-term follow-up study that involved ACL_{PT}, ACL_R , and CTRL. Clinically, the OLH is employed to assess functional capacities after ACL injuries (Ageberg et al., 2007), as it demonstrates

moderate correlations with patient-reported outcomes (Reinke et al., 2011; Sernert et al., 1999) and is reliable for assessing function post-ACL reconstruction (Sernert et al., 1999). Furthermore, since the OLH represents a biomechanically challenging task, it is often used to assess readiness to return to sports activities (Barber-Westin and Noyes, 2011). Using ITP to analyze knee kinematics during a functionally demanding task in CTRL, ACL_{PT}, and ACL_{R} might assist in further understanding the extent and mechanisms underlying longstanding dysfunctions, as well as the movement compensation strategies which follow these injuries.

Hence, our aims were to employ FDA on knee-joint kinematic data collected during OLHs, and compare kinematic curves between ACL_{R}, ACL_{PT}, and CTRL subject groups. On the basis of existing research and conventional analyses on these kinematic data (Tengman et al., 2013), we hypothesized that FDA would identify between-group and between-leg differences in the ACL-ruptured groups, with the involved leg of ACL-injured subjects exhibiting lesser knee flexion during take-off and landing (Gokeler et al., 2010; Orishimo et al., 2010), as well as greater knee abduction (Paterno et al., 2010) and external rotation (Deneweth et al., 2010) during landing. Furthermore, we anticipated that the ITP would identify time-intervals during which groups differed, providing novel information not available when using more conventional statistical approaches.

4.2 Method

4.2.1 Subjects

The research protocol used, which adhered to the Declaration of Helsinki, was approved by the Regional Ethical Review Board (Umeå, Sweden) before subject recruitment. To meet inclusion, subjects had to be in good self-reported health with no contra-indication to completing the study protocol. Omitting unilateral ACL injuries in the ACL_{R} and ACL_{PT} groups, individuals were excluded when presenting with a current or prior traumatic musculoskeletal injury to the knee, inflammatory or rheumatic disease, neurological condition, or history of bilateral ACL injury.

All subjects provided written informed consent to participate in this study, which was part of the larger-scale KACL20 project that addressed several aspects of knee function (Tengman et al., 2014). The demographic characteristics of subjects presenting a complete kinematic data set for the OLH task are reported in TABLE 1, along with hoping distances. All ACL_{R} and ACL_{PT} subjects had been treated $\sim$ 20 years ago (range: 17 to 28 y) in two separate hospitals that had been following different treatment protocols, as described in (Tengman et al., 2014). Although the CTRL subjects were purposefully selected to match as closely as possible the demographic characteristics of the ACL-ruptured subjects, non-parametric analyses revealed between -group differences in body mass index ($p = 0.0005$), with CTRL differing from both ACL_{R} ($p = 0.0140$) and ACL_{PT} ($p = 0.0007$).

Subjects (n = 95)	**Age (y)**	**BMI (kg/m^2)**	**Men (%)**	**One-leg hop distance (m)**	
				Injured non-dominant	**Non-injured dominant**
ACL_R ($n = 31$)	46 ± 5	27 ± 3^a	64.5	1.13 ± 0.27	1.20 ± 0.26
ACL_{PT} ($n = 33$)	48 ± 6	28 ± 4^a	63.3	1.00 ± 0.22	1.10 ± 0.26
CTRL ($n = 31$)	47 ± 5	$25 \pm 3^{b,c}$	64.5	1.08 ± 0.23	1.07 ± 0.25

Table 4.1: Demographic characteristics and maximal one-leg hopping distances of subjects presented by group. Injured and non-injured legs are in ACL_R and ACL_{PT}, whereas non-dominant and dominant legs are in CTRL. Significant differences between-groups were derived from non-parametric analyses (Kruskal-Wallis and Mann-Whitney with Bonferroni corrections). ^{a,b,c}: significantly different ($p < 0.05$) from the CTRL, ACL_R, and ACL_{PT} values, respectively.

4.2.2 Experimental procedures

Each subject was familiarized with the experimental protocol and tested in a single session. After recording height and body mass; subjects warmed-up for 6-min on an ergometer bicycle (Monark AB, Sweden) at a fairly light intensity, i.e., 11 on the 20-point Borg scale (Borg, 1982). Subjects then practiced the OLH sub-maximally under supervision and guidance from an investigator. The familiarization period was followed by 2 min of rest, after which testing was initiated.

One-leg hops for maximal distance were performed barefoot beginning in single-legged upright stance over a custom-made force-plate sampling at 1200 Hz (Department of Biomedical Engineering and Informatics, University Hospital of Umeå, Sweden). Force-plate data were time synchronized with the motion analysis system, and used to determine hop take-off. During hopping, subjects were requested to hop forward as far as possible, landing on the same leg without losing balance. Arms were held across the chest to limit contribution of arm swing to performance and occlusion of lower-body markers. Hops were performed three times on each leg, starting on the non-injured leg for the ACL-injured groups and dominant leg for controls (i.e., preferred leg to kick a ball), and followed by the contralateral leg. When subjects failed to perform a hop appropriately (e.g., lost balance during landing), it was replaced with an additional trial after rest. On the basis of the horizontal displacement of the lateral malleolus, only data from the longest hop on each leg were kept for analysis, with corresponding distances presented in Table 1.

4.2.3 Motion capture

Body motion was monitored during OLHs at a 240-Hz sampling rate using a calibrated 8-camera 3D motion analysis system (Oqus 300+, Qualisys Medical AB®, Gothenburg, Sweden) and QTM software version 2.7 (Qualisys Medical AB®, Gothenburg, Sweden). Forty-two retro-reflective markers were taped onto the skin of individuals over anatomical landmarks following standard marker placement guidelines used previously (Grip and Häger, 2013). From the reference markers, an 8-segment full-body biomechanical model with 6-degrees-of-freedom was constructed in Visual3D Professional™ Software v.4.96.7 (C-Motion Inc., Germantown, Maryland, USA), with the local coordinates of all body segments derived from a static measurement captured prior to the dynamic hopping trials.

4.2.4 Data processing

Biomechanical data were exported in the c3d format and processed in Visual3D. Marker data were routinely interpolated in QTM using a B-spline interpolation, allowing a maximum of 30 frames for gap filling. Both marker and force-plate data were then filtered with a 6-Hz bi-directional second order low-pass Butterworth filter.

Take-off event was defined from the kinetic data as the instance when the vertical ground reaction forces reached minimal values. Touch-down event was determined from the kinematic data as the instance when the vertical velocities of the lateral malleolus marker reached minimal values. On the basis of these events, all hops were divided into three phases: 1) take-off, 0.7 ms prior to and including take-off events; 2) flight, between take-off and touch-down events; and 3) landing, 0.7 ms following and including touch-down events.

Kinematic parameters were calculated using rigid-body analysis and Euler angles obtained from the static calibration. Knee-joint angles (°) were computed using an $x - y - z$ Cardan sequence equivalent to the Joint Coordinate System (Grood and Suntay, 1983), with the conventions used for interpreting motion directions illustrated in Figure 4.1. The kinematic curves of knee-joint angles in the three planes of motions were extracted for FDA. Although angular velocity and acceleration data were also investigated, only joint-angle curves are reported here as preliminary analysis revealed similar trends in results for the other two parameters.

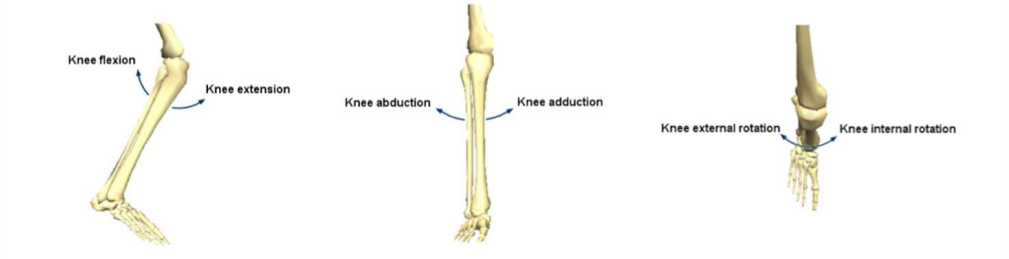


Figure 4.1: Knee-joint motion in the sagittal (left figure), coronal (middle figure), and transverse (right figure) planes of motion with conventions used to interpret direction of motion.

4.2.5 Statistical method

An inferential analysis on the knee-joint kinematic curves was performed for the data in the sagittal, coronal, and transverse planes and for the three hopping phases independently from one another using ITP, described in Chapter 2. Given the various durations of flight (range: 150 to 350 ms), data from this phase were time normalized and expressed as a percentage.

In ITP, time-dependent data are decomposed in the abscissa domain (i.e., time domain) into smaller parts using B-spline methods. The procedure then uses non-parametric permutation tests to provide adjusted p-values for each part of the domain via an interval-wise control of the family wise error rate (Pini and Vantini, 2013), thus enabling identification of time intervals wherein data from

different groups statistically differ. The process ensures that the probability of wrongly rejecting any time interval (i.e., false positive) is below the chosen significance level, herein set at 5%.

To compare motion curves between groups, an ITP-based ANOVA using Scheffé-based pair-wise comparisons was applied following previously described procedures (Abramowicz et al., 2014) (see Appendix 4.A for more details). The ITP was first applied to data from the injured legs of ACL_R and ACL_{PT} subjects and non-dominant legs of CTRL subjects, and then to the data from the non-injured and dominant legs of these respective groups. Finally, an ITP-based t-test was also applied to the point-wise difference of curves between legs within each group, calculated as the injured minus the non-injured leg in ACL_R and ACL_{PT} and the non-dominant minus the dominant leg in CTRL subjects. All computations were performed in R version 3.03 (R Core Team, 2012).

4.3 Results

4.3.1 Between-group comparisons

Statistical comparisons of the kinematic curves involving the injured and dominant legs of subjects (Figure 4.2) revealed no marked differences between ACL_R and CTRL. In contrast, the performance of ACL_{PT} in the sagittal plane substantially differed from the two other groups, with the most pronounced differences spanning the maximum knee flexion time-points during both take-off and landing. Specifically, ACL_{PT} exhibited lesser knee flexion than ACL_R from 636 to 207 ms prior to take-off, and from 621 to 122 ms compared to CTRL. Similarly, during landing, ACL_{PT} demonstrated lesser knee flexion than CTRL and ACL_R from 36 to 464 and 22 to 464 ms following ground contact, respectively. In addition, ACL_{PT} had greater external rotation than CTRL in vicinity of 10 ms post-landing.

Statistical comparisons of the kinematic curves involving non-injured and dominant legs (Figure 4.3) showed similar, but less pronounced, differences than those involving injured and non-dominant legs. Between-group differences were only significant in the sagittal plane during take-off, with ACL_{PT} again demonstrating lesser knee flexion than ACL_R and CTRL from 464 to 136 ms prior to take-off.

4.3.2 Within-group comparisons

No significant between-leg point-wise differences in the knee angle curves were identified within the CTRL (Figure 4.4). In the other hand, both ACL_{PT} and ACL_R showed between-leg disparities, wherein the injured leg in the correspondent groups exhibited more pronounced knee flexion throughout the first third (0% to 34%) and third-quarter (50% to 77%) of flight. Moreover, the injured leg of ACL_{PT} also exhibited increased external rotation at several instances during the landing.

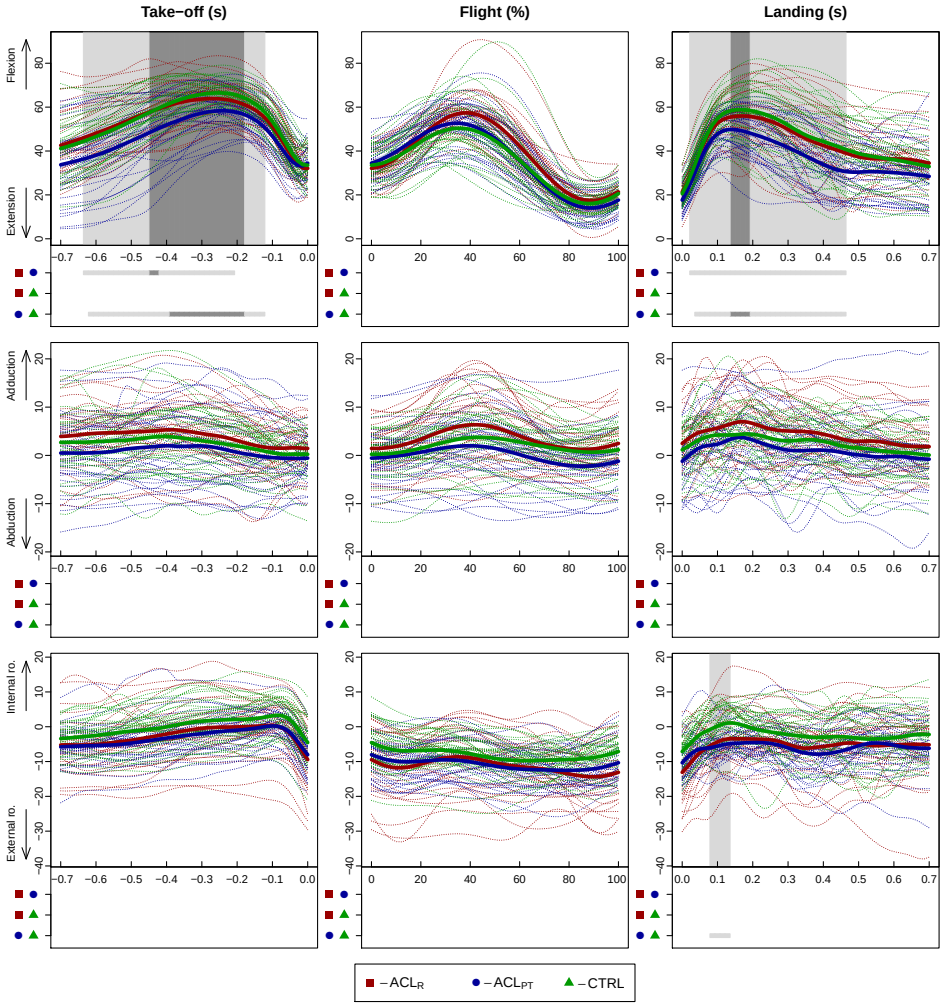


Figure 4.2: Between-group comparisons of knee angles in the sagittal (top row), coronal (middle row), and transverse (bottom row) planes of motion during the take-off (left column), flight (middle column), and landing (right column) phases of one-leg hops performed using the injured leg in ACL_R and ACL_{PT} subjects and non-dominant leg in CTRL subjects. The bold solid lines correspond to group means and the dashed lines represent the data for a given individual (ACL_R in red, ACL_{PT} in blue, and CTRL in green). The grey areas within the plots indicate significant between-group differences detected using Interval Testing Procedure based ANOVAs. Results from the pairwise group comparisons are underlined in grey in the panels below the plots, with the pairs indicated by the color-coded symbols (ACL_R by red rectangles, ACL_{PT} by blue circles, and CTRL by green triangles). Significant differences at 5% and 1% are represented in light and dark grey, respectively. ACL_R, ACL-ruptured subjects treated surgically with reconstruction and physiotherapy; ACL_{PT}, ACL-ruptured subjects treated conservatively with physiotherapy only; CTRL, knee-healthy controls.

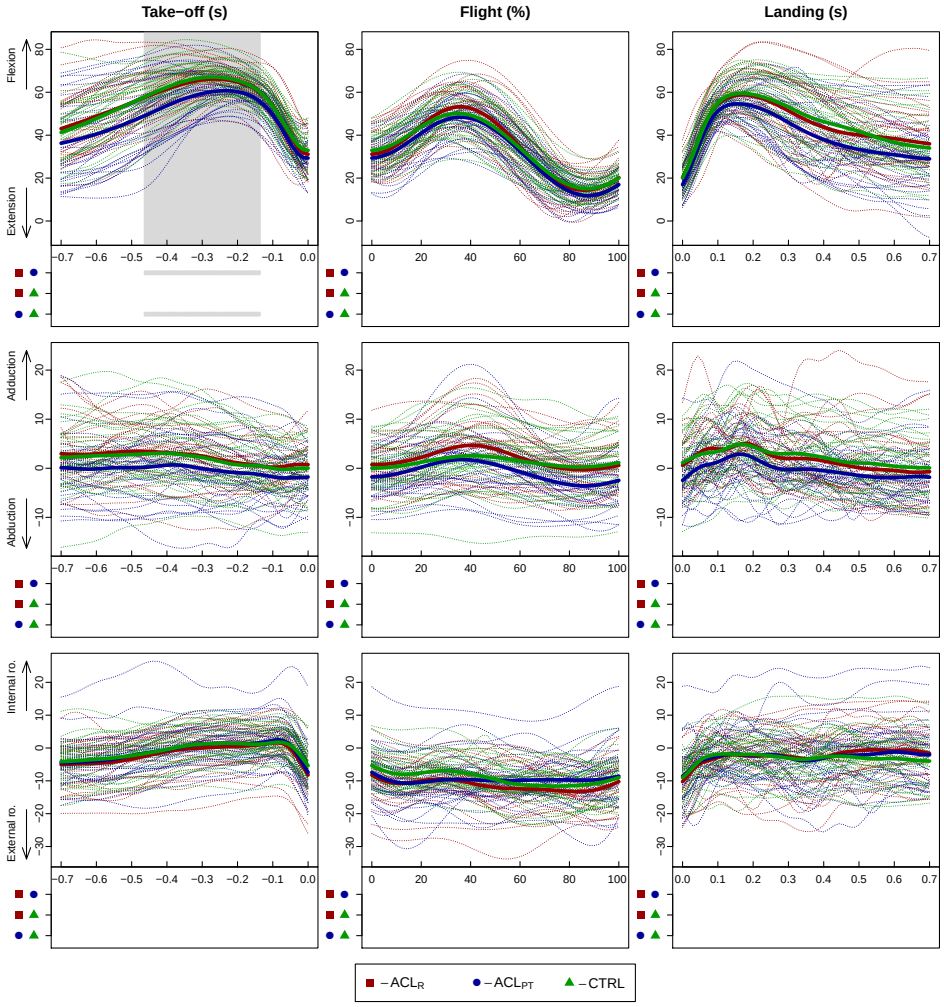


Figure 4.3: Between-group comparisons of knee angles in the sagittal (top row), coronal (middle row), and transverse (bottom row) planes of motion during the take-off (left column), flight (middle column), and landing (right column) phases of one-leg hops performed using the non-injured leg in ACL_R and ACL_{PT} subjects and dominant leg in CTRL subjects. The bold solid lines correspond to group means and the dashed lines represent the data for a given individual (ACL_R in red, ACL_{PT} in blue, and CTRL in green). The grey areas within the plots indicate significant between-group differences detected using Interval Testing Procedures based ANOVAs. Results from the pairwise group comparisons are underlined in grey in the panels below the plots, with the pairs indicated by the color-coded symbols (ACL_R by red rectangles, ACL_{PT} by blue circles, and CTRL by green triangles). Significant differences at 5% are represented in light grey. ACL_R, ACL-ruptured subjects treated surgically with reconstruction and physiotherapy; ACL_{PT}, ACL-ruptured subjects treated conservatively with physiotherapy only; CTRL, knee-healthy controls.

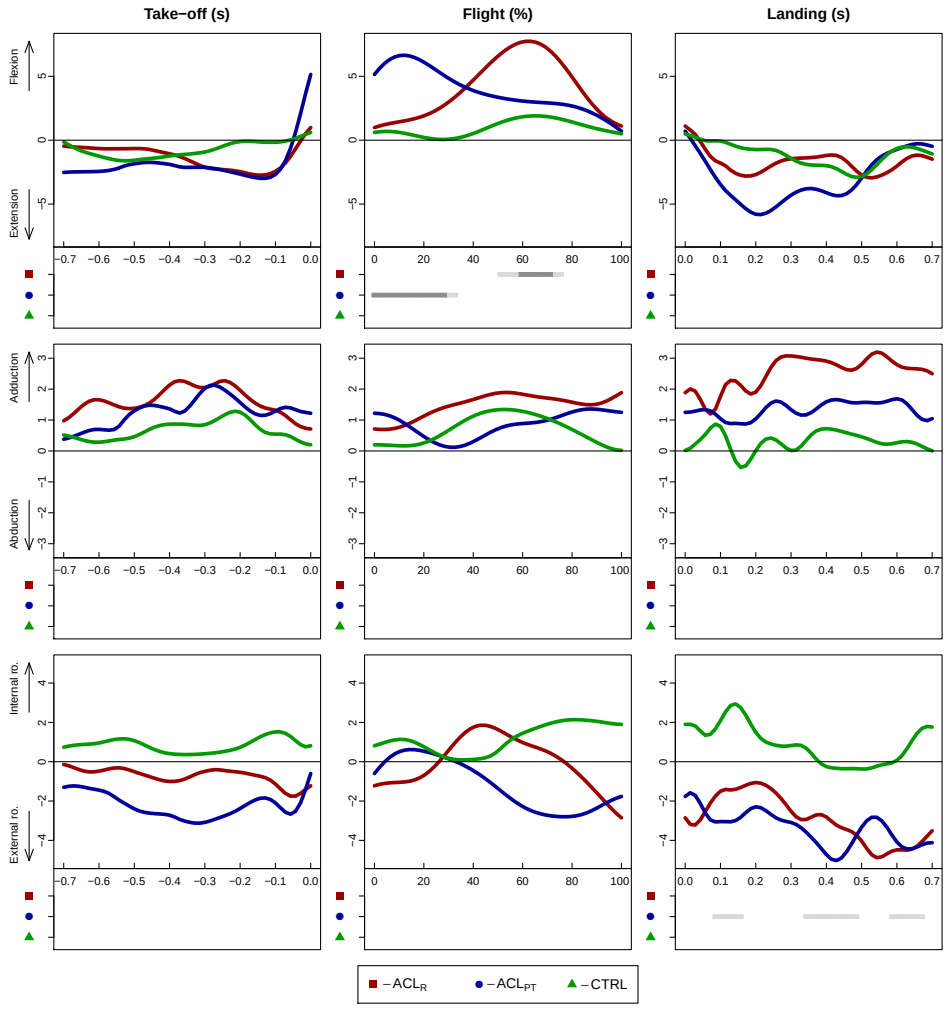


Figure 4.4: Within-group comparisons of knee angles in the sagittal (top row), coronal (middle row), and transverse (bottom row) planes of motion during the take-off (left column), flight (middle column), and landing (right column) phases of one-leg hops. The bold solid lines correspond to group means of between-leg differences (i.e., injured minus non-injured in ACL_R and ACL_{PT}, and non-dominant minus dominant leg in CTRL). For clarity, no individual curves are presented. Results from the Interval Testing Procedure based paired -t-tests are underlined in grey in the panels below the plots, with the pairs indicated by the color-coded symbols (ACL_R by red rectangles, ACL_{PT} by blue circles, and CTRL by green triangles). Significant between-leg differences at 5% and 1% are represented in light and dark grey, respectively. ACL_R, ACL-ruptured subjects treated surgically with reconstruction and physiotherapy; ACL_{PT}, ACL-ruptured subjects treated conservatively with physiotherapy only; CTRL, knee-healthy controls.

4.4 Discussion

Despite interventions, long-term ($\sim$ 20 years) knee-joint movement discrepancies in ACL-ruptured individuals were detected here using a novel FDA approach, which statistically compared the entire knee-kinematic curves of OLHs between and within three different population groups. By employing ITP, we not only identified between-group and between-leg differences in knee-displacement curves of distinct populations during a challenging task, but also isolated the time-interval ranges during which movement patterns differed.

Use of ITP is a step forward from previous reports investigating similar hop performances, wherein the main focus was on a few discrete variables, such as knee angles at given time-points (Gokeler et al., 2010; Orishimo et al., 2010; Paterno et al., 2010). Indeed, our ITP approach analyzed the entire time-varying structure of the kinematic data rather than a single registered event, providing a more comprehensive overview of knee-joint function and ability to identify differences. Such information can be useful for clinicians during OLH evaluations and assist in detecting deviations in movement patterns from the anticipated norm. For example, our data informs practitioners that ACL-deficient subjects without full functional recovery are likely to exhibit lesser knee flexion from $\sim$640 to 200 ms prior to take-off and from $\sim$40 to 460 ms post-landing compared to ACL_R and CTRL Additionally, ACL_PT might also exhibit greater external rotation in vicinity of 10 ms post-landing compared to CTRL, even several years after injury. In contrast, the flight phase and abduction/adduction kinematic curves provided little evidence of movement discrepancies in our ACL-injured cohort, suggesting that these particular aspects may be of lesser diagnostic value in the long term during OLH evaluation. Given the key differences identified, individuals with conservative treatment would not only benefit from post-injury rehabilitation, but also a more consistent and long-term program. Such a program could involve feedback training targeting their functional deficits during landing (Gokeler et al., 2013), knee extensor strengthening in presence of deficits (Schmitt et al., 2012), as well as core exercise training (Jamison et al., 2013) to enhance lower-body dynamic stabilization.

As hypothesized, between-group and between-leg differences in ACL-ruptured groups were identified, although mainly in ACL_PT. In terms of between-group analyses and consistent with previous investigations, the injured leg of ACL_PT demonstrated lesser knee flexion during take-off and landing (Gokeler et al., 2010; Orishimo et al., 2010) and greater external rotation during landing (Deneweth et al., 2010) than the non-dominant leg of controls. Our findings extend on previous work employing more traditional analyses by detecting that differences in knee movement patterns occur not only at specific instances, such as the maximal knee flexion during take-off, but rather throughout large portions of the take-off. These latter results, however, did not apply to ACL_R, as the kinematics of their reconstructed-ACL knee did not significantly differ from CTRL at any time during the OLH. As such, our results suggest superior long-term knee-joint functional stability in ACL_R (reconstruction with physiotherapy) than in ACL_PT (physiotherapy only), supporting recent investigations in this area (Roos et al., 2014). Potential mediators of long-term knee dysfunctions in

ACL-injured individuals include reduced relative strength, particularly of knee extensors (Oberländer et al., 2013; Roos et al., 2014); diminished feed-forward strategies (Bryant et al., 2009); increased reliance on hip and ankle joints for dynamic stabilization (Oberländer et al., 2013); lower activity levels (Tengman et al., 2014) incurring deconditioning; and impaired sensory functions (Ageberg and Fridén, 2008), predominantly involving an increase in the threshold for detecting passive motion of injured knees (Lee et al., 2009). The ability to detect passive motion partly relies on the stimulation of Ruffini type mechanoreceptors (Lephart et al., 1998) that, after one year, can be absent in untreated ruptured ACLs (Denti et al., 1994), supporting Barretts (Barrett, 1991) earlier findings of poorer detection of passive motion in ACL_{PT} than ACL_{R}, both inferior to CTRL.

The fact that the time-interval of significant differences between-groups in the sagittal plane involved the event of maximum knee flexion during take-off and landing substantiates the use of the latter measures when more traditional statistical approaches are employed. However, it is also clear from our findings that the event of maximum knee flexion is not the only characteristic distinguishing OLH performances between ACL-deficient individuals against matched controls, thereby encouraging scientists to employ more inclusive statistical procedures to movement analysis. Concurrently, referring to Figures 4.2 and 4.4, our results indicate that using the uninjured ACL_{PT} leg as a reference for comparisons (i.e., normative data) provides rather different results than when contrasted directly against the non-dominant leg of matched controls. For example, the noted discrepancies during take-off in the sagittal plane between ACL_{PT} and CTRL (Figure 4.2) were not present when comparing between legs of ACL_{PT} (Figure 4.4). In this instance, the time of maximum knee flexion was not able to detect differing movement patterns between injured and non-injured legs, and was not sensitive in identifying knee-joint movement differences. Therefore, although previously concluded that comparisons to the uninvolved leg can serve as an appropriate reference for normal OLH performance in ACL-injured individuals (Petschnig and Baron, 2009; Van der Harst et al., 2007), our analyses caution that such an approach may obscure functional impairments.

Considering the novelty of employing ITP on clinically relevant data, our group piloted various approaches prior to selecting the most suitable for data reporting and interpretation. As stated in our methods, ITP-based ANOVA was also applied during preliminary analyses to the angular velocity- and acceleration-time curves. However, analysis of knee-joint position was the most effective at detecting deficits herein, and provided the most practically relevant information to clinicians. Inherently, observing angular displacements during functional activities is more viable in a clinical environment than estimating velocities and accelerations. Similarly, to complement our analyses, FDA was also applied on landmark-registered aligned data (Ramsay et al., 2009) during take-off and landing by linearly transforming time to align maximum knee flexion angles of individuals to an identical relative time point. Doing so reduced the phase variability around the maximum knee flexion event, with a greater analytical focus on curve amplitudes. Since similar results were obtained between aligned and unaligned data, only the latter analyses were presented to maintain temporal

features of the OLH. Lastly, alternative methods than interval-wise control for adjusting P-value statistics were piloted. Again, comparable outcomes were obtained, emphasizing the presence of the between- and within-group differences detected within our sampled populations, strengthening the rigor of the proposed statistical approach.

Another advantage of our novel application of ITP on clinically-relevant data is that each statistical result was derived using an interval-wise control of the family wise error rate, implying a stern control of probabilities of accepting false-positive intervals. Furthermore, the technique used here, albeit only applied to knee-joint data, is generalizable to a range of continuous data, such as other joint-kinematic, kinetic, and electromyographic data. In the future, we foresee generalizing FDA use to a range of statistical questions, incorporating additional covariates within FDA, and extending its use to ANCOVA-based models. Indeed, various factors may have contributed to the differences observed, such as age and sex (Ageberg et al., 2001), as well as take-off angles and jumping technique (Wakeling, 2009). Accounting for hopping distance might further elucidate whether similar or dissimilar mechanisms are employed to fulfill this biomechanically challenging task in distinct populations.

One limitation of this study is the pure focus on knee-joint angle as opposed to whole body, which was a conscious choice to test the use of FDA on kinematic data. When landing during OLH, Roos et al. (2014) reported lower center of mass velocities just prior to landing, ranges of knee motion, travelled distances, and knee movement fluency (i.e., a measure representing medial-lateral control) in ACL_{PT} compared to ACL_{R}, with both significantly differing from controls. Applying FDA to center of mass displacements and velocities could complement and extend their study findings, describing the temporal profile of differences in movement patterns; not just at landing, but throughout the hop. Identifying the onset of such discrepancies might aid in further understanding the underlying mechanisms associated with impaired motion, and highlight the most challenging sections of OLHs, or other tasks, in ACL-injured individuals. Thereby, key information could be gained, which could help tailor rehabilitation programs to individuals, with future studies planned in this area.

4.5 Conclusions

Our novel implementation of FDA on knee-joint kinematic curves highlighted knee-joint dysfunctions persisting $\sim$20 years post-ACL ruptures following intervention that were more consistently identified in ACL_{PT} than ACL_{R}, consistent with previous studies. Given the key differences identified, individuals with conservative treatment would not only benefit from post-injury rehabilitation, but also a more consistent and long-term follow-up with ongoing rehabilitation. Although similar results may be observed using traditional statistics, our analytical approach has the advantage of considering the entire time-varying structure of the kinematic data set, identifies time-intervals in which compromised knee movement patterns are evident, and provides a more comprehensive and detailed descriptive of human motion. With future refinement of our model, it would be possible to take into account various covariates, such as age, sex, and hopping

distances. Embracing such an approach could further explain the persistence of dysfunctions in previously injured individuals, assist in analyzing a range of functional movements, and guide rehabilitation programs.

Appendix

4.A Statistical methodology for domain selection in functional ANOVA

Suppose to observe a collection of curves $\mathbf{y}=\{\mathbf{y}_{11},...,\mathbf{y}_{n_1 1},...,\mathbf{y}_{1g},...,\mathbf{y}_{n_g g}\}$ from $g > 2$ functional populations, and that the functions associated to the same group are independent and identically distributed observation of the same random function, i.e., $\{y_{i\tau}\}_{i=1,...,n_\tau} \overset{iid}{\sim} Y_\tau, \forall \tau \in \{1,...,g\}$, and that Y_{τ_1} and Y_{τ_2} are independent random functions, for each $\tau_1 \neq \tau_2$. We aim at testing the equality in distribution of all functional populations against the difference in distribution of at least one population from the other ones:

$$H_0 : Y_1 \overset{d}{=} Y_2 \overset{d}{=} ... \overset{d}{=} Y_g \quad \text{vs.} \quad H_1 : \exists\, \tau_1, \tau_2 \; s.t. \; Y_{\tau_1} \neq Y_{\tau_2}. \tag{4.1}$$

In order to perform the test, we apply the ITP proposed in Pini and Vantini (2013). The ITP is a testing procedure for functional data constituted by the following steps: (i) a high-dimensional functional basis is selected, and data are represented by means of the coefficients of the basis expansion (in this case, we will use a B-spline basis expansion $y_{ij}(t) = \sum_{k=1}^{p} c_{ij}^{(k)} b_m^{(k)}(t)$); (ii) the family of all possible intervals of hypotheses pertaining ordered basis components (including the univariate hypotheses themselves) is tested; (iii) the p-values associated to each component is adjusted by calculating the maximum among the p-values of each test containing that hypothesis. Details on the implementation of the ITP are reported in Pini and Vantini (2013). To build the family of tests used in the second phase, we propose employing the Non Parametric Combination (NPC) procedure Pesarin and Salmaso (2010), which is based on joint univariate permutation tests. Indeed, the NPC is a procedure that enables to perform a multivariate test for each possible dimension of the vector to be tested and for each sample size, and it is distribution-free, meaning that the normality assumption is not needed.

As a final result, the procedure provides a family of p adjusted p-values $\pi^{(k)}$, $k = 1,...,p$, one for each basis function used in the first step. The ITP is provided with an interval-wise control of the FWER that, since the B-spline basis used is local, results in a control on each possible interval in terms of the common abscissa along which the data are observed. In particular, while rejecting the marginal null hypotheses with associated adjusted p-value $\pi^{(k)} < \alpha$, we have that the probability of wrongly rejecting any interval of the domain is less than α. The result of the test is, therefore, a selection of the intervals of the domain that are significantly different among groups.

If one is just interested in selecting the intervals that presents differences

among the g groups, without performing pairwise group comparisons, it would be possible to construct an ITP that uses all possible permutations of the observed values over the g groups, and a permutationally-invariant version of the Fisher test statistic (Basso et al., 2009) to construct the univariate tests of the second phase:

$$T^* = \sum_{\tau=1}^{g} n_\tau \left(\bar{c}_\tau^{(k)^*} - \bar{c}^{(k)^*} \right)^2, \tag{4.2}$$

where $\bar{c}^{(k)^*}$ is the sample mean of all permuted coefficients (which is identical to the original sample mean), and $\bar{c}_\tau^{(k)^*}$ is the sample mean of the permuted coefficients associated the group τ. Then, the multivariate tests can be built by applying a combining function to the joint univariate results, as suggested in the NPC methodology.

However, once the significance of each coefficients of the basis expansion is tested in the ANOVA framework, if the null hypothesis is rejected, we also want to investigate which groups are identically distributed and which are not. To perform this kind of analysis, usually a two-group test is applied to all possible pairs of groups, that is, $g(g-1)/2$ comparisons of the form $H_0^{\tau_1,\tau_2} : Y_{\tau_1} \stackrel{d}{=} Y_{\tau_2}$ versus $H_1^{\tau_1,\tau_2} : Y_{\tau_1} \stackrel{d}{\neq} Y_{\tau_2}$, for each $\tau_1 \neq \tau_2$ are tested in a Scheffé-like framework. Indeed, all the two-group comparison are performed simultaneously to provide a result that is adjusted with respect to the multiplicity (e.g., in the parametric framework, Scheffe 1959 and Kramer 1957).

The method that we propose for the pairwise group comparisons is based on the application of the Closed Testing Procedure (CTP) Marcus et al. (1976). We perform an ANOVA-based ITP based on the statistic (4.2) for any possible subselections of at least two groups, that is, the two-group comparison $H_0^{\tau_1,\tau_2}$, the three-group comparison $H_0^{\tau_1,\tau_2,\tau_3} = \{Y_{\tau_1} \stackrel{d}{=} Y_{\tau_2} \stackrel{d}{=} Y_{\tau_3}\}$, the four-group comparison, and so on, up to the final g-group comparison, that coincides with the starting ANOVA test (4.1). Then, the p-values of the two-group comparisons are adjusted, for each basis coefficient k, by taking the maximum among all p-values associated to the same coefficient:

$$\bar{\pi}^{(k)^{\tau_1,\tau_2}} = \max \left\{ \pi^{(k)^{\tau_1,\tau_2}}, \pi^{(k)^{\tau_1,\tau_2,\tau_3}}, \pi^{(k)^{\tau_1,\tau_2,\tau_4}} ..., \pi^{(k)^{\tau_1,\tau_2,\tau_3,\ldots,\tau_g}} \right\}.$$

Note that, as the test statistics of each test of the closure family have possibly a dependent structure, as suggested in Basso et al. (2009), we apply for all tests, pooled permutations involving the whole data vector. Thus, it is possible to use the exact same permutations to test all the hypotheses $H_0^{\tau_1,\tau_2}, H_0^{\tau_1,\tau_2,\tau_3}, ...$, and for the ANOVA hypothesis H_0 in (4.1), and perform all tests jointly.

By applying this procedure, based on both the ITP (provided with an interval-wise control on the abscissa level), and the CTP (provided with a strong FWER control on the group level), we are able to provide the following error control: *if in any interval of the domain (or its complementary set) all the populations have the same distribution, the probability of rejecting at least part of the interval (or the complementary set) in at least one of the two-group comparisons is always controlled.*

Chapter 5

Functional ANCOVA: Analysis of Human Movement

Abstract

We here extend the analysis of knee movements presented in the previous chapter, by considering in the tests the effects of other covariates. In detail, an ANCOVA model is considered, in which the functional data (i.e., the time-varying flexion of the knee during a one-leg hop) are modeled by means of a linear model with fixed scalar covariates and time-varying fixed regression coefficients, in the framework described in Chapter 3. ITP-based tests on the regression coefficients of the model are performed, and the intervals presenting significant effects are selected.

Keywords: Biomechanics, Curve Analysis, Rehabilitation, Analysis of Variance, Functional Linear Model

5.1 Introduction and data description

Anterior Cruciate Ligament (ACL) injuries are common worldwide, and are typically treated either with conservative physiotherapy or with surgery and rehabilitative exercises. As discussed in Chapter 4, there is a controversy regarding whether all ACL-ruptured individuals should undergo surgery and the extent to which the long-term functional outcomes differ between conservatively and surgically treated patients (Delincé and Ghafil, 2012). Hence, it is of great importance to further study the long-term outcome after ACL-rupture. More specifically, as in the previous chapter, we focus on knee-joint kinematics data during a one-leg-hop, comparing individuals from the surgery and physiotherapy groups (ACL_{R} and ACL_{PT}, respectively) with age and gender matched knee-healthy control subjects (CTRL). We here focus on the differences between the three groups on the injured (non-dominant for CTRL) leg.

This motivating data set originates from the larger-scale KACL20 project that addressed several aspects of knee function (Tengman et al., 2014). Traditional analysis of kinematic data typically reports results for landmarks of the curves, such as maximal knee angle during take-off. Basic characteristics and

results from traditional analysis of the one-leg-hop are presented by Tengman et al. (2014). Previous univariate results performed on some features of the jump (e.g., maximum and minimum knee flexion) show a lower knee flexion among the individuals that are treated with physiotherapy.

The functional ANOVA for comparing the three groups reported in Chapter 4 indicates, coherently with previous results, that the physiotherapy group has different knee-joint kinematics during specific parts of the jump, detected especially on the flexion/extension angle during take-off and landing (Chapter 4). However, no covariates are taken into consideration into that study.

Here we are extending parts of that analysis by including the information coming from some additional covariates. In detail, we want to investigate if the difference between the movements of ACL_{PT} subjects and the ones of the other two groups resulting from an ANOVA analysis is really due to the treatment that individuals had (i.e., physiotherapy), or it can be explained by means of other variable characterizing the individuals (e.g., gender, BMI, age). We only present here results for between-group comparisons considering the flexion-extension angle, only including covariates available for all three groups.

Descriptive statistics for all included covariates are presented in Table 5.1. Figure 5.1 displays instead a scatter plot of the covariates. The different colors and point characters are associated to the three groups (CTRL: green triangle; ACL_{PT}: blue circle; ACL_{R}: red square).

Variable	Surgery	Physiotherapy	Control
Jump length (m)	1.13 (0.27)	1.00 (0.22)	1.08 (0.23)
BMI (kg/m^2)	27 (3)	28 (4)	25 (3)
Weight (kg)	83 (16)	86 (15)	78 (15)
Height (m)	1.74 (0.09)	1.74 (0.08)	1.77 (0.10)
Gender ($male/female$)	20/11	21/12	20/11
Age ($years$)	46 (5)	48 (6)	47 (5)

Table 5.1: Means and standard deviations for all variables considered as covariates in the presented analysis; for gender, the frequencies are presented.

5.2 Methodology

An inferential analysis on the knee-joint kinematic curves was performed for the data in the sagittal, plane and for the three hopping phases independently from one another. The analysis is performed by applying the ITP to perform tests on a functional-on-scalar linear model, according to the methodology described in Chapter 3.

In detail, let y_{ij} denote the functional data representing the knee flexion during the jump over time. The index i denotes the groups ($i = 1, 2, 3$, corresponding to ACL_{R}, ACL_{PT} and CTRL, respectively), and the index $j = 1, ..., n_i$ denotes the individuals. We are interested in testing the following model:

$$y_{ij}(t) = \beta_0(t) + \tau_i(t) + \sum_{l=1}^{L} \beta_l(t) x_{ijl} + \epsilon_{ij}(t), \quad i = 1, 2, 3,\ j = 1, ..., n_i, \qquad (5.1)$$

where $n_1 = 31$ ACL_{R} individuals, $n_2 = 33$ ACL_{PT} individuals and $n_3 = 31$ CTRL individuals. In model (5.1) we introduce functional coefficients: $\beta_0(t)$

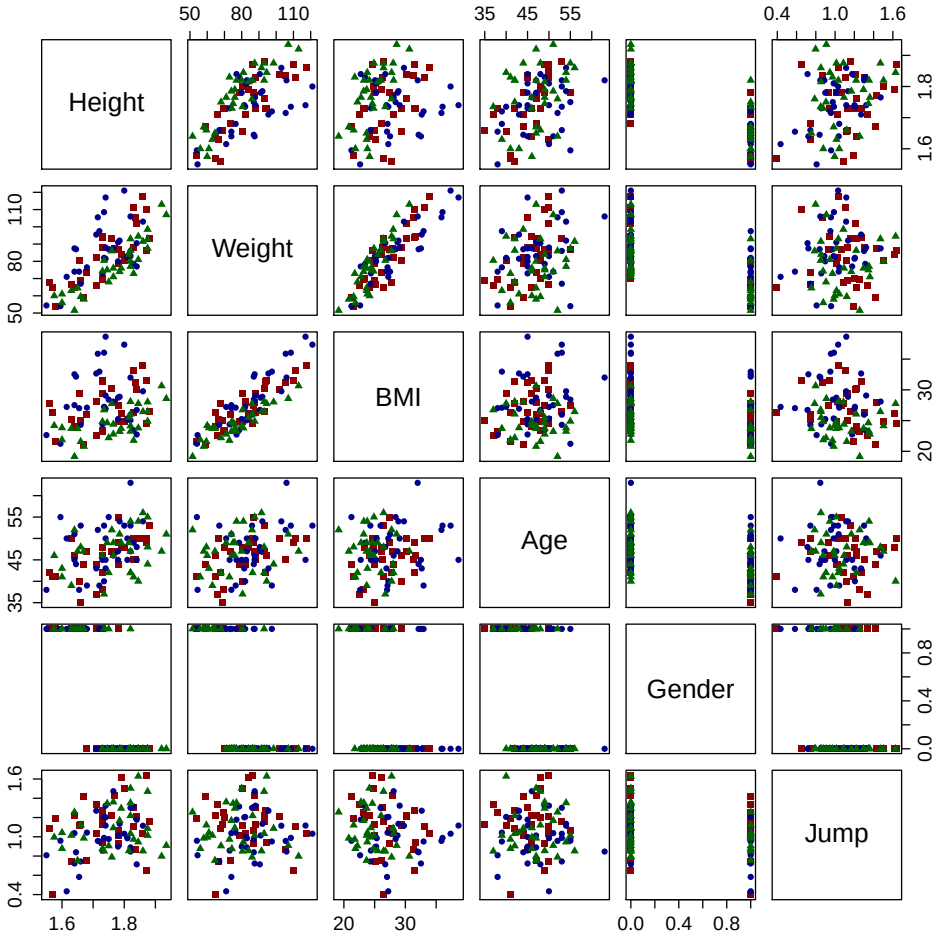


Figure 5.1: Scatter plot of covariates. The different colors and point characters correspond to the three groups: CTRL: green triangle; ACL_{PT}: blue circle; ACL_{R}: red square.

is the fixed common mean, $\tau_i(t)$ is the fixed treatment effect, $\beta_l(t)$ is the fixed coefficient associated to covariate l, and $\epsilon_{ij}(t)$ are random zero-mean $i.i.d$ errors with finite total variance ($\mathbb{E}[\|\epsilon_{ij}\|^2_{L^2}] < \infty$). Finally, we have a constraint for the treatment effects: $\sum_{i=1}^{3} \tau_i(t) = 0$. Note that the model (5.1) can be written as a functional-on-scalar linear model presented in Chapter 3, by introducing two dichotomic covariates expressing the membership of subjects to the three groups.

We are interested in developing different types of functional tests. First of all, we can perform a functional F-test on the regression model, to test is at least one covariate (including the groups) has a significant effect:

$$\begin{cases} H_{0,Model} & : \tau_i(t) = \beta_l(t) = 0 \; \forall i = 1, ..., 3, \; \forall l = 1, ..., L; \\ H_{1,Model} & : \exists i \text{ s.t. } \tau_i(t) \neq 0 \text{ or } \exists l \text{ s.t. } \beta_l(t) \neq 0. \end{cases} \tag{5.2}$$

Then, we can perform functional t-tests on the regression coefficients. In particular, we have three tests on the differences between the effects of the three groups:

$$H_{0,i_1 i_2} : \tau_{i_1}(t) = \tau_{i_2}(t); \quad H_{1,i_1 i_2} : \tau_{i_1}(t) \neq \tau_{i_2}(t), \quad \begin{matrix} \forall i_1, i_2 = 1, 2, 3, \\ i_1 \neq i_2, \end{matrix} \tag{5.3}$$

and $L + 1$ tests on the effects of the other covariates:

$$H_{0,l} : \beta_l(t) = 0; \quad H_{1,l} : \beta_l(t) \neq 0, \quad \forall l = 0, ..., L. \tag{5.4}$$

The functional tests (5.2), (5.3), and (5.4) are performed by applying the ITP, based on a B-spline basis expansion. The procedure selects, for each test (5.2) to (5.4), the intervals of the domain presenting significant effects. The resulting adjusted p-values are provided with an exact interval-wise control of the Family Wise Error Rate in the case of the F-test (5.2), and with an asymptotic interval-wise control of the Family Wise Error Rate in the case of the t-tests (5.3) and (5.4).

5.3 Results

The analysis is performed individually on three different phases of the one-leg-hop: take-off, flight, and landing. As in the previous ANOVA analysis, discussed in Chapter 4, the data of the flight phase are aligned, in order to have the same take-off and landing instant for all individuals. For each phase of the jump, we tested a full model including all six covariates and all groups' effects. Then, we started reducing the model by removing, with a step-wise backward procedure, the non-significant covariates. As the final aim of this analysis is to test for differences between the groups, the covariates indicating the membership of individuals to the groups are never removed, even in the case of non-significant effects.

This model selection (performed independently for each phase) provides the same reduced model for all phases. In detail, excluding the groups' effects, the only significant covariate is the jump length. All other covariates are on the contrary excluded from the model, due to not being significant. We here present

only the results of this reduced model on the three phases, i.e., an ANCOVA model including the effects of groups and of the jump length. The plots of the results of the three full models (for take-off, flight, and landing) are reported in Appendix 5.A.

Figure 5.2 displays the curves, and the results of the tests on this model, for the take-off, flight, and landing phases (left, middle, and right panels, respectively). In detail, the upper panels of the figure represent the results of the ITP-based functional F-test. The grey areas indicate the parts of the domain where we detect significance effects at 10% (light grey) and 5% (dark grey) significance levels. The other panels represents instead the results of functional t-tests (from top to bottom: intercept $\beta_0(t)$, effect of the jump length, difference between surgery and physiotherapy, difference between control and physiotherapy, and difference between control and surgery). The curves represented in each panel correspond to the OLS estimates of the functional coefficients of the linear model, and, as for the functional F-test, grey areas indicate the presence of a significant effect of the corresponding covariate.

The F-test indicates the presence of at least one significant effect in all three phases, throughout a great part of the domain, confirming the validity of the model. The intercept is significant throughout the entire domain, and the estimate of the coefficient shows that it is catching the common mean of all curves. The jump length has, not surprisingly, a significant effect throughout all three phases, in a great part of the time domain. In detail, the associated coefficient is positive during take-off and landing, indicating that to jump longer, individuals have to perform a higher flexion of the knee during both take-off and landing. During the flight, the coefficient associated to jump length switches its sign. The sign is negative in the two intervals corresponding to the minimum of flexion and positive sign in an interval corresponding to the maximum of flexion. This indicates that, to jump longer, the movement of the knee during flight has to be more pronounced (with higher maximal flexion and lower minimal flexion).

The physiotherapy group is significantly different with respect to the other two groups both during take-off and during landing, whereas three groups do not differ significantly during flight. The coefficient associated to the differences between the effects of the other two groups and the physiotherapy one is positive. This means that patients treated with physiotherapy have, in mean, a lower flexion during these two phases with respect to individuals in surgery and control groups. These findings are in line with the results of the simpler ANOVA model reported in Chapter 4. Even after having discounted for the jump length, physiotherapy group remains significantly different with respect to the other two groups.

5.4 Discussion

In this paper, data from a long-term follow-up study after anterior cruciate ligament injury are analyzed, applying the functional-on-scalar linear model previously described. Knee kinematics of individuals treated with physiotherapy or surgery and healthy controls were compared during a one-leg hop. The comparison between the three groups was carried out by taking into account individual-

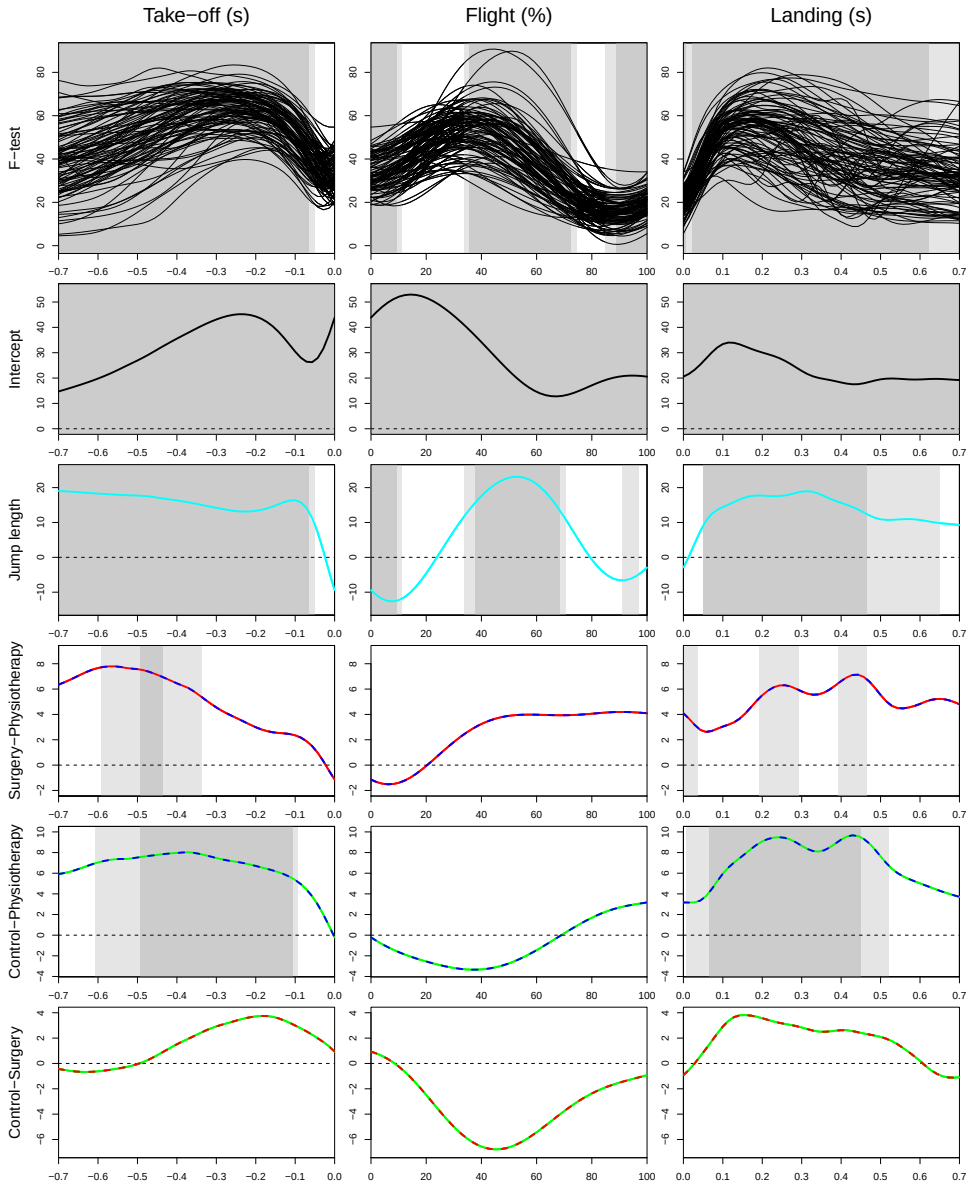


Figure 5.2: Results of the tests on the ANCOVA model on knee flexion angle for take-off (left), flight (middle), and landing (right). Upper panels: functional data and ITP-based functional F-test. Other panels: OLS estimates of functional coefficients and functional t-tests (from top to bottom: intercept, jump length, difference between surgery and physiotherapy, difference between control and physiotherapy, and difference between control and surgery). Grey areas indicate a significant result of the test at 10% (light grey) and 5% (dark grey) levels.

specific covariates, such as the jump length, BMI, age and gender.

The analysis of these data showed that the effect of jump length on knee kinematics is significantly different from zero, while the effects of BMI and age are not. In line with the results of the functional one-way ANOVA reported in Chapter 4, even after having discounted for the jump length, physiotherapy group remains significantly different with respect to the other two groups.

It is worth mentioning that the knee motion data analyzed in this chapter and in the previous one, are part of a larger data set, which also includes two more replicates for each jump. An interesting future development of this work would be the extension of the analysis to take into account replicates. To this purpose, the methodology presented in Chapter 3 can be extended to the case of a functional mixed-effects model, in which the three replicates are considered and individuals are included as random-effects. Another interesting future development of the analysis reported here is its extension to other knee angles (i.e., the abduction/adduction and rotation angles reported in Chapter 4), and other body parts during the same jump.

Appendix

5.A Results of the full models

We here report the results of the tests on the ANCOVA model on the knee flexion data including all covariates (full model) reported in Table 5.1 and Figure 5.1. The result of the tests on the full model for take-off, flight, and landing phases are reported in Figures 5.3, 5.4, and 5.5, respectively.

Each panel displays the result of a different test (from top to bottom and left to right: functional F-test; t-tests for: intercept; jump length; BMI; weight; gender; height; age; difference between surgery and physiotherapy; difference between control and physiotherapy; and difference between control and surgery.) Grey areas displayed in each panel indicate the parts of the domain where we detect significance effects at 10% (light grey) and 5% (dark grey) significance levels on the corresponding tests.

In all three phases, the results of the full models are consistent with the ones of the reduced models that include only jump length and the groups' effects. Jump length is significant in all three phases, and the movements of ACL_{PT} group differ from the ones of other two groups during take-off and landing, but not during flight. The estimate of the functional regression coefficients that are significantly different from zero obtained in this full model is consistent with the one obtained in the reduced model presented in Figure 5.2.

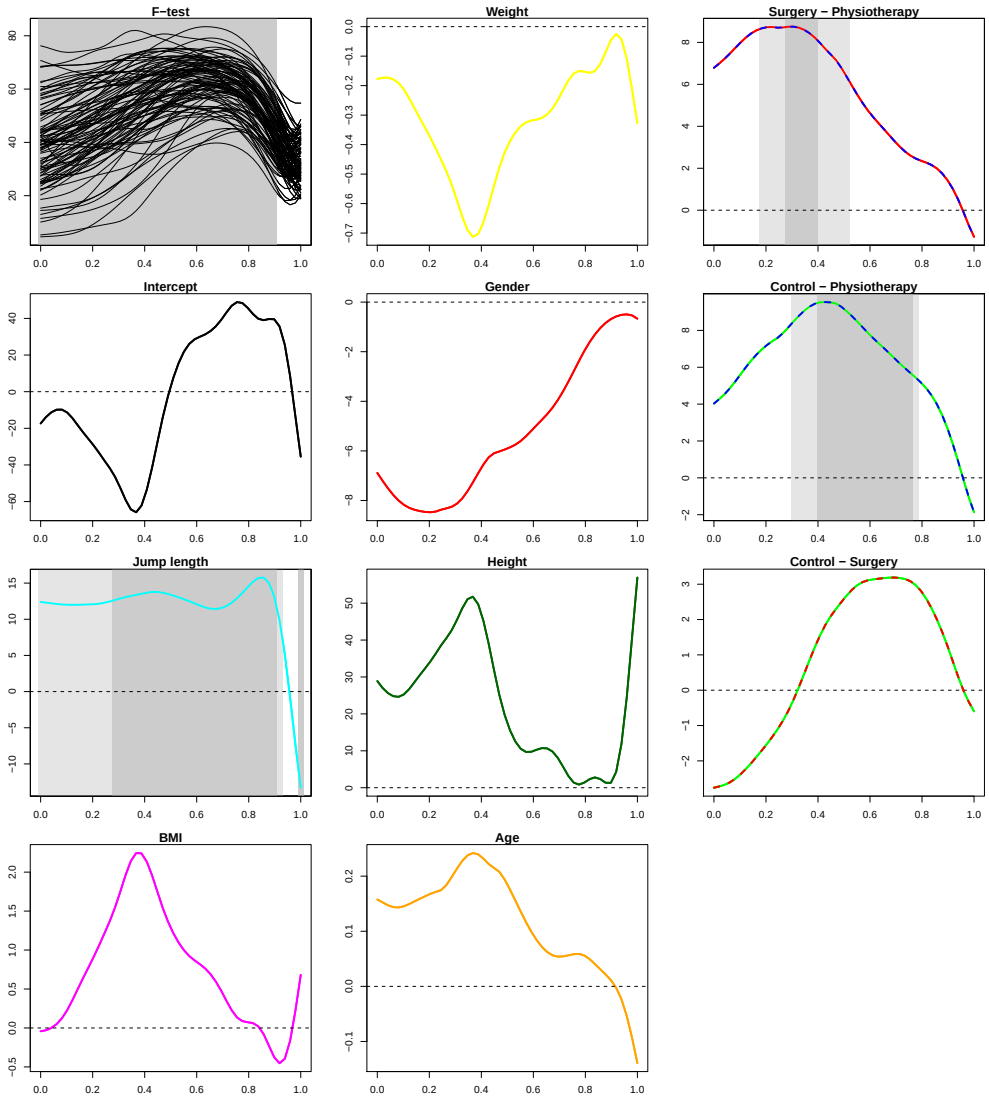


Figure 5.3: Results of the tests on the ANCOVA model on knee flexion angle including all covariates for take-off. From top to bottom and left to right: functional data and ITP-based functional F-test; OLS estimates of functional coefficients and ITP-based functional t-tests for: intercept; jump length; BMI; weight; gender; height; age; difference between surgery and physiotherapy; difference between control and physiotherapy; and difference between control and surgery. Grey areas indicate a significant result of the test at 10% (light grey) and 5% (dark grey) levels.

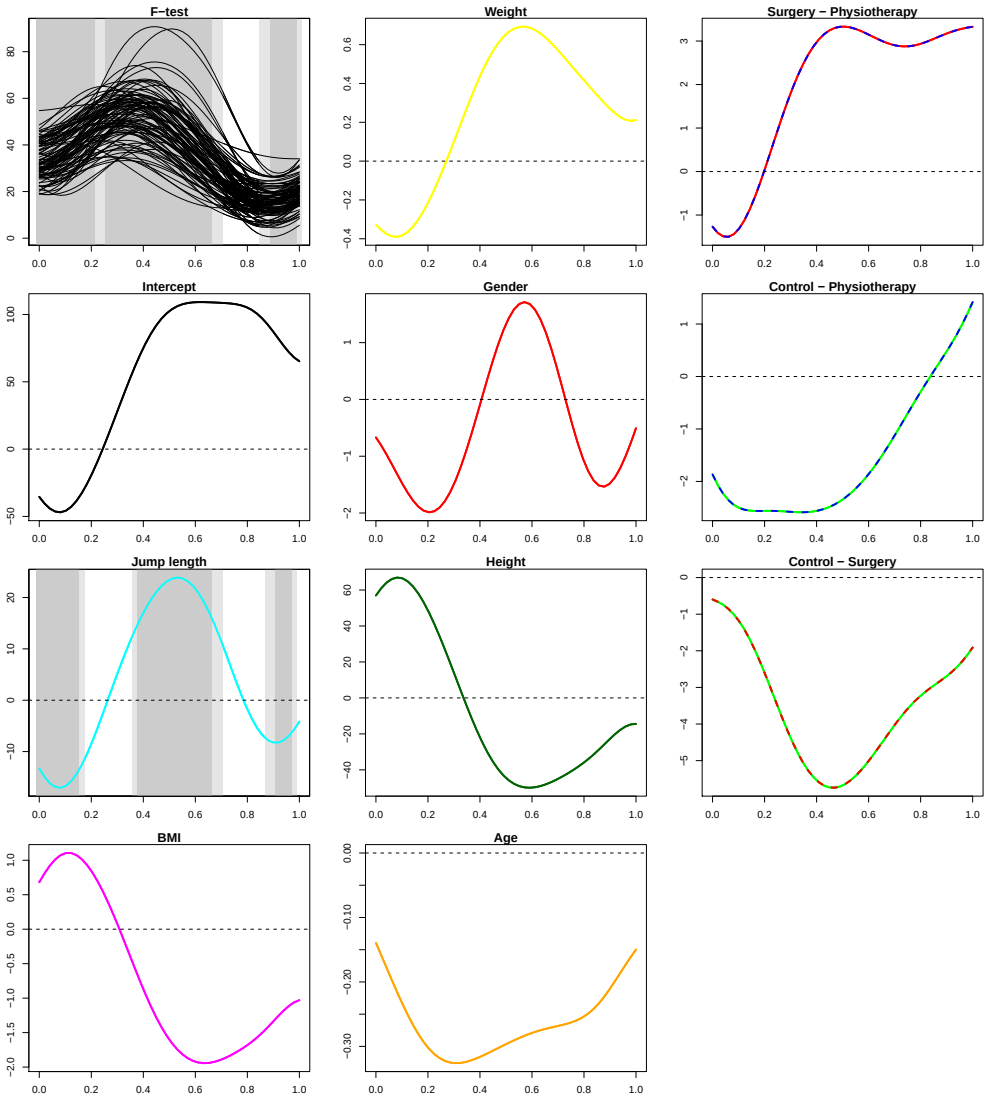


Figure 5.4: Results of the tests on the ANCOVA model on knee flexion angle including all covariates for flight. From top to bottom and left to right: functional data and ITP-based functional F-test; OLS estimates of functional coefficients and ITP-based functional t-tests for: intercept; jump length; BMI; weight; gender; height; age; difference between surgery and physiotherapy; difference between control and physiotherapy; and difference between control and surgery. Grey areas indicate a significant result of the test at 10% (light grey) and 5% (dark grey) levels.

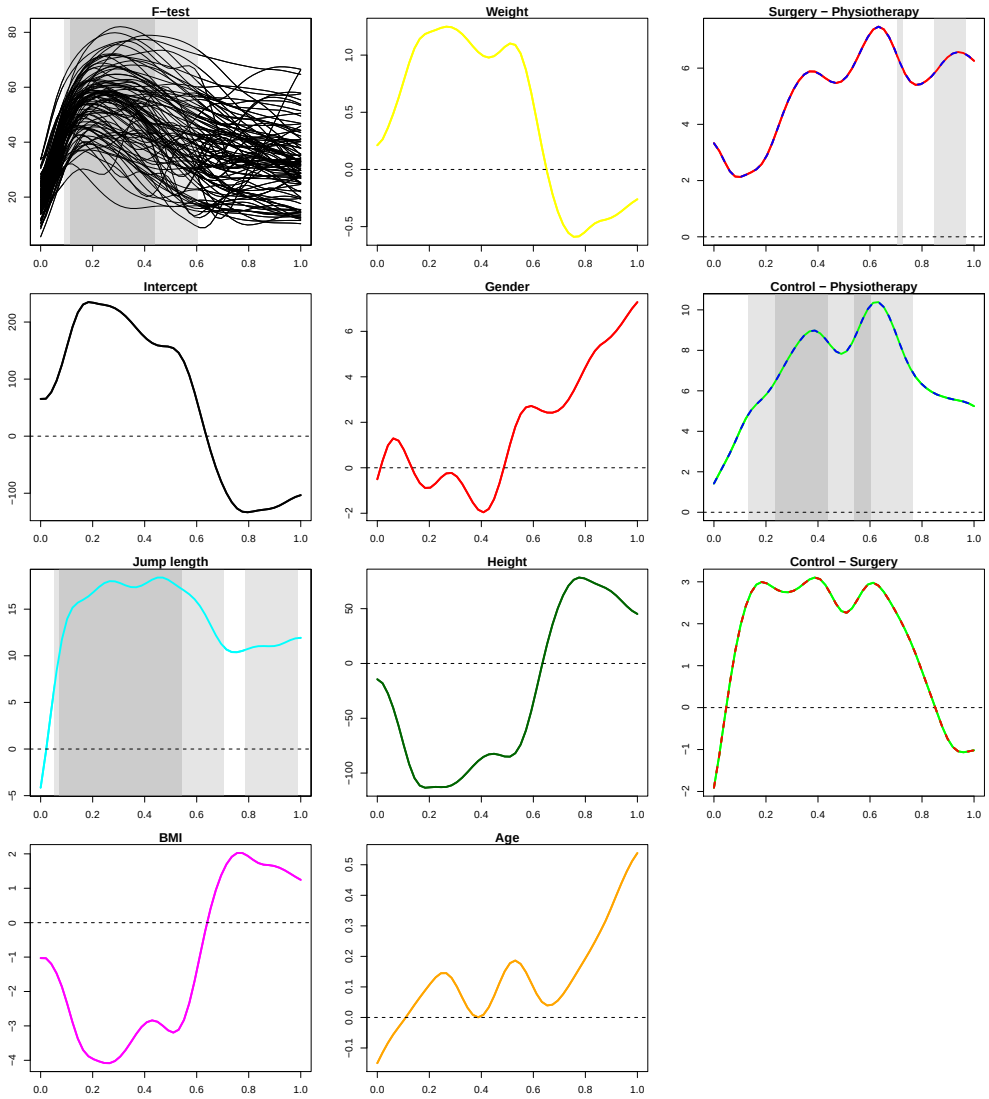


Figure 5.5: Results of the tests on the ANCOVA model on knee flexion angle including all covariates for landing. From top to bottom and left to right: functional data and ITP-based functional F-test; OLS estimates of functional coefficients and ITP-based functional t-tests for: intercept; jump length; BMI; weight; gender; height; age; difference between surgery and physiotherapy; difference between control and physiotherapy; and difference between control and surgery. Grey areas indicate a significant result of the test at 10% (light grey) and 5% (dark grey) levels.

Chapter 6

Two-Way Functional ANOVA: Analysis of Laser Spectra

Abstract

Remote monitoring of remote laser welding is an important technological issue in many industrial fields. Traditionally, many monitoring approaches are possible, and most of them involve the statistical analysis of a set of discrete variables summarizing the entire emission spectrum. Some examples are the monitoring of the overall visible emission, or of emission in separated wavelength ranges, defined by physical evaluations of the welding process. Although these methods are promising, they can provide practitioners with different results, giving an unclear overall description of the welding process. Moreover, the analyzed parameters cannot describe entirely the emission spectra, as they only summarize some of their features. In this paper, we propose to analyze the whole functional shape of emission data, by applying a two-way functional ANOVA model. In the model the effects of both the gap between welded plates and the location of the laser beam are estimated throughout all the recorded wavelength range, and wavelength bands presenting significant effects are inferentially selected. The selected bands can be used to remotely monitor the gap between the plates during the welding process.

Keywords: Remote Laser Welding, Remote Monitoring, Spectroscopy, Functional Two-Way ANOVA

6.1 Introduction

Laser welding technologies are quickly replacing, for instance in the automotive industry, conventional welding processes. Furthermore, nowadays the laser welding is often used in remote configurations, i.e., configurations in which the laser beam is moved along the seam with the help of a laser scanner.

One of the most common applications of such process is the welding of zinc-coated steel in the lap-joint configuration. However, this particular configuration and materials present a lot of technical issues, since the boiling temperature of the zinc is significantly lower than the one of steel (approx. 906°C and 1500°C,

respectively). Consequently, during the welding, highly pressurized zinc vapors are often produced at the interface of the two metals, and may cause defects in the welded material, such as spatters and porosities, that can compromise its quality.

The classical solution applied in industrial processes to prevent these defects is to leave a small gap (order of hundreds nano meters) between the two metal sheets, to facilitate the degassing (Akhter et al., 1991; Steen et al., 2003). One of the methods used to produce such gap is laser dimpling, which uses a remote pulsed laser to generate protuberances on one of the plates (AG, 2005; Schwoerer, 2008; Gu, 2010).

However, the variance of the height of laser-dimples can cause errors in the final gap dimension, which can cause defects on the welded material, compromising its quality and causing variations in the mechanical properties of the weld bead.

In Colombo et al. (2013), a method to remotely monitor the gap in remote laser welding, avoiding destructive off-line tests, is proposed. According to this technique, optical emissions are monitored during remote laser welding by a spectroscope. Then, from the acquired spectra, different indicators, or summarizing variables, are evaluated, such as the overall emission across the considered range, and the emissions in separated wavelength ranges (defined by physical evaluations of the welding process). For each of the obtained variables, univariate analysis of variance is performed, and the statistical significance of the effects of the gap value is used to compare the tested methods. The optical emission recorded during the experiment can also depend on the location of the laser beam on the welded surface. That is why, in Colombo et al. (2013), a two-way analysis of variance is performed, to evaluate the effect of the gap on the emission taking into account the different locations.

The former analysis is a valid instrument to assess how the emission is influenced by the gap, and the results can be used to provide an indicator to evaluate the gap remotely. However, the choice of the better indicator to evaluate the gap effect is difficult to perform, and in any case the obtained summarizing variables cannot describe the whole emission spectrum.

The aim of this paper is to extend the analysis performed in Colombo et al. (2013) in a functional data analysis framework (Ramsay and Silverman, 2005). Indeed the acquired emission signals are continuous functions of wavelength, and can be analyzed by taking into account the whole functional shape. Hence, we here perform a functional analysis of variance, to assess whether the whole emission function is influenced by the gap and the location. The analysis will take into account the functional shape of the signals, instead of only some discrete indicators that summarize the signal.

Furthermore, the final aim of the functional ANOVA that we present here is to select the wavelength bands presenting significant effects of the gap and the location on the emission, controlling the probability of false discoveries. The result of such technique will be the selection of a band of wavelengths that can be used to remotely monitor the gap between the plates during the welding process.

6.2 Experimental procedure and data acquisition

A Through Optical Combiner Monitoring architecture (Capello et al., 2008; Colombo and Previtali, 2009, 2010) is used to perform and monitor the laser welding. According to this technique, the monitoring of laser emission is performed remotely. Indeed, far from the work area, the optical emissions from the welding process are directly observed inside the laser source through the optical combiner of the fiber laser source with a spectroscope. An extensive description of the experimental welding procedure and the monitoring technology goes beyond the scope of this paper, and can be found in Colombo et al. (2013).

The main objective of this study is to assess the effects of both gap and location on the emission data. To analyze these effects, the emission is acquired in correspondence of different levels of gap and location, in a repeated factorial design. Three values of gap, corresponding to 100 nm, 200 nm and 300 nm are explored. For each of the analyzed gap values, three replicates are produced, for a total of nine welded specimens. Inside each specimen, five emission spectra are acquired at five different locations, for a total number of 45 acquired spectra. The emission data are described in detail in Colombo et al. (2013).

To record the optical emission in the visible range, optical emission spectroscopy, i.e., the analysis of emitted light with high-wavelength resolution, is used. The laser emission is acquired at 703 discrete wavelengths between 400.521 nm and 800.030 nm. The acquired data, as well as the three means, corresponding to the three different values of the gap, are represented in Figure 6.1.

Note that in the explored wavelength range, it is possible to distinguish between three different types of emission:

- between 400 nm and 530 nm, it is observed the emission related to electronic transition, i.e., the plasma emission;
- around 535 nm we observe a strong emission line, corresponding to the laser emission;
- above 540 nm we observe the emission due to the thermal black-body radiation, i.e. the thermal emission.

The aim of the following analysis is to assess whether the gap and the location have some effects on the emission, by taking into account the whole functional shape of the spectrum, controlling the probability of false discoveries. Finally, we want to locate possible wavelength bands in which the significant effects are detected.

6.3 Domain-selective functional two-way ANOVA

To analyze the laser emission data we will apply a functional two-way ANOVA model with interaction. In our model, the functional response (i.e., emission spectrum) will be expressed as the result of two main effects (i.e., gap and location), and of an interaction term between the main effects. The aim of the analysis is to test for the significance of every term in the model (i.e., interaction and main effects).

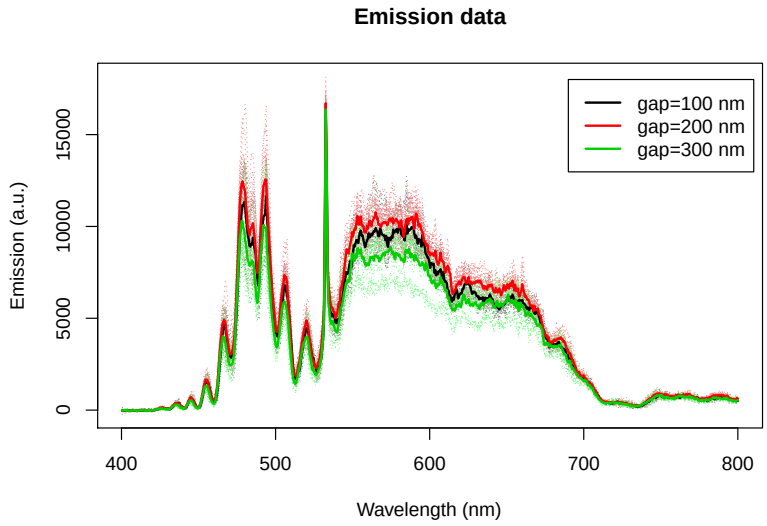


Figure 6.1: Emission data captured in the factorial laser welding experiment (dashed lines), colored according to the three different gaps; and means of the three groups, corresponding to the three different gaps (solid lines).

So, let $I_{ijl}(\lambda)$ be the functional data (i.e., the emission as function of the wavelength λ) associated with gap level i ($i = 1, 2, 3$), location level j ($j = 1, ..., 5$) and replicate l ($l = 1, 2, 3$). In our case, $\lambda \in (400.521\text{nm}, 800.030\text{nm})$. The functional ANOVA model that we want to study is the following:

$$I_{ijl}(\lambda) = \mu(\lambda) + \alpha_i(\lambda) + \beta_j(\lambda) + \gamma_{ij}(\lambda) + \epsilon_{ijl}(\lambda), \quad \begin{array}{l} i = 1, 2, 3, \\ j = 1, ..., 5, \\ l = 1, 2, 3. \end{array} \tag{6.1}$$

In model (6.1), $\mu(\lambda)$ is the common mean, $\alpha_i(\lambda)$ the gap effect, $\beta_j(\lambda)$ the location effect, and $\gamma_{ij}(\lambda)$ the gap-location interaction effect (with the classical constraints: $\sum_{i=1}^{3} \alpha_i(\lambda) = 0$; $\sum_{j=1}^{5} \beta_j(\lambda) = 0$; $\sum_{i=1}^{3} \sum_{j=1}^{5} \gamma_{ij}(\lambda) = 0$). The errors $\epsilon_{ijl}(\lambda)$ (with $i = 1, 2, 3$; $j = 1, ..., 5$; and $l = 1, 2, 3$) are assumed to be independent and identically distributed zero-mean random functions.

In our model, all effects (as well as the errors) are functions of the wavelength. Indeed, we assume that the effects can change through frequency, and in particular some of them may be expressed only in some wavelength bands.

The aim of our analysis is to test for the significance of all coefficients of model (6.1). In particular, we want to perform the following tests, i.e, the functional counterparts of classical ANOVA tests:

- a functional F-test on the model, jointly for all factors:

$$\begin{cases} H_{0,Model} & : \alpha_i(\lambda) = \beta_j(\lambda) = \gamma_{ij}(\lambda) = 0 \ \forall i, j, \\ H_{1,Model} & : \exists i, j \text{ s.t. } \alpha_i(\lambda) \neq 0, \text{ or } \beta_j(\lambda) \neq 0, \text{ or } \gamma_{ij}(\lambda) \neq 0; \end{cases} \tag{6.2}$$

- a functional test for the effect of each factor, separately:

$$H_{0,Gap}: \ \alpha_i(\lambda) = 0 \ \forall i; H_{1,Gap}: \exists i \text{ s.t. } \alpha_i(\lambda) \neq 0 \tag{6.3}$$

$$H_{0,Loc}: \ \beta_j(\lambda) = 0 \ \forall j; H_{1,Loc}: \exists j \text{ s.t. } \beta_j(\lambda) \neq 0 \tag{6.4}$$

$$H_{0,Gap*Loc}: \ \gamma_{ij}(\lambda) = 0 \ \forall i,j; H_{1,Gap*Loc}: \exists i,j \text{ s.t. } \gamma_{ij}(\lambda) \neq 0. \tag{6.5}$$

Note that, exactly as in a classical two-way ANOVA, (6.2) is a test of significance of the whole functional model, whereas the three tests (6.3-6.5) allow to perform a model selection, i.e., to reduce the model by considering only the factors with a significant effect on the emission.

The main difference that we have here with respect to a classical ANOVA framework is that, being the response neither scalar nor multivariate, but functional, tests (6.2-6.5) involve functional coefficients, i.e., the null hypothesis is rejected whenever there is a significant difference between the corresponding groups in at least one wavelength band.

The problem of testing a functional ANOVA model has been widely discussed in the literature of this last decades, and it can be addressed in several ways Cuevas et al. (2004); Cuesta-Albertos and Febrero-Bande (2010); Abramovich and Angelini (2006); Antoniadis and Sapatinas (2007). In detail, the problem of inference is approached from two different perspectives: parametric and non-parametric inference. The former approach commonly relies on distributional assumptions on functional data and on asymptotic results, while the latter relies instead on permutation or bootstrap computational intensive techniques. A common feature of all these works, is that the final result from the ANOVA testing determines whether the hypothesis of equality of the distributions (or of the mean values) of the groups of functions is globally accepted or rejected. In particular, by applying these tests on our model, we would be only able to answer the question "Are there any statistically significant effect of gap and/or location on the emission spectrum?". In the case of a positive answer, these tests are not able to select the wavelength bands in which the effects are detected, making the results of the analysis nearly useless for the remote monitoring of the gap. Indeed, in our application, it is of great importance to identify the significantly different bands, as we would like to use the (possibly) selected band to monitor the gap between plates during the welding.

In detail, for each one of the tests, in case of rejection of the null hypothesis, we want to select the wavelength intervals where significant differences are detected, and classical functional inferential tools are unable to provide this information. That is why we apply the Interval Testing Procedure (ITP, citealtpini2013), that is a testing procedure for functional data that enables to select the intervals of the domain (in our case: wavelength bands) presenting significant effects. Another advantage of the ITP is that it is a non-parametric procedure. In particular, we neither need to assume the normality of the residuals of the model, nor to specify the covariance structure of the residuals (that can both be difficult to assess).

The procedure is based on the following steps:

1. **Basis Expansion**: functional data are represented through the coefficients of a truncated ordered basis expansion $I_{ijl}(\lambda) = \sum_{k=1}^{p} I_{ijl}^{(k)} \phi^{(k)}(\lambda)$ (in this application localized in space, i.e., B-splines);

2. **Interval-Wise Testing**: a suitable test is performed on each interval of ordered basis coefficients $I_{ijl}^{(k)}, ..., I_{ijl}^{(k+h)}$ (in this case associated to different intervals of the domain);

3. **Multiple Correction**: for each component of the basis expansion, an adjusted p-value is computed from the p-values of the tests performed in the previous step, as detailed in Pini and Vantini (2013).

The adjusted p-values provided by the ITP enable a selection of the significant basis components, for each corresponding test. If a local basis is used as in our case, this selection translates into a selection of statistically significant intervals of the domain, i.e., statistically significant wavelength bands.

In the functional ANOVA framework hereby depicted, the basis expansion applied to model (6.1) leads to a classical ANOVA model for each coefficient of the basis expansion, i.e.,:

$$I_{ijl}^{(k)} = \mu^{(k)} + \alpha_i^{(k)} + \beta_j^{(k)} + \gamma_{ij}^{(k)} + \epsilon_{ijl}^{(k)}, \quad \begin{aligned} i &= 1, 2, 3, \\ j &= 1, ..., 5, \\ l &= 1, 2, 3, \\ k &= 1, ..., p. \end{aligned}$$

The procedure provides as a result an adjusted p-value for each component of the basis expansion used in the first step, that is in our case, the 1+3 families of adjusted p-values corresponding to the F-test on the model and the three tests on single effects:

$$\begin{cases} H_{0,Model}^{(k)} & : \alpha_i^{(k)} = \beta_j^{(k)} = \gamma_{ij}^{(k)} = 0 \ \forall i, j, \\ H_{1,Model}^{(k)} & : \exists i, j \text{ s.t. } \alpha_i^{(k)} \neq 0, \text{ or } \beta_j^{(k)} \neq 0, \text{ or } \gamma_{ij}^{(k)} \neq 0; \end{cases}$$

$$\begin{aligned} H_{0,Gap}^{(k)} &: \alpha_i^{(k)} = 0 \ \forall i; H_{1,Gap}^{(k)} : \exists i \text{ s.t. } \alpha_i^{(k)} \neq 0 \\ H_{0,Loc}^{(k)} &: \beta_j^{(k)} = 0 \ \forall j; H_{1,Loc}^{(k)} : \exists j \text{ s.t. } \beta_j^{(k)} \neq 0 \\ H_{0,Gap*Loc}^{(k)} &: \gamma_{ij}^{(k)} = 0 \ \forall i, j; H_{1,Gap*Loc}^{(k)} : \exists i, j \text{ s.t. } \gamma_{ij}^{(k)} \neq 0. \end{aligned}$$

To perform the ITP, we just need to specify how to perform the interval-wise tests of the second phase. As in Pini and Vantini (2013), we approach this problem by means of multivariate Non-Parametric Combination (NPC, Pesarin and Salmaso 2010) of permutation tests. The NPC is a procedure that enables to build multivariate permutation tests by means of combining univariate joint permutation tests, and is initialized with joint univariate permutation tests. For the two-way ANOVA tests (6.2-6.5) that we here want to perform, we use asymptotically exact tests based on the permutations of residuals under the reduced model, described by Freedman and Lane (1983).

In detail, we perform each test by randomly permuting the residuals of the reduced model, i.e., the model under the corresponding null hypothesis. As test statistics, we use the two-way ANOVA statistics of the corresponding classical F-tests. This provides exact tests for $H_{0,Model}$, and approximated (asymptotically exact) tests for $H_{0,Gap}$, $H_{0,Loc}$ and $H_{0,Gap*Loc}$.

The adjusted p-values of the test of $H_{0,Model}$ are provided by an Interval-Wise control of the Family Wise Error Rate, which in our framework states that, if the significant wavelength bands are detected as the ones associated to an adjusted p-value lower than α: given any wavelength band in which the emission is not influenced by any factor, the probability that this band is (wrongly) selected as significant is lower than α. The adjusted p-values of the tests of $H_{0,Gap}$, $H_{0,Loc}$ and $H_{0,Gap*Loc}$ are instead provided with an asymptotic Interval-Wise control of the Family Wise Error Rate (see Chapter 3).

6.4 Results

6.4.1 Results of the tests

We applied the two-way ANOVA to test for gap, location, and interaction effects on the emission data collected in the laser welding experiment. The ITPs that we apply to test the factors of the model are based on a B-spline basis expansion of order 2 (degree 1) with 200 evenly spaced knots (i.e., one knot every two nm). This means that the support of every basis function used for the expansion is four nm. Figure 6.2 (on the right) shows a representation of emission data after the B-spline expansion.

Results consistent to the ones reported here can be found also with a different number of knots. A discussion about the robustness of the method with respect to this parameter is reported in the next subsection.

The presence of at least one significant factor in the model is confirmed by a highly significant F-test (the minimum adjusted p-value over all basis components is equal to zero). On the other hand, the interaction term $\gamma_{ij}(\lambda)$ is not significant for any wavelength. Indeed, the minimum adjusted p-value associated to the interaction term is equal to 0.92, meaning that the interaction term is not significant for any basis component.

After removing the interaction term from model (6.1), we found significant effects for both the gap and the location. Hence, we only report here the results of the tests on this second model, i.e., the additive functional ANOVA model without interaction:

$$y_{ijl}(\lambda) = \mu(\lambda) + \alpha_i(\lambda) + \beta_j(\lambda) + \epsilon_{ijl}(\lambda), \quad \begin{matrix} i = 1,2,3, \\ j = 1,...,5, \\ l = 1,2,3. \end{matrix}$$

Note that, in this case, the F-test on the model becomes:

$$\begin{cases} \tilde{H}_{0,Model} & : \ \alpha_i(\lambda) = \beta_j(\lambda) = 0 \ \forall i,j; \\ \tilde{H}_{1,Model} & : \exists i,j \text{ s.t. } \alpha_i(\lambda) \neq 0, \text{ or } \beta_j(\lambda) \neq 0. \end{cases}$$

Figure 6.2 reports the results of the additive two-way functional ANOVA of emission data. In particular, the left panels of the figure report the adjusted p-values associated to the 200 B-spline basis functions used to describe the data. On the top panel, we report the adjusted p-values of $\tilde{H}_{0,Model}$ (testing in the reduced model the significance of at least one factor between gap and location), on the middle panel the adjusted p-values of $H_{0,Gap}$ and on the bottom panel

the adjusted p-values of $H_{0,Loc}$. For ease of visualization of the test results, the right panels of the figure report the significant intervals at 5% and 1% levels (areas colored in light and dark grey, respectively) on the emission data for the three tests. The B-spline representation of data is reported on the same figure, and to appreciate differences between groups data are colored in the three panels according to the corresponding tests: on the top panel data are colored differently according to the different levels of both gap and location (i.e., only the three replicates for each experimental condition are plotted with the same color); on the middle panel data are colored according to the different levels of gap; on the bottom panel data are colored according to the different levels of location.

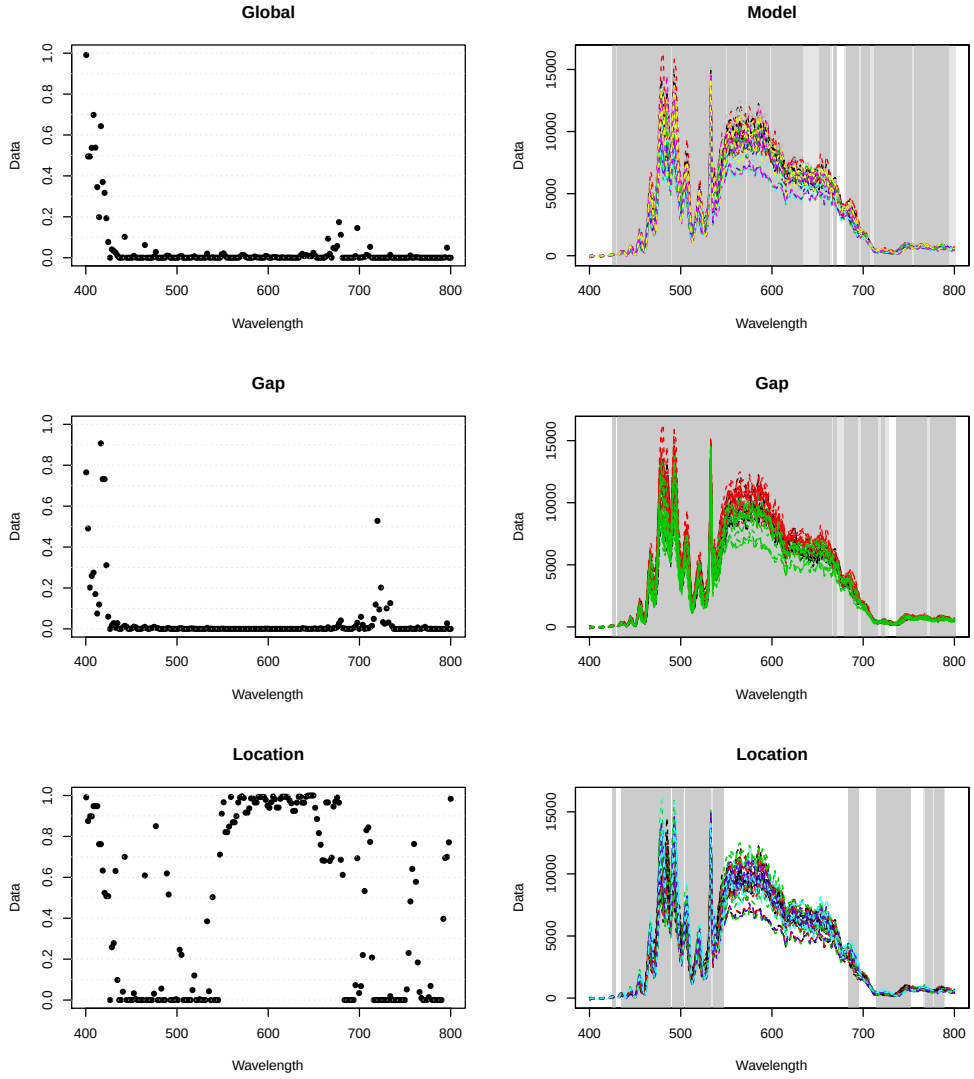


Figure 6.2: Left: ITP adjusted p-values for the tests on $\tilde{H}_{0,Model}$ (top), $H_{0,Gap}$ (middle) and $H_{0,Loc}$ (bottom). Right: B-spline representation of emission data colored according to gap and location levels (top), gap levels (middle) and location levels (bottom). The gray areas represent significant intervals at 5% and 1% levels (light and dark grey, respectively)

We find a significant effect of at least one factor among gap and location in nearly the entire wavelength domain, both at 5% and 1% significance levels (test

of $\tilde{H}_{0,Model}$, top panels of Figure 6.2). This result suggests that, as expected, the welding conditions have a significant effect on the spectrum for all three types of emission (plasma, laser and thermal). The same result was found in Colombo et al. (2013). What is new in the analysis here reported is the fact that, by considering the whole functional shape of data, we are able to precisely locate the wavelength bands presenting significant effects of the two factors.

In particular, the gap has a significant effect in a great part of the wavelength domain, both at 5% and 1% significance levels (test of $H_{0,Gap}$, middle panels of Figure 6.2), suggesting that, consistently with previous results, the emission is significantly influenced by the gap on all three types of emission.

The location has a significant effect, both at 5% and 1% significance levels, for some wavelength intervals, located mainly in the plasma and laser emission ranges (test of $H_{0,Loc}$, bottom panels of Figure 6.2). This suggests that plasma emission is influenced by the location.

The most important result highlighted by this analysis, and completely new with respect to the literature in this field, is that, looking at all results together, we detect a band (i.e., the band $\lambda \in (547 \text{ nm}, 681 \text{ nm})$ corresponding to the thermal emission), in which the gap effect is significant and the location one is not. This suggests the use of emission data on this band to monitor the gap between the plates during the welding process at any possible location. Indeed, in this band the emission is significantly influenced by the gap and not by the location. Thanks to the interval-wise control of the FWER, we know that, if the emission would not be effected by neither the gap nor the location in the band $\lambda \in (547 \text{ nm}, 681 \text{ nm})$, we would have selected it as significant with a probability lower that 5%. Furthermore, if the gap would not affect the emission in the same band, we would have selected it as significant with a probability approximately lower that 1%.

As a final detail, note that, coherently with the results reported in (Colombo et al., 2013) the gap effect is not monotonic. Indeed, the black curves corresponding to the lower value of gap (i.e., 100nm) are mainly located in the middle between green curves (300nm) and red curves (200nm), meaning that the gap effect is not linear. For this reason we here choose to consider the gap as a factor and not a numeric variable. If one would be interested in considering the gap as a numeric variable, its value would necessarily need to be added in the model in a non-linear way.

6.4.2 Robustness analysis with respect to the number of knots

To appreciate how the results of the analysis that we reported are robust with respect to the number of knots, we performed the test by varying this parameter. We considered the cases summarized in Table 6.1, based on B-splines expansions with a different number of knots. In particular, we started from the maximum possible resolution (i.e., 703 knots), and reduced the resolution until a minimum value of 100 knots (still sufficiently high to have a good description of data). The bandwidth corresponding to each B-spline basis function in each case is reported in the table. In all cases we took piecewise-linear B-splines with evenly spaced knots.

The test on the interaction $H_{0,Gap*Loc}$ is found not significant in all explored cases. Hence, we remove the interaction from the model, testing a simpler additive model on the gap and the location. The results of the tests are summarized in Table 6.1, and reported in Figure 6.3. In the figure, the results of tests on $\tilde{H}_{0,Model}$, $H_{0,Gap}$ and $H_{0,Loc}$ at 1% and 5% levels are represented by means of the gray bands below each graphic. The axis on the left indicates the number of knots used for the B-spline basis expansion, which varies between 100 and 703 knots, according to the values reported in Table 6.1.

Knots	Bandwidth (nm)	$\tilde{H}_{0,Model}$	$H_{0,Gap}$	$H_{0,Loc}$
703	1.14	Significant (P,L,T)	Significant (P,L,T)	Significant (P,high T)
400	2	Significant (P,L,T)	Significant (P,L,T)	Significant (P,high T)
200	4	Significant (P,L,T)	Significant (P,L,T)	Significant (P,L,high T)
100	8	Significant (P,L,T)	Significant (P,L,T)	Significant (P,L,high T)

Table 6.1: Results of tests on $\tilde{H}_{0,Model}$, $H_{0,Gap}$ and $H_{0,Loc}$, varying the number of knots on plasma (P), laser (L) and thermal (T) emission (level: 5%). The bandwidth of a single B-spline basis function for each case is reported.

The results seem to be robust with respect to the number of knots. Both the F-test on the model $\tilde{H}_{0,Model}$ and the test on the gap $H_{0,Gap}$ report significant differences at both 5% and 1% in almost all the wavelength domain. In particular, the gap has a significant effect on all three types of emission. On the other hand, the location has a significant effect on the emission only on the high-wavelength band of thermal emission (where, anyway, the signal is very low), and in some intervals on the plasma and laser emissions. In all cases, the thermal band detected in the previous subsection, presenting a significant effect of the gap but not of the location is preserved.

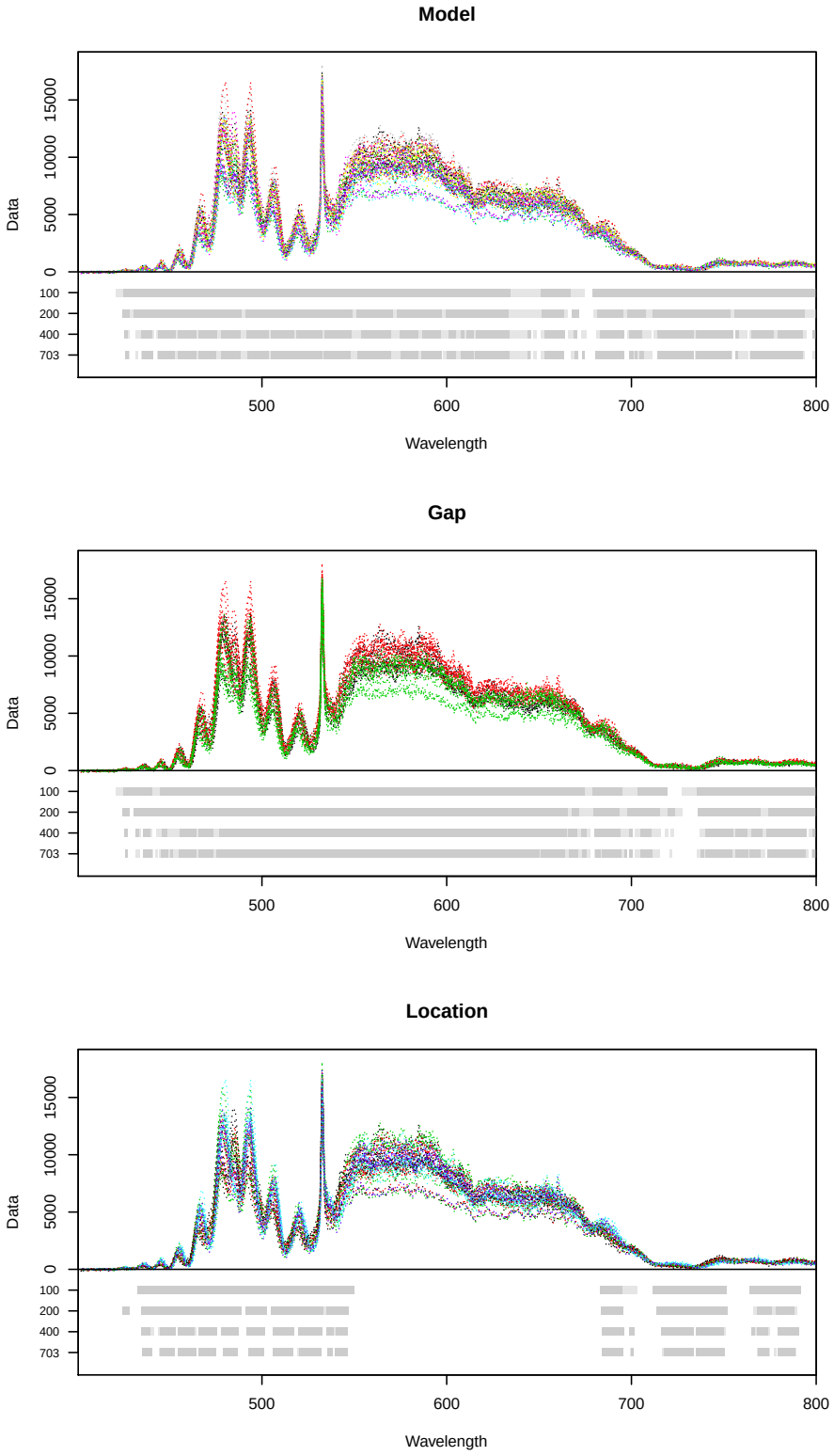


Figure 6.3: Emission data colored according to gap and location levels (top), gap levels (middle) and location levels (bottom). The gray bands below each graphic represent the results of tests on $\tilde{H}_{0,Model}$, $H_{0,Gap}$ and $H_{0,Loc}$, resp., varying the number of knots, at 5% and 1% levels (light and dark grey, resp.)

Chapter 7

Functional One-Population Test: Analysis of Climatic Data

Abstract

In this paper, aleatory and epistemic uncertainties in energy generation systems are investigated. The former are described by probability distributions, whereas the latter by possibility distributions. In particular, given the evolution in time of the aleatory uncertainty, time-varying probability distributions are considered and they are elicited by Functional Data Analysis. Then, the joint propagation of both types of uncertainty is performed by Monte Carlo simulation and Fuzzy Interval Analysis. The method is applied to a model of an energy system made of a solar panel, a storage energy system and the loads. As a quantitative indicator of the analysis we evaluate the Expected Energy Not Supplied.

Keywords: Photo-Voltaic Energy, Irradiation, Functional Data Analysis, Fuzzy Interval Analysis, Monte Carlo Simulation

7.1 Introduction

Renewable energy is getting more and more important as a solution for the climate change concerns. However, it is affected by large uncertainties, due to the intermittent nature of the energy source (regardless from the type of renewable source, the amount of energy daily available can present high variations from one day to another, at the same site) Borges (2012). In addition, a long-term prediction of the energy that is daily available is a difficult problem, due to the complexity of the system at hand. These issues mine the reliability of the renewable energy, making it difficult to completely rely on it.

In this respect, we propose a methodology able to evaluate the amount of energy not supplied from renewable sources. In particular, we consider a system composed by a renewable energy source (e.g., photovoltaic energy, wind energy), an end-user demanding energy on a daily basis, and a battery able to store the -possibly generated- energy that is not required by the end-user. We consider two types of uncertainty: randomness due to inherent variability in the system

behavior (aleatory uncertainty) and imprecision due to lack of knowledge and information on the system (epistemic uncertainty) as typically distinguished in system risk analysis Helton and Oberkampf (2004). As illustrated in recent works of risk analysis Li and Zio (2012), we address the co-existence of aleatory and epistemic uncertainties in the reliability assessment of a distributed generation system, representing the aleatory variables as probabilistic and the epistemic ones as possibilistic, and apply a hybrid propagation approach of both types of uncertainties. We embrace the Monte Carlo Simulation and Fuzzy Interval Analysis approach for the joint propagation of uncertainties Li and Zio (2012); Baraldi and Zio (2008).

Traditionally, aleatory uncertainty of an energy distribution system is represented by a unique probability density function that is inferred from historical data of one fixed period Li and Zio (2012); Baraldi and Zio (2008). Nevertheless, the data distribution evolve through time in a continuous way. Here we propose to consider that time variation within a Functional Data Analysis (FDA) framework Ramsay and Silverman (2002, 2005); Ferraty and Vieu (2006), where data are represented as functions of a continuous variable, which in our application is time. By applying FDA methods, it is then possible to model the entire time evolution of data. We propose to analyze this time evolution in order to obtain more realistic results from the uncertainty analysis. If we neglect long-term climatic changes, we may assume that this time evolution is one-year periodic. Hence, in this work, we model climatic data as random samples from parametric distributions with one-year periodic parameters.

To exemplify the methodology, we analyze the aleatory and epistemic uncertainties of a model of a photovoltaic energy distribution system made of a solar panel, a storage energy system and loads (power demanded by the end-users). As a quantitative indicator of the analysis we evaluate the Expected Energy Not Supplied, a reliability index commonly used in this field Billinton et al. (1984). The results of the uncertainty propagation in the case study are compared with: *i*) the pure probabilistic uncertainty propagation approach based on the same time-varying distributions Marseguerra and Zio (2002); and *ii*) the Monte Carlo Simulation and Fuzzy Interval Analysis approach considering the random variables constant in time, i.e. described by a unique probability density function.

7.2 Methodology

The methodology that we propose to evaluate uncertainties in renewable energy generation is based on the joint modeling and propagation of all the uncertainties of the model, that can be either aleatory or epistemic. The first step in order to evaluate the uncertainties in renewable energy generation consists in modeling the system of energy generation, listing all sources of uncertainty in the model inputs (e.g., electricity demand) that propagate to the model output (e.g. electricity supply). These sources of uncertainty can be distinguished into two types: epistemic uncertainties (due to due to lack of knowledge, and for which no historical data are available) and aleatory uncertainties (due to the intrinsic variability of the system, typically modeled by means of large amounts of historical data). The former type of uncertainty is also referred to as “reducible”

uncertainty to highlight that a gain of information about the system can lead to a reduction of epistemic uncertainty. In renewable energy applications, epistemic uncertainty typically characterizes the parameters of the devices due to i) the lack of information provided by the manufacturers for commercial reasons and ii) the limited quantity of data available for each house for private issues Izquierdo et al. (2011).

Aleatory uncertainties can instead be due to the variability of the energy source (e.g., wind speed and direction, solar irradiation) and the loads (i.e., power demanded by the end-users). In the current risk assessment practice, both types of uncertainties are represented by means of probability distributions with fixed parameters. However, potential limitations are associated to a probabilistic representation of epistemic uncertainty under limited information Helton and Oberkampf (2004) and a number of alternative representation frameworks have been proposed, e.g., fuzzy set theory, evidence theory, possibility theory and interval analysis Klir and Yuan (1995); Aven and Zio (2011). Given the representation power of possibility theory and its relative mathematical simplicity, we adopt it to describe the epistemic uncertainty in renewable energy applications. In addition, we represent the aleatory uncertainty by means of time-varying probability distributions since the variables associated with renewable energy systems can vary with time (e.g., day and night, seasons, etc.). For instance, the solar irradiation in summer is higher than in winter; as a consequence, the mean of its probability distribution should change with seasons (i.e., it should be higher in summer and lower in winter). The parameters of these time-varying probability distributions have been evaluated from historical data by applying FDA techniques.

Finally, the aleatory and epistemic uncertainties are jointly propagated on the renewable energy system by combining Monte Carlo simulation and Fuzzy Interval Analysis. Actually, the possibilistic representation of uncertainty can both be combined with and transformed into the traditional probabilistic representation. In the following subsections we present the details of the methodology adopted to model the time-varying data distribution, and propagate the aleatory and epistemic uncertainties from the input variables of the system to the output variable of interest.

7.2.1 Uncertainty modeling of time-varying data

Suppose to observe n realizations of the quantity of interest ξ (e.g., irradiation, wind speed) for the chosen location through time, during the year: for each time unit t_q (i.e., a day), we observe n different samples of functional data $\xi_i(t_q)$, where $i = 1, .., n$ denotes the sample units, and $q = 1, ..., Q$ denotes the different time units. We suppose that, for a fixed time t_q, the observed data $\xi_i(t_q)$ is a random independent sample from a fixed parametric distribution $F_{\boldsymbol{\eta}}$, described by a set of unknown parameters $\boldsymbol{\eta}(t_q) \in \mathbb{R}^r$:

$$\xi_i(t_q) \sim F_{\boldsymbol{\eta}}, \quad \forall i = 1, ..., n,\ q = 1, ..., Q. \tag{7.1}$$

The distribution $F_{\boldsymbol{\eta}}$ can be chosen in different ways, according to the data that we are modeling. For instance, a Beta distribution is typically used to model

solar irradiation, while a Weibull distribution is used to model wind speed Atwa et al. (2010); Li and Zio (2012); Salameh et al. (1995).

We assume that observations on different time units are conditionally independent, given the values of the parameters $\boldsymbol{\eta}$. In particular, this implies that the dependence structure of the solar irradiation on different days is entirely expressed by means of the time-varying structure of its parameters, which we suppose can be modeled as smooth and regular functions of time, due to the intrinsic regularity of data.

To estimate the time-varying parameters we adopt, for each time unit t_q, the method of moments. So, to elicit a time-varying estimate for the distribution of data, we only need to find time-varying estimates for the first r moments of the distribution of data. Furthermore, we suppose that the moments of the distribution are regular one year-periodic functions. Since the sample daily moments are extremely non-regular functions, we consider a method to regularize data, estimating a proper low dimensional functional space in which they are defined, by exploiting the procedure proposed in Pini and Vantini (2013). In the following, we describe the smoothing procedure applied to estimate the rth moment of data distribution.

Estimate of time-varying moments

Suppose that we want to estimate the moment of order r of the data distribution, $r \geq 1$. Let $\varphi_i(t) = \xi_i(t)^r$. We need to estimate the mean of functions $\varphi_i(t)$. We apply a Fourier-based Interval Testing Procedure (ITP) described in Pini and Vantini (2013). The method consists in the following three steps:

1. **Basis Expansion**: functional data (in our case: time-varying irradiation data) are represented through the coefficients of a truncated ordered basis expansion;
2. **Interval-Wise Testing**: a suitable test is performed on each interval of ordered basis coefficients;
3. **Multiple Correction**: for each component of the basis expansion, an adjusted p-value is computed from the p-values of the tests performed in the previous step.

The final result of the procedure is a family of adjusted p-values, one for each basis function used in the expansion. This result can be used to select the basis components that are statistically significant to describe the mean function of data (for instance, selecting all the components with associated adjusted p-value lower than the 5% level).

In our application, data are assumed to be one-year periodic functions. Hence, a natural choice for the basis used to describe data is the one-year periodic Fourier basis. In detail, we use an interpolating Fourier expansion:

$$\varphi_i(t_q) = \frac{a_i^{(0)}}{2} + \sum_{h=1}^{(Q-1)/2} a_i^{(h)} \cos\left(\frac{2\pi}{Q} h t_q\right) + b_i^{(h)} \sin\left(\frac{2\pi}{Q} h t_q\right), \quad Q = 365. \quad (7.2)$$

In the case of a Fourier basis expansion, intervals of basis components are frequency bands. In detail, equation (7.2) associates at each data, and for each

frequency $h > 0$, a bivariate vector of coefficients $(a_i^{(h)}, b_i^{(h)})$, and for the frequency $h = 0$ a coefficient $a_0^{(h)}$. Denote as $\left(A^{(h)}, B^{(h)}\right)'$ the bivariate distribution of coefficients $(a_i^{(h)}, b_i^{(h)})$ (and $A^{(0)}$ the distribution of the 0th frequency). For each frequency $h > 0$, the ITP can be applied to associate an adjusted p-value to each of the following bivariate tests:

$$H_0^{(h)} : \mathbb{E}\left[\left(A^{(h)}, B^{(h)}\right)'\right] = (0,0)' \text{ vs. } H_1^{(h)} : \mathbb{E}\left[\left(A^{(h)}, B^{(h)}\right)'\right] \neq (0,0)', \quad (7.3)$$

while for the 0th frequency, an adjusted p-value to the univariate test:

$$H_0^{(0)} : \mathbb{E}\left[A^{(0)}\right] = 0 \text{ vs. } H_1^{(h)} : \mathbb{E}\left[A^{(0)}\right] = 0, \quad (7.4)$$

In particular, by means of tests (7.3)-(7.4), we aim at selecting the frequencies that are significantly different from zero in the expansion of the mean signal. The final time-varying estimate of the mean function will then be evaluated as the Fourier basis expansion on these frequencies.

In order to apply the ITP, we then need to specify how to perform the interval-wise tests of the second phase of the procedure, that is in our case, how to perform a test on each frequency band. In detail, we need a test for each frequency (single-frequency tests), and a test for each interval of frequencies (multiple-frequency test). The approach that we use to perform such tests is a non-parametric approach based on the suitable combination of joint permutation tests on each frequency. We start by performing each single-frequency test, by means of permutation tests Pesarin and Salmaso (2010). For each frequency $h > 0$ we perform a bivariate test to test the null hypothesis $\mathbb{E}[(A^{(h)}, B^{(h)})] = (0,0)$, based on the joint changes of the signs of vectors $(a_i^{(h)}, b_i^{(h)})$ and on the Hotelling T^2 statistic $T(\mathbf{a}^{(h)^*}, \mathbf{b}^{(h)^*}) = (\bar{a}^{(h)^*}, \bar{b}^{(h)^*})' S_{h,h}^* (\bar{a}^{(h)^*}, \bar{b}^{(h)^*})$, where $(\mathbf{a}^{(h)^*}, \mathbf{b}^{(h)^*})$ denote the permuted data, and $S_{h,h}^* \in \mathbb{R}^{(2\times 2)}$ is the covariance matrix of permuted data at frequency h. For the 0th frequency, we perform a univariate permutation test based on the squared of the univariate Student t statistic and on the change of the signs of the coefficients $a_i^{(0)}$. It is important to note here that the permutations used to build the single-frequency tests are the same across frequency. This aspect will be key to build the multiple-frequency tests.

To perform multiple-frequency tests, we combine the results of the single-frequency tests by means of the non-parametric combination (NPC) methodology, based on the Fisher combining function Pesarin and Salmaso (2010). The NPC is a method able to build multiple-feature permutation tests by means of combining joint single-feature permutation tests.

According to the ITP, when the tests on each interval of basis components (in this case: each frequency band) is performed, the adjusted p-value associated to the tests (7.3) on frequency h is computed as the maximum among all p-values of tests pertaining that frequency. Once the adjusted p-values are computed, we can select as significant all the frequencies with an associated adjusted p-value lower than 5%. This final selection is provided with an interval-wise control of the family wise error rate. In detail, this control means that the probability of wrongly rejecting any frequency band is lower than 5%.

Once selected the significant frequencies, the estimate of the functional moment will then be the one-year periodic function obtained by means of the Fourier

expansion of the sample mean coefficients of functions φ_i, restricted to the selected frequencies. That is, if $\mathbf{v} = (v_i, ..., v_{(Q-1)/2})$ is the index vector identifying the final selection of significant frequencies, ($v_h = 0$ if the result of the h-th test is $H_0^{(h)}$, $v_h = 1$ if the result of the h-th test is $H_1^{(h)}$), the final estimate of the functional moment is given by:

$$\hat{\mu}_\varphi(t_q) = \frac{\bar{a}^{(0)}}{2} + \sum_{h \in \mathbf{v}} \bar{a}^{(h)} \cos\left(\frac{2\pi}{Q} h t_q\right) + \bar{b}^{(h)} \sin\left(\frac{2\pi}{Q} h t_q\right), \tag{7.5}$$

where $\bar{a}^{(h)} = 1/n \sum_{i=1}^{n} a_i^{(h)}$, and $\bar{b}^{(h)} = 1/n \sum_{i=1}^{n} b_i^{(h)}$.

7.2.2 Joint propagation of aleatory and epistemic uncertainties

When both the aleatory and epistemic uncertainties are described by probability distributions, a pure probabilistic approach can be adopted for their propagation to the model output. This approach consists on the Monte Carlo sampling of possible values of all the input variables from the corresponding probability distributions and the subsequent computation of the model output in correspondence of the input values sampled Marseguerra and Zio (2002)). Random realizations of the model output can be obtained repeating a large number of times this procedure considering each time new samples of the input variables.

Instead, when the epistemic uncertainty is represented in possibilistic terms, the joint propagation of the aleatory and epistemic uncertainty can be performed by combining the Monte Carlo technique and the extension principle of fuzzy set theory by means of the following two main steps Baudrit et al. (2006): (*i*) repeated Monte Carlo sampling of the random variables to process aleatory uncertainty; and (*ii*) fuzzy interval analysis to process epistemic uncertainty.

In this work, the random variables are represented by time-varying probability distributions; therefore these two steps have to be repeated for all the time steps in the period of interest. Details of possibility theory are not reported here for brevity sake, the interested reader is referred to Dubois (2006). The operative steps of the procedure for the case study under analysis are illustrated in Appendix 7.A.

7.3 Case Study

The case study that we present here concerns the design of a solar panel that provides electrical energy to a house located in the south of Spain. The size and number of the panels is a trade-off between their performance to satisfy the demand of energy and the high costs of construction and maintenance. To perform this evaluation we consider the demand of power requested by the end-users and the possibility of storing the generated exceedance power in a battery, that is necessary when the power from the solar energy is not sufficient (e.g. during cloudy days) or it is completely absent (e.g. during nights). This case study deals with a big amount of uncertainty due to the stochasticity of the behavior of the end-users, the variability of the solar irradiation, the lack of knowledge about some operation parameters of the solar panels.

The system consists of three different parts: the solar panel, the load and the battery, as illustrated in Figure 7.1.

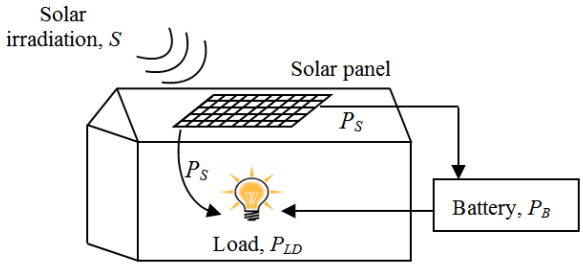


Figure 7.1: Scheme of the system of a solar panel, load and battery considered in the case study

The power generated by the solar panel, P_S[kW], is a function of the solar irradiation, S, the number of solar cells, N, and a vector of operation parameters, $\boldsymbol{\theta} = (I_{MPP}, V_{MPP}, V_{OC}, I_{SC}, N_{ot}, k_c, T_a)$ Li and Zio (2012):

$$P_S = N \cdot FF \cdot V_y \cdot I_y, \tag{7.6}$$

where $I_y = S \cdot I_{SC} + k_c(T_c - 25)$, $V_y = V_{OC} - k_v \cdot T_c$, $T_c = T_a + S(N_{ot} - 20)/(0.8)$, $FF = (V_{MPP} \cdot I_{MPP})/(V_{OC} \cdot I_{SC})$. I_{MPP}[A] and V_{MPP}[V] are the current and voltage at maximum power point, respectively, V_{OC}[V] is the open circuit voltage, I_{SC}[A] is the short circuit current, N_{ot}[°C] is the nominal operating temperature, k_c[A/°C] is the current temperature coefficient, k_v[V/°C] is the voltage temperature coefficient, T_a[°C] is the ambient temperature, and the load, P_{LD}[kW], is the power demanded by the end-users.

The output model of the battery is the power, P_B[kW], that can be stored in the battery when the solar panel produces more power than the demand, i.e. when $P_{Diff} = P_S - P_{LD} > 0$, and can be given to the end-users when the opposite occurs, i.e. when $P_{Diff} = P_S - P_{LD} < 0$. In the present study we have adopted a dynamic model Chen et al. (2011) to represent the level of charge of the battery, calculating the difference between stored energies of two consecutive steps. The following equations describe the model of the battery when it is charging, i.e. $\Delta P_B(t) = -P_{Diff} < 0$ (7.7)-(7.8), when it is discharging, i.e. $\Delta P_B(t) = -P_{Diff} > 0$ (7.9)-(7.10) and when it is idle, i.e. $\Delta P_B(t) = P_{Diff} = 0$ (7.11).

$$-\eta_c \Delta P_B(t) \Delta t_{min} \leq K_c Q_{max}; \tag{7.7}$$

$$Q(t+1) = Q(t) - \eta_c \Delta P_B(t) \Delta t_{min}; \tag{7.8}$$

$$\Delta P_B(t) \Delta t_{min}/\eta_d \leq K_d Q_{max}; \tag{7.9}$$

$$Q(t+1) = Q(t) - \Delta P_B(t) \Delta t_{min}/\eta_d; \tag{7.10}$$

$$Q(t+1) = Q(t) - W_{hourly}. \tag{7.11}$$

In equations (7.7-7.11), $Q(t)$[kWh] is the capacity of the battery at hour t, η_c and η_d are the charging and discharging efficiency, respectively, K_c and K_d are the maximum portion of rated capacity that can be added to and withdraw from storage in an hour, respectively, Q_{max} is the rated maximum stored energy, W_{hourly}[kWh] is the battery hourly discharged energy, Δt_{min} is the scheduling interval. The parameter values adopted in the model are: $\eta_c = \eta_d = 0.85$, $K_c = K_d = 0.3$, $Q_{max} = 40$, $W_{hourly} = 0.5$ kWh and $\Delta t_{min} = 1$ h. In this work, the initial level in the battery has been assumed to be equal to zero.

7.3.1 Uncertainty representation

In the model of the solar panel (7.6) the inputs can be classified in *i*) aleatory variable, i.e. the solar irradiation and the load, *ii*) epistemic variables, i.e. the operation parameters of the vector $\boldsymbol{\theta}$, and *iii*) constant, i.e. the number of solar cells N that in the present simulation has been taken equal to 30.

Operation parameters

The operation parameters $\boldsymbol{\theta}$ are classified into parameters provided by the manufacturers, e.g. $I_{MPP}, V_{MPP}, V_{OC}, I_{CS}, N_{ot}, k_c, k_v$, and by the end-users, e.g. T_a. Both are associated with epistemic uncertainty, and we represent them by trapezoidal possibility distributions ($\pi^{I_{MPP}}$, $\pi^{V_{MPP}}$, $\pi^{V_{OC}}$, $\pi^{I_{CS}}$, $\pi^{N_{ot}}$, π^{k_c}, π^{k_v}) as proposed in Li and Zio (2012).

Solar irradiation

Solar irradiation $S[\mathrm{kW/m^2}]$ depends on the variability of the weather. It is typically described by a probabilistic distribution, e.g. a Beta distribution, whose parameters, α and β, are inferred from sufficient historical data and are fixed for a given period Li and Zio (2012). In the present paper, coherently with the literature, we represent the solar irradiation with the Beta distribution. The main novelty of our approach with respect to the literature is that we consider the evolution of solar irradiation through time, estimating different values of the parameters α and β for each day of the year, according the method explained in Subsection 7.2.1.

The historical data used to elicit the parameters are daily irradiations in a geographical close area near Seville, Spain, (the square with latitude in the interval $[37, 38]$ and longitude in $[-6, -5]$), registered from July 1983 to June 2005 and stored in the database NASA: Earth Surface Meteorology for Solar Energy NASA (2008)[1]. By way of example, Figure 7.2 shows an histogram of the historical data recorded and the correspondent Beta distribution of the solar irradiation in four different days in July and August (1$^{\text{st}}$ and 21$^{\text{st}}$ July, 11$^{\text{th}}$ and 31$^{\text{st}}$ August, respectively). The figure compares a Beta distribution characterized by constant parameters (green line), and one characterized by time-varying parameters estimated with the method that we will present in this Section (red line). We observe from the histograms that the distribution of solar irradiation is changing through time even in the relatively small period of two months. Hence, a correct approach to model such irradiation data is to consider its time-varying distribution, rather than a constant one.

We suppose that, for a fixed day t_q, the observed data $S_i(t_q)$ is a random independent sample from a beta distribution of parameters $\alpha(t_q)$ and $\beta(t_q)$:

$$S_i(t_q) \sim Beta(\alpha(t_q), \beta(t_q)), \quad \forall i = 1, ..., 22,\ q = 1, ..., 365. \tag{7.12}$$

To estimate the time-varying parameters we adopt the methodology discussed in Subsection 7.2.1. Since we need to estimate, for each t_q, the two parameters

[1]These data were obtained from the NASA Langley Research Center Atmospheric Science Data Center Surface meteorological and Solar Energy (SSE) web portal supported by the NASA LaRC POWER Project. Data are freely available at: NASA Surface Meteorology and Solar Energy, A Renewable Energy Resource web site (release 6.0): http://eosweb.larc.nasa.gov

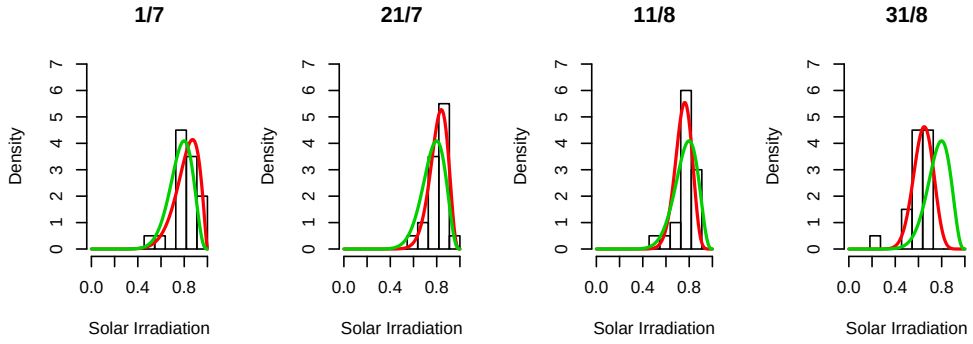


Figure 7.2: Histogram of the recorded data and the correspondent Beta distribution for the 1st and 21st of July, 11th and 31st of August, and corresponding Beta distributions characterized by: constant parameters (green line); time-varying parameters (red line).

$\alpha(t_q)$ and $\beta(t_q)$, we need to estimate the mean and variance of the data distribution, for each time unit t_q.

Load

The load, P_{LD}, is affected by aleatory uncertainty since its value depends on the behavior of the end-users. Typically it is modeled by a normal probabilistic distribution Liu et al. (2011), with parameters inferred from the large amount of historical data available. In this work, we use a normal distribution, estimating two different time-varying mean values for days and nights, μ_{day}, and μ_{night}, respectively, following the procedure explained in Section 7.2.1, and maintaining a same standard deviation σ.

As well as the solar irradiation, also the load P_{LD}[kW] has a time varying structure. In particular, we suppose that, for each day of the year t_q, the load has two normal distributions for days and nights, with the same constant standard deviation ($\sigma = 0.25$ kW) and two time-varying means ($\mu_{P_{LD},day}(t_q)$ and $\mu_{P_{LD},night}(t_q)$, respectively). The model assumed for the day and night load, for each time t_q is then the following:

$$P_{LD,day/night}(t_q) \sim N(\mu_{P_{LD},day/night}(t_q), \sigma^2), \quad q = 1, ..., Q. \tag{7.13}$$

To estimate the day and night mean functions, it is not possible to proceed applying the ITP to the daily load data, as they are not directly available. The daily mean electrical consumption of a house in the south of Spain is about 24.54 kWh Sech-Spahousec (n.d.) and in the night the demand of electricity is the half than during the day Omie (2012). Thus, the estimated means of the hourly load for days and nights are 1.363 kW and 0.682 kW, respectively. Since these data are aggregated through the entire year, it is not possible to infer a time varying distribution. Consequently, a different approach is here necessary.

Most of the usual household electrical devices (e.g. washing machine, refrigerator, TV) are approximately used in the same way in summer and winter, and, thus, their electrical consumption can be assumed to follow a constant distribution throughout the year. The only devices that may have a time-varying load are the air conditioning systems (whose load varies in the warm months depend-

ing on the external temperature) and the lighting (whose load changes through the year depending on the variation of daylight time). Since for the former, the load is higher than for the latter, we consider only the air conditioning systems (AC) as a device with a time varying load. Since the AC load depends on the external temperature, we first apply the ITP to minimum and maximum daily temperatures, to find a smooth time-varying estimate of their mean functions. Then, we use this estimate to calculate the time-varying day and night means of the load due to the AC. The steps of the procedure applied to calculate the load are detailed in Appendix 7.B.

7.3.2 Uncertainty propagation

The joint propagation of aleatory and epistemic uncertainties represented by probability and possibility distributions, respectively, is carried out by Monte Carlo simulation and Fuzzy Interval Analysis. Since the analysis is time-varying, the procedure is repeated for each time steps in the period of interest.

In the present case study, the aleatory variables are the solar irradiation and the loads that vary during days and nights. As a consequence of their variation, the level of energy in the storage system varies too. We assume that at the first time step it is day (i.e., there is solar irradiation) and the level of energy in the storage system is equal to zero.

When the power generated by the solar panel is higher than the demands of the end-users, the level of energy in the storage system increases and the end-users are satisfied (the energy not supplied (ENS) is equal to zero); instead when the power generated is lower than the demands of the end-users two cases can occur: 1) there is enough energy in the storage system to supply the end-users, so the level in the storage decreases but the end-users are satisfied (the ENS is equal to zero); 2) there is not enough energy in the storage system, so the level in the storage decreases to zero (if it is not already zero) and the ENS to the end-users is positive.

The following time steps have been considered in our case:

- $\Delta t_{min} = 1$[h] is the smallest time step of the system model. The total number of hours in the period of interest is defined by the variable N_{steps};
- $\Delta t_{max} = 12$[h] is the time interval in which the power generated by the solar panel, P_S, and the one demanded by the end-users, P_{LD}, can be considered constant. This assumption has been introduced to reduce the computational time of the simulation and to distinguish only between day and night, and is coherent with the calculation of P_{LD} proposed in Section 3.2. Therefore, the total number of different values considered for those variables is $N_{steps}/\Delta t_{max}$.

The joint uncertainty propagation that consists in combining Monte Carlo technique with the extension principle of fuzzy set theory is illustrated in details in Appendix 7.A with respect to the case study considered in this work. At the end of the procedure an ensemble of m fuzzy random realizations (fuzzy intervals) π^k_{EENS}, $k = 1, \ldots, m$, of the Expected Energy Not Supplied (EENS) index is obtained.

On the basis of the rule of the possibility theory Baudrit et al. (2006), these possibilistic distributions can be aggregated. As a result, two cumulative distribution functions (cdfs), called belief and plausibility (i.e., the lower and upper cdfs, respectively), of the Expected Energy Not Supplied are obtained. They can be interpreted as bounding cumulative distribution functions Baudrit et al. (2006) and they contain all the possible cumulative distribution functions that can be generated by a pure probabilistic approach that considers all the inputs variables as probabilistic. For the sake of comparison, we have embraced also this method with m = 10000 samples of the probabilistic variables: in this case, the possibilistic distributions of the input variables are transformed into probabilistic distributions by the normalization method given in Flage et al. (2008).

7.3.3 Results

The adjusted p-values for the first and second moments are reported in the top panels of Figure 7.3, where the 5% level is indicated as a horizontal red line. For the first moment, the p-values associated to the first three frequencies are lower than the chosen significance level (and are, furthermore, lower than every typically-used significance level), whereas all other p-values are higher. Hence, we have a rejection corresponding to the mean value (zero-frequency) and the sine and cosine functions of period one year. For the second moment we have instead a rejection on the first four frequencies, corresponding to the mean value and the sine and cosine functions of period one year, six and three months. The final estimates are periodic functions fully described by the sample means coefficients on these frequencies.

To appreciate the result of the test, the lower panels of Figure 7.3 show, for the first two moments, the two estimates of the mean: ITP estimate (red) and daily estimate (black). Gray lines are the solar irradiation data in southern Spain for the first moment (left), and squared solar irradiations for the second moment (right). Comparing the ITP and daily estimates, it can be seen that the first method gives smooth curves, which follow the yearly fluctuations of the quantity of interest, whereas the second one gives extremely irregular functions.

The results of the analysis on the min-max temperature data, and the subsequent results of the load parameters are presented in Figure 7.4. On the top, the ITP-adjusted p-values for the minimum (left) and maximum (right) temperatures are reported. In this case, the ITP selects as significant the mean value and the first two frequencies, both for the min and the max temperatures. On the middle, the daily minimum and maximum temperatures data in southern Spain (light blue and red lines, respectively) are shown, together with the ITP estimates of the two functional means, evaluated according to the ITP results. The horizontal line indicates the threshold temperature at which the AC is turned on, $T_{thres} = 26°\text{C}$. On the bottom panel, the estimates of the time-varying means of the load, for days and nights (yellow and black lines, respectively) are reported. To appreciate how the modeling of the load data is related to the test results, the figure indicates the densities of the simulated day and night load $P_{LD,day}(t_q)$ and $P_{LD,night}(t_q)$ for a summer and a winter day. Note that the variability of the normal distribution remains constant, whereas its mean level changes from days and nights, and from winter to summer.

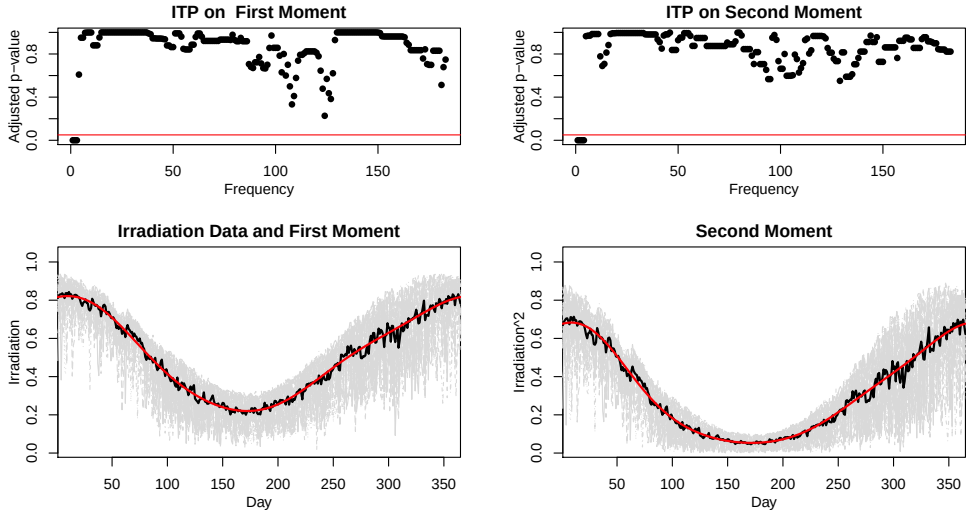


Figure 7.3: Top: ITP-adjusted p-values for each frequency of the Fourier expansion for the first (left) and second (right) moment. Bottom: ITP (red) and daily (black) estimates of the first two moments. Gray lines: solar irradiation (left); and squared solar irradiation (right).

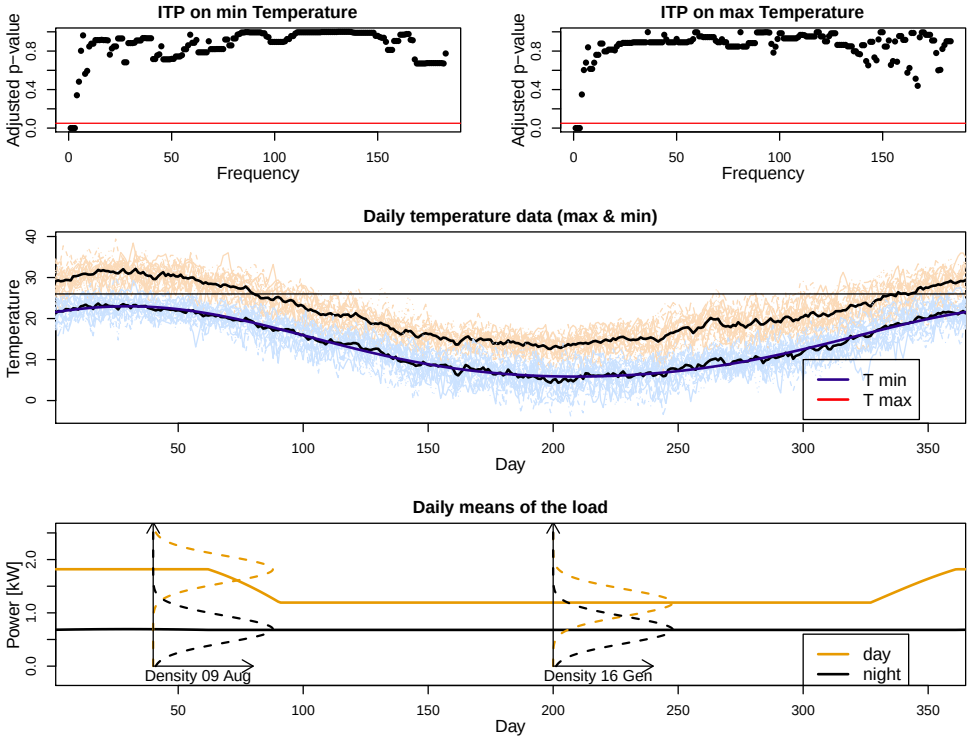


Figure 7.4: Top: ITP-adjusted p-values for each frequency of the Fourier expansion for the min (left) and max (right) temperatures. Middle: daily min (light blue) and max (light red) temperatures data and ITP estimates for the means (bold blue and red lines). Bottom: estimates of the time-varying means of the load for days (yellow) and nights (black), and densities of simulated data in a summer and winter day.

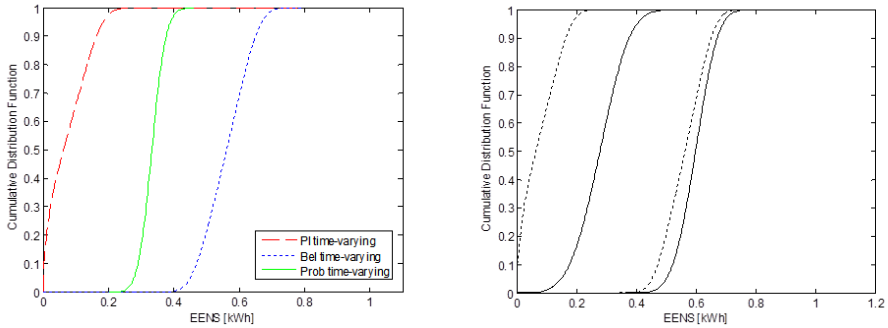


Figure 7.5: Left: comparison of the cumulative distribution functions of the EENS [kWh] obtained by the pure probabilistic approach (solid line) with the belief (dotted line) and plausibility (dashed line) functions obtained by the Monte Carlo and Fuzzy Interval Analysis. Right: comparison of the lower and upper cumulative distribution functions of the EENS obtained by the Monte Carlo and Fuzzy Interval Analysis approach considering constant (solid line) and time-varying (dotted line) parameters of the probabilistic distribution.

The computation of the EENS index has been performed by applying the method of Section 3.3 considering the time-varying parameters of the probabilistic distribution of the solar irradiation and of the loads determined above. The analysis has been carried out with respect to the month of July that is a critical period for the high demand of power by the end-users. In fact, the hot temperature reached in the south of Spain gives rise to a large use of air conditioners. Figure 7.5 reports on the left panel a comparison of the cumulative distribution functions of the EENS index obtained by the probabilistic uncertainty propagation approach (solid lines) with the belief (lower curves) and plausibility (upper curves) functions obtained by the Monte Carlo and Fuzzy Interval Analysis approach described in Section 3.3.

The Monte Carlo and Fuzzy Interval Analysis method explicitly propagates the aleatory and epistemic uncertainty: the separation between the belief and plausibility functions reflects the imprecision in the knowledge of the possibilistic variables and the slope pictures the variability of the probabilistic variables. Instead, the uncertainty in the output distribution of the pure probabilistic approach is given only by the slope of the cumulative distribution. As expected, the cumulative distribution of the EENS obtained by the pure probabilistic method is within the belief and plausibility functions obtained by the Monte Carlo and Fuzzy Interval Analysis approach.

Figure 7.5 (on the right) compares the previous results, carried out with the Monte Carlo and Fuzzy Interval Analysis approach, with those obtained by the same method but by considering constant the parameters of the probabilistic distributions of the solar irradiation S, and the loads P_{LD}. A conservative measure of the EENS distribution to be used to evaluate the size of the panel can be chosen as the $99th$ percentile of the distribution. The lower and upper values of this measure, obtained with time-varying or constant parameters, are reported in Table 7.1. The value obtained by considering time-varying parameters (with a model that describes with a higher precision the real climatic conditions) are lower than the ones obtained with constant parameters.

It can be seen that the lower and upper cumulative distributions functions

	EENS	
	Lower value [kWh]	Upper value [kWh]
Time-varying	0.22	0.70
Constant	0.45	0.78

Table 7.1: Lower and upper values of the $99th$ percentile of the EENS distribution, evaluated with time-varying and constant parameters.

obtained by considering time-varying parameters are always lower than those resulted by keeping constant those parameters. This means that a time-varying analysis allows designing the solar panel with smaller dimension. Furthermore, the gap between the cumulative distributions functions obtained by considering time-varying parameters is higher than that between the curves obtained by keeping constant those parameters. In particular, by considering time-varying parameters, we introduce a higher variability on the EENS estimation, due to the fact that the distribution of data changes daily. The higher variability allows considering within our model the situation in which the solar panel fully support the load demand, including the zero value in the EENS distribution.

7.4 Conclusions

We illustrated a methodology to represent and joint propagate the aleatory and epistemic uncertainties of renewable energy generation systems. We represented the former ones by probability distributions and the second ones by possibility distributions. In particular, we focused on the aleatory variables that present a time-varying behavior (e.g., solar irradiation and loads) and we elicited time-varying probability distributions from historical climatic data.

Once all uncertainties have been represented, we proceeded to evaluate the output of interest (e.g., the Expected Energy Not Supplied) by propagating the uncertainties through the model of the energy distribution system. The results that can be obtained from this analysis can provide a support in the decision process for the dimensioning of the energy generation system.

In this work, we applied the methodology to a model of an energy system made of a solar panel, a storage energy system and the loads. In particular, we considered the variations in time of the solar irradiation and the loads, describing them by probabilistic distributions with time-varying parameters. We evaluated the Expected Energy Not Supplied as a quantitative indicator of the analysis.

Two main results have to be highlighted: (i): the uncertainty propagation method divides the contribution of the aleatory and epistemic uncertainty, identifying an upper and a lower bound of values for the EENS, i.e. an interval of values of the EENS index is determined. This can be of interest in the decision making process to identify the proper size of the solar panel; (ii) accounting for the time-varying parameters in the distributions of the solar irradiation and of the loads leads to more realistic results that allows to reduce the dimension of the solar panel. Thus, considering constant parameters an overestimation of the size of the solar panel can be done.

Appendix

7.A Joint uncertainty propagation

The operative steps of the procedure for the joint uncertainty propagation by Monte Carlo simulation and Fuzzy Interval Analysis are here detailed with respect to the case study presented in Section 7.3. As a quantitative indicator of the analysis, the Expected Energy Not Supplied index is computed.

1. Set $k = 1$ (outer loop processing aleatory uncertainty).
2. Sample the solar irradiations $\tilde{S}_l^k$, $l = 1, ..., N_{steps}/\Delta t_{max}$ from Beta distribution (equation (7.1)) if l is an odd number (i.e. when it is day), otherwise, set $\tilde{S}_l^k = 0$ (i.e. when it is night). Then, sample the loads $\tilde{P}_{LD,l}^k$, $l = 1, ..., N_{steps}/\Delta t_{max}$ from equation (7.13) taking into account the different distributions associated with that variable during the days and nights. The vectors $[\tilde{\mathbf{S}}^k]_l$ and $[\tilde{\mathbf{P}}_{LD}^k]_l$, are transformed into $[\tilde{\mathbf{S}}^k]_j$ and $[\tilde{\mathbf{P}}_{LD}^k]_j$, $j = 1, ..., N_{steps}$, respectively, repeating each value Δt_{max} times, to obtain values of solar irradiations and loads for each hour in all the period of interest.
3. Set $\alpha = 0$ (middle loop processing epistemic uncertainty).
4. Set $j = 1$ (inner loop processing the time variation).
5. Select the corresponding α-cuts of the possibility distributions ($\pi^{I_{MPP}}$, $\pi^{V_{MPP}}$, $\pi^{V_{OC}}$, $\pi^{I_{CS}}$, $\pi^{N_{ot}}$, π^{k_c}, π^{k_v}) as intervals of possible values of the possibilistic variables I_{MPP}, V_{MPP}, V_{OC}, I_{CS}, N_{ot}, k_c, k_v.
6. Calculate the smallest and largest values of the solar power generated, $\underline{P}_{S,j,\alpha}^k$ and $\overline{P}_{S,j,\alpha}^k$, respectively, by equation (7.6) considering the fixed values S_j^k sampled for the random variables S and all values of the possibilistic variables I_{MPP}, V_{MPP}, V_{OC}, I_{CS}, N_{ot}, k_c, k_v in the α-cuts of their possibility distributions.
7. Compute the value $\underline{P}_{Diff,j,\alpha}^k = \underline{P}_{S,j,\alpha}^k - P_{j,k}^L$: if $\underline{P}_{Diff,j,\alpha}^k > 0$, go to step 7.a.; if $\underline{P}_{Diff,j,\alpha}^k < 0$ go to step 7.b., else go to step 7.c.:
    a. set to zero the Energy Not Supplied index, $\overline{ENS}_{j,\alpha}^k = 0$, and increase the level of energy in the battery by equation (7.8), $\underline{Q}_{j+1,\alpha}^k = f(Q_{j,\alpha}^k, \underline{P}_{B,j,\alpha}^k, \eta_c)$, where $\underline{P}_{B,j,\alpha}^k = -\underline{P}_{Diff,j,\alpha}^k$ if the constraint defined in equation (7.7) is verified, otherwise it is computed by equation (7.7). If the level of energy in the battery at the step $j+1$ is higher than its maximum capacity, i.e. $\underline{Q}_{j+1,\alpha}^k > Q_{max}$, then, set $\underline{Q}_{j+1,\alpha}^k = Q_{max}$;

b. decrease the level of energy in the battery by equation (7.10), $\underline{Q}^k_{j+1,\alpha} = f(Q^k_{j,\alpha}, \underline{P}^k_{B,j,\alpha}, \eta_d)$; if the constraint defined in equation (7.9) is verified $\underline{P}^k_{B,j,\alpha} = -\underline{P}^k_{Diff,j,\alpha}$ (case i.), otherwise $\underline{P}^k_{B,j,\alpha}$ is computed by equation (7.9) (case ii.). If the level of energy in the battery at the step $j+1$ is higher than zero, the Energy Not Supplied index is computed as $\overline{ENS}^k_{j,\alpha} = 0$ for the case i., and $\overline{ENS}^k_{j,\alpha} = -\underline{P}^k_{Diff,j,\alpha} - \underline{P}^k_{B,j,\alpha}$ for the case ii.; otherwise, set, $\underline{Q}^k_{j+1,\alpha} = 0$ and $\overline{ENS}^k_{j,\alpha} = -\underline{P}^k_{Diff,j,\alpha}$;

c. set $\overline{ENS}^k_{j,\alpha} = 0$, and decrease the level of the battery by equation (7.11), $\underline{Q}^k_{j+1,\alpha} = f(Q^k_{j,\alpha}, W_{hourly})$. If the level of energy in the battery at the step $j+1$ is lower than zero, then set $\underline{Q}^k_{j+1,\alpha} = 0$.

8. Repeat step 7. for the evaluation of the lower bounds of $\underline{ENS}^k_{j,\alpha}$, computing the upper values of $\overline{P}^k_{Diff,j,\alpha}$, $\overline{P}^k_{B,j,\alpha}$ and $\overline{Q}^k_{j,\alpha}$.

9. If $j \leq N_{steps}$, then set $j = j+1$ and return to step 5.; otherwise go to step 10.

10. Compute the total lower and upper bounds of the ENS index in the period under analysis as $\underline{ENS}^k_\alpha = \sum_{j=1}^{N_{steps}} \underline{ENS}^k_{j,\alpha}$, $\overline{ENS}^k_\alpha = \sum_{j=1}^{N_{steps}} \overline{ENS}^k_{j,\alpha}$; the lower and upper bounds of EENS, $\underline{EENS}^k_\alpha$ and $\overline{EENS}^k_\alpha$, are obtained by performing the means of $\underline{ENS}^k_\alpha$ and $\overline{ENS}^k_\alpha$, respectively.

11. Take the extreme values, $\underline{EENS}^k_\alpha$ and $\overline{EENS}^k_\alpha$, found in 10. as the lower and upper limit of the α-cut of the Expected Energy Not Supplied.

12. If $\alpha \neq 1$, then set $\alpha = \alpha + \Delta\alpha$ and return to step 4. to compute the EENS for another α-cut; otherwise a fuzzy random realization, π^k_{EENS}, of the EENS has been identified. If $k \neq m$, where m is the number of simulations, then set $k = k+1$ and return to step 2.; else stop the algorithm.

At the end of the procedure the fuzzy random realizations (fuzzy intervals) π^k_{EENS}, $k = 1, ..., m$ of the Expected Energy Not Supplied index is constructed as the collection of the values $\underline{EENS}^k_\alpha$ and $\overline{EENS}^k_\alpha$, found at step 10. (in other words, π^k_{EENS} is defined by all its α-cut intervals $(\underline{EENS}^k_\alpha, \overline{EENS}^k_\alpha)$.

7.B Time-varying estimate of the load

The operative steps of the procedure applied to find time-varying estimates of the load is briefly described here. Starting from the daily minimum and maximum temperatures in the Seville area, stored in the NASA data base NASA (2008), we calculate the time-varying mean of the load of an AC with some fixed characteristics. We consider a class "A" device, with an Energy Efficiency Ratio (EER) equal to 3.5. The number of AC installed in the house is set equal to the mean number of conditioners in Spanish homes in Andalusia, which is 1.623 INE (2008). The nominal power of the AC is calculated as $P^{AC}_N = Surf \cdot Ceiling \cdot 25$ ENEA (2006), where $Surf = 20m^2$ is the surface of the room and $Ceiling = 2.7m$ is the height of the ceiling. All data are chosen to indicate a representative Spanish house. Finally, since the proportion of Spanish that leave the AC turned

on at night is equal to 7.6% INE (2008), we multiply the AC load at nights by this proportion.

In order to calculate the mean load of such AC system, first of all, we find functional estimates for the mean tendency of the daily minimum and maxi-mum temperature for the given location ($T_{min}(t_q)$ and $T_{max}(t_q)$ [°C], respectively), by means of the ITP on min and max temperatures, as shown in Section 3.1.1. Then, for each day t_q, we perform the following calculation:

- We fix a threshold temperature $T_{thres} = 26$°C, and suppose that the AC is turned on when the external temperatures exceed the threshold, as in Izquierdo et al. (2011).

- We estimate the daily lapse of time in which the AC is turned on $h_{on}(t_q)$[h], supposing for each day a linear temperature profile between $T_{min}(t_q)$ and $T_{max}(t_q)$:

$$h_{on}(t_q) = 24\left(\frac{T_{max}(t_q) - T_{thres}}{T_{max}(t_q) - T_{min}(t_q)}\right). \tag{7.14}$$

  This approximation is justified by the comparison of our results with a daily temperature profile estimated from hourly data FreeMeteo (2012).

- The quantity $h_{on}(t_q)$ is then divided into daily (10.00 a.m. - 10.00 p.m.) and nightly (10.00 p.m. - 10.00 a.m.) hours of switching on ($h_{on}^{day}(t_q)$ and $h_{on}^{night}(t_q)$, respectively), assuming that $T_{max}(t_q)$ is attained at 4.00 p.m. and $T_{min}(t_q)$ at 6.00 a.m. FreeMeteo (2012).

- The mean power load on days of the AC is then calculated as:

$$\mu_{P_{LD},day}(t_q) = P_N^{AC} n_{room} h_{on}^{day}(t_q)/(12EER) \tag{7.15}$$

  The mean load on nights, is:

$$\mu_{P_{LD},night}(t_q) = P_N^{AC} n_{room} h_{on}^{night} \cdot 0.076(t_q)/(12EER) \tag{7.16}$$

  Note that both quantities are divided by 12[h] in order to found an estimate of the hourly power.

- The quantities $\mu_{P_{LD},day}$ and $\mu_{P_{LD},night}$ are finally added to the day and night fixed averages (mean load without AC), calculated in order to maintain the values of 1.363 kW and 0.682 kW as yearly means.

Part III

Software

Chapter 8

The package `fdatest`

The implementation of the ITP for one- and two-population tests, F- and t-tests for functional-on-scalar linear models, and the test of significance of the factors of a functional ANOVA model, have been made available for users in the R-package `fdatest` (Pini and Vantini, 2014), downloadable on CRAN.

The current version of the package requires functional data evaluated on a uniform grid; it automatically projects each function on a chosen functional basis (B-splines or Fourier in the current version); it performs the entire family of interval-wise tests; and, finally, it provides the matrix of the p-values of the previous tests and the vector of the adjusted p-values. The functional basis and the kind of test (including paired or unpaired scenario in the two-population framework) can be chosen by the user.

The main functions included in the current version of the package are the following:

- `ITP1bspline`: B-spline-based ITP for testing the mean of one population;
- `ITP1fourier`: Fourier-based ITP for testing the mean of one population;
- `ITP2bspline`: B-spline-based ITP for comparing the means of two populations;
- `ITP2fourier`: Fourier-based ITP (sine-cosine decomposition) for comparing the means of two populations;
- `ITP2pafourier`: Fourier-based ITP (amp-phase decomposition) for comparing the means of two populations;
- `ITPlmbspline`: B-spline-based ITP for testing functional-on-scalar linear models (F-test and t-tests);
- `ITPaovbspline`: B-spline-based ITP for testing the factors on a functional ANOVA;

Each function takes as input the point-wise evaluations of the functional data on a uniform grid, the mean (or mean difference) under the null hypothesis (in the case of one- or two- population tests), the covariates and factors (in the case of tests on linear models and ANOVA), the parameters characterizing the basis expansion (e.g., the order of the basis functions and the number of knots for the B-spline basis), and the number of random permutations to be used in the permutation tests. In addition, for the two-population case, the user can choose whether the test is paired or not.

A graphical output of the tests, displaying functional data, adjusted p-values and significant intervals at a chosen significance level is obtained applying the `plot` method to the output of each function. In addition, with the `ITPimage` function, it is possible to visualize the p-value heatmap.

The package also contains the `NASAtemp` dataset, providing the data analyzed in the NASA daily temperature case study, reported in the Appendix of Chapter 2. In the following, we report the manual of the `fdatest` package, with a detailed description of each function of the package, and some examples of how they are used to perform the ITP.

Package 'fdatest'

November 24, 2014

Type Package

Title Interval Testing Procedure for functional data

Version 2.0

Date 2014-22-10

Depends fda

Author Alessia Pini, Simone Vantini

Maintainer Alessia Pini <alessia.pini@mail.polimi.it>

Description The package implements the Interval Testing Procedure for functional data in different frameworks (i.e., one or two-population frameworks, functional linear models) by means of different basis expansions (i.e., B-spline, Fourier, and phase-amplitude Fourier). The current version of the package requires functional data evaluated on a uniform grid; it automatically projects each function on a chosen functional basis; it performs the entire family of multivariate tests; and, finally, it provides the matrix of the p-values of the previous tests and the vector of the corrected p-values. The functional basis, the coupled or uncoupled scenario, and the kind of test can be chosen by the user. The package provides also a plotting function creating a graphical output of the procedure: the p-value heat-map, the plot of the corrected p-values, and the plot of the functional data.

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2014-11-24 12:11:14

R topics documented:

`fdatest-package` *Interval Testing Procedure for functional data*

Description

The package implements the Interval Testing Procedure for functional data in different frameworks (i.e., one or two-population frameworks) by means of different basis expansions (i.e., B-spline, Fourier, and phase-amplitude Fourier). The current version of the package requires functional data evaluated on a uniform grid; it automatically projects each function on a chosen functional basis; it performs the entire family of multivariate tests; and, finally, it provides the matrix of the p-values of the previous tests and the vector of the corrected p-values. The functional basis, the coupled or uncoupled scenario, and the kind of test can be chosen by the user. The package provides also a plotting function creating a graphical output of the procedure: the p-value heat-map, the plot of the corrected p-values, and the plot of the functional data.

Details

Package:	fdatest
Type:	Package
Version:	1.0
Date:	2013-02-27
License:	GPL-2

Author(s)

Alessia Pini, Simone Vantini

Maintainer: Alessia Pini <alessia.pini@mail.polimi.it>

References

A. Pini and S. Vantini (2013). The Interval Testing Procedure: Inference for Functional Data Controlling the Family Wise Error Rate on Intervals. *MOX-report 13/2013*, Politecnico di Milano.

See Also

See also `ITP1bspline`, `ITP1fourier`, `ITP2bspline`, `ITP2fourier`, `ITP2pafourier`, `ITPlmbspline`, and `ITPimage`.

Examples

```
data(NASAtemp)
# Performing the ITP for one population with the Fourier basis
ITP.result <- ITP1fourier(NASAtemp$milan,maxfrequency=15,B=1000)
# Plotting the results of the ITP
## Not run:

ITPimage(ITP.result)


## End(Not run)
# Selecting the significant coefficients
which(ITP.result$corrected.pval < 0.05)

# Performing the ITP for two populations with the B-spline basis
ITP.result <- ITP2bspline(NASAtemp$milan,NASAtemp$paris,nknots=30,B=1000)
# Plotting the results of the ITP
## Not run:

ITPimage(ITP.result,abscissa.range=c(0,12))


## End(Not run)
# Selecting the significant components for the radius at 5% level
which(ITP.result$corrected.pval < 0.05)
```

ITP1bspline *One population Interval Testing Procedure with B-spline basis*

Description

The function implements the Interval Testing Procedure for testing the center of symmetry of a functional population evaluated on a uniform grid. Data are represented by means of the B-spline expansion and the significance of each basis coefficient is tested with an interval-wise control of the Family Wise Error Rate. The default parameters of the basis expansion lead to the piece-wise interpolating function.

Usage

```
ITP1bspline(data, mu = 0, order = 2, nknots = dim(data)[2], B = 10000)
```

Arguments

data	Pointwise evaluations of the functional data set on a uniform grid. data is a matrix of dimensions c(n,J), with J evaluations on columns and n units on rows.
mu	The center of symmetry under the null hypothesis: either a constant (in this case, a constant function is used) or a J-dimensional vector containing the evaluations on the same grid which data are evaluated. The default is mu=0.
order	Order of the B-spline basis expansion. The default is order=2.
nknots	Number of knots of the B-spline basis expansion. The default is nknots=dim(data)[2].
B	The number of iterations of the MC algorithm to evaluate the p-values of the permutation tests. The defualt is B=10000.

Value

ITP1bspline returns an object of class "ITP1".

An object of class "ITP1" is a list containing at least the following components:

basis	String vector indicating the basis used for the first phase of the algorithm. In this case equal to "B-spline".
test	String vector indicating the type of test performed. In this case equal to "1pop".
mu	Center of symmetry under the null hypothesis (as entered by the user).
coeff	Matrix of dimensions c(n,p) of the p coefficients of the B-spline basis expansion. Rows are associated to units and columns to the basis index.
pval	Uncorrected p-values for each basis coefficient.
pval.matrix	Matrix of dimensions c(p,p) of the p-values of the multivariate tests. The element (i,j) of matrix pval.matrix contains the p-value of the joint NPC test of the components (j,j+1,...,j+(p-i)).
corrected.pval	Corrected p-values for each basis coefficient.
labels	Labels indicating the population membership of each data (in this case always equal to 1).
data.eval	Evaluation on a fine uniform grid of the functional data obtained through the basis expansion.
heatmap.matrix	Heatmap matrix of p-values (used only for plots).

Author(s)

Alessia Pini, Simone Vantini

References

A. Pini and S. Vantini (2013). The Interval Testing Procedure: Inference for Functional Data Controlling the Family Wise Error Rate on Intervals. *MOX-report 13/2013*, Politecnico di Milano.

See Also

See also ITP1fourier, ITP2bspline, ITP2fourier, ITP2pafourier, and ITPimage.

Examples

```
# Importing the NASA temperatures data set
data(NASAtemp)
# Performing the ITP for two populations with the B-spline basis
ITP.result <- ITP1bspline(NASAtemp$paris,mu=4,nknots=50,B=1000)
# Plotting the results of the ITP
## Not run:

ITPimage(ITP.result,abscissa.range=c(0,12))


## End(Not run)
# Selecting the significant components for the radius at 5% level
which(ITP.result$corrected.pval < 0.05)
```

ITP1fourier *One population Interval Testing Procedure with Fourier basis*

Description

The function implements the Interval Testing Procedure for testing the center of symmetry of a functional population evaluated on a uniform grid. Data are represented by means of the Fourier expansion and the significance of each basis coefficient is tested with an interval-wise control of the Family Wise Error Rate.

Usage

```
ITP1fourier(data, mu = 0, maxfrequency=floor(dim(data)[2]/2), B = 10000)
```

Arguments

data	Pointwise evaluations of the functional data set on a uniform grid. data is a matrix of dimensions c(n,J), with J evaluations on columns and n units on rows.
mu	The center of symmetry under the null hypothesis: either a constant (in this case, a constant function is used) or a J-dimensional vector containing the evaluations on the same grid which data are evaluated. The default is mu=0.
maxfrequency	The maximum frequency to be used in the Fourier basis expansion of data. The default is floor(dim(data)[2]/2), leading to an interpolating expansion.
B	The number of iterations of the MC algorithm to evaluate the p-values of the permutation tests. The defualt is B=10000.

Value

ITP1fourier returns an object of class "ITP1".

An object of class "ITP1" is a list containing at least the following components:

basis	String vector indicating the basis used for the first phase of the algorithm. In this case equal to "Fourier".
test	String vector indicating the type of test performed. In this case equal to "1pop".
mu	Center of symmetry under the null hypothesis (as entered by the user).
coeff	Matrix of dimensions c(n,p) of the p coefficients of the Fourier basis expansion. Rows are associated to units and columns to the basis index: the first column is a0, the following (p-1)/2 columns are the ak coefficients (sine coefficients) and the last (p-1)/2 columns the bk coefficients (cosine coefficients).
pval	Uncorrected p-values for each frequency.
pval.matrix	Matrix of dimensions c(p,p) of the p-values of the multivariate tests. The element (i,j) of matrix pval.matrix contains the p-value of the joint NPC test of the components (j,j+1,...,j+(p-i)).
corrected.pval	Corrected p-values for each frequency.
labels	Labels indicating the population membership of each data (in this case always equal to 1).
data.eval	Evaluation on a fine uniform grid of the functional data obtained through the basis expansion.
heatmap.matrix	Heatmap matrix of p-values (used only for plots).

Author(s)

Alessia Pini, Simone Vantini

References

A. Pini and S. Vantini (2013). The Interval Testing Procedure: Inference for Functional Data Controlling the Family Wise Error Rate on Intervals. *MOX-report 13/2013*, Politecnico di Milano.

See Also

See also ITP1bspline, ITP2bspline, ITP2fourier, ITP2pafourier, and ITPimage.

Examples

```
# Importing the NASA temperatures data set
data(NASAtemp)
# Performing the ITP
ITP.result <- ITP1fourier(NASAtemp$milan,maxfrequency=20,B=1000)
# Plotting the results of the ITP
#\dontrun{

plot(ITP.result,main='NASA data',xrange=c(1,365),xlab='Day')
```

```
#}
# Selecting the significant coefficients
which(ITP.result$corrected.pval < 0.05)
```

`ITP2bspline` *Two populations Interval Testing Procedure with B-spline basis*

Description

The function implements the Interval Testing Procedure for testing the difference between two functional populations evaluated on a uniform grid. Data are represented by means of the B-spline basis and the significance of each basis coefficient is tested with an interval-wise control of the Family Wise Error Rate. The default parameters of the basis expansion lead to the piece-wise interpolating function.

Usage

```
ITP2bspline(data1, data2, mu = 0,
            order = 2, nknots = dim(data1)[2], B = 10000, paired = FALSE)
```

Arguments

`data1`	Pointwise evaluations of the first population's functional data set on a uniform grid. `data1` is a matrix of dimensions `c(n1,J)`, with `J` evaluations on columns and `n1` units on rows.
`data2`	Pointwise evaluations of the second population's functional data set on a uniform grid. `data2` is a matrix of dimensions `c(n2,J)`, with `J` evaluations on columns and `n2` units on rows.
`mu`	The difference between the first functional population and the second functional population under the null hypothesis. Either a constant (in this case, a constant function is used) or a `J`-dimensional vector containing the evaluations on the same grid which `data` are evaluated. The default is `mu=0`.
`order`	Order of the B-spline basis expansion. The default is `order=2`.
`nknots`	Number of knots of the B-spline basis expansion. The default is `nknots=dim(data1)[2]`.
`B`	The number of iterations of the MC algorithm to evaluate the p-values of the permutation tests. The defualt is `B=10000`.
`paired`	A logical indicating whether the test is paired. The default is `FALSE`.

Value

`ITP2bspline` returns an object of `class` `"ITP2"`.

An object of class `"ITP2"` is a list containing at least the following components:

`basis`	String vector indicating the basis used for the first phase of the algorithm. In this case equal to `"B-spline"`.

test	String vector indicating the type of test performed. In this case equal to "2pop".
mu	Difference between the first functional population and the second functional population under the null hypothesis (as entered by the user).
paired	Logical indicating whether the test is paired (as entered by the user).
coeff	Matrix of dimensions c(n,p) of the p coefficients of the B-spline basis expansion, with n=n1+n2. Rows are associated to units and columns to the basis index. The first n1 rows report the coefficients of the first population units and the following n2 rows report the coefficients of the second population units
pval	Uncorrected p-values for each basis coefficient.
pval.matrix	Matrix of dimensions c(p,p) of the p-values of the multivariate tests. The element (i,j) of matrix pval.matrix contains the p-value of the joint NPC test of the components (j,j+1,...,j+(p-i)).
corrected.pval	Corrected p-values for each basis coefficient.
labels	Labels indicating the population membership of each data.
data.eval	Evaluation on a fine uniform grid of the functional data obtained through the basis expansion.
heatmap.matrix	Heatmap matrix of p-values (used only for plots).

Author(s)

Alessia Pini, Simone Vantini

References

A. Pini and S. Vantini (2013). The Interval Testing Procedure: Inference for Functional Data Controlling the Family Wise Error Rate on Intervals. *MOX-report 13/2013*, Politecnico di Milano.

See Also

See also summary.ITPlm for summaries. For different types of ITP-based tests, see ITP1bspline, ITP1fourier, ITP2fourier, ITP2pafourier, and ITPimage.

Examples

```
# Importing the NASA temperatures data set
data(NASAtemp)
# Performing the ITP
ITP.result <- ITP2bspline(NASAtemp$milan,NASAtemp$paris,nknots=50,B=1000)
# Plotting the results of the ITP
#\dontrun{

#ITPimage(ITP.result,abscissa.range=c(0,12))
plot(ITP.result,main='NASA data',xrange=c(1,365),xlab='Day')

#}
# Selecting the significant components for the radius at 5% level
which(ITP.result$corrected.pval < 0.05)
```

ITP2fourier *Two populations Interval Testing Procedure with Fourier basis*

Description

The function implements the Interval Testing Procedure for testing the difference between two functional populations evaluated on a uniform grid. Data are represented by means of the Fourier basis and the significance of each basis coefficient is tested with an interval-wise control of the Family Wise Error Rate.

Usage

```
ITP2fourier(data1, data2, mu = 0,
            maxfrequency=floor(dim(data1)[2]/2), B = 10000, paired = FALSE)
```

Arguments

data1	Pointwise evaluations of the first population's functional data set on a uniform grid. data1 is a matrix of dimensions c(n1,J), with J evaluations on columns and n1 units on rows.
data2	Pointwise evaluations of the second population's functional data set on a uniform grid. data2 is a matrix of dimensions c(n2,J), with J evaluations on columns and n2 units on rows.
mu	The difference between the first functional population and the second functional population under the null hypothesis. Either a constant (in this case, a constant function is used) or a J-dimensional vector containing the evaluations on the same grid which data are evaluated. The default is mu=0.
maxfrequency	The maximum frequency to be used in the Fourier basis expansion of data. The default is floor(dim(data1)[2]/2), leading to an interpolating expansion.
B	The number of iterations of the MC algorithm to evaluate the p-values of the permutation tests. The defualt is B=10000.
paired	A logical indicating whether the test is paired. The default is FALSE.

Value

ITP2fourier returns an object of class "ITP2".

An object of class "ITP2" is a list containing at least the following components:

basis	String vector indicating the basis used for the first phase of the algorithm. In this case equal to "Fourier".
test	String vector indicating the type of test performed. in this case equal to "2pop".
mu	Difference between the first functional population and the second functional population under the null hypothesis (as entered by the user).
paired	Logical indicating whether the test is paired (as entered by the user).

coeff	Matrix of dimensions c(n,p) of the p coefficients of the Fourier basis expansion. Rows are associated to units and columns to the basis index: the first n1 rows report the coefficients of the first population units and the following n2 rows report the coefficients of the second population units; the first column is a0, the following (p-1)/2 columns are the ak coefficients (sine coefficients) and the last (p-1)/2 columns the bk coefficients (cosine coefficients).
pval	Uncorrected p-values for each frequency.
pval.matrix	Matrix of dimensions c(p,p) of the p-values of the multivariate tests. The element (i,j) of matrix pval.matrix contains the p-value of the joint NPC test of the frequencies (j,j+1,...,j+(p-i)).
corrected.pval	Corrected p-values for each frequency.
labels	Labels indicating the population membership of each data.
data.eval	Evaluation on a fine uniform grid of the functional data obtained through the basis expansion.
heatmap.matrix	Heatmap matrix of p-values (used only for plots).

Author(s)

Alessia Pini, Simone Vantini

References

A. Pini and S. Vantini (2013). The Interval Testing Procedure: Inference for Functional Data Controlling the Family Wise Error Rate on Intervals. *MOX-report 13/2013*, Politecnico di Milano.

See Also

See also ITP2pafourier, ITP2bspline, ITP1fourier, ITP1bspline, and ITPimage.

Examples

```
# Importing the NASA temperatures data set
data(NASAtemp)
# Performing the ITP
ITP.result <- ITP2fourier(NASAtemp$milan,NASAtemp$paris,maxfrequency=20,B=1000,paired=TRUE)
# Plotting the results of the ITP
#\dontrun{

plot(ITP.result,main='NASA data',xrange=c(1,365),xlab='Day')

#}
# Selecting the significant coefficients
which(ITP.result$corrected.pval < 0.05)
```

`ITP2pafourier` *Two populations Interval Testing Procedure with Fourier basis (phase-amplitude decomposition)*

Description

The function implements the Interval Testing Procedure for testing the difference between two functional populations evaluated on a uniform grid. Data are represented by means of the Fourier basis expansion with the phase-amplitude decomposition and the significance of the amplitude and phase of each frequency is tested with an interval-wise control of the Family Wise Error Rate.

Usage

```
ITP2pafourier(data1, data2,
              maxfrequency=floor(dim(data1)[2]/2), B = 10000, paired = FALSE)
```

Arguments

`data1`	Pointwise evaluations of the first population's functional data set on a uniform grid. `data1` is a matrix of dimensions `c(n1,J)`, with `J` evaluations on columns and `n1` units on rows.
`data2`	Pointwise evaluations of the second population's functional data set on a uniform grid. `data2` is a matrix of dimensions `c(n2,J)`, with `J` evaluations on columns and `n2` units on rows.
`maxfrequency`	The maximum frequency to be used in the Fourier basis expansion of data. The default is `floor(dim(data1)[2]/2)`, leading to an interpolating expansion.
`B`	The number of iterations of the MC algorithm to evaluate the p-values of the permutation tests. The defualt is `B=10000`.
`paired`	A logical indicating whether the test is paired. The default is `FALSE`.

Value

`ITP2pafourier` returns an object of `class` `"ITP2"`.

An object of class `"ITP2"` is a list containing at least the following components:

`basis`	String vector indicating the basis used for the first phase of the algorithm. Equal to `"paFourier"`.
`test`	String vector indicating the type of test performed. Equal to `"2pop"`.
`paired`	Logical indicating whether the test is paired (as entered by the user).
`coeff_phase`	Matrix of dimensions `c(n,p)` of the `p` phases of the Fourier basis expansion. Rows are associated to units and columns to frequencies: the first `n1` rows report the coefficients of the first population units and the following `n2` rows report the coefficients of the second population units.

`coeff_amplitude`
: Matrix of dimensions `c(n,p)` of the `p` amplitudes of the Fourier basis expansion. Rows are associated to units and columns to frequencies: the first `n1` rows report the coefficients of the first population units and the following `n2` rows report the coefficients of the second population units.

`pval_phase`
: Uncorrected p-values of the phase tests for each frequency.

`pval_amplitude`
: Uncorrected p-values of the amplitude tests for each frequency.

`pval.matrix_phase`
: Matrix of dimensions `c(p,p)` of the p-values of the multivariate tests on phase. The element `(i,j)` of matrix `pval.matrix_phase` contains the p-value of the joint NPC test of the frequencies `(j,j+1,...,j+(p-i))`.

`pval.matrix_amplitude`
: Matrix of dimensions `c(p,p)` of the p-values of the multivariate tests on amplitude. The element `(i,j)` of matrix `pval.matrix_amplitude` contains the p-value of the joint NPC test of the frequencies `(j,j+1,...,j+(p-i))`.

`corrected.pval_phase`
: Corrected p-values of the phase tests for each frequency.

`corrected.pval_amplitude`
: Corrected p-values of the amplitude tests for each frequency.

`labels`
: Labels indicating the population membership of each data.

`data.eval`
: Evaluation on a fine uniform grid of the functional data obtained through the basis expansion.

`heatmap.matrix_phase`
: Heatmap matrix of p-values for phase (used only for plots).

`heatmap.matrix_amplitude`
: Heatmap matrix of p-values for amplitude (used only for plots).

Author(s)

Alessia Pini, Simone Vantini

References

A. Pini and S. Vantini (2013). The Interval Testing Procedure: Inference for Functional Data Controlling the Family Wise Error Rate on Intervals. *MOX-report 13/2013*, Politecnico di Milano.

See Also

See also `ITP2fourier`, `ITP2bspline`, `ITP1fourier`, `ITP1bspline`, and `ITPimage`.

Examples

```
# Importing the NASA temperatures data set
data(NASAtemp)
# Performing the ITP
ITP.result <- ITP2pafourier(NASAtemp$milan,NASAtemp$paris,maxfrequency=20,B=1000,paired=TRUE)
# Plotting the results of the ITP
```

```
#\dontrun{

plot(ITP.result,main='NASA data',xrange=c(1,365),xlab='Day')

#}
# Selecting the significant coefficients
which(ITP.result$corrected.pval < 0.05)
```

`ITPaovbspline` *Interval Testing Procedure for testing Functional analysis of variance with B-spline basis*

Description

`ITPaovbspline` is used to fit and test functional analysis of variance. The function implements the Interval Testing Procedure for testing for significant differences between several functional population evaluated on a uniform grid. Data are represented by means of the B-spline basis and the significance of each basis coefficient is tested with an interval-wise control of the Family Wise Error Rate. The default parameters of the basis expansion lead to the piece-wise interpolating function.

Usage

```
ITPaovbspline(formula, order = 2,
              nknots = dim(model.response(model.frame(formula)))[2],
              B = 10000, method = "residuals")
```

Arguments

`formula`	An object of class "`formula`" (or one that can be coerced to that class): a symbolic description of the model to be fitted.
`order`	Order of the B-spline basis expansion. The default is `order=2`.
`nknots`	Number of knots of the B-spline basis expansion. The default is `dim(model.response(model.frame(formula)))[2]`.
`B`	The number of iterations of the MC algorithm to evaluate the p-values of the permutation tests. The defualt is `B=10000`.
`method`	Permutation method used to calculate the p-value of permutation tests. Choose "`residuals`" for the permutations of residuals under the reduced model, according to the Freedman and Lane scheme, and "`responses`" for the permutation of the responses, according to the Manly scheme.

Value

`ITPaovbspline` returns an object of `class` "`ITPaov`".

The function `summary` is used to obtain and print a summary of the results.

An object of class "`ITPlm`" is a list containing at least the following components:

`call` The matched call.

`design.matrix` The design matrix of the functional-on-scalar linear model.

`basis` String vector indicating the basis used for the first phase of the algorithm. In this case equal to `"B-spline"`.

`coeff` Matrix of dimensions `c(n,p)` of the `p` coefficients of the B-spline basis expansion. Rows are associated to units and columns to the basis index.

`coeff.regr` Matrix of dimensions `c(L+1,p)` of the `p` coefficients of the B-spline basis expansion of the intercept (first row) and the `L` effects of the covariates specified in `formula`. Columns are associated to the basis index.

`pval.F` Uncorrected p-values of the functional F-test for each basis coefficient.

`pval.matrix.F` Matrix of dimensions `c(p,p)` of the p-values of the multivariate F-tests. The element `(i,j)` of matrix `pval.matrix` contains the p-value of the joint NPC test of the components `(j,j+1,...,j+(p-i))`.

`corrected.pval.F`
Corrected p-values of the functional F-test for each basis coefficient.

`pval.factors` Uncorrected p-values of the functional F-tests on each factor of the analysis of variance, separately (rows) and each basis coefficient (columns).

`pval.matrix.factors`
Array of dimensions `c(L+1,p,p)` of the p-values of the multivariate F-tests on factors. The element `(l,i,j)` of array `pval.matrix` contains the p-value of the joint NPC test on factor `l` of the components `(j,j+1,...,j+(p-i))`.

`corrected.pval.factors`
Corrected p-values of the functional F-tests on each factor of the analysis of variance (rows) and each basis coefficient (columns).

`data.eval` Evaluation on a fine uniform grid of the functional data obtained through the basis expansion.

`coeff.regr.eval`
Evaluation on a fine uniform grid of the functional regression coefficients.

`fitted.eval` Evaluation on a fine uniform grid of the fitted values of the functional regression.

`residuals.eval` Evaluation on a fine uniform grid of the residuals of the functional regression.

`R2.eval` Evaluation on a fine uniform grid of the functional R-squared of the regression.

`heatmap.matrix.F`
Heatmap matrix of p-values of functional F-test (used only for plots).

`heatmap.matrix.factors`
Heatmap matrix of p-values of functional F-tests on each factor of the analysis of variance (used only for plots).

Author(s)

Alessia Pini, Simone Vantini

References

D. Freedman and D. Lane (1983). A Nonstochastic Interpretation of Reported Significance Levels. *Journal of Business & Economic Statistics* 1.4, 292-298.

B. F. J. Manly (2006). Randomization, *Bootstrap and Monte Carlo Methods in Biology*. Vol. 70. CRC Press.

A. Pini and S. Vantini (2013). The Interval Testing Procedure: Inference for Functional Data Controlling the Family Wise Error Rate on Intervals. *MOX-report 13/2013*, Politecnico di Milano.

See Also

See `summary.ITPaov` for summaries and `plot.ITPaov` for plotting the results.

See also `ITPlmbspline` to fit and test a functional-on-scalar linear model applying the ITP, and `ITP1bspline`, `ITP2bspline`, `ITP2fourier`, `ITP2pafourier` for one-population and two-population tests.

Examples

```
# Importing the NASA temperatures data set
data(NASAtemp)

temperature <- rbind(NASAtemp$milan,NASAtemp$paris)
groups <- c(rep(0,22),rep(1,22))

# Performing the ITP
## Not run:

ITP.result <- ITPaovbspline(temperature ~ groups,B=1000,nknots=20,order=3)

# Summary of the ITP results
summary(ITP.result)

# Plot of the ITP results
layout(1)
plot(ITP.result)

# All graphics on the same device
layout(matrix(1:4,nrow=2,byrow=FALSE))
plot(ITP.result,main='NASA data', plot.adjpval = TRUE,xlab='Day',xrange=c(1,365))


## End(Not run)
```

`ITPimage` *Plot of the Interval Testing Procedure results*

Description

Plotting function creating a graphical output of the ITP: the p-value heat-map, the plot of the corrected p-values, and the plot of the functional data.

Usage

```
ITPimage(ITP.result, alpha = 0.05, abscissa.range = c(0, 1), nlevel = 20)
```

Arguments

ITP.result	Results of the ITP, as created by ITP1bspline, ITP1fourier, ITP2bspline, ITP2fourier, and ITP2pafourier.
alpha	Level of the hypothesis test. The default is alpha=0.05.
abscissa.range	Range of the plot abscissa. The default is c(0,1).
nlevel	Number of desired color levels for the p-value heatmap. The default is nlevel=20.

Value

No value returned. The function produces a graphical output of the ITP results: the p-value heatmap, a plot of the corrected p-values and the plot of the functional data. The basis components selected as significant by the test at level alpha are highlighted in the plot of the corrected p-values by a gray area.

Author(s)

Alessia pini, Simone Vantini

References

A. Pini and S. Vantini (2013). The Interval Testing Procedure: Inference for Functional Data Controlling the Family Wise Error Rate on Intervals. *MOX-report 13/2013*, Politecnico di Milano.

See Also

See also ITP1bspline, ITP1fourier, ITP2bspline, ITP2fourier, and ITP2pafourier.

Examples

```
# Importing the NASA temperatures data set
data(NASAtemp)
# Performing the ITP for one population with the Fourier basis
ITP.result <- ITP1fourier(NASAtemp$milan,maxfrequency=15,B=1000)
# Plotting the results of the ITP
## Not run:

ITPimage(ITP.result)
```

```
## End(Not run)
# Selecting the significant coefficients
which(ITP.result$corrected.pval < 0.05)

# Performing the ITP for two populations with the B-spline basis
ITP.result <- ITP2bspline(NASAtemp$milan,NASAtemp$paris,nknots=30,B=1000)
# Plotting the results of the ITP
## Not run:

ITPimage(ITP.result,abscissa.range=c(0,12))


## End(Not run)
# Selecting the significant components for the radius at 5% level
which(ITP.result$corrected.pval < 0.05)
```

`ITPlmbspline` *Interval Testing Procedure for testing Functional-on-Scalar Linear Models with B-spline basis*

Description

`ITPlmbspline` is used to fit and test functional linear models. It can be used to carry out regression, and analysis of variance. The function implements the Interval Testing Procedure for testing the significance of the effects of scalar covariates on a functional population evaluated on a uniform grid. Data are represented by means of the B-spline basis and the significance of each basis coefficient is tested with an interval-wise control of the Family Wise Error Rate. The default parameters of the basis expansion lead to the piece-wise interpolating function.

Usage

```
ITPlmbspline(formula, order = 2,
             nknots = dim(model.response(model.frame(formula)))[2],
             B = 10000, method = "residuals")
```

Arguments

`formula`	An object of class "`formula`" (or one that can be coerced to that class): a symbolic description of the model to be fitted.
`order`	Order of the B-spline basis expansion. The default is `order=2`.
`nknots`	Number of knots of the B-spline basis expansion. The default is `dim(model.response(model.frame(formula)))[2]`.
`B`	The number of iterations of the MC algorithm to evaluate the p-values of the permutation tests. The defualt is `B=10000`.
`method`	Permutation method used to calculate the p-value of permutation tests. Choose "`residuals`" for the permutations of residuals under the reduced model, according to the Freedman and Lane scheme, and "`responses`" for the permutation of the responses, according to the Manly scheme.

Value

`ITPlmbspline` returns an object of `class` "`ITPlm`".

The function `summary` is used to obtain and print a summary of the results.

An object of class "`ITPlm`" is a list containing at least the following components:

`call`	The matched call.
`design.matrix`	The design matrix of the functional-on-scalar linear model.
`basis`	String vector indicating the basis used for the first phase of the algorithm. In this case equal to "`B-spline`".
`coeff`	Matrix of dimensions `c(n,p)` of the `p` coefficients of the B-spline basis expansion. Rows are associated to units and columns to the basis index.
`coeff.regr`	Matrix of dimensions `c(L+1,p)` of the `p` coefficients of the B-spline basis expansion of the intercept (first row) and the `L` effects of the covariates specified in `formula`. Columns are associated to the basis index.
`pval.F`	Uncorrected p-values of the functional F-test for each basis coefficient.
`pval.matrix.F`	Matrix of dimensions `c(p,p)` of the p-values of the multivariate F-tests. The element `(i,j)` of matrix `pval.matrix` contains the p-value of the joint NPC test of the components `(j,j+1,...,j+(p-i))`.
`corrected.pval.F`	Corrected p-values of the functional F-test for each basis coefficient.
`pval.t`	Uncorrected p-values of the functional t-tests for each partial regression coefficient including the intercept (rows) and each basis coefficient (columns).
`pval.matrix.t`	Array of dimensions `c(L+1,p,p)` of the p-values of the multivariate t-tests. The element `(l,i,j)` of array `pval.matrix` contains the p-value of the joint NPC test on covariate `l` of the components `(j,j+1,...,j+(p-i))`.
`corrected.pval.t`	Corrected p-values of the functional t-tests for each partial regression coefficient including the intercept (rows) and each basis coefficient (columns).
`data.eval`	Evaluation on a fine uniform grid of the functional data obtained through the basis expansion.
`coeff.regr.eval`	Evaluation on a fine uniform grid of the functional regression coefficients.
`fitted.eval`	Evaluation on a fine uniform grid of the fitted values of the functional regression.
`residuals.eval`	Evaluation on a fine uniform grid of the residuals of the functional regression.
`R2.eval`	Evaluation on a fine uniform grid of the functional R-squared of the regression.
`heatmap.matrix.F`	Heatmap matrix of p-values of functional F-test (used only for plots).
`heatmap.matrix.t`	Heatmap matrix of p-values of functional t-tests (used only for plots).

Author(s)

Alessia Pini, Simone Vantini

References

D. Freedman and D. Lane (1983). A Nonstochastic Interpretation of Reported Significance Levels. *Journal of Business & Economic Statistics* 1.4, 292-298.

B. F. J. Manly (2006). Randomization, *Bootstrap and Monte Carlo Methods in Biology*. Vol. 70. CRC Press.

A. Pini and S. Vantini (2013). The Interval Testing Procedure: Inference for Functional Data Controlling the Family Wise Error Rate on Intervals. *MOX-report 13/2013*, Politecnico di Milano.

See Also

See `summary.ITPlm` for summaries and `plot.ITPlm` for plotting the results.

See also `ITPaovbspline` to fit and test a functional analysis of variance applying the ITP, and `ITP1bspline`, `ITP2bspline`, `ITP2fourier`, `ITP2pafourier` for one-population and two-population tests.

Examples

```
# Importing the NASA temperatures data set
data(NASAtemp)

temperature <- rbind(NASAtemp$milan,NASAtemp$paris)
groups <- c(rep(0,22),rep(1,22))

# Performing the ITP
## Not run:

ITP.result <- ITPlmbspline(temperature ~ groups,B=1000,nknots=20)
# Summary of the ITP results
summary(ITP.result)

# Plot of the ITP results
layout(1)
plot(ITP.result,main='NASA data', plot.adjpval = TRUE,xlab='Day',xrange=c(1,365))

# All graphics on the same device
layout(matrix(1:6,nrow=3,byrow=FALSE))
plot(ITP.result,main='NASA data', plot.adjpval = TRUE,xlab='Day',xrange=c(1,365))


## End(Not run)
```

`NASAtemp` *NASA daily temperatures data set*

Description

It contains the daily mean temperatures registered from July 1983 to June 2005 and stored in the NASA database Earth Surface Meteorology for Solar Energy of two different geographical locations: the region (45-46 North, 9-10 East), including the city of Milan (Italy), and the region (48-49 North, 2-3 East), including the city of Paris (France).

Usage

```
data(NASAtemp)
```

Format

List of 2 elements:

- `milan` Matrix of dimensions `c(22,365)` containing the daily mean temperatures of the region (45-46 North, 9-10 East), including the city of Milan (Italy) registered from July 1983 to June 2005 (22 years).
- `paris` Matrix of dimensions `c(22,365)` containing the daily mean temperatures of the region (48-49 North, 2-3 East), including the city of Paris (France) registered from July 1983 to June 2005 (22 years).

Source

These data were obtained from the NASA Langley Research Center Atmospheric Science Data Center Surface meteorological and Solar Energy (SSE) web portal supported by the NASA LaRC POWER Project. Data are freely available at: NASA Surface Meteorology and Solar Energy, A Renewable Energy Resource web site (release 6.0): http://eosweb.larc.nasa.gov

Examples

```
data(NASAtemp)
## Not run:

matplot(t(NASAtemp$milan),type='l')
matplot(t(NASAtemp$paris),type='l')


## End(Not run)
```

`plot.ITP1` *Plotting ITP results for one-population tests*

Description

`plot` method for class `"ITP1"`. Plotting function creating a graphical output of the ITP for the test of the mean of one population: functional data and ITP-adjusted p-values are plotted.

Usage

```
## S3 method for class 'ITP1'
plot(x, xrange = c(0, 1), alpha1 = 0.05, alpha2 = 0.01,
     ylab = "Functional Data", main = NULL, lwd = 1, col = 1,
     pch = 16, ylim = range(object$data.eval), ...)
```

Arguments

x	The object to be plotted. An object of class "ITP1", that is, a result of an ITP for comparison between two populations. Usually a call to ITP1bspline or ITP1fourier.
xrange	Range of the x axis.
alpha1	First level of significance used to select and display significant differences. Default is alpha1 = 0.05.
alpha2	Second level of significance used to select and display significant differences. Default is alpha1 = 0.01. alpha1 and alpha2 are s.t. alpha2 < alpha1. Otherwise the two values are switched.
ylab	Label of y axis of the plot of functional data. Default is "Functional Data".
main	An overall title for the plots (it will be pasted to "Functional Data" for the first plot and "adjusted p-values" for the second plot).
lwd	Line width for the plot of functional data.
col	Color used to plot the functional data.
pch	Point character for the plot of adjusted p-values.
ylim	Range of the y axis.
...	Additional plotting arguments that can be used with function plot, such as graphical parameters (see par).

Value

No value returned. The function produces a graphical output of the ITP results: the plot of the functional data and the one of the adjusted p-values. The basis components selected as significant by the test at level alpha1 and alpha2 are highlighted in the plot of the corrected p-values and in the one of functional data (in case the test is based on a local basis, such as B-splines) by gray areas (light and dark gray, respectively).

Author(s)

Alessia Pini, Simone Vantini

References

A. Pini and S. Vantini (2013). The Interval Testing Procedure: Inference for Functional Data Controlling the Family Wise Error Rate on Intervals. *MOX-report 13/2013*, Politecnico di Milano.

See Also

See also ITP1bspline and ITP1fourier to perform the ITP to test for the mean of a functional populations. See plot.ITP2 and plot.ITPlm for the plot method applied to the ITP results of two-population tests and linear models, respectively.

Examples

```
# Importing the NASA temperatures data set
data(NASAtemp)

# Performing the ITP for one population with the B-spline basis
ITP.result.bspline <- ITP1bspline(NASAtemp$paris,mu=4,nknots=50,B=1000)
# Plotting the results of the ITP
plot(ITP.result.bspline,xlab='Day',xrange=c(0,365),main='NASA data')
# Selecting the significant components for the radius at 5% level
which(ITP.result.bspline$corrected.pval < 0.05)

# Performing the ITP for one population with the Fourier basis
ITP.result.fourier <- ITP1fourier(NASAtemp$milan,maxfrequency=20,B=1000)

# Plotting the results of the ITP
layout(1)
plot(ITP.result.fourier,xlab='Day',xrange=c(1,365),main='NASA data')

# Selecting the significant components for the radius at 5% level
which(ITP.result.fourier$corrected.pval < 0.05)
```

---

plot.ITP2 *Plotting ITP results for two-population tests*

Description

plot method for class "ITP2". Plotting function creating a graphical output of the ITP for the test of comparison between two populations: functional data and ITP-adjusted p-values are plotted.

Usage

```
## S3 method for class 'ITP2'
plot(x, xrange = c(0, 1), alpha1 = 0.05, alpha2 = 0.01,
     ylab = "Functional Data", main = NULL, lwd = 1,
     col = c(1, 2), pch = 16, ylim = range(object$data.eval), ...)
```

Arguments

x	The object to be plotted. An object of class "ITP2", that is, a result of an ITP for comparison between two populations. Usually a call to ITP2bspline, ITP2fourier or ITP2pafourier.

xrange	Range of the x axis.
alpha1	First level of significance used to select and display significant differences. Default is alpha1 = 0.05.
alpha2	Second level of significance used to select and display significant differences. Default is alpha1 = 0.01. alpha1 and alpha2 are s.t. alpha2 < alpha1. Otherwise the two values are switched.
ylab	Label of y axis of the plot of functional data. Default is "Functional Data".
main	An overall title for the plots (it will be pasted to "Functional Data" for the first plot and "adjusted p-values" for the second plot).
lwd	Line width for the plot of functional data.
col	Color used to plot the functional data.
pch	Point character for the plot of adjusted p-values.
ylim	Range of the y axis.
...	Additional plotting arguments that can be used with function plot, such as graphical parameters (see par).

Value

No value returned. The function produces a graphical output of the ITP results: the plot of the functional data and the one of the adjusted p-values. The basis components selected as significant by the test at level alpha1 and alpha2 are highlighted in the plot of the corrected p-values and in the one of functional data (in case the test is based on a local basis, such as B-splines) by gray areas (light and dark gray, respectively). In the case of a Fourier basis with amplitude and phase decomposition, two plots of adjusted p-values are done, one for phase and one for amplitude.

Author(s)

Alessia Pini, Simone Vantini

References

A. Pini and S. Vantini (2013). The Interval Testing Procedure: Inference for Functional Data Controlling the Family Wise Error Rate on Intervals. *MOX-report 13/2013*, Politecnico di Milano.

See Also

See also ITP2bspline, ITP2fourier, ITP2pafourier to perform the ITP to test for differences between two populations. See plot.ITP1 and plot.ITPlm for the plot method applied to the ITP results of one-population tests and a linear models, respectively.

Examples

```
# Importing the NASA temperatures data set
data(NASAtemp)

# Performing the ITP for two populations with the B-spline basis
```

```
ITP.result.bspline <- ITP2bspline(NASAtemp$milan,NASAtemp$paris,nknots=30,B=1000)
# Plotting the results of the ITP
plot(ITP.result.bspline,xlab='Day',xrange=c(1,365),main='NASA data')

# Selecting the significant components for the radius at 5% level
which(ITP.result.bspline$corrected.pval < 0.05)

# Performing the ITP for two populations with the Fourier basis
ITP.result.fourier <- ITP2fourier(NASAtemp$milan,NASAtemp$paris,
                      maxfrequency=20,B=1000,paired=TRUE)
# Plotting the results of the ITP
layout(1)
plot(ITP.result.fourier,xlab='Day',xrange=c(1,365),main='NASA data')
# Selecting the significant components for the radius at 5% level
which(ITP.result.fourier$corrected.pval < 0.05)
```

`plot.ITPaov` *Plotting ITP results for functional analysis of variance testing*

Description

`plot` method for class "`ITPaov`". Plotting function creating a graphical output of the ITP for the test on a functional analysis of variance: functional data, and ITP-adjusted p-values of the F-tests on the whole model and on each factor are plotted.

Usage

```
## S3 method for class 'ITPaov'
plot(x,xrange=c(0,1), alpha1=0.05, alpha2=0.01,
     plot.adjpval=FALSE,ylim=range(x$data.eval),col=1,
     ylab='Functional Data',main=NULL,lwd=1,pch=16,...)
```

Arguments

`x`	The object to be plotted. An object of class "`ITPaov`", usually, a result of a call to `ITPaovbspline`.
`xrange`	Range of the `x` axis.
`alpha1`	First level of significance used to select and display significant effects. Default is `alpha1 = 0.05`.
`alpha2`	Second level of significance used to select and display significant effects. Default is `alpha1 = 0.01`. `alpha1` and `alpha2` are s.t. `alpha2 < alpha1`. Otherwise the two values are switched.
`plot.adjpval`	A logical indicating wether the plots of adjusted p-values have to be done. Default is `plot.adjpval = FALSE`.
`col`	Colors for the plot of functional data. Default is `col = 1`.
`ylim`	Range of the `y` axis. Default is `ylim = range(x$data.eval)`.

ylab	Label of y axis of the plot of functional data. Default is "Functional Data".
main	An overall title for the plots (it will be pasted to "Functional Data and F-test" for the first plot and "factor" for the other plots).
lwd	Line width for the plot of functional data. Default is lwd=16.
pch	Point character for the plot of adjusted p-values. Default is pch=16.
...	Additional plotting arguments that can be used with function plot, such as graphical parameters (see par).

Value

No value returned. The function produces a graphical output of the ITP results: the plot of the functional data, functional regression coefficients, and ITP-adjusted p-values of the F-tests on the whole model and on each factor. The basis components selected as significant by the tests at level alpha1 and alpha2 are highlighted in the plot of the corrected p-values and in the one of functional data by gray areas (light and dark gray, respectively). The first plot reports the gray areas corresponding to a significant F-test on the whole model. The remaining plots report the gray areas corresponding to significant F-tests on each factor (with colors corresponding to the levels of the factor).

Author(s)

Alessia Pini, Simone Vantini

References

A. Pini and S. Vantini (2013). The Interval Testing Procedure: Inference for Functional Data Controlling the Family Wise Error Rate on Intervals. *MOX-report 13/2013*, Politecnico di Milano.

See Also

See also ITPaovbspline to fit and test a functional analysis of variance applying the ITP, and summary.ITPaov for summaries. See plot.ITPlm, plot.ITP1, and plot.ITP2 for the plot method applied to the ITP results of functional-on-scalar linear models, one-population and two-population, respectively.

Examples

```
# Importing the NASA temperatures data set
data(NASAtemp)

temperature <- rbind(NASAtemp$milan,NASAtemp$paris)
groups <- c(rep(0,22),rep(1,22))

# Performing the ITP
## Not run:

ITP.result <- ITPaovbspline(temperature ~ groups,B=1000,nknots=20,order=3)

# Summary of the ITP results
summary(ITP.result)
```

```
# Plot of the ITP results
layout(1)
plot(ITP.result)

# All graphics on the same device
layout(matrix(1:4,nrow=2,byrow=FALSE))
plot(ITP.result,main='NASA data', plot.adjpval = TRUE,xlab='Day',xrange=c(1,365))


## End(Not run)
```

`plot.ITPlm` *Plotting ITP results for functional-on-scalar linear model testing*

Description

`plot` method for class "`ITPlm`". Plotting function creating a graphical output of the ITP for the test on a functional-on-scalar linear model: functional data, functional coefficients and ITP-adjusted p-values for the F-test and t-tests are plotted.

Usage

```
## S3 method for class 'ITPlm'
plot(x, xrange = c(0, 1), alpha1 = 0.05, alpha2 = 0.01,
     plot.adjpval = FALSE, col = c(1, rainbow(dim(x$corrected.pval.t)[1])),
     ylim = range(x$data.eval), ylab = "Functional Data",
     main = NULL, lwd = 1, pch = 16, ...)
```

Arguments

`x`	The object to be plotted. An object of class "`ITPlm`", usually, a result of a call to `ITPlmbspline`.
`xrange`	Range of the `x` axis.
`alpha1`	First level of significance used to select and display significant effects. Default is `alpha1 = 0.05`.
`alpha2`	Second level of significance used to select and display significant effects. Default is `alpha1 = 0.01`. `alpha1` and `alpha2` are s.t. `alpha2 < alpha1`. Otherwise the two values are switched.
`plot.adjpval`	A logical indicating wether the plots of adjusted p-values have to be done. Default is `plot.adjpval = FALSE`.
`col`	Vector of colors for the plot of functional data (first element), and functional coefficients (following elements). Default is `col = c(1, rainbow(dim(x$corrected.pval.t)[1]))`.
`ylim`	Range of the `y` axis. Default is `ylim = range(x$data.eval)`.
`ylab`	Label of `y` axis of the plot of functional data. Default is "`Functional Data`".

main	An overall title for the plots (it will be pasted to "Functional Data and F-test" for the first plot and "t-test" for the other plots).
lwd	Line width for the plot of functional data. Default is lwd=16.
pch	Point character for the plot of adjusted p-values. Default is pch=16.
...	Additional plotting arguments that can be used with function plot, such as graphical parameters (see par).

Value

No value returned. The function produces a graphical output of the ITP results: the plot of the functional data, functional regression coefficients, and ITP-adjusted p-values for the F-test and t-tests. The basis components selected as significant by the tests at level alpha1 and alpha2 are highlighted in the plot of the corrected p-values and in the one of functional data by gray areas (light and dark gray, respectively). The plot of functional data reports the gray areas corresponding to a significant F-test. The plots of functional regression coefficients report the gray areas corresponding to significant t-tests for the corresponding covariate.

Author(s)

Alessia Pini, Simone Vantini

References

A. Pini and S. Vantini (2013). The Interval Testing Procedure: Inference for Functional Data Controlling the Family Wise Error Rate on Intervals. *MOX-report 13/2013*, Politecnico di Milano.

See Also

See also ITPlmbspline to fit and test a functional-on-scalar linear model applying the ITP, and summary.ITPlm for summaries. See plot.ITPaov, plot.ITP1, and plot.ITP2 for the plot method applied to the ITP results of functional analysis of variance, one-population and two-population, respectively.

Examples

```
# Importing the NASA temperatures data set
data(NASAtemp)

data <- rbind(NASAtemp$milan,NASAtemp$paris)
lab <- c(rep(0,22),rep(1,22))

# Performing the ITP
## Not run:

ITP.result <- ITPlmbspline(data ~ lab,B=1000,nknots=20)
# Summary of the ITP results
summary(ITP.result)

# Plot of the ITP results
layout(1)
```

```
plot(ITP.result,main='NASA data',xlab='Day',xrange=c(1,365))

# Plots of the adjusted p-values
plot(ITP.result,main='NASA data', plot.adjpval = TRUE,xlab='Day',xrange=c(1,365))

# To have all plots in one device
layout(matrix(1:6,nrow=3,byrow=FALSE))
plot(ITP.result,main='NASA data', plot.adjpval = TRUE,xlab='Day',xrange=c(1,365))

## End(Not run)
```

`summary.ITPaov` *Summarizing Functional Analysis of Variance Fits*

Description

`summary` method for class "`ITPaov`".

Usage

```
## S3 method for class 'ITPaov'
summary(object, ...)
```

Arguments

`object`	An object of class "`ITPaov`", usually, a result of a call to `ITPaovbspline`.
`...`	Further arguments passed to or from other methods.

Value

The function `summary.ITPaov` computes and returns a list of summary statistics of the fitted functional analysis of variance given in `object`, using the component "`call`" from its arguments, plus:

`factors`	A `L x 1` matrix with columns for the factors of ANOVA, and corresponding (two-sided) ITP-adjusted minimum p-values of the corresponding tests of significance (i.e., the minimum p-value over all `p` basis components used to describe functional data).
`R2`	Range of the functional R-squared.
`ftest`	ITP-adjusted minimum p-value of functional F-test.

Author(s)

Alessia Pini, Simone Vantini

References

A. Pini and S. Vantini (2013). The Interval Testing Procedure: Inference for Functional Data Controlling the Family Wise Error Rate on Intervals. *MOX-report 13/2013*, Politecnico di Milano.

See Also

See also `ITPlmbspline`, `ITP1bspline`, `ITP2bspline`, `ITP2fourier`, `ITP2pafourier`, and `ITPimage`.

Examples

```
# Importing the NASA temperatures data set
data(NASAtemp)

temperature <- rbind(NASAtemp$milan,NASAtemp$paris)
groups <- c(rep(0,22),rep(1,22))

# Performing the ITP
## Not run:

ITP.result <- ITPaovbspline(temperature ~ groups,B=1000,nknots=20,order=3)

# Summary of the ITP results
summary(ITP.result)


## End(Not run)
```

`summary.ITPlm` *Summarizing Functional-on-Scalar Linear Model Fits*

Description

`summary` method for class "`ITPlm`".

Usage

```
## S3 method for class 'ITPlm'
summary(object, ...)
```

Arguments

`object`	An object of class "`ITPlm`", usually, a result of a call to `ITPlmbspline`.
`...`	Further arguments passed to or from other methods.

Value

The function summary.ITPlm computes and returns a list of summary statistics of the fitted functional-on-scalar linear model given in object, using the component "call" from its arguments, plus:

ttest	A L+1 x 1 matrix with columns for the functional regression coefficients, and corresponding (two-sided) ITP-adjusted minimum p-values of t-tests (i.e., the minimum p-value over all p basis components used to describe functional data).
R2	Range of the functional R-squared.
ftest	ITP-adjusted minimum p-value of functional F-test.

Author(s)

Alessia Pini, Simone Vantini

References

A. Pini and S. Vantini (2013). The Interval Testing Procedure: Inference for Functional Data Controlling the Family Wise Error Rate on Intervals. *MOX-report 13/2013*, Politecnico di Milano.

See Also

See also ITPaovbspline, ITP1bspline, ITP2bspline, ITP2fourier, ITP2pafourier, and ITPimage.

Examples

```
# Importing the NASA temperatures data set
data(NASAtemp)

temperature <- rbind(NASAtemp$milan,NASAtemp$paris)
groups <- c(rep(0,22),rep(1,22))

# Performing the ITP
## Not run:

ITP.result <- ITPlmbspline(temperature ~ groups,B=1000,nknots=20)

# Summary of the ITP results
summary(ITP.result)


## End(Not run)
```

Index

Bibliography

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

Abramowicz, K., Häger, C., Hébert-Losier, K., Pini, A., Schelin, L., Strandberg, J. and Vantini, S. (2014), An inferential framework for domain selection in functional anova, *in* E. G. Bongiorno, E. Salinelli, A. Goia and P. Vieu, eds, 'Contributions in infinite-dimensional statistics and related topics', Società Editrice Esculapio, pp. 13–18.

AG, D. C. (2005), 'Patent no. de10241593'.

Ageberg, E. and Fridén, T. (2008), Normalized motor function but impaired sensory function after unilateral non-reconstructed acl injury: patients compared with uninjured controls, *Knee Surg. Sport. Tr. A.* **16**(5), 449–456.

Ageberg, E., Pettersson, A. and Fridén, T. (2007), 15-year follow-up of neuromuscular function in patients with unilateral nonreconstructed anterior cruciate ligament injury initially treated with rehabilitation and activity modification a longitudinal prospective study, *Am. J. Sport Med.* **35**(12), 2109–2117.

Ageberg, E., Thomeé, R., Neeter, C., Silbernagel, K. G. and Roos, E. M. (2008), Muscle strength and functional performance in patients with anterior cruciate ligament injury treated with training and surgical reconstruction or training only: A two to five-year followup, *Arthrit. Care Res.* **59**(12), 1773–1779.

Ageberg, E., Zätterström, R., Fridén, T. and Moritz, U. (2001), Individual factors affecting stabilometry and one-leg hop test in 75 healthy subjects, aged 15–44 years, *Scand. J. Med. Sci. Sport.* **11**(1), 47–53.

Ait Si Selmi, T., Fithian, D. and Neyret, P. (2006), The evolution of osteoarthritis in 103 patients with acl reconstruction at 17 years follow-up, *Knee* **13**(5), 353–358.

Akhter, R., Steen, W. M. and Watkins, K. G. (1991), Welding zinc-coated steel with a laser and the properties of the weldment, *J. Laser Appl.* **3**(2), 9–20.

Anderson, M. J. and Legendre, P. (1999), An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model, *J. Stat. Comput. Sim.* **62**(3), 271–303.

Anderson, M. J. and Robinson, J. (2001), Permutation tests for linear models, *Aust. NZ J. Stat.* **43**(1), 75–88.

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

Ardern, C. L., Webster, K. E., Taylor, N. F. and Feller, J. A. (2011), Return to sport following anterior cruciate ligament reconstruction surgery: a systematic review and meta-analysis of the state of play, *Brit. J. Sport. Med.* p. bjsports76364.

Arveson, W. (2002), *A Short Course on Spectral Theory*, Vol. 209 of *Graduate Texts in Mathematics*, first edn, Springer, New York.

Atwa, Y. M., El-Saadany, E. F., Salama, M. M. A. and Seethapathy, R. (2010), Optimal renewable resources mix for distribution system energy loss minimization, *Power Systems, IEEE Transactions on* **25**(1), 360–370.

Aven, T. and Zio, E. (2011), Some considerations on the treatment of uncertainties in risk assessment for practical decision making, *Reliability Engineering & System Safety* **96**(1), 64–74.

Baraldi, P. and Zio, E. (2008), A combined monte carlo and possibilistic approach to uncertainty propagation in event tree analysis, *Risk Anal.* **28**(5), 1309–1326.

Barber-Westin, S. D. and Noyes, F. R. (2011), Objective criteria for return to athletics after anterior cruciate ligament reconstruction and subsequent reinjury rates: a systematic review, *Phys. Sport. Med.* **39**(3), 100–110.

Barrett, D. S. (1991), Proprioception and function after anterior cruciate reconstruction, *J. Bone Joint Surg. Br.* **73**(5), 833–837.

Basso, D., Pesarin, F., Salmaso, L. and Solari, A. (2009), *Permutation Tests for Stochastic Ordering and ANOVA*, Springer.

Baudrit, C., Dubois, D. and Guyonnet, D. (2006), Joint propagation and exploitation of probabilistic and possibilistic information in risk assessment, *IEEE T. Fuzzy Syst.* **14**(5), 593–608.

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.

Billinton, R., A., R. N. and Allan, R. N. (1984), *Reliability evaluation of power systems*, Vol. 2, Plenum press New York.

Boogaart, K. G., Egozcue, J. J. and Pawlowsky-Glahn, V. (2014), Bayes hilbert spaces, *Aust. NZ J. Stat.* **56**(2), 171–194.

Borg, G. A. (1982), Psychophysical bases of perceived exertion, *Med. Sci. Sport. Exerc.* **14**(5), 377–381.

Borges, C. L. T. (2012), An overview of reliability models and methods for distribution systems with renewable energy distributed generation, *Renew. Sustain. Energ. Rev.* **16**(6), 4008–4015.

Bosq, D. (2000), *Linear processes in function spaces: theory and applications*, Vol. 149, Springer.

Box, J. F. (1981), Gosset, fisher, and the t distribution, *Am. Stat.* **35**(2), 61–66.

Box, J. F. (1987), Guinness, gosset, fisher, and small samples, *Stat. Sci.* pp. 45–52.

Brophy, R. H., Wright, R. W. and Matava, M. J. (2009), Cost analysis of converting from single-bundle to double-bundle anterior cruciate ligament reconstruction, *Am. J. Sport. Med.* **37**(4), 683–687.

Bryant, A. L., Newton, R. U. and Steele, J. (2009), Successful feed-forward strategies following acl injury and reconstruction, *J. Electromyogr. Kines.* **19**(5), 988–997.

Capello, E., Colombo, D. and Previtali, B. (2008), Monitoring through the optical combiner in fibre laser welding, *in* 'ICALEO 2008 Congress proceedings', pp. 75–84.

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, C., Duan, S., Cai, T., Liu, B. and Hu, G. (2011), Optimal allocation and economic analysis of energy storage system in microgrids, *IEEE T. Power Electr.* **26**(10), 2762–2773.

Colombo, D., Colosimo, B. M. and Previtali, B. (2013), Comparison of methods for data analysis in the remote monitoring of remote laser welding, *Opt. Laser. Eng.* **51**(1), 34–46.

Colombo, D. and Previtali, B. (2009), Fiber laser welding of titanium alloys and its monitoring through the optical combiner, *in* 'ICALEO 2009 Congress proceedings', pp. 620–629.

Colombo, D. and Previtali, B. (2010), Through optical combiner monitoring of fiber laser processes, *Int. J. Mater. Form.* **3**(1), 1123–1126.

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

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

Davison, A. C. and Hinkley, D. V. (1997), *Bootstrap methods and their application*, Cambridge university press.

De Moivre, A. (1733), Approximatio ad summam terminorum binomii $(a + b)^n$ in seriem expansi.

Delaigle, A. and Hall, P. (2010), Defining probability density for a distribution of random functions, *Ann. Stat.* pp. 1171–1193.

Delincé, P. and Ghafil, D. (2012), Anterior cruciate ligament tears: conservative or surgical treatment? a critical review of the literature, *Knee Surg. Sport. Tr. A.* **20**(1), 48–61.

Deneweth, J. M., Bey, M. J., McLean, S. G., Lock, T. R., Kolowich, P. A. and Tashman, S. (2010), Tibiofemoral joint kinematics of the anterior cruciate ligament-reconstructed knee during a single-legged hop landing, *Am. J. Sport. Med.* **38**(9), 1820–1828.

Denti, M., Monteleone, M., Berardi, A. and Panni, A. (1994), Anterior cruciate ligament mechanoreceptors: histologic studies on lesions and reconstruction, *Clin. Orthop. Relat. R.* **308**, 29–32.

Dubois, D. (2006), Possibility theory and statistical reasoning, *Comput. Statist. Data Anal.* **51**(1), 47–69.

Egozcue, J. J., Díaz-Barrero, J. L. and Pawlowsky-Glahn, V. (2006), Hilbert space of probability density functions based on aitchison geometry, *Acta Mathematica Sinica* **22**(4), 1175–1182.

Egozcue, J. J. and Pawlowsky-Glahn, V. (2006), Simplicial geometry for compositional data, *Geological Soc. London, Special Publications* **264**(1), 145–159.

ENEA (2006), 'I condizionatori dell'aria: raffrescatori e pompe di calore', Ente per le Nuove Tecnologie, l'Energia e l'Ambiente.

Fan, J. and Lin, S. (1998), Test of significance when data are curves, *J. Amer. Statist. Assoc.* **93**(443), 1007–1021.

Fan, J. and Zhang, J. T. (2000), Two-step estimation of functional linear models with applications to longitudinal data, *J. R. Stat. Soc. Ser. B Stat. Methodol.* **62**(2), 303–322.

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

Fisher, R. A. (1925*a*), Applications of 'student's' distribution, *Metron* **5**(3), 90–104.

Fisher, R. A. (1925*b*), *Statistical methods for research workers*, Genesis Publishing Pvt Ltd.

Fisher, R. A. (1936), The coefficient of racial likeness and the future of craniometry, *J. Anthropol. Inst. of Great Britain and Ireland* pp. 57–63.

Flage, R., Aven, T. and Zio, E. (2008), Alternative representations of uncertainty in system reliability and risk analysis-review and discussion, *in* 'European Safety and Reliability Conference (ESREL), Valencia, SPAIN', Vol. 43, pp. 22–25.

Freedman, D. and Lane, D. (1983), A nonstochastic interpretation of reported significance levels, *J. Bus. Econ. Stat.* **1**(4), 292–298.

FreeMeteo (2012), 'Hourly weather history for barcelona, spain. station reporting: Sevilla/san pablo'.
URL: *http://freemeteo.com*

Frobell, R. B. (2011), Change in cartilage thickness, posttraumatic bone marrow lesions, and joint fluid volumes after acute acl disruptiona two-year prospective mri study of sixty-one subjects, *J. Bone Joint Surg.* **93**(12), 1096–1103.

Frobell, R. B., Roos, E. M., Roos, H. P., Ranstam, J. and Lohmander, L. S. (2010), A randomized trial of treatment for acute anterior cruciate ligament tears, *New Eng. J. Med.* **363**(4), 331–342.

Gauss, C. F. (1809), *Theoria motus corporum coelestium in sectionibus conicis solem ambientium.*

Gertheiss, J., Goldsmith, J., Crainiceanu, C. and Greven, S. (2013), Longitudinal scalar-on-functions regression with application to tractography data, *Biostatistics* **14**(3), 447–461.

Godwin, A., Takahara, G., Agnew, M. and Stevenson, J. (2010), Functional data analysis as a means of evaluating kinematic and kinetic waveforms, *Theor. Issues Ergonom. Sci.* **11**(6), 489–503.

Gokeler, A., Benjaminse, A., Hewett, T. E., Paterno, M. V., Ford, K. R., Otten, E. and Myer, G. D. (2013), Feedback techniques to target functional deficits following anterior cruciate ligament reconstruction: Implications for motor control and reduction of second injury risk, *Sport. Med.* **43**(11), 1065–1074.

Gokeler, A., Hof, A. L., Arnold, M. P., Dijkstra, P. U., Postema, K. and Otten, E. (2010), Abnormal landing strategies after acl reconstruction, *Scand. J. Med. Sci. Sport.* **20**(1), e12–e19.

Good, P. I. (2005), *Permutation, parametric and bootstrap tests of hypotheses*, Vol. 3, Springer, New York.

Gosset, W. S. S. (1908), The probable error of a mean, *Biometrika* **6**, 1–25.

Grip, H. and Häger, C. (2013), A new approach to measure functional stability of the knee based on changes in knee axis orientation, *J. Biomech.* **46**(5), 855–862.

Grood, E. S. and Suntay, W. J. (1983), A joint coordinate system for the clinical description of three-dimensional motions: application to the knee, *J. Biomech. Eng.* **105**(2), 136–144.

Gu, H. (2010), Laser lap welding of zinc coated steel sheet with laser-dimple technology, *J. Laser Appl.* **22**(3), 87–91.

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*a*), Two-sample tests in functional data analysis starting from discrete data, *Stat. Sinica* **17**(4), 1511.

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

Hastie, T., Tibshirani, R. and Friedman, J. (2009), *The elements of statistical learning*, Vol. 2, Springer.

Helton, J. C. and Oberkampf, W. L. (2004), Alternative representations of epistemic uncertainty, *Reliab. Eng. Syst. Safe.* **85**(1), 1–10.

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.

Hotelling, H. (1931), The generalization of student's ratio, *Ann. Math. Stat.* **2**(3), 360–378.

Hron, K., Menafoglio, A., Templ, M., Hruzova, K. and Filzmoser, P. (2014), Simplicial principal component analysis for density functions in bayes spaces, Technical Report 25/2014, MOX, Politecnico di Milano.

INE (2008), 'Instituto nacional de estadisticas: Encuesta de hogares y medio ambiente 2008', http://www.ine.es/jaxi.

Izquierdo, M., Moreno-Rodríguez, A., González-Gil, A. and García-Hernando, N. (2011), Air conditioning in the region of madrid, spain: An approach to electricity consumption, economics and co2 emissions, *Energy* **36**(3), 1630–1639.

Jamison, S. T., McNally, M. P., Schmitt, L. C. and Chaudhari, A. M. W. (2013), The effects of core muscle activation on dynamic trunk position and knee abduction moments: implications for acl injury, *J. Biomech.* **46**(13), 2236–2241.

Johnson, R. A. and Wichern, D. W. (2007), *Applied multivariate statistical analysis*, 6th edn, Pearson Education Inc.

Klir, G. J. and Yuan, B. (1995), *Fuzzy sets and fuzzy logic, theory and applications*, Upper Saddle River, NJ: Prentice-Hall.

Kramer, C. Y. (1957), Extension of multiple range tests to group correlated adjusted means, *Biometrics* **13**(1), 13–18.

Lai, T. L., Robbins, H. and Wei, C. Z. (1979), Strong consistency of least squares estimates in multiple regression ii, *J. Multivariate Anal.* **9**(3), 343–361.

Laplace, P. S. (1820), *Théorie analytique des probabilités*, Courcier.

Lee, H., Cheng, C. and Liau, J. (2009), Correlation between proprioception, muscle strength, knee laxity, and dynamic standing balance in patients with chronic anterior cruciate ligament deficiency, *Knee* **16**(5), 387–391.

Lehmann, E. L. and Romano, J. P. (2006), *Testing statistical hypotheses*, Springer, New York.

Lephart, S. M., Pincivero, D. M. and Rozzi, S. L. (1998), Proprioception of the ankle and knee, *Sport. Med.* **25**(3), 149–155.

Li, Y. and Zio, E. (2012), Uncertainty analysis of the adequacy assessment model of a distributed generation system, *Renew. Energ.* **41**, 235–244.

Liu, Z., Wen, F. and Ledwich, G. (2011), Optimal siting and sizing of distributed generators in distribution systems considering uncertainties, *IEEE T. Power Deliver.* **26**(4), 2541–2551.

Lyapunov, A. M. (1901), Nouvelle forme du théoreme sur la limite de probabilité, *Acad. Imp. Sci. St. Petersbourg, Classe Phys. Math* **12**, 12.

Mahalanobis, P. C. (1927), Analysis of race mixture in bengal, *Jour. and Proc. Asiat. Soc., Bengal,(New Series) Vol. XXIII* (3), 301–333.

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

Marron, J. S. and Alonso, A. M. (2014), Overview of object oriented data analysis, *Biomet. J.* .

Marseguerra, M. and Zio, E. (2002), *Basics of the Monte Carlo method with application to system reliability*, LiLoLe - Verlag GmbH, Hagen, Germany.

Menafoglio, A., Guadagnini, A. and Secchi, P. (2013), A kriging approach based on aitchison geometry for the characterization of particle-size curves in heterogeneous aquifers, *Stochastic Environmental Research and Risk Assessment* pp. 1–17.

Mihelic, R., Jurdana, H., Jotanovic, Z., Madjarevic, T. and Tudor, A. (2011), Long-term results of anterior cruciate ligament reconstruction: a comparison with non-operative treatment with a follow-up of 17–20 years, *Int. Orthop.* **35**(7), 1093–1097.

NASA (2008), 'Surface meteorology and solar energy, a renewable energy resource web site (release 6.0)', http://eosweb.larc.nasa.gov/sse.

Oberländer, K. D., Brüggemann, G.-P., Hoeher, J. and Karamanidis, K. (2013), Altered landing mechanics in acl-reconstructed patients., *Med. Sci. Sport. Ex.* **45**(3), 506–513.

Omie (2012), 'Precio orario del mercado diario', http://www.omie.es/files/flash/ResultadosMercado.swf.

Orishimo, K. F., Kremenic, I. J., Mullaney, M. J., McHugh, M. P. and Nicholas, S. J. (2010), Adaptations in single-leg hop biomechanics following anterior cruciate ligament reconstruction, *Knee Surg. Sport. Tr. A.* **18**(11), 1587–1593.

Passerini, T., Sangalli, L. M., Vantini, S., Piccinelli, M., Bacigaluppi, S., Antiga, L., Boccardi, E., Secchi, P. and Veneziani, A. (2012), An integrated statistical investigation of internal carotid arteries of patients affected by cerebral aneurysms, *Cardiovascu. Eng. and Tech.* pp. 1–15.

Pataky, T. C., Robinson, M. A. and Vanrenterghem, J. (2013), Vector field statistical analysis of kinematic and force trajectories, *J. Biomech.* **46**(14), 2394–2401.

Paterno, M. V., Schmitt, L. C., Ford, K. R., Rauh, M. J., Myer, G. D., Huang, B. and Hewett, T. E. (2010), Biomechanical measures during landing and postural stability predict second anterior cruciate ligament injury after anterior cruciate ligament reconstruction and return to sport, *Am. J. Sport. Med.* **38**(10), 1968–1978.

Pernin, J., Verdonk, P., Selmi, T. A. S., Massin, P. and Neyret, P. (2010), Long-term follow-up of 24.5 years after intra-articular anterior cruciate ligament reconstruction with lateral extra-articular augmentation, *Am. J. Sport. Med.* **38**(6), 1094–1102.

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

Petschnig, R. and Baron, R. (2009), Functional limitation of the lower limb following anterior cruciate ligament reconstruction up to 72 weeks, *Sport-und Präventivmedizin* **39**(2), 30–33.

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.

Pini, A. and Vantini, S. (2014), *fdatest: Interval Testing Procedure for Functional Data*. R package version 2.0.
URL: *http://CRAN.R-project.org/package=fdatest*

R Core Team (2012), *R: a language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: *http://www.R-project.org/*

Ramsay, J. O., Hooker, G. and Graves, S. (2009), Registration: Aligning features for samples of curves, *in* 'Functional Data Analysis with R and MATLAB', Springer, pp. 117–130.

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.

Rao, C. R. and Mitra, S. K. (1971), *Generalized inverse of matrices and its applications*, Wiley Series in Probability and Statistics, John Wiley and Sons Inc.

Reinke, E. K., Spindler, K. P., Lorring, D., Jones, M. H., Schmitz, L., Flanigan, D. C., An, A. Q., Quiram, A. R., Preston, E., Martin, M. et al. (2011), Hop tests correlate with ikdc and koos at minimum of 2 years after primary acl reconstruction, *Knee Surg. Sport. Tr. A.* **19**(11), 1806–1816.

Reiss, P. T., Huang, L. and Mennes, M. (2010), Fast function-on-scalar regression with penalized basis expansions, *Int. J. Biostat.* **6**(1).

Roos, P. E., Button, K., Sparkes, V. and van Deursen, R. W. M. (2014), Altered biomechanical strategies and medio-lateral control of the knee represent incomplete recovery of individuals with injury during single leg hop, *J. Biomech.* **47**(3), 675–680.

Røtterud, J. H., Sivertsen, E. A., Forssblad, M., Engebretsen, L. and Årøen, A. (2011), Effect of gender and sports on the risk of full-thickness articular cartilage lesions in anterior cruciate ligament–injured knees a nationwide cohort study from sweden and norway of 15 783 patients, *Am. J. Sport. Med.* **39**(7), 1387–1394.

Ryan, W., Harrison, A. and Hayes, K. (2006), Functional data analysis of knee joint kinematics in the vertical jump, *Sport. Biomech.* **5**(1), 121–138.

Salameh, Z. M., Borowy, B. S. and Amin, A. R. A. (1995), Photovoltaic module-site matching based on the capacity factors, *IEEE Trans. on Energy Conversion* **10**(2), 326–332.

Sangalli, L. M., Secchi, P., Vantini, S. and Veneziani, A. (2009), A case study in exploratory functional data analysis: geometrical features of the internal carotid artery, *J. Amer. Statist. Assoc.* **104**(485), 37–48.

Scheffe, H. (1959), *The analysis of variance*, Vol. 72, John Wiley & Sons.

Schmitt, L. C., Paterno, M. V. and Hewett, T. E. (2012), The impact of quadriceps femoris strength asymmetry on functional performance at return to sport following anterior cruciate ligament reconstruction, *J. Orthop. Sport. Phys. Ther.* **42**(9), 750–759.

Schott, J. R. (2007), Some high-dimensional tests for a one-way manova, *J. Multivariate Anal.* **98**(9), 1825–1839.

Schwoerer, T. (2008), Robot-guided remote laser scanner welding for highly-productive welding applications, *in* 'ICALEO 2008 Congress proceedings', pp. 392–398.

Secchi, P., Stamm, A. and Vantini, S. (2013), Inference for the mean of large p small n data: A finite-sample high-dimensional generalization of hotellings theorem, *Electron. J. Stat.* **7**, 2005–2031.

Sech-Spahousec, P. (n.d.), 'Análisis del consumo energético del sector residencial en españa'.

Sernert, N., Kartus, J., Köhler, K., Stener, S., Larsson, J., Eriksson, B. I. and Karlsson, J. (1999), Analysis of subjective, objective and functional examination tests after anterior cruciate ligament reconstruction a follow-up of 527 patients, *Knee Surg. Sport. Tr. A.* **7**(3), 160–165.

Silvers, H. J. and Mandelbaum, B. R. (2011), Acl injury prevention in the athlete, *Sport. Orthop. Traumat.* **27**(1), 18–26.

Spitzner, D. J., Marron, J. S. and Essick, G. K. (2003), Mixed-model functional anova for studying human tactile perception, *J. Amer. Statist. Assoc.* **98**(462), 263–272.

Srivastava, M. S. (2007), Multivariate theory for analyzing high dimensional data, *J. Japan Statist. Soc* **37**(1), 53–86.

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.* .

Steen, W. M., Mazumder, J. and Watkins, K. G. (2003), *Laser material processing*, Springer.

Stensdotter, A., Tengman, E., Olofsson, L. B. and Häger, C. (2013), Deficits in single-limb stance more than 20 years after acl injury, *Eur. J Physiotherapy* **15**(2), 78–85.

Tarpey, T. (2003), Clustering Functional Data, *J. Classif.* **20**, 93–114.

Tengman, E., Brax Olofsson, L., Nilsson, K. G., Tegner, Y., Lundgren, L. and Häger, C. (2014), Anterior cruciate ligament injury after more than 20 years: I. physical activity level and knee function, *Scand. J. Med. Sci. Sport.* .

Tengman, E., Grip, H. and Häger, C. (2013), One-leg hop - kinematic analysis ∼20 years after anterior cruciate ligament injury, *Gait Posture* **38**(Supplement 1), S25–S26.

Van der Harst, J. J., Gokeler, A. and Hof, A. L. (2007), Leg kinematics and kinetics in landing from a single-leg hop for distance. a comparison between dominant and non-dominant leg, *Clin. Biomech.* **22**(6), 674–680.

von Porat, A., Henriksson, M., Holmström, E., Thorstensson, C. A., Mattsson, L. and Roos, E. M. (2006), Knee kinematics and kinetics during gait, step and hop in males with a 16 years old acl injury compared with matched controls, *Knee Surg. Sport. Tr. A.* **14**(6), 546–554.

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.

Wakeling, J. M. (2009), Patterns of motor recruitment can be determined using surface emg, *J. Electromyogr. Kines.* **19**(2), 199–207.

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.

Xergia, S. A., Pappas, E., Zampeli, F., Georgiou, S. and Georgoulis, A. D. (2013), Asymmetries in functional hop tests, lower extremity kinematics, and isokinetic strength persist 6 to 9 months following anterior cruciate ligament reconstruction, *J. Orthop. Sport. Phys. Ther.* **43**(3), 154–162.

Yamaguchi, S., Sasho, T., Tsuchiya, A., Wada, Y. and Moriya, H. (2006), Long term results of anterior cruciate ligament reconstruction with iliotibial tract: 6-, 13-, and 24-year longitudinal follow-up, *Knee Surg. Sport. Tr. A.* **14**(11), 1094–1100.

Zeng, C., Pan, Z., MaWhinney, S., Barón, A. E. and Zerbe, G. O. (2011), Permutation and f distribution of tests in the multivariate general linear model, *Am. Stat.* **65**(1), 31–36.

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.