# Non-parametric methods for functional data

Johan Strandberg


Department of Mathematics
and Mathematical Statistics
Umeå University, 2020

Doctoral dissertation
Department of Mathematics and Mathematical Statistics
Umeå University
SE-901 87 Umeå
Sweden


Dissertation for PhD
ISBN 978-91-7855-374-7 (print)
ISBN 978-91-7855-375-4 (pdf)
ISSN 1653-0829
Electronic version available at: http://umu.diva-portal.org/
Printed by: Cityprint AB i Norr AB
Umeå, Sweden 2020

*To my family*

# Contents

# List of papers

The thesis is based on the following papers:

**I**. Abramowicz, K., Häger, C., Hébert-Losier, K., Pini, A., Schelin, L., Strandberg, J., & Vantini, S. (2014). An inferential framework for domain selection in functional ANOVA. E.G. Bongiorno, E. Salinelli, A. Goia, P. Vieu (Eds.) *Contributions in infinite-dimensional statistics and related topics*, 13–18.

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

**III**. Strandberg, J., Pini, A., Häger, C. K., & Schelin, L. (2020). Analysis choices provide contrasting conclusions when evaluating jump performance: A multi-aspect inferential method applied to kinematic curves from the one-leg vertical hop in knee-injured and asymptomatic persons. *Manuscript.*

**IV**. Abramowicz, K., Schelin, L., Sjöstedt de Luna, S., & Strandberg, J. (2019). Multiresolution clustering of dependent functional data with application to climate reconstruction. *Stat, 8*(1), e240.

**V**. Abramowicz, K., Sjöstedt de Luna, S., & Strandberg, J. (2020). Nonparametric clustering methods to identify latent structures from a sequence of dependent categorical data. *Manuscript.*

**VI**. Strandberg, J., de Luna, S. S., & Mateu, J. (2019). Prediction of spatial functional random processes: comparing functional and spatio-temporal kriging approaches. *Stochastic Environmental Research and Risk Assessment, 33*(10), 1699-1719.

# Abstract

In this thesis we develop and study non-parametric methods within three major areas of functional data analysis: testing, clustering and prediction. The thesis consists of an introduction to the field, a presentation and discussion of the three areas, and six papers.

In Paper I, we develop a procedure for testing for group differences in functional data. In case of significant group differences, the test procedure identifies which of the groups that significantly differ, and also the parts of the domain they do so, while controlling the type I error of falsely rejecting the null hypothesis. In Paper II, the methodology introduced in Paper I is applied to knee kinematic curves from a one-leg hop for distance to test for differences within and between three groups of individuals (with and without knee deficits). It was found that two of the groups differed in their knee kinematics. We also found that the individual kinematic patterns differed between the two legs in one of the groups. In Paper III, we test for group differences in three groups with respect to joint kinematics from a vertical one-leg hop using a novel method that allows accounting for multiple joints at the same time. The aim of Paper III, as one of few within the field of biomechanics, is to illustrate how different choices prior to the analysis can result in different contrasting conclusions. Specifically, we show how the conclusions depend on the choice of type of movement curve, the choice of leg for between-group comparisons and the included joints.

In Paper IV, we present a new non-parametric clustering method for dependent functional data, the double clustering bagging Voronoi method. The objective of the method is to identify latent group structures that slowly vary over domain and give rise to different frequency patterns of functional data object types. The method uses a bagging strategy based on random Voronoi tessellations in which local representatives are formed and clustered. Combined with the clustering method, we also propose a multiresolution approach which allows identification of latent structures at different scales. A simulated dataset is used to illustrate the method's potential in finding stable clusters at different scales. The method is also applied to varved lake sediment data with the aim of reconstructing the climate over the past 6000 years, at different resolutions. In Paper V, we expand and modify the bagging strategy used in Paper IV, by considering different methods of generating the tessellations and clustering the local representatives of the tessellations. We propose new methods for clustering dependent categorical data (e.g., labelled functional data) along a one-dimensional domain, which we also compare in a simulation study.

In Paper VI, two kriging approaches to predict spatial functional pro-

cesses are compared, namely functional kriging and spatio-temporal kriging. A simulation study is conducted to compare their prediction performance and computational times. The overall results show that prediction performance is about the same for stationary spatio-temporal processes while functional kriging works better for non-stationary spatio-temporal processes. Furthermore, the computational time for (ordinary) kriging for functional data, was considerably lower than spatio-temporal kriging. Conditions are also formulated under which it is proved that the two functional kriging methods: ordinary kriging for functional data and pointwise functional kriging coincide.

*Keywords:* functional data analysis, testing, clustering, prediction, inference, bagging Voronoi strategy, kriging, dependency

# Acknowledgements

First of all, I would like to thank my supervisor Sara Sjöstedt de Luna for the opportunity of embarking on this journey and taking me on under your wings. Throughout the years, together with my assistant supervisors Lina Schelin, Konrad Abramowicz and Charlotte Häger, you provided me with mentoring, support and inspiration. The devotion and professionalism of all four of you truly inspires me and I could not have asked for any better supervisory team than you.

I would also like to thank all the people that I have met and worked with over the years at the university. In particular, I would like to thank all the staff at the department and my co-authors Alessia Pini, Simone Vantini, Jorge Mateu and Kim Hébert-Losier. It has been a pleasure working with you. Linda Vidman and Therese Kellgren, I am glad to have shared the journey as doctoral students with you. You have made it much easier. To the (dys)-functional family and the research group of the U-motion Lab, thanks for the inspiration and all the fun throughout the years.

I want to thank my parents, brothers and friends for your support and being there for me. To my parents-in-law, thank you for your understanding and your willingness to help, whenever needed.

Last but not least, my deepest gratitude goes to my wife Linnea. Thank you for your love, patience and being my rock. Special thank you for taking care of everything at home during the last steps of this journey - without you, I would have not finished on time. You and the boys are the ones who have given me the energy to reach the finish line. I love you with all my heart!

Umeå, October 2020
Johan Strandberg

# 1 Introduction

In various research fields, such as economics, climatology, meteorology and biomechanics, continuous data series that can be expressed in the form of functions (curves), are continuously being observed and monitored. With the increasing amount of data being observed and stored, the demand to handle and analyse such data has also increased. Functional data analysis (FDA) is a relatively new branch within statistics that provides tools to work with such objects, treating each function in a sample as an entity (see, e.g., Ramsay, 1982; Ramsay and Silverman, 2005). In this thesis, we focus on three areas of statistical analysis within FDA.

The first topic we study relates to a practical application where we analyse human movement strategies after an anterior cruciate ligament (ACL) injury. In clinical biomechanics, it has often been common to reduce continuously observed data series into a few summary measures of interest. However, different curve types can yield similar summary measures, and thus, possibly relevant information contained in the functions may be lost. More recently, the biomechanical community have started to use FDA techniques, including statistical parametric mapping (Pataky et al., 2013), bootstrap prediction bands (Lenhoff et al., 1999), pointwise functional analysis of variance (ANOVA) (Godwin et al., 2010), and functional principal components (Ryan et al., 2006).

In Paper I, we develop an ANOVA testing procedure to investigate if there are differences between functions from different groups, and if so, on which parts of the domain the differences occur. In Paper II we apply and generalise this methodology to a multivariate analysis of variance (MANOVA) procedure to analyse data from a long-term follow-up study of ACL injuries. The methods are used to examine and compare the entire time-domain of knee-joint kinematic curves from a one-leg hop for distance of individuals belonging to three groups. The objective is to test for significant differences in knee motion between and within the groups, and also to identify in which parts of the hop they differ. In Paper III, we introduce another novel FDA testing procedure. It allows for testing between-groups differences, identification of significant regions, while simultaneously accounting for covariates and different aspects of recorded movement. Using the method, we analyse joint kinematic curves obtained from a vertical one-leg hop. In this dataset, each individual is represented by functions describing the kinematics from three trials performed on each leg and records from various joints. Our aim is to illustrate how analysis of different aspects may affect the clinical conclusions.

The inferential procedures developed in Papers I - III are based on nonparametric permutation tests and result in a selection of significant domain

regions, while controlling the type I error of falsely rejecting the null hypotheses. Under conditions on the individual tests used, the methods are provided with so called interval-wise control of the type I error. In particular, if the null hypothesis is true on a specific interval, the probability that it will be falsely rejected in this region is controlled.

The second area of interest in this thesis regards clustering dependent functional data. The interest in such functional data methods is inspired by the paleolimnology research of the sediment, which was sourced from the bottom of Lake Kassjön in northern Sweden (Petterson, 1999). The sediment consist of thousands of annual layers whose seasonal patterns can be described using functional forms and may be associated to the weather condition the year the layer was deposited. In climatology, climate is being defined as frequencies of weather types over longer time periods and it is often of interest to study the climate and its changes at different time scales. Previous methodological works on this dataset includes the works of Arnqvist et al. (2016) and Arnqvist and Sjöstedt de Luna (2019). However, in contrast to those, where the focus is on clustering the annual layers into homogenous groups, our focus is on the latent process, i.e., the climate, that give rise to the annual layers. In particular, in Paper IV, based on the ideas introduced in Secchi et al. (2013), we develop a new clustering method, the double clustering bagging Voronoi (DCBV) method, which consists of two steps and is suitable for both temporal and spatial dependent data. We also propose a multiresolution approach that is designed to identify stable clusters at different scales. Together with the DCBV method it is applied to the sediment data with the purpose of reconstructing the climate regimes in northern Sweden over the past six thousand years, at different resolutions. In Paper V, we study how the mechanism of the second level clustering in the DCBV method, being very general and flexible to its construction, can be modified in order to handle different types of dependency structures.

In Paper VI, we still work with dependent functional data but turn our attention to prediction of spatial functional random processes. Given a sample of curves observed at different locations, a curve at a new location can be predicted by a functional kriging approach, i.e., as a linear combination of the observed curves. A spatial functional process can also be viewed as a spatio-temporal (Sp.T.) process and, thus, a Sp.T. kriging approach can also be used to predict a curve (at a dense grid of values over its time domain), see, e.g., Cressie and Wikle (2015) and Montero et al. (2015). However, it is still unclear which of these approaches that is preferable for the analysis of a particular dataset, as pointed out in Delicado et al. (2010). In Paper VI, we compare these two approaches with respect to prediction performance and computational time, mainly through a simulation study but also on a real

dataset. We also provide conditions under which we show that two of the functional kriging methods coincide.

The remaining part of this thesis is organised as follows. In Section 2, an introduction to FDA is given, including a discussion about functional representations and misalignment of functional data. In Section 3, I present and discuss: 1) the functional tests that have been used in connection to the ACL application, 2) the clustering methods as well as the multiresolution approach that have been developed and used in connection to the sediment application and 3) the different prediction methods that have been compared and evaluated. In Section 4, the papers included in this thesis are summarised and in Section 5, I present my conclusions and possible suggestions for future research.

# 2 Functional data analysis

The field of functional data analysis (FDA) can be described as a collection of statistical methods that deals with analysis of samples of entities that are in the form of functions. The field, although some of its mathematical foundations are from much earlier, was introduced in the 1980s by Jim Ramsey whom coined the term 'functional data analysis' (Ramsay, 1982). The books by Ramsay and Silverman (2005), Ferraty and Vieu (2006) and Hsing and Eubank (2015) provides a broad overview of the field covering its fundamental concepts, many different methodologies and real data applications. The advantage of FDA in comparison to traditional statistical multivariate methods is the additional information that can be extracted from data by exploiting its natural ordering. Additional understanding can be gained by, e.g., studying the derivatives and primitives of the functional objects. For an extensive overview of recent advances within the field, see, e.g., Wang et al. (2016).

## 2.1 Functional representation

Although the objects of the analysis are functions, in practice, functional data are observed as discrete pairs of observations $\{(t_{ij}, y_{ij})\}$, $i = 1, ..., n$, $j = 1, ..., n_i$. The most common way to represent these observations in the form of real-valued functions $X_1(t), ..., X_n(t)$, $t \in T \subset \mathbb{R}$, is through a linear combination of a set of $K$ known basis functions,

$$X_i(t) = \Phi(t)\mathbf{c}_i, \quad i = 1, ..., n, \tag{1}$$

where $\Phi(t) = \{\phi_1(t), ..., \phi_K(t)\}$ and $\mathbf{c}_i = (c_{i1}, ..., c_{iK})^\intercal$ are the basis functions and coefficients, respectively. Two typical choices of basis functions are Fourier bases and b-spline bases, where Fourier bases are suitable for periodic data while b-spline bases may be used for non periodic data.

We assume the following representation for the observed data:

$$\mathbf{y}_i = X_i(\mathbf{t}_i) + \boldsymbol{\epsilon}_i, \quad i = 1, ..., n,$$

where $\mathbf{y}_i = (y_{i1}, ..., y_{in_i})^\intercal$, $\mathbf{t}_i = (t_{i1}, ..., t_{in_i})^\intercal$ and where $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, ..., \epsilon_{in_i})^\intercal$ is an unobservable error term with zero mean and constant variance. The coefficients may be estimated by the least squares criterion,

$$\hat{\mathbf{c}}_i = \underset{\mathbf{c}_i}{\operatorname{argmin}} \{\mathbf{y}_i - \boldsymbol{\Phi}(\mathbf{t}_i)\mathbf{c}_i\}^\intercal \{\mathbf{y}_i - \boldsymbol{\Phi}(\mathbf{t}_i)\mathbf{c}_i\}, \quad i = 1, ..., n,$$

where

$$\Phi(\mathbf{t}_i) = \{\phi_1(\mathbf{t}_i), ..., \phi_K(\mathbf{t}_i)\} = \begin{pmatrix} \phi_1(t_{i1}) & \cdots & \phi_K(t_{i1}) \\ \vdots & \ddots & \vdots \\ \phi_1(t_{in_i}) & \cdots & \phi_K(t_{in_i}) \end{pmatrix}.$$

The explicit solution of the optimisation problem is given by

$$\hat{\mathbf{c}}_i = \{\Phi(\mathbf{t}_i)^\intercal \Phi(\mathbf{t}_i)\}^{-1} \Phi(\mathbf{t}_i)^\intercal \mathbf{y}_i.$$

Sometimes, we want to control the smoothness of the estimated function such that it neither overfits nor becomes too rough. This can be attained by choosing the number of basis functions. Alternatively, we can do it by introducing a penalty term in the criterion,

$$\hat{\mathbf{c}}_i = \underset{\mathbf{c}_i}{\operatorname{argmin}} \{\mathbf{y}_i - \boldsymbol{\Phi}(\mathbf{t}_i)\mathbf{c}_i\}^\intercal \{\mathbf{y}_i - \boldsymbol{\Phi}(\mathbf{t}_i)\mathbf{c}_i\} + \lambda_i \cdot PEN(X_i),$$

where $\lambda_i$ is a smoothing parameter controlling the roughness of the fit, and $PEN$ is the penalty term. Usually the integrated second square derivative is penalised,

$$PEN(X_i) = \int_T \{D^2 X_i(t)\}^2 dt = \int_T \mathbf{c}_i^\intercal \{D^2\Phi(t)\}^\intercal \{D^2\Phi(t)\}\mathbf{c}_i dt = \mathbf{c}_i^\intercal \mathbf{R} \mathbf{c}_i,$$

where $\mathbf{R} = \int_T \{D^2\Phi(t)\}^\intercal \{D^2\Phi(t)\} dt$ is the penalty matrix. The criterion now has the solution

$$\hat{\mathbf{c}}_i = \{\Phi(\mathbf{t}_i)^\intercal \Phi(\mathbf{t}_i) + \lambda_i \mathbf{R}\}^{-1} \Phi(\mathbf{t}_i)^\intercal \mathbf{y}_i,$$

where the smoothing parameter $\lambda_i$ may be determined, e.g., by generalised cross-validation (see, e.g., Wahba, 1990).

## 2.2 Functional alignment

Functional data may suffer from being misaligned, in the sense that timings of features between functional objects in a sample occur at different times. For example, if we consider Figure 1, in which five simulated sample curves before and after alignment are presented, we may notice that the mean curve after alignment serves as a better representative of the amplitude variation of the sample of curves than the mean taken before alignment. In general, disregarding misalignment may affect the intended analysis and the issue of misalignment is a well known problem within FDA (Marron et al., 2015).

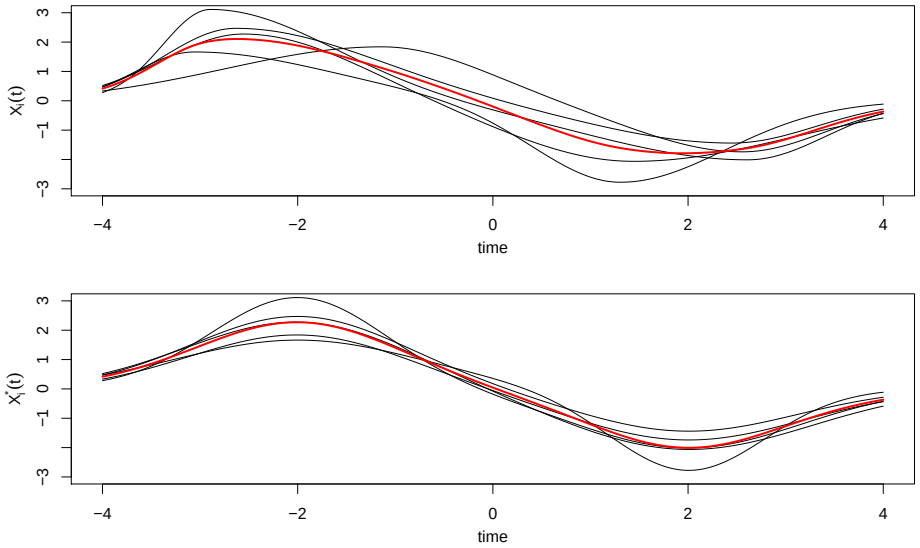


Figure 1: The top and bottom panel illustrates five simulated curves (black lines) before and after alignment, respectively. The red bold lines are the respective (pointwise) means of the sample curves.

Features in the data may be certain characteristics such as peaks or valleys as in the simulated example (see Figure 1) but they may also be due to other certain well defined restrictions. For example, in the one-leg hops described in Paper II and III, we account for individual differences in the speed and duration of hops such that, e.g., events as the take off and landing occur at the same relative time point for all participants. In the sediment application, described in Paper IV, we also perform alignment, here with respect to the different sedimentation rates within and between years in order to make the functional representations of the annual seasonal patterns more comparable. In both these works, as well as in the simulated example, we used landmark

registration to align the functions (Kneip and Gasser, 1992). Landmark registration is a method that involves defining a piecewise linear monotone time transformation $g_i(\cdot)$ such that the registered functions $X_i^*(\cdot) = X_i(g_i(\cdot))$, $i = 1, ..., n$, have their identified features, or landmarks, aligned at their average location. An obvious drawback with the method is that we must have all landmarks to be identifiable in order for the method to work. Another is the fact that it ignores what happens in between the landmarks. However, there are other methods that does not require the identification of landmarks, e.g., methods that uses some family of warping functions together with a similarity measure to align functions. For a general review on the topic, including many methods, I refer to Marron et al. (2015) and the references therein.

# 3 The topics of study in this thesis

In this section I present the three main topics of this thesis. First, I introduce and discuss the different testing procedures that have been developed and applied in connection to the ACL application. After that, I discuss different bagging (Voronoi) tessellation clustering strategies, suitable for clustering dependent functional data such as the sediment data with the aim to recover underlying latent signals. Lastly, I discuss different prediction approaches for spatially dependent functional data.

## 3.1 Testing procedures for functional data

A hypothesis test is used to examine the plausibility of a statement regarding a characteristic of a population, based on sample data. In this thesis, we consider methods for performing hypothesis testing when the sample consists of functional data objects. As in classical univariate and multivariate frameworks, such methods for functional data can be divided into two types: parametric and non-parametric tests. Parametric tests (Cuevas et al., 2004; Abramovich and Angelini, 2006; Staicu et al., 2014; Zhang and Liang, 2014) rely on distributional assumptions (e.g., normality) while non-parametric tests often typically rely on computationally intensive resampling techniques such as, e.g., permutation or bootstrap techniques (Davison and Hinkley, 1997; Pesarin and Salmaso, 2010). In testing functional data, we also have the different aspect of extent that the hypothesis is tested on. A large part of the existing tests for functional data are global testing procedures, i.e., tests that provide general statements about the hypotheses, e.g., if the mean functions differ or not between two groups of functional data, without specifying

on which part of the domain (Hall and Tajvidi, 2002; Cardot et al., 2004; Corain et al., 2014). In case of a significant global test, it is however often also of interest to know in which parts of the domain the null hypothesis is rejected, performing so called local inference.

Local inference for functional data has been discussed in, e.g., Ramsay and Silverman (2005) and Reiss et al. (2010), in which confidence band methods with a pointwise control of the type I error are proposed. However, a pointwise control is not equivalent to having a control of the type I error over complete regions. Another group of procedures that addresses this problem are based on adjustments of pointwise and setwise p-values. Vsevolozhskaya et al. (2014) proposed a method based on a closed testing procedure (CTP) (Marcus et al., 1976) applied on tests on preselected regions of the domain. This method can provide localised significant regions at the resolution specified by the partitioning selected prior to the inferential procedure. The method can provide a control of the probability of falsely selecting any set of the regions. Another approach, proposed by Pini and Vantini (2016) is the Interval Testing Procedure (ITP), that builds on coefficients of a basis expansion of the functional data. ITP can provide a so called interval-wise control of the type I error, and is presented in the framework of testing the equality of mean functions for two functional populations. Both these methods rely on initial choices, such as the partitioning and the type of basis expansion used, which may have an effect on the conclusions drawn. Pini and Vantini (2017) introduced the Interval-Wise Testing procedure (IWT), which neither rely on a basis expansion nor on an apriori partitioning. In theory, IWT is based on an infinite number of tests on all possible subintervals of the domain and results in a continuous adjusted p-value function which can be used for domain selection by thresholding it at a given significance level. Further, it can provide an interval-wise control of the type I error. More recently, the IWT procedure has been shown to provide interval-wise error control for more general testing situations than the two-sample inferential setting studied in the original paper by Pini and Vantini (2017). In Abramowicz et al. (2018) it was extended and applied to functional-on-scalar linear models while in Pini et al. (2019) it was studied in a two-sample inferential setting with additional focus on detection of significance along multiple functional aspects of the data, e.g., in different dimensions in multivariate functional data.

In Paper I, we extend ITP to the context of ANOVA while in Paper II we further extend it to MANOVA and apply them to draw new conclusions about knee kinematics data. Kinematics data is also studied in Paper III, where the multi-aspect IWT presented in Pini et al. (2019) is extended to functional-on-scalar linear models.

We now present the IWT procedure as well as the multi-aspect IWT

procedure in the context of functional-on-scalar linear models in more detail. This is followed by a short discussion about the main ideas behind the ITP.

### 3.1.1 Interval-wise testing

Suppose we have a sample of functional data $X_i(t)$, $t \in D = [0,1]$, $X_i(\cdot) \in \mathcal{C}^0(D)$, for individuals $i = 1, ..., n$ and $L$ observed scalar covariates. Consider the functional-on-scalar model defined by,

$$X_i(t) = \beta_0(t) + \sum_{l=1}^{L} \beta_l(t) x_{li} + \epsilon_i(t), \tag{2}$$

where $x_{li} \in \mathbb{R}$, $l = 1, ..., L$, are the scalar covariates, $\beta_l(t)$ the time dependent coefficients and $\epsilon_i(t)$ independent and identically distributed functional errors with zero mean and finite total variance, i.e., $\int_D E[\epsilon_i(t)]^2 dt < \infty$.

To test statements on the covariates $\boldsymbol{\beta}(t) = (\beta_0(t), ..., \beta_L(t))^\intercal$ in a general setting at a given time point $t$, we may consider the following hypothesis

$$H_0^t\text{: } \mathbf{C}\boldsymbol{\beta}(t) = \mathbf{c}_0(t) \text{ against } H_1^t\text{: } \mathbf{C}\boldsymbol{\beta}(t) \neq \mathbf{c}_0(t), \tag{3}$$

where $\mathbf{C} \in \mathbb{R}^{q \times (L+1)}$ is a design matrix specifying the number of linear combinations $q$ of the functional coefficients to jointly be tested and $\mathbf{c}_0(t) = (c_{01}(t), ..., c_{0q}(t))^\intercal$ is a vector of fixed functions. The hypothesis can be tested, e.g., using the test statistic

$$T(t) = \{\mathbf{C}\hat{\boldsymbol{\beta}}(t) - \mathbf{c}_0(t)\}^\intercal \{\mathbf{C}\hat{\boldsymbol{\beta}}(t) - \mathbf{c}_0(t)\}, \tag{4}$$

where the estimate $\hat{\beta}(t)$ of $\beta(t)$ can be estimated (pointwise for a given $t$) using, e.g., ordinary least squares (OLS), see Abramowicz et al. (2018) for details.

As the objects we investigate are of functional nature, we are not only interested in investigating the hypothesis at a single time point. One alternative is to perform simultaneously inference on the whole domain, i.e., testing the global hypothesis,

$$H_0^D\text{: } \{\forall t \in D,\ H_0^t \text{ is true}\} \text{ against } H_1^D\text{: } \{\exists t \in D : H_0^t \text{ is not true}\}. \tag{5}$$

To do that we may use an integrated version of the test statistic in (4),

$$T^D = \int_D T(t) dt.$$

In case of rejection of the global null hypothesis, it is often of interest to identify the intervals on the domain on which the significant differences lie.

This may naturally be done by testing the hypothesis in (3) for each $t \in D$. The challenge is however to control the family-wise error rate (FWER), i.e, the rate of false significant discoveries (Type I errors) committed in the uncountable infinite number of (dependent) hypothesis tests. The interval-wise testing (IWT) procedure introduced in Pini and Vantini (2017), was developed for testing functional hypotheses and selecting significant intervals. The procedure relies on an adjusted p-value function that can be thresholded to select significant sets where the null hypothesis is rejected. The adjusted p-value function builds on p-values obtained through testing all intervals $\mathcal{I} \subseteq D$ for significance. In particular, for all $\mathcal{I}$ we test the hypotheses,

$$H_0^{\mathcal{I}}\text{: } \{\forall t \in \mathcal{I},\ H_0^t \text{ is true}\} \text{ against } H_1^{\mathcal{I}}\text{: } \{\exists t \in \mathcal{I} : H_0^t \text{ is not true}\}, \qquad (6)$$

which may be tested with the test statistic

$$T^{\mathcal{I}} = \int_{\mathcal{I}} T(t)dt.$$

The adjusted p-value function $\tilde{p}(t)$ at time point $t$ is then computed as the supremum of the p-values $p_D^{\mathcal{I}}$ obtained from all the tests of hypothesis (6) for which $t \in \mathcal{I}$,

$$\tilde{p}_D(t) = \sup_{\{\mathcal{I} \subseteq D : t \in \mathcal{I}\}} p_D^{\mathcal{I}}.$$

The IWT procedure is provided with an intermediate control that lies between weak control of the FWER (assuming that the null hypothesis is true on the entire domain, the probability of detecting false significant discoveries is controlled) and strong control of the FWER (allowing any configuration of sets of the domain where the null hypothesis is true, the probability of detecting false significant discoveries in any of these sets is controlled), given that the tests used are exact. In detail, given the exactness of all tests on intervals, the procedure provides a control of the interval-wise error rate (IWER), i.e, for all $\alpha \in (0, 1)$,

$$\forall \mathcal{I} \subseteq D : H_0^{\mathcal{I}} \text{ true } \implies P(\exists t \in \mathcal{I} : \tilde{p}_D(t) \leq \alpha) \leq \alpha.$$

The above states that when the null hypothesis is true on an interval of the domain, the probability of wrongly rejecting any part of it is controlled. Abramowicz et al. (2018) proved that the IWT procedure using asymptotic exact tests are provided with asymptotic control of the IWER. The results are discussed in the context of functional-on-scalar linear models.

### 3.1.2 Multi-aspect IWT

In Pini et al. (2019), the IWT procedure is extended to a multi-aspect IWT procedure that is used to test for differences in (multivariate) functional data while jointly accounting for several aspects of the data. The aspects could, e.g., relate to different order of derivatives of the functional data or to different dimensions in the multivariate functional data. The multi-aspect IWT procedure, in case of rejection of a global null hypothesis, provides adjusted multi-aspect p-value functions that are used to identify the significant intervals and the corresponding aspects they occur at.

In Paper III, we consider an extension of the model in (2), where for each individual we consider multiple aspects of data simultaneously. Formally, we define,

$$X_{ij}(t) = \beta_{0j}(t) + \sum_{l=1}^{L} \beta_{lj}(t) x_{li} + \epsilon_{ij}(t),\ t \in D, \tag{7}$$

for individuals $i = 1, ..., n$ with aspects $j = 1, ..., J$. In the same way, as described in the previous subsection we may for each aspect $j$ compute an adjusted p-value function $\tilde{p}_{D^j}(t)$ that provides control of the IWER of aspect $j$, on the basis of a test statistic $T_j^{\mathcal{I}} = \int_{\mathcal{I}} T_j(t)dt$ corresponding to the $j$th aspect.

Now, let $\mathscr{K}$ be the family of all possible non-empty subsets of $\{1, ..., J\}$ ($2^J - 1$ in total). For any $\mathbf{k} \in \mathscr{K}$, let $T_{\mathbf{k}}^{\mathcal{I}} = \sum_{j \in \mathbf{k}} T_j^{\mathcal{I}}$, and let $\tilde{p}_{D^{\mathbf{k}}}$ denote the adjusted p-value corresponding to the test $T_{\mathbf{k}}^{\mathcal{I}}$. We construct the adjusted multi-aspect p-value functions $\check{p}_{D^j}(t)$, $j = 1, ..., J$, as follows,

$$\check{p}_{D^j}(t) = \max_{\{\mathbf{k} \in \mathscr{K} : j \in \mathbf{k}\}} \sup_{\{\mathcal{I} \subseteq D : t \in \mathcal{I}\}} \tilde{p}_{D^{\mathbf{k}}}^{\mathcal{I}}.$$

Pini et al. (2019) study the performance of the method based on such adjusted p-value functions for the two-sample inference. They provide exactness conditions on the individual tests, under which different types of the IWER control are provided along the multiple aspects of the functional data.

In Paper III, we combine the multi-aspect IWT procedure introduced in Pini et al. (2019) together with the functional-on scalar linear model introduced in Abramowicz et al. (2018). The resulting procedure retains the properties of the IWER (asymptotically).

### 3.1.3 Interval testing procedure

An alternative to the IWT procedure for performing local inference is the interval testing procedure (ITP), presented by Pini and Vantini (2016). The

main difference between ITP and IWT is that ITP uses basis expansions to represent the data, reducing its dimensionality.

Instead of performing tests on all the intervals of the domain, tests are made on all subsets of consecutive basis coefficients. For each basis coefficient $q$, $q = 1, ..., K$, an adjusted p-value $\lambda^{(q)}$ is computed as the maximum from the p-values obtained from testing all the consecutive subsets containing coefficient $q$, i.e, for all sets $\mathbf{q} \in \mathcal{S}$ where $\mathcal{S} = \{(i, i+1, ..., j) : i \leq j, \text{ and } i, j \in \{1, ..., K\}\}$ such that $q \in \mathbf{q}$. Given that for all $\mathbf{q} \in \mathcal{S}$ the tests used are exact, ITP is provided with an interval-wise control of the FWER, i.e., for all $\alpha \in (0, 1)$,

$$\forall \mathbf{q} \in \mathcal{S} : H_0^{\mathbf{q}} \text{ true } \implies P(\exists q \in \mathbf{q} : \lambda^{(q)} \leq \alpha) \leq \alpha.$$

A disadvantage of this method compared to IWT is that this method relies on initial choices of number and type of basis functions that may affect the conclusion. An advantage is, however, that selection of significant components (using the adjusted p-values) has potential to provide different insight depending on the basis used. A selection of, e.g., B-splines allows detection of significant intervals of the domain, as B-splines have local support. A selection of, e.g., Fourier basis would on the other hand allow detection of significant intervals in the frequency domain.

In Paper II, we apply the ITP in the framework of ANOVA (developed in Paper I) and MANOVA using functional permutation tests. In particular, we first applied the MANOVA-based ITP to investigate if there were any intervals with group differences on any of the dimensions in the multivariate functional data. To identify the dimensions with significant differences, we applied the ANOVA-based ITP on all of the dimensions independently and separately from each other. In this case, a Bonferroni correction was used to adjust for the multiplicity of analyses made. In case of a rejected null hypothesis, we also investigated which groups that differed by testing all pairs of groups in a Scheffé-like framework to account for the multiplicity, see Paper I for the details.

### 3.1.4 Practical aspects of the introduced procedures

Throughout this thesis we use functional permutation tests. Permutation tests for testing the equality of two (or more) group mean functions are exact and therefore applicable to the results in Pini and Vantini (2016) and Pini and Vantini (2017). For functional-on-scalar models where covariates are not only group indicators, permutation tests may be performed using estimated residuals based on (e.g., OLS) estimated $\boldsymbol{\beta}(\cdot)$ coefficients. In this case the permutation tests are only asymptotically exact. In the context of

the ANOVA- and MANOVA-based ITPs used in Paper I and II, we randomly permute the functional objects under the null hypothesis. In the context of the functional-on-scalar model used in Paper III, the test statistic is evaluated for a large number of randomly selected rearrangements (permutations) of the estimated residuals under the null hypothesis. The p-value ($p_D^{\mathcal{I}}$) of the permutation test is then obtained by calculating the proportion of permutations leading to at least as extreme value of the test statistic as the value of the test statistic applied on the original data. For the specific details on permutation tests, including a study of their properties, we refer to Pesarin and Salmaso (2010). For details on applying the permutation test in the context of functional linear models (2), see Appendix B in Abramowicz et al. (2018).

Another practical aspect of the tests performed is related to the infinite cardinality of the set of all intervals in the IWT procedure. In practical implementation, we discretise the domain into atoms and perform the tests on all possible intervals formed by the atoms. For example of such implementation for functional linear models, we refer to Abramowicz et al. (2018).

## 3.2 Clustering dependent functional data

The objective of clustering is to partition a set of observations into distinct groups (clusters) such that the similarity between the objects are high within groups and low in-between them. For independent functional objects, many clustering methods within the field of FDA have been developed, e.g., different variants of the k-means method (Abraham et al., 2003; Sangalli et al., 2010), hierarchical methods (Ferraty and Vieu, 2006; Ferreira and Hitchcock, 2009), and model based approaches (James and Sugar, 2003; Chiou and Li, 2007). Clustering methods that deals with spatially dependent functional data have also been proposed, see, e.g., Giraldo et al. (2012), Secchi et al. (2013) and Abramowicz et al. (2017).

In Paper IV, we assume that the underlying dependence between the functional objects is coupled to a latent structure that slowly varies over the domain and causes different frequency patterns of the curve types to occur for the different states of the latent variable. See Figure 2 for an illustrative one-dimensional example, where labels $c_1, \ldots, c_N$ correspond to latent signal and $l_1, \ldots, l_N$ are related to the functional object types observed in the data. Observe, that in contrary to latent signal labels which change slowly, the neighbouring observe functional object types are not necessarily alike and may varying substantially. An example of model that could be use for data generation is the hidden Markov model (Rabiner, 1989).

In Paper IV, we introduce a new clustering method, the double clustering

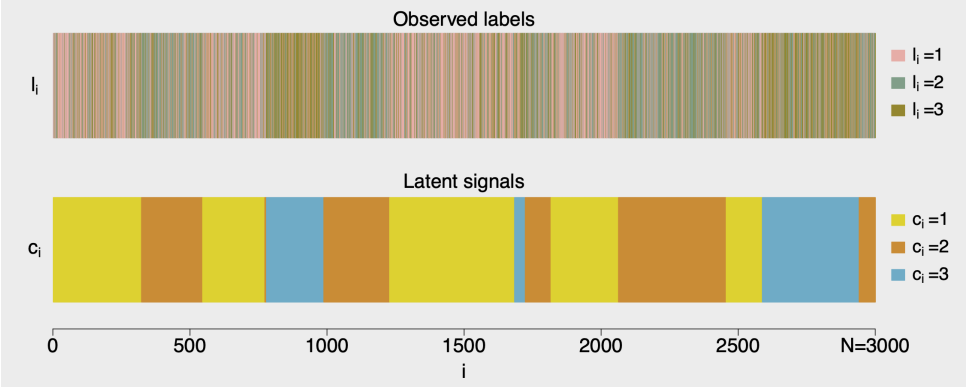


Figure 2: A simulated example illustrating the dynamics of the observed functional object type labels (top) and the true latent signal labels (bottom). (source: Paper V)

bagging Voronoi (DCBV) method, that aims to recover the slowly varying latent structure generating the observable different frequency pattern of the functional object types. It consists of two levels of clustering. At the first level, the functional data are clustered (and labelled) into groups of curves with similar shapes using, e.g., k-means. At the second level, the spatial dependency is taken into account by grouping relative frequency representatives of the labels obtained from the first level clustering, using a bagging Voronoi strategy. Further, we also propose a multiresolution approach which in combination with the DCBV method allows to search for stable clusters at different spatial resolutions, see Section 3.2.1 for more details.

In Paper V, we explore new methods to improve on the second level clustering in the DCBV method. We propose and compare several new clustering methods for grouping observed categorical data, such as labelled functional data, along a one-dimensional domain, aiming to recover the latent signal.

In this section, we present the main ideas behind the DCBV method, with focus on the second level clustering, as well as the multiresolution approach. The ideas behind the clustering methods introduced in Paper V are also presented.

### 3.2.1 The double clustering bagging Voronoi method

Assume that we have observed a sample of functional objects $X_{s_i}(t)$, $i = 1, ..., N$, at $N$ different locations, $s_1, ..., s_N \in D \subset \mathbb{R}^d$. The first step of the DCBV method is to project the functional objects onto the space of labels $l_{s_i} \in \{1, ..., k\}$, clustering the different curves into different types with

respect to their shapes. The second level of clustering is based on a bagging Voronoi strategy introduced in Secchi et al. (2013), consisting of two phases: a bootstrap and an aggregation phase. The whole procedure of the DCBV method is illustrated in Figure 3 in the case of a one-dimensional domain, and also described in more detail below.

**Bootstrap phase** In the bootstrap phase, for $b = 1, ..., B$, we perform the following three steps on each bootstrap replicate $b$:

(i) Generate a random Voronoi tessellation of the considered domain (a partition of random intervals in the case of a one-dimensional domain) into $n$ cells $\{V_j^b\}_{j=1}^n$.

(ii) Compute the local representative for each element of the tessellation. For $j = 1, \ldots, n$, we let the local representative $\mathbf{x}_j^b$, corresponding to the $j$-th cell $V_j^b$ of the $b$-th partition, be the $d$-dimensional relative frequency vector:

$$\mathbf{x}_j^b = \left[ \frac{|\{s_i \in D : s_i \in V_j^b \text{ and } l_{s_i} = 1\}|}{|V_j^b|}, \cdots, \frac{|\{s_i \in D : s_i \in V_j^b \text{ and } l_{s_i} = k\}|}{|V_j^b|} \right],$$

where $|A|$ denotes the number of elements in set $A$.

(iii) Cluster the representatives $\{\mathbf{x}_1^b, \ldots, \mathbf{x}_n^b\}$ into $K$ clusters (i.e., the number of states in the latent signal) using a suitable clustering method. Assign the cluster label obtained for each representative $\mathbf{x}_i^b$, to all sites $s_i$ belonging to the same tessellation element as its representative. Save the cluster labels for each of the $N$ sites.

**Aggregation phase** After the bootstrap phase, we are provided with a frequency distribution of cluster labels for each site. The aggregation phase uses this information to determine the final clustering. First, a matching procedure is applied in order to match the cluster labels across the bootstrap replicates (see, e.g., Stephens, 2000). Then, for each site $s_i$, a relative frequency vector $\{\pi_i^1, ..., \pi_i^K\}$ of cluster assignments to each of the $K$ clusters over the bootstrap replicates is computed. The final classification $\hat{c}_i$ of each site $s_i$, is then determined by the label that corresponds to the the mode of the frequency vector, i.e.,

$$\hat{c}_i = \arg \max_{q=1,...,K} \pi_i^q.$$

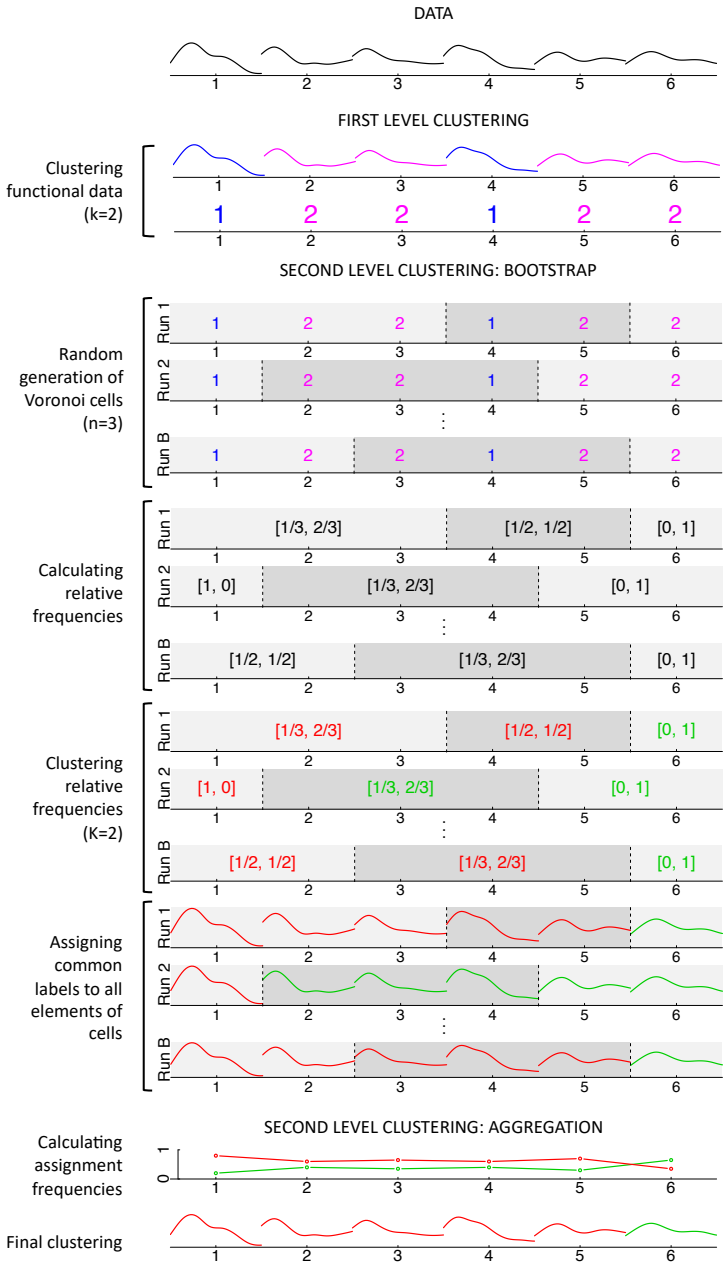


Figure 3: A graphical scheme describing the steps of the DCBV method with two clusters at both clustering levels (k=2, K=2) and n=3. (source: Paper IV)

**Entropy and parameter selection** In the DCBV method, it is needed to decide upon three parameters for the second level clustering, namely, $B$, $n$ and $K$. The parameter $B$ is the number of bootstrap replicates, normally given a sufficiently large value such that the solution of the problem obtains a reasonable level of accuracy. The choice of the parameter $n$, i.e., the number of cells in the tessellation, is related to the dependency structure of the latent field and its frequency distributions. Inspired by Secchi et al. (2013) and Abramowicz et al. (2017), the optimal value of $n$ is determined by measuring the stability of the final classification for each site $s_i$ using the normalised Shannon entropy $\eta(i)$ and averaging it over all sites to obtain a measure of the overall stability of the clustering for a given $n$,

$$\bar{\eta} = \frac{1}{N}\sum_{i=1}^{N}\eta(i) := \frac{1}{N}\sum_{i=1}^{N} -\frac{1}{\log(K)}\sum_{q=1}^{K}\pi_i^q \log(\pi_i^q).$$

A low value of $\eta(i)$ (zero as lowest) indicates a stable and precise clustering of the site $s_i$, while a high value (one as highest) is attained if the distribution of cluster assignments are close to uniform. We choose the value of $n$ to be the one that minimises $\bar{\eta}$, for a given $K$, in order to minimise the average uncertainty of the clustering. The average normalised entropy can also be used to decide the value of the third parameter $K$, i.e., the number of latent clusters, along with an analysis of the obtained cluster structures, see Paper IV or Abramowicz et al. (2017) for a discussion.

**Multiresolution** In some situations different underlying latent structures act simultaneously over a region at different resolutions, which in turn can have affects on the frequency pattern of the functional object types. In Paper IV, we introduce a multiresolution approach coupled to the DCBV method that searches for several latent structures at different scales. The approach is based on the parameter $n$ as it is connected to the strength of the underlying dependence. The average normalised entropy as function of $n$ is used to search, for each $K$, minima that could lead to discoveries of distinct underlying latent patterns. In Paper IV, the method is applied to lake sediment data with the aim of identifying climate, being described as frequencies of different weather types (yearly sediment layers), at different (temporal) scales. We were able to detect two climate signals varying at two different scales as seen in Figure 4.

### 3.2.2 Extensions of the bagging Voronoi strategy

The bagging Voronoi strategy has been successfully used in the context of clustering dependent functional data in Paper IV, but also in earlier works,

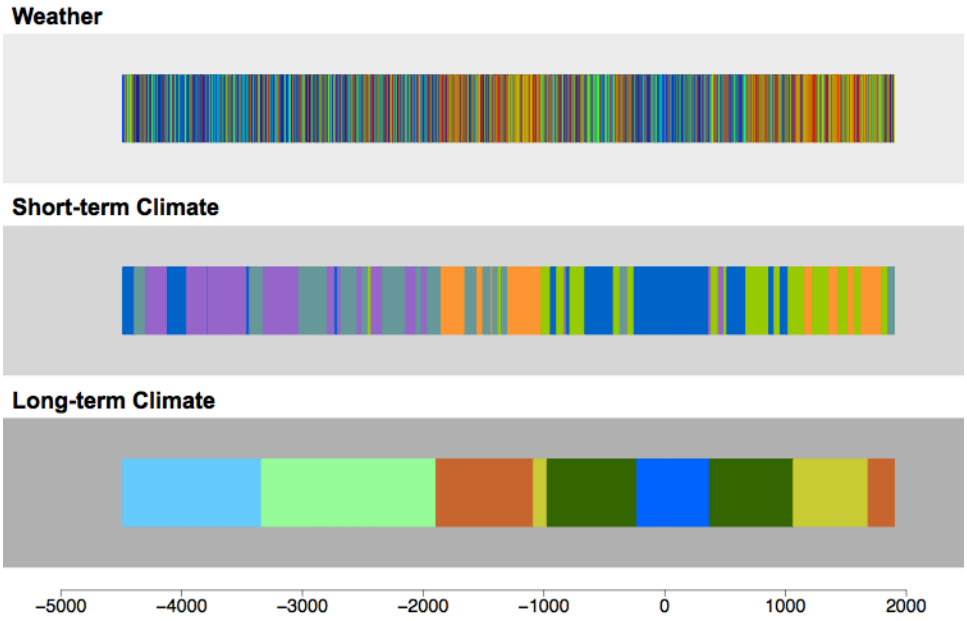


Figure 4: The temporal dynamics of weather (yearly sediment layers), short-term climate and long-term climate. The colours correspond to different types of 'weather' and 'climates'. (source: Paper IV)

e.g., in Secchi et al. (2013) and Abramowicz et al. (2017). In the two latter, the Voronoi cells are represented by functional objects rather than frequencies as in the DCBV case.

In general, the principles of the bagging strategy are broad and can be applied to obtain a spectrum of new clustering methods suitable for different types of underlying data generating mechanisms. In particular, by considering different methods for generating the tessellations, forming the tessellation cell representatives, and grouping the representatives, additional clustering methods for dependent categorical data can be constructed. In Paper V, we introduce and study some variants for the case of temporal dependent categorical data generated by a latent signal. For example, a new way of forming the tessellation is introduced, leading to a different distribution of individual cell lengths than the one obtained with a Voronoi tessellation. We also introduce a two step procedure which involves the normalised entropy function in the construction of the tessellation. Further, we consider clustering the representatives in two ways, by a classical k-means method and its weighted version, taking into account the length of the individual cells.

## 3.3 Prediction of functional data

In the third part of this thesis we study the problem of predicting spatial dependent functional data. In the presence of dependence, kriging is a suitable and probably the most known method, and has a long history of being used to predict objects at new locations based on observed values on other locations, especially for objects that are real- or vector valued, see e.g., Cressie (2015), Cressie and Wikle (2015) and references therein. A kriging predictor is defined as the best linear unbiased predictor (BLUP) minimising the mean squared prediction error (MSPE). The kriging predictor is thus a weighted sum of the observed data, where the objects closer to the location of prediction is typically given a higher weight.

In Paper VI, two kriging approaches to predict spatial functional random processes are presented and compared, namely functional kriging and spatio-temporal kriging. A spatial functional random process $\{X_s(t), s \in D \subset \mathbb{R}^d,\ t \in T \subset \mathbb{R}\}$ is defined as a process with stochastic functional objects $X_s(t)$, $t \in T \subset \mathbb{R}$ at any spatial location $s \in D \subset \mathbb{R}^d$. A spatial functional random process can also be viewed as a spatio-temporal (Sp.T.) random process $\{Z(s,t) = X_s(t),\ (s,t) \in D \times T\}$. To predict a curve at a new location $s_0$, given a set of curves observed at $n$ different spatial locations (see Figure 5), a functional kriging approach would predict it by a linear combination of the observed curves. As an alternative, it can also be predicted with a Sp.T. kriging approach, i.e., for a dense grid of values over $T$, it can be predicted based on linear combinations of a time-grid of values over the observed curves, see, e.g., Cressie and Wikle (2015) and Montero et al. (2015).

An important issue raised in Delicado et al. (2010), is the question which of these approaches (functional kriging or Sp.T. kriging) is to prefer for an analysis of a particular dataset. In Paper VI, we conduct a simulation study to compare their prediction performance and computational times. We also formulate conditions under which we show that two of the functional kriging methods coincide. In this section, we present an introduction to the two approaches.

### 3.3.1 Functional kriging

Since the pioneering work by Goulard and Voltz (1993), the development of functional kriging predictors has received increased interest. The case of ordinary functional kriging, i.e., when the mean of the random function is assumed to be independent of the spatial location has, e.g., been considered by Delicado et al. (2010), Giraldo et al. (2010, 2011) and Nerini et al. (2010). The case of universal functional kriging, i.e., when the mean function may

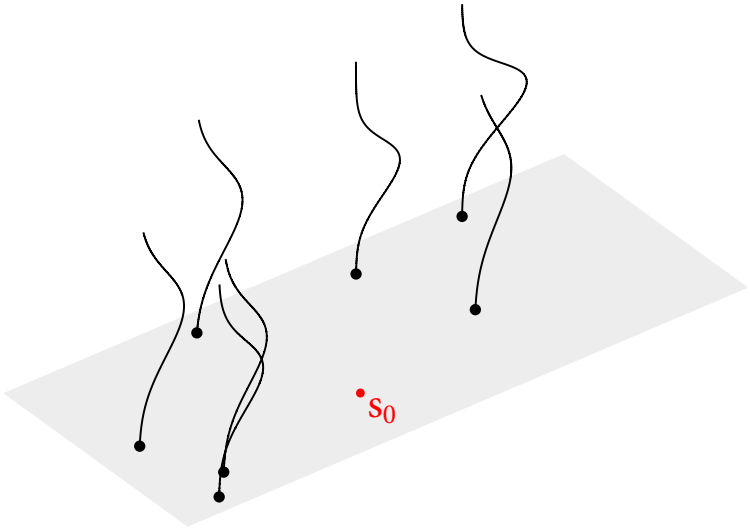


Figure 5: An example of observed curves and a location $s_0$ where it is of interest to predict a curve.

also depend on the location, has more recently been considered by Caballero et al. (2013), Menafoglio et al. (2013), Ignaccolo et al. (2014), and Reyes et al. (2015). In this thesis, the focus is on the ordinary kriging case. In particular, we consider kriging predictiors for second-order isotropic and stationary spatial functional random processes $\{X_s(t), s \in D \subset \mathbb{R}^d,\ t \in T \subset \mathbb{R}\}$, i.e., processes satisfying,

$$\text{(i)}\ \ E[X_s(t)] = m(t) \text{ and } Var[X_s(t)] = \sigma^2(t)\ \ \forall s \in D \text{ and } \forall t \in T, \tag{8}$$

$$\text{(ii)}\ \ Cov[X_s(r), X_v(t)] = C(\|s - v\|, r, t)\ \ \forall s, v \in D \text{ and } \forall r, t \in T,$$

where $\|\cdot\|$ denotes the (Euclidean) distance measure.

Goulard and Voltz (1993) proposed one of the first functional kriging predictors of a curve $X_{s_0}(t)$, $t \in T$, given the observed $n$ curves $X_{s_i}(t)$, $t \in T$, $i = 1, ..., n$, of the form

$$\hat{X}_{s_0}(t) = \sum_{i=1}^{n} \lambda_i X_{s_i}(t), \quad t \in T,\ \lambda_1, ..., \lambda_n \in \mathbb{R}.$$

It was later studied by Giraldo et al. (2007, 2011) and there named ordinary kriging for functional data (OKFD). In order to find the BLUP, the optimal kriging weights $\lambda_1, ..., \lambda_n$, are given by the solution of the following

optimisation problem, minimising the MSPE,

$$\min_{\lambda_1,\ldots,\lambda_n} E\Big[\int_T \{\hat{X}_{s_0}(t) - X_{s_0}(t)\}^2\, dt\Big], \tag{9}$$

subject to the unbiasedness condition of the predictor, $\sum_{i=1}^n \lambda_i = 1$. In Giraldo et al. (2011), it is shown that the optimal kriging weights are functions of the dependence structure, in terms of the (isotropic) trace-variogram,

$$\gamma(\|s-v\|) = \frac{1}{2}E\Big[\int_T \{X_s(t) - X_v(t)\}^2\, dt\Big], \;\; \forall s, v \in D.$$

In practice, the trace-variogram is not known and thus has to be estimated. A common way to estimate it is by fitting a parametric variogram model $\gamma(h \mid \theta)$, e.g., the spherical, Gaussian, exponential, stable or Matérn variogram model to the estimated empirical trace-variogram

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{i,j \in N(h)} \int_T \{X_{s_i}(t) - X_{s_j}(t)\}^2 dt,$$

where $N(h) = \{(s_i, s_j) : \|s_i - s_j\| \in (h-\epsilon, h+\epsilon)\}$, for some $\epsilon > 0$, and $|N(h)|$ denotes the number of distinct elements in $N(h)$, by a least squares method, cf. Cressie (2015). Typically, the observed random functions, $X_{s_i}(t)$, $i = 1, ..., n$, are observed only at a finite raster of values and Goulard and Voltz (1993) suggested to fit parametric models to $X_{s_i}(t)$'s whereas Giraldo et al. (2007, 2011) represented them by a linear combination of known basis functions (cf. Section 2.1).

Giraldo et al. (2008, 2010) proposed another kriging predictor, the pointwise functional kriging predictor (PWFK),

$$\hat{X}_{s_0}(t) = \sum_{i=1}^n \lambda_i(t) X_{s_i}(t), \;\; t \in T,$$

which allows the $\lambda_i$'s to depend on $t$. The infinite dimensional problem of finding the optimal $\lambda_i(t)$ functions that minimises the MSPE in (9) subject to the unbiasedness constraint of the predictor, $\sum_{i=1}^n \lambda_i(t) = 1$, for all $t \in T$, is there solved by letting the $\lambda_i(t)$'s and the $X_{s_i}(t)$'s be represented as linear combinations of known basis functions. Consequently the optimisation problem is reduced into a solvable multivariate geostatistics problem. A full derivation of the solution of the PWFK optimisation problem, when the same basis system is used for both the $X_{s_i}(t)$ and the $\lambda_i(t)$ functions, is given in Giraldo et al. (2010), while a solution for general (possibly different)

basis systems is given in Paper VI. However, in both cases, the optimal $\lambda_i(t)$ functions depend on the covariances between the basis coefficients of the $X_{s_i}(t)$ functions, which in practice need to be estimated. Giraldo et al. (2010) suggest to estimate these covariances via a linear model of coregionalisation (Goulard and Voltz, 1992).

In Giraldo (2009, 2014), and independently in Nerini et al. (2010), a third kriging alternative was proposed, the functional kriging total model (FKTM),

$$\hat{X}_{s_0}(t) = \sum_{i=1}^{n} \int_T \lambda_i(t,v) X_{s_i}(v) dv, \quad t \in T,$$

where the weight $\lambda_i(t,v)$ determines how much influence the *i-th* observed function at time $v$ should have on the predicted function at time $t$. The $\lambda_i(t,v)$'s are chosen to minimise (9) subject to the unbiasedness constraint. A solution based on representing both $\lambda_i(t,v)$'s and $X_{s_i}(t)$'s through a linear combinations of known basis functions, is given in Giraldo (2014). The solution of determining the optimal basis coefficients for the $\lambda_i(t,v)$'s involves functions of the covariances between the basis coefficients of the $X_{s_i}(t)$ functions, which in practice need to be estimated, see Giraldo (2014) for the details.

In some cases, the predictors of the introduced functional kriging methods can be shown to coincide. Menafoglio and Petris (2016) showed that if the realisations of $X_s(t)$ belong to the Hilbert space of square integrable functions on $T$, and the functional second-order stationary random process is Gaussian, the kriging weights of FKTM and OKFD agree a.s. for any orthonormal base. In Paper VI, we formulate the conditions under which we prove that OKFD and PWFK coincide. The class of processes satisfying the conditions includes stationary spatial functional random processes as well as some non-stationary processes.

### 3.3.2 Spatio-temporal kriging

Since a spatial functional process can also be viewed as a Sp.T. random process $\{Z(s,t) = X_s(t), (s,t) \in D \times T\}$, it could also be predicted by Sp.T. kriging methods. A Sp.T. process is said to be second-order stationary and spatially isotropic if

(i) $E[Z(s,t)] = m$ and $Var[Z(s,t)] = \sigma_Z^2$ $\forall s \in D$ and $\forall t \in T$,

(ii) $Cov[Z(s,r), Z(v,t)] = C_Z(\|s-v\|, \mid r-t \mid)$ $\forall s, v \in D$ and $\forall r, t \in T$.

Note that the class of stationary Sp.T processes is a subset of the class of stationary functional random processes.

For a set of observed values $\{Z(s_1, t_{11}), \ldots, Z(s_1, t_{1m_1}), \ldots, Z(s_n, t_{n1}), \ldots, Z(s_n, t_{nm_n})\}^\intercal \in \mathbb{R}^N$, $N = \sum_{i=1}^n m_i$, the Sp.T. ordinary kriging predictor, at location $s_0$ and time point $t \in T$, is given by,

$$\hat{Z}(s_0, t) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \lambda_{ij}^t Z(s_i, t_{ij}),$$

see, e.g., Cressie and Wikle (2015). In order to find the BLUP, the optimal kriging weights $\lambda_{ij}^t$'s are here determined by the solution of the following optimisation problem, minimising the MSPE,

$$\min_{\lambda_{11}^t, \ldots, \lambda_{1m_1}^t, \ldots, \lambda_{n1}^t, \ldots, \lambda_{nm_n}^t} E[(\hat{Z}(s_0, t) - Z(s_0, t))^2],$$

subject to the unbiasedness condition, $\sum_{i=1}^n \sum_{j=1}^{m_i} \lambda_{ij}^t = 1$. As in the functional case, the solution results in that the optimal kriging weights are functions of the dependence structure, which may be described by the (spatially isotropic) Sp.T. semivariogram,

$$\gamma_Z(\|s - v\|, |r - t|) = \frac{1}{2} E[(Z(s, r) - Z(v, t))^2], \;\; s, v \in D \text{ and } r, t \in T.$$

In practise, the Sp.T. semivariogram is not known and typical it needs to be estimated by fitting a parametric variogram model to the empirical variogram,

$$\hat{\gamma}_Z(h, u) = \frac{1}{2|N(h, u)|} \sum_{(i,j,k,l) \in N(h,u)} \{Z(s_i, t_{ik}) - Z(s_j, t_{jl})\}^2,$$

where $N(h, u) = \{(s_i, t_{ik}), (s_j, t_{jl}) : \|s_i - s_j\| \in (h - \epsilon, h + \epsilon), \text{ and } |t_{ik} - t_{jl}| \in (u - \delta, u + \delta)\}$, for some $\epsilon, \delta > 0$.

It is often challenging to identify and estimate a suitable Sp.T. variogram model, especially when no prior knowledge of the dependence structure is known. The functional kringing approach is easier in this sense, since only the spatial dimension needs to be considered for the (trace-)variogram estimate.

# 4 Summary of papers

In this section, we summarise and highlight the key results of the six papers included in the thesis. The papers can be divided into three parts. In Papers I-III, we develop and apply methods for testing human movement strategies after an ACL injury. In Papers IV and V, we present novel clustering methods suitable for dependent functional data such as the lake sediment data. In Paper VI, we compare and evaluate different kriging methods for prediction of dependent functional data.

## Paper I: An inferential framework for domain selection in functional ANOVA

We present an ANOVA testing procedure for testing functional data, including a Scheffé-like framework for pairwise group comparisons. The testing procedure relies on the Interval Testing Procedure (ITP) that builds on testing coefficients of basis expansions of the functional data. In case of significant group differences, the test procedure identifies which of the groups that significantly differ, and also the parts of the domain they do so, while controlling the type I error. More specifically, the procedure is provided with an interval-wise error control that controls the probability of falsely rejecting any part of the (coefficient) interval in at least one of the two-group comparisons, given that all populations have the same distribution in any of the intervals.

## Paper II: One-leg hop kinematics 20 years following anterior cruciate ligament rupture: Data revisited using functional data analysis

The ITP-based ANOVA testing procedure introduced in Paper I is applied to investigate and compare the one-leg hops of individuals long after ACL rupture. In particular, we investigate if differences are present in knee-kinematic movements between three groups: a surgical treated group, a group treated with physiotherapy only and a control group of knee-healthy individuals. The results shows that the movements of the physiotherapy group differ compared to the movements in the control group; significant differences were found in the take-off (at 1% level) and the landing (at 5% level) phases. In addition, significant between-leg differences for individuals are also present in the physiotherapy group during the flight (at 5% level) and landing (at 1% level) phases.

## Paper III: Analysis choices provide contrasting conclusions when evaluating jump performance: A multi-aspect inferential method applied to kinematic curves from the one-leg vertical hop in knee-injured and asymptomatic persons

A novel multi-aspect statistical methodology in the context of functional-on-scalar linear models is proposed to address how common choices made prior to the analysis affect the clinical conclusions in the field of biomechanics.

The introduced method allows us to test for group differences in multivariate functional data while accounting for covariates. In presence of significant between-group differences, the inferential procedure selects the significant domain regions and the specific dimensions where the groups differ. The method is provided with a so called interval-wise control of the type I error, under conditions on the individual tests used. We apply the method on kinematics curves obtained from a vertical one-leg hop performed by knee-injured and asymptomatic individuals. The practical aim is to illustrate how choices made prior to the analysis may affect the clinical conclusions. The choices we consider are: 1) based on repeated hops, how should we represent an individual: by an average, maximal or a variability curve? 2) based on kinematics from both legs, which one should be included in the analysis? 3) based on kinematics from several joints, which ones to include in the analysis? The paper provides a full exploration of the results obtained from the different choices. In summary, many significant group differences were found when individuals were represented with a mean or by a max curve, while fewer were found when represented with a variability curve. When the injured leg was compared to the dominant leg in the asymptomatic persons, we found group differences for the ankle and trunk kinematics. If instead the injured leg was compared to the non-dominant leg in the controls, we found differences in the knee and hip joint kinematics.

## Paper IV: Multiresolution clustering of dependent functional data with application to climate reconstruction

We introduce a new non-parametric clustering method for dependent functional data, the double clustering bagging Voronoi (DCBV) method. The frequency pattern of the types of functions are assumed to be generated by a slowly varying latent signal(s). The aim of the DCBV method is to recover the hidden latent signal(s). The method consists of two levels of clustering. In the first level, the functional data are clustered into different curve types, with respect to their shapes. In the second level, the dependency is taken into account through a bagging Voronoi procedure. In particular, neighbourhoods are randomly generated (Voronoi tessellation) in which local representatives are formed (on the basis of the first level cluster types) and then clustered. This is repeated many times and the result of the second level clustering are then bagged together to form a final clustering. Further, combined with the DCBV method, we propose a multiresolution approach that allows to identify stable clusters at different scales. A simulated dataset is used to illustrate the method's efficiency in detecting stable clusters at

different resolutions. The method is also applied to varved lake sediment data with the aim of reconstructing the past climate in the northern part of Sweden over the past 6000 years, at different time resolutions. We identified two different resolutions, with stable climate clusters characterised by different frequencies of sediment (weather) types. The mean sojourn times for the clusters on the two resolutions were estimated to be approximately 100 and 700 years, respectively.

## Paper V: Nonparametric clustering methods to identify latent structures from a sequence of dependent categorical data

We construct and study a number of new methods, within the framework of bagging strategies, to cluster dependent categorical data. The frequencies of the categorical data are assumed to depend on a latent signal which slowly varies over time. The aim of the methods are to recover the sequence of states of the latent signal on the basis of the observed categorical data. The methods are constructed using different strategies of generating random tessellations, and clustering tessellation cell representatives. Specifically, three ways of constructing a tessellation are considered: a centroid-based (Voronoi) tessellation, an edge-based tessellation and an adaptive-edge-based tessellation. We also consider clustering the cell representatives in two ways, using a classical k-means method and a weighted k-means method. Thus, the combination of generating a tessellation and clustering cell representatives leads to six methods. Further, a simulation study is performed in order to compare the methods suitability in identifying the latent signal for several different dependency structures. The study shows that the methods based on adaptive-edge tessellations in general works best. We also use the normalised entropy function as a tool for combining methods into composite methods, showing that they have the potential to further improve the clustering results.

## Paper VI: Prediction of spatial functional random processes: comparing functional and spatio-temporal kriging approaches

We compare two kriging approaches to predict spatial functional random processes, functional kriging and spatio-temporal (Sp.T.) kriging. We formulate conditions under which we show that the two functional kriging methods ordinary kriging for functional data (OKFD) and pointwise functional kriging (PWFK) coincide. The performance of the two approaches is mainly

assessed via a simulation study but also on a real dataset. The comparisons are restricted to Sp.T. kriging versus OKFD as the two functional kriging methods PWFK and functional kriging total model (FKTM) coincide with OKFD in several situations. The study shows that the two approaches are rather equal for stationary Sp.T. processes while OKFD is to prefer for stationary functional, but non-stationary Sp.T. processes. Furthermore, in all simulated cases, the computational time was substantially shorter for OKFD compared to Sp.T. kriging.

# 5 Concluding remarks and future research

In Papers I-III, the aim of the methods used was to infer whether group differences in curve data existed or not, and in case of existence to find where in the domain they did so. The multi-aspect method introduced in Paper III allow us to compare curve data from multiple aspects simultaneously, while controlling the type I error of falsely rejecting the null hypothesis on any of the multiple aspects. However, the p-value function that adjusts for several aspects is typically larger than the ones that solely adjusts for the domain. Thus, a limitation of using the multi-aspect analysis is that the power of the test decreases when more aspects are included in the analysis. It would be of interest to study and further develop suitable and more powerful inferential methods for complex assessments such like the vertical one-leg hop analysed in Paper III.

In Papers IV and V, non-parametric clustering methods based on bagging strategies are introduced and studied. They all show good potential in recovery of the hidden structures that give rise to different frequency patterns of observable objects that can be categorised. In Paper V, we noticed that the methods based on an adaptive tessellation in general improve the results. In our work, the adaptive tessellation was created by non-homogeneous sampling of cell edges on a one-dimensional domain. Possible future work is to develop and study the adaptivity for Voronoi tessellations, which would correspond to non-uniform sampling of the cell nuclei instead. That approach could be then generalised naturally to higher dimensions, which is not the straightforward case for the edge tessellation methods.

The purpose of the study in Paper VI was to compare and shed light over the functional and Sp.T. kriging methods' strengths and weaknesses. In general, we found that the performances of the two approaches were quite equal for stationary Sp.T. processes while functional kriging was to prefer in cases of non-stationary Sp.T. processes. An advantage with Sp.T. kriging is the fact that it may also be used in cases where the functional objects

only have been partially recorded over different subintervals of the domain. Moreover, in general Sp.T. kriging methods demands more work, e.g., in specifying and estimating a suitable spatio-temporal variogram model, and a mean function model, making it more vulnerable. An advantage with the functional approach is that it also can incorporate other embeddings for the data, i.e., other geometries than the Euclidean one. In Paper VI, stationary functional processes were considered. A direct extension of the work would be to develop and compare functional and Sp.T. kriging methods that deals with non-stationary functional processes.

# References


Abraham, C., Cornillon, P.-A., Matzner-Løber, E., and Molinari, N. (2003). Unsupervised curve clustering using B-splines. *Scandinavian Journal of Statistics*, 30(3):581–595.

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

Abramowicz, K., Arnqvist, P., Secchi, P., Sjöstedt de Luna, S., Vantini, S., and Vitelli, V. (2017). Clustering misaligned dependent curves applied to varved lake sediment for climate reconstruction. *Stochastic Environmental Research and Risk Assessment*, 31(1):71–85.

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

Arnqvist, P., Bigler, C., Renberg, I., and Sjöstedt de Luna, S. (2016). Functional clustering of varved lake sediment to reconstruct past seasonal climate. *Environmental and Ecological Statistics*, 23(4):513–529.

Arnqvist, P. and Sjöstedt de Luna, S. (2019). Model based functional clustering of varved lake sediments. *arXiv e-prints*.

Caballero, W., Giraldo, R., and Mateu, J. (2013). A universal kriging approach for spatial functional data. *Stochastic Environmental Research and Risk Assessment*, 27(7):1553–1563.

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

Chiou, J.-M. and Li, P.-L. (2007). Functional clustering and identifying substructures of longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4):679–699.

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

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.

Cressie, N. and Wikle, C. K. (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.

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

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge university press.

Delicado, P., Giraldo, R., Comas, C., and Mateu, J. (2010). Statistics for spatial functional data: some recent contributions. *Environmetrics: The official journal of the International Environmetrics Society*, 21(3-4):224–239.

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

Ferreira, L. and Hitchcock, D. B. (2009). A comparison of hierarchical methods for clustering functional data. *Communications in Statistics-Simulation and Computation*, 38(9):1925–1949.

Giraldo, R. (2009). *Geostatistical analysis of functional data*. PhD thesis, Ph. D. thesis, Universitat Politecnica de Catalunya.

Giraldo, R. (2014). Cokriging based on curves, prediction and estimation of the prediction variance. *InterStat*, 2:1–30.

Giraldo, R., Delicado, P., and Mateu, J. (2007). Geostatistics for functional data: An ordinary kriging approach. Technical report, Universitat Politècnica da Catalunya, http://hdl.handle.net/2117/1099.

Giraldo, R., Delicado, P., and Mateu, J. (2008). Continuous time-varying kriging for spatial prediction of functional data: an environmental application. Technical report, Universitat Politècnica da Catalunya, http://hdl.handle.net/2117/2167.

Giraldo, R., Delicado, P., and Mateu, J. (2010). Continuous time-varying kriging for spatial prediction of functional data: An environmental application. *Journal of Agricultural, Biological, and Environmental Statistics*, 15(1):66–82.

Giraldo, R., Delicado, P., and Mateu, J. (2011). Ordinary kriging for function-valued spatial data. *Environmental and Ecological Statistics*, 18(3):411–426.

Giraldo, R., Delicado, P., and Mateu, J. (2012). Hierarchical clustering of spatially correlated functional data. *Statistica Neerlandica*, 66(4):403–421.

Godwin, A., Takahara, G., Agnew, M., and Stevenson, J. (2010). Functional data analysis as a means of evaluating kinematic and kinetic waveforms. *Theoretical Issues in Ergonomics Science*, 11(6):489–503.

Goulard, M. and Voltz, M. (1992). Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, 24(3):269–286.

Goulard, M. and Voltz, M. (1993). Geostatistical interpolation of curves: a case study in soil science. In *Geostatistics Tróia'92*, pages 805–816. Springer.

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

Hsing, T. and Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*, volume 997. John Wiley & Sons.

Ignaccolo, R., Mateu, J., and Giraldo, R. (2014). Kriging with external drift for functional data for air quality monitoring. *Stochastic Environmental Research and Risk Assessment*, 28(5):1171–1186.

James, G. M. and Sugar, C. A. (2003). Clustering for sparsely sampled functional data. *Journal of the American Statistical Association*, 98(462):397–408.

Kneip, A. and Gasser, T. (1992). Statistical tools to analyze data representing a sample of curves. *The Annals of Statistics*, pages 1266–1305.

Lenhoff, M. W., Santner, T. J., Otis, J. C., Peterson, M. G., Williams, B. J., and Backus, S. I. (1999). Bootstrap prediction and confidence bands: a superior statistical method for analysis of gait data. *Gait & Posture*, 9(1):10–17.

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., Ramsay, J. O., Sangalli, L. M., and Srivastava, A. (2015). Functional data analysis of amplitude and phase variation. *Statistical Science*, pages 468–484.

Menafoglio, A. and Petris, G. (2016). Kriging for hilbert-space valued random fields: The operatorial point of view. *Journal of Multivariate Analysis*, 146:84–94.

Menafoglio, A., Secchi, P., Dalla Rosa, M., et al. (2013). A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space. *Electronic Journal of Statistics*, 7:2209–2240.

Montero, J.-M., Fernández-Avilés, G., and Mateu, J. (2015). *Spatial and spatio-temporal geostatistical modeling and kriging*. John Wiley & Sons.

Nerini, D., Monestiez, P., and Manté, C. (2010). Cokriging for spatial functional data. *Journal of Multivariate Analysis*, 101(2):409–418.

Pataky, T. C., Robinson, M. A., and Vanrenterghem, J. (2013). Vector field statistical analysis of kinematic and force trajectories. *Journal of Biomechanics*, 46(14):2394–2401.

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

Petterson, G. (1999). *Image analysis, varved lake sediments and climate reconstruction*. PhD thesis, Umeå University.

Pini, A., Spreafico, L., Vantini, S., and Vietti, A. (2019). Multi-aspect local inference for functional data: Analysis of ultrasound tongue profiles. *Journal of Multivariate Analysis*, 170:162–185.

Pini, A. and Vantini, S. (2016). The interval testing procedure: a general framework for inference in functional data analysis. *Biometrics*, 72(3):835–845.

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

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

Ramsay, J. (1982). When the data are functions. *Psychometrika*, 47(4):379–396.

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

Reiss, P. T., Huang, L., and Mennes, M. (2010). Fast function-on-scalar regression with penalized basis expansions. *The International Journal of Biostatistics*, 6(1).

Reyes, A., Giraldo, R., and Mateu, J. (2015). Residual kriging for functional spatial prediction of salinity curves. *Communications in Statistics-Theory and Methods*, 44(4):798–809.

Ryan, W., Harrison, A., and Hayes, K. (2006). Functional data analysis of knee joint kinematics in the vertical jump. *Sports Biomechanics*, 5(1):121–138.

Sangalli, L. M., Secchi, P., Vantini, S., and Vitelli, V. (2010). K-mean alignment for curve clustering. *Computational Statistics & Data Analysis*, 54(5):1219–1233.

Secchi, P., Vantini, S., and Vitelli, V. (2013). Bagging Voronoi classifiers for clustering spatial functional data. *International Journal of Applied Earth Observation and Geoinformation*, 22:53–64.

Staicu, A.-M., Li, Y., Crainiceanu, C. M., and Ruppert, D. (2014). Likelihood ratio tests for dependent data with applications to longitudinal and functional data analysis. *Scandinavian Journal of Statistics*, 41(4):932–949.

Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):795–809.

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

Wahba, G. (1990). *Spline models for observational data*. SIAM.

Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, 3:257–295.

Zhang, J.-T. and Liang, X. (2014). One-way ANOVA for functional data via globalizing the pointwise F-test. *Scandinavian Journal of Statistics*, 41(1):51–71.