POLITECNICO DI MILANO

Department of Mathematics
Doctoral Program in Mathematical Models and Methods in Engineering

Covariance operators as object data: statistical methods and applications

Doctoral Dissertation of:
Davide Pigoli

Supervisor:
Prof. Piercesare Secchi

The Chair of the Doctoral Program:
Prof. Roberto Lucchetti

Year 2013 - XXV cycle

Abstract

In recent years, new technologies have provided data more complex than numbers or vectors, such as high dimensional arrays, curves, shapes, diffusion tensors... These kinds of complex data can be analysed in the framework of Object Oriented Data Analysis. This work deals with a particularly interesting example of complex data: those belonging to a Riemannian manifold. These data are particularly interesting both from a mathematical and from a practical point of view. In particular, we focus on the case of covariance operators.

First, a framework is developed for the analysis of covariance operators of functional random processes, where the covariance operator itself is the object of interest. Distances for comparing positive definite covariance matrices are either extended or shown to be inapplicable for functional data. In particular, an infinite dimensional analogue of the Procrustes size and shape distance is developed. The proposed distances are used to address important inferential problems, namely, the point estimation of covariance operators and the comparison of covariance operators between two population of curves. These techniques are applied to two problems where inference concerning the covariance is of interest. Firstly, in data arising from a study into cerebral aneurysms, it is necessary to investigate the covariance structures of radius and curvature curves among different groups of patients. Secondly, in a philological study of cross-linguistic dependence, the use of covariance operators has been suggested as a way to incorporate quantitative phonetic information. It is shown that distances between languages derived from phonetic covariance functions can provide insight into relationships between the Romance languages.

A second contribution lies in the introduction of spatial dependence among Riemannian data. We consider both the modeling of the dependence on the manifold, generalizing the definition of covariance in linear spaces through the expected values of square distances, and the possibility to approximate non Euclidean data in the appropriate tangent space, where traditional statistical techniques can be used. First, the Riemannian semivariogram of a field of covariance matrices is defined. Then, we propose an estimator for the mean which considers both the non Euclidean nature of the data and their spatial correlation. Simulated data are used to evaluate the performance of the proposed estimator: taking into account spatial dependence leads to better estimates when observations are irregularly spaced in the region of interest. This allows to address a meteorological problem, namely, the estimation of the covariance matrix between temperature and precipitation for the province of Quebec in Canada. Finally, a kriging estimator based on a tangent space model is proposed for covariance fields. This allows to deal with non stationary fields, the deterministic drift being handled in the tangent space with traditional spatial statistics techniques.

Contents

Introduction

This thesis is part of a line of research which deals with the statistical analysis of data belonging to a non Euclidean space. In recent years, attention to the statistical analysis of non Euclidean data has been growing. The conceptual framework is that of Object Oriented Data Analysis, as defined in Wang and Marron (2007). This approach focuses on the atoms of the statistical analysis. While this is usually a number or a vector, new technologies have provided different kinds of data, such as high dimensional arrays, curves, shapes, diffusion tensors... Many of these can be seen as elements of a non Euclidean space. Indeed, non Euclidean data are mathematical objects more complex than numbers or vectors and they do not belong to a linear space. Thus, even the most simple statistical operations, such as finding a centerpoint for the data distribution or evaluating variability about this center, represent a challenge. Statistical analysis needs to carefully consider the mathematical properties of the data at hand and consequently to reformulate traditional methods in this new setting.

Data belonging to a Riemannian manifold are particularly interesting both from a mathematical and from a practical point of view. Studies in this field have been motivated by many applications: for example Shape Analysis (see, e.g, Jung et al., 2011), Diffusion Tensor Imaging (see Dryden et al., 2009, and references therein) and estimation of covariance structures. The general aim of these studies is the extension to Riemannian data of traditional statistical methods developed for Euclidean data, such as point estimation of mean and variance (Pennec et al., 2006; Dryden et al., 2009), exploratory data analysis, dimensional reduction (Fletcher et al., 2004), testing hypothesis among different populations (Schwartzman et al., 2010) and smoothing (Yuan et al., 2012). The common idea is to find the correct distance to compare two elements in the non Euclidean space and to build statistical methods based on that distance.

This thesis proposes a twofold contribution in this research field. First, we consider the generalization of the above methods to infinite dimensional covariance operators. This problem arises within Functional Data Analysis, where the interest is on the second order structure of the functional random process. Second, the problem of spatial dependence among Riemannian data is considered, with a

particular attention to the case of covariance matrices.

The choice to address these specific problems is guided both by their mathematical and statistical interest and by the considered applications. Indeed, we illustrate two problems where the analysis of covariance operators of functional random variables brings insight to the statistical analysis. The first one is the comparison of the covariance functions between different groups of patients of the well known AneuRisk dataset (see Sangalli et al., 2009a) aiming at supporting some choices of the previously published analysis. The second analysis addresses a linguistic problem, namely the exploration of relationships among Romance languages. Here, the statistical units are the covariance structures among frequencies for speakers of different languages. Data come from speech recordings provided by Prof. J. Coleman (Phonetic Laboratory, University of Oxford) and preprocessed by P.Z. Hadjipantelis (University of Warwick).

A meteorological problem is the most immediate application for spatial statistics for Riemannian data. In particular, we show the analysis of covariance matrices between temperature and precipitations measured in different stations in Quebec, Canada.

Organization of the thesis

This thesis is organized in two parts. Part I deals with covariance operators of functional random processes. Under non restrictive assumption on the random process which generates the data, the covariance operator is a trace class operator with non negative eigenvalues. Thus, it belongs to an infinite dimensional non Euclidean space. Our first concern is to understand which distance is the most appropriate to compare two elements of this space. Chapter 1 recalls the main properties of these operators and introduces some possible distances. In Chapter 2, we illustrate two inferential problems - the point estimation of covariance operators and the comparison of covariance operators between two population of curves- that can be addressed relying on a distance-based approach. Chapter 3 describes two applicative problems where the analysis of covariance operators is needed.

Part II consists of the introduction of spatial dependence among Riemannian data. It is not a trivial problem to define stochastic dependence in non linear spaces. Here, we consider both the modeling of the dependence on the manifold, generalizing the definition of covariance in linear spaces through the expected values of square distances, and the possibility to approximate non Euclidean data in the appropriate tangent space, where traditional statistical techniques can be used.

Chapter 4 illustrates the mathematical properties of the space of covariance

matrices and the existing statistical techniques for dealing with covariance matrices as object data. The problems of estimating the Fréchet mean from a sample of spatially dependent Riemannian data is addressed in Chapter 5. Indeed, we develop our methods explicitly for the case of positive definite symmetric matrices, with the aim to deal with covariance matrices between meteorological variables, but they are valid in general for data belonging to Riemannian manifold. Following the terminology introduced in Wang and Marron (2007), these data are mildly non Euclidean, being locally Euclidean and admitting a tangent space. However, the proposed ideas can be helpful also in a strongly non Euclidean setting, where only a definition of distance is available. Finally, in Chapter 6 a kriging procedure, i.e. the estimation of the field in an unobserved location, is developed for covariance matrix fields. Here, the deterministic drift is handled in the tangent space, where linear methods can be applied.

Part I

Statistical methods for covariance operators

Data sets are increasingly becoming available that are best described as being functional. In recent years, research in this field has provided many statistical techniques to deal with this kind of data (Ramsay and Silverman, 2005; Ferraty and Vieu, 2006; Horváth and Kokoszka, 2012). However, this work mainly focused on mean functions and on the use of associated bases to provide insight into these means, while little attention has been paid to the explicit analysis of the covariance operator. In many applied problems, the covariance operator is either directly or indirectly intrinsically interesting in its own right. This paper is focused on the inference for the covariance operator of a functional random process and primarily in providing a better understanding of which kinds of metric could be appropriate for this inference.

Some recent works (Panaretos et al., 2010; Fremdt et al., 2012) examined testing the equality of covariance structures from two groups of functional curves by defining a test statistic through the Karhunen-Loève expansions of the two covariance structures. This method are therefore based on the Hilbert-Schmidt metric, exploiting the immersion of the space of covariance operators in the Hilbert-Schmidt space. However, this extrinsic approach ignores the geometry of the space of covariance operators.

Here, we first consider the problem of the definition of possible metrics for covariance operators. While it will be seen that some finite dimensional distances for positive definite covariance matrices (Dryden et al., 2009) naturally lend themselves to functional analogues, others do not have natural extensions. Then, it will be shown that the extended metrics can be effectively used for inference, both in estimation and in testing the underlying covariance structure between different groups.

Analysis of the covariance operator arises in many applied contexts, two of which will be detailed. Firstly, in data associated with brain aneurysms, several patient populations are routinely considered similar enough to be treated as one population. By means of the proposed permutation test, we will explore their covariance structures to understand whether they can indeed be so combined. Secondly, in the linguistic analysis of human speech, the overall mean structure of the data produced is often not of interest, but rather the variations that can be found within the language. Here we will show that different languages can be compared and even predicted through functional distances, allowing, for the first time, a quantitative analysis of comparative philological relations based on speech recordings rather than discrete textual analysis.

Chapter 1

On the definitions of distance between covariance operators

This chapter is devoted to consider covariance operator on infinite dimensional Hilbert spaces as object data. The principal mathematical properties of these objects are recalled, with the aim to understand which distances are suitable to compare two covariance operators. It will be seen that not all matrix based distances are extendable to the infinite dimensional case.

1.1 Some remarks on operators on $L^2(\Omega)$

In this section we focus on properties and definitions that will be useful below. More details and proofs can be found, e.g., in Zhu (2007).

Definition 1.1.1 *Let B_1 be the closed ball of unitary radius in $L^2(\Omega)$, consisting of all $f \in L^2(\Omega)$ such that $||f||_{L^2(\Omega)} \leq 1$, where $L^2(\Omega)$ is the Hilbert space of square-integrable functions on $\Omega \subseteq \mathbb{R}$. A bounded linear operator $T : L^2(\Omega) \to L^2(\Omega)$ is compact if $T(B_1)$ is compact in the norm of $L^2(\Omega)$.*

An important property of a compact operator on $L^2(\Omega)$ is the existence of a canonical decomposition. This decomposition implies the existence of two orthonormal bases $\{u_k\}, \{v_k\}$ for $L^2(\Omega)$ such that

$$Tf = \sum_k \sigma_k \langle f, v_k \rangle u_k,$$

or, equivalently,

$$Tv_k = \sigma_k u_k,$$

where $\langle ., . \rangle$ indicates the scalar product in $L^2(\Omega)$. The sequence $\{\sigma_k\} \in \mathbb{R}$ is called the sequence of singular values for T.

Definition 1.1.2 *A bounded linear operator T is self-adjoint if $T = T^*$.*

If the operator is compact and self-adjoint, there exists a basis $\{v_k\}$ such that

$$Tf = \sum_k \lambda_k \langle f, v_k \rangle v_k,$$

or, equivalently,

$$Tv_k = \lambda_k v_k$$

and $\{\lambda_k\} \in \mathbb{R}$ is called the sequence of eigenvalues for T.

A compact operator T is said to be *trace class* if

$$\text{trace}(T) := \sum_k \langle Te_k, e_k \rangle < +\infty$$

for every orthonormal basis $\{e_k\}$. For non negative definite operators, it can be seen that the definition is independent of the choice of the basis. In the case of self-adjoint non negative definite operators the trace is thus the sum of the eigenvalues and this is finite for trace class self-adjoint operator. We indicate with $S(L^2(\Omega))$ the space of the trace class operators on $L^2(\Omega)$.

A compact operator T is said to be Hilbert-Schmidt if its Hilbert-Schmidt norm is bounded, i.e.

$$||T||^2_{HS} = \text{trace}(T^*T) < +\infty.$$

This is a generalisation of the Frobenius norm for finite-dimensional matrices.

These properties are crucial in the context of the statistical analysis of functional data. Indeed, let $\mathbf{f}$ be a random function which takes values in $L^2(\Omega)$, $\Omega \subseteq \mathbb{R}$, such that $\mathbb{E}[||\mathbf{f}||^2_{L^2(\Omega)}] < +\infty$. The covariance operator is

$$C_{\mathbf{f}} g(t) = \int_\Omega c_{\mathbf{f}}(s,t) g(s) \mathrm{d}s,$$

for $g \in L^2(\Omega)$, where

$$c_{\mathbf{f}}(s,t) = \text{cov}(\mathbf{f}(s), \mathbf{f}(t)) = \mathbb{E}\left[(\mathbf{f}(s) - \mathbb{E}\left[\mathbf{f}(s)\right])\left(\mathbf{f}(t) - \mathbb{E}\left[\mathbf{f}(t)\right]\right)\right].$$

Then, $C_{\mathbf{f}}$ is a trace class, self-adjoint, compact operator on $L^2(\Omega)$ with non negative eigenvalues (see, e.g., Bosq, 2000, Section 1.5). The space of covariance operators is therefore strictly included in the linear spaces of trace class operators and Hilbert-Schmidt operators. This fact is often neglected in the development of statistical methods for covariance operators, linear methods being used that would be appropriate for generic trace-class operators. In the following, we introduce possible transformations to map the space of covariance operators to the space of Hilbert-Schmidt operators, where linear operations are allowed.

Finally, we recall the definition of unitary operator on $L^2(\Omega)$, which will be needed for the definition of Procrustes distance in the functional setting.

Definition 1.1.3 *A bounded linear operator R on $L^2(\Omega)$ is said to be unitary if*

$$||Rf||_{L^2(\Omega)} = ||f||_{L^2(\Omega)} \quad \forall f \in L^2(\Omega).$$

We indicate with $O(L^2(\Omega))$ the space of unitary operators on $L^2(\Omega)$.

We take now advantage of these tools from functional analysis to introduce possible distances that can be used to measure the difference between two covariance operators.

1.2 Distances between covariance operators

In this section we propose several distances that can be used to compare the covariance operators of two random functions taking values in $L^2(\Omega)$. These are a generalisation to the functional setting of metrics that have been proved useful for the case of positive semi-definite matrices (Dryden et al., 2009). However, not all matrix based distances are extendable to the functional case.

Two popular metrics for finite dimensional covariance matrix analysis are the log Euclidean metric and the affine invariant Riemannian metric. While both would appear to be natural candidates for generalisation to covariance operators, in both cases, this is not straightforward due to the natural trace class structure of the covariance operator. The trace class property implies that the (descending) ordered eigenvalues λ_i are summable, i.e.

$$\sum_i \lambda_i < \infty \Rightarrow \lambda_i \to 0 \text{ as } i \to \infty$$

The log Euclidean distance for two positive definite matrices, M_1 and M_2, is defined as

$$d_{\log}(M_1, M_2) = ||\log(M_1) - \log(M_2)||.$$

with the $\log()$ indicating the matrix logarithm. This is not well defined for trace class operators as this quantity tends to infinity. The affine invariant Riemannian metric for positive definite matrices is defined as

$$d_{\text{Riem}}(M_1, M_2) = ||\log(M_1^{-\frac{1}{2}} M_2 M_1^{-\frac{1}{2}})||$$

which requires consideration of the inverse. For a compact operator, even when it is positive definite, the inverse is not well defined (see, e.g., Zhu, 2007, Section 1.3).

Even though in applications only finite dimensional representations are available, these are usually not full rank (i.e. they have zero eigenvalues), meaning that these metrics present the same difficulties than in the infinite dimensional case.

This problem can be handled projecting the observation in a subspace but in general different subspaces should be needed to best approximate the two operators and the choice of the correct level of truncation is often problematic.

Moreover, since the distance between infinite dimensional operators is not well defined, it is not clear how to interpret the asymptotic behaviour of the distance computed between finite dimensional representations. This could be an issue when the dimensionality of the problem is high and different choices are possible for the projected space.

We thus resort to some alternative distances which are well defined for self-adjoint trace class operators with nonnegative eigenvalues.

1.2.1 Distance between kernels in $L^2(\Omega \times \Omega)$

Distances between covariance operators can be naturally defined using the distance between their integral kernels in $L^2(\Omega \times \Omega)$. Let S_1 and S_2 be two covariance operators and

$$S_i f(t) = \int_\Omega s_i(s,t) f(s) \mathrm{d}s, \quad \forall \ f \in L^2(\Omega).$$

Then, we can define the distance

$$d_L(S_1, S_2) = ||s_1 - s_2||_{L^2(\Omega\times\Omega)} = \sqrt{\int_\Omega \int_\Omega (s_1(x,y) - s_2(x,y))^2 \mathrm{d}x\mathrm{d}y}.$$

This distance is correctly defined, since it inherits all the properties of the distance in the Hilbert space $L^2(\Omega \times \Omega)$. However, it does not exploit in any way the particular structure of covariance operators and therefore it need not to be useful for highlighting significant differences between covariance structures. Indeed, this is the distance induced by the Hilbert-Schmidt norm, since for the case of Hilbert-Schmidt kernel operators $||S_1 - S_2||_{HS} = ||s_1 - s_2||_{L^2(\Omega\times\Omega)}$. Thus, it exploits the immersion of the space of covariance operators in the Hilbert-Schmidt space, being an extrinsic metric that ignores the geometry of the space of interest.

1.2.2 Spectral distance

A second possibility is to regard the covariance operator as an element of $\mathfrak{L}(L^2(\Omega))$, the space of the linear bounded operators on $L^2(\Omega)$. It follows that the distance between S_1 and S_2 can be defined as the operator norm of the difference. We recall that the norm of a self-adjoint bounded linear operator on $L^2(\Omega)$ is defined as

$$||T||_{\mathfrak{L}(L^2(\Omega))} = \sup_{v \in L^2(\Omega)} \frac{|\langle Tv, v\rangle|}{\langle v, v\rangle}$$

and for a covariance operator it coincides with the absolute value of the first (i.e. largest) eigenvalue. Thus,

$$d_{\mathfrak{L}}(S_1, S_2) = ||S_1 - S_2||_{\mathfrak{L}(L^2(\Omega))} = |\widetilde{\lambda}_1|$$

where $\widetilde{\lambda}_1$ is the first eigenvalue of the operator $S_1 - S_2$. The distance $d_{\mathfrak{L}}(.,.)$ generalises the matrix spectral norm which is often used in the finite dimensional case (see, e.g., El Karoui, 2008). This distance takes into account the spectral structure of the covariance operators, but it appears restrictive in that it focuses only on the behaviour on the first mode of variation.

1.2.3 Square root operator distance

Since covariance operators are trace class, we can generalise the square root matrix distance (see Dryden et al., 2009). Indeed, S being a self-adjoint trace class operator, there exists a Hilbert-Schmidt self adjoint operator $(S)^{\frac{1}{2}}$ defined as

$$(S)^{\frac{1}{2}} f = \sum_k \lambda_k^{\frac{1}{2}} \langle f, v_k\rangle v_k, \tag{1.1}$$

where λ_k are eigenvalues and v_k eigenfunctions of S. We can therefore define the square root distance between two covariance operators S_1 and S_2 as

$$d_R(S_1, S_2) = ||(S_1)^{\frac{1}{2}} - (S_2)^{\frac{1}{2}}||_{HS}.$$

Inspiration for this kind of distance comes from the log-Euclidean distance for positive definite matrices. There, the logarithmic transformation allows to map the non Euclidean space in a linear space. As mentioned above, a logarithmic map for covariance operators is not available. Thus, we choose a different transformation, namely, the square root transformation. This has been shown to behave in a similar way in the finite dimensional setting (see Dryden et al., 2009) but it is also well defined for trace class operators.

Any power greater than $1/2$ would be a possible candidate distance. For trace class operators in general, the square root operator is the smallest power for which the transformed operator is guaranteed to be Hilbert-Schmidt, meaning that it is the closest available to the log-Euclidean distance. In addition, it can be interpreted as a distance which takes into account the full eigenstructure of the covariance operator, both eigenfunctions and eigenvalues.

1.2.4 Procrustes reflection size-and-shapes distance

The square root operator distance looks at the distance between the square root operators $(S_1)^{\frac{1}{2}}$ and $(S_2)^{\frac{1}{2}}$ in the space of Hilbert-Schmidt operators. However, this is only a particular choice of a broad family of distances, which are based on the mapping of the two operators S_1 and S_2 from the space of covariance operators to the space of Hilbert-Schmidt operators. We can consider in general a transformation $S_i \to L_i$, so that $S_i = L_i L_i^*$ and define the distance as the Hilbert-Schmidt norm of $L_1 - L_2$. Considering this more general framework, it is easy to see that any of this transformation is defined up to a unitary operator R:

$$(L_i R)(L_i R)^* = L_i R R^* L_i^* = L_i L_i^* = S_i.$$

To avoid the arbitrariness of the transformation, it is meaningful to use a Procrustes approach which looks for the unitary operator R which best matches the two operators L_1 and L_2, however they are defined.

In Dryden et al. (2009), a Procrustes reflection size-and-shape distance is proposed to compare two positive definite matrices. Our aim is to generalise this distance to the case of covariance operators on $L^2(\Omega)$. Let S_1 and S_2 be two covariance operators on L^2. We define the square of the Procrustes distance between S_1 and S_2 as

$$d_P(S_1, S_2)^2 = \inf_{R \in O(L^2(\Omega))} ||L_1 - L_2 R||^2_{HS} = \inf_{R \in O(L^2(\Omega))} \text{trace}((L_1 - L_2 R)^*(L_1 - L_2 R)),$$

where L_i are such that $S_i = L_i L_i^*$ for $i = 1, 2$.

As mentioned above, the decomposition $S_i = L_i L_i^*$ can be seen as a general form of transformation, mapping S_i to a space where a linear metric is appropriate. In particular, a good choice could be the square root transformation.

Proposition 1.2.1 *Let σ_k be the singular values of the compact operator $L_2^* L_1$. Then*

$$d_P(S_1, S_2)^2 = ||L_1||^2_{HS} + ||L_2||^2_{HS} - 2 \sum_{k=1}^{+\infty} \sigma_k.$$

Proof. Note that

$$\begin{aligned} d_P(S_1, S_2)^2 &= \inf_{R \in O(L^2(\Omega))} \text{trace}((L_1 - L_2 R)^*(L_1 - L_2 R)) \\ &= \inf_{R \in O(L^2(\Omega))} \{\text{trace}(L_1^* L_1) + \text{trace}(L_2^* L_2) - 2\text{trace}(R^* L_2^* L_1)\} \\ &= ||L_1||^2_{HS} + ||L_2||^2_{HS} - 2 \sup_{R \in O(L^2(\Omega))} \text{trace}(R^* L_2^* L_1). \end{aligned}$$

We therefore look for the unitary operator R which maximises $\mathrm{trace}(R^* L_2^* L_1)$. Exploiting the definition of the trace operator and the singular value decomposition for the compact operator $L_2^* L_1$ (which is trace class - see Bosq (2000, Section 1.5)),

$$L_2^* L_1 v_k = \sigma_k u_k \ \text{ for } \ k = 1, \ldots, +\infty,$$

we obtain

$$\mathrm{trace}(R^* L_2^* L_1) = \sum_{k=1}^{+\infty} \langle R^* L_2^* L_1 e_k, e_k \rangle = \sum_{k=1}^{+\infty} \langle R^* L_2^* L_1 v_k, v_k \rangle =$$

$$= \sum_{k=1}^{+\infty} \sigma_k \langle R^* u_k, v_k \rangle \leq \sum_{k=1}^{+\infty} \sigma_k ||R^* u_k||_{L^2(\Omega)} ||v_k||_{L^2(\Omega)} = \sum_{k=1}^{+\infty} \sigma_k ||u_k||_{L^2(\Omega)} ||v_k||_{L^2(\Omega)} = \sum_{k=1}^{+\infty} \sigma_k.$$

Thus, the maximum is reached for any operator $\widetilde{R}$ such that

$$\widetilde{R}^* u_k = v_k \ \ \forall k = 1, \ldots, +\infty,$$

or, equivalently,

$$\widetilde{R} v_k = u_k \ \ \forall k = 1, \ldots, +\infty.$$

Substituting this optimal transformation in the definition of the distance,

$$\begin{aligned} d_P(S_1, S_2)^2 &= ||L_1||_{HS}^2 + ||L_2||_{HS}^2 - 2\mathrm{trace}(\widetilde{R}^* L_2^* L_1) \\ &= ||L_1||_{HS}^2 + ||L_2||_{HS}^2 - 2 \sum_{k=1}^{+\infty} \langle \widetilde{R}^* L_2^* L_1 e_k, e_k \rangle \\ &= ||L_1||_{HS}^2 + ||L_2||_{HS}^2 - 2 \sum_{k=1}^{+\infty} \langle \widetilde{R}^* L_2^* L_1 v_k, v_k \rangle \\ &= ||L_1||_{HS}^2 + ||L_2||_{HS}^2 - 2 \sum_{k=1}^{+\infty} \sigma_k \underbrace{\langle \widetilde{R}^* u_k, v_k \rangle}_{=1}. \end{aligned}$$

□

The Procrustes distance takes into account the arbitrariness in the definition of the map $S_i \to L_i$. It is worth noticing that the unitary transformation allows the operator L_i to become non self-adjoint. Thus, this extends the analysis in the Hilbert-Schmidt space to go beyond symmetric operators.

1.2.5 Finite dimensional approximation

In practical applications, we observe only a finite dimensional representation of the operators of interest. Therefore, ideally we would require square root distance

and Procrustes size-and-shape distance between two finite dimensional representations to be a good approximation of the distance between the infinite dimensional operators. We show this fact for the more general case of Procrustes distance, with the square root distance being a special case where $L_i = (S_i)^{\frac{1}{2}}$ and R is constrained to be the identity operator.

Let $\{e_k\}_{k=1}^{+\infty}$ be a basis for $L^2(\Omega)$, $V_p = \text{span}\{e_1, \ldots, e_p\}$ and S_i^p be the restriction of S_i on V_p, i.e.

$$S_i^p g = \sum_{k=1}^{p} \langle g, e_k \rangle S_i e_k \quad \forall \ g \in V_p.$$

In practical situations, V_p will be the subspace which contains the finite dimensional representation of the functional data. Let us assume that, for $p \to +\infty$, $L_i^p \to L_i$ with respect to the Hilbert-Schmidt norm, where $S_i^p = L_i^p L_i^{p*}$. This is not restrictive, since we can choose for instance $L_i = (S_i)^{\frac{1}{2}}$, but every choice which guarantees this convergence is suitable. Then, the distance between the two restricted operators is

$$d_P(S_1^p, S_2^p)^2 = ||L_1^p||_{HS}^2 + ||L_2^p||_{HS}^2 - 2 \sum_{k=1}^{p} \langle \widetilde{R}^p L_2^{p*} L_1^p e_k, e_k \rangle.$$

Since $V_p \subset L^2(\Omega)$, we can choose a subset $v_1^p, \ldots, v_p^p$, $v_k^p \in \{v_k\}_{k=1}^{+\infty}$ which is an orthonormal basis for V_p. However, they need not be the first p elements of the basis coming from the canonical decomposition of $L_2^* L_1$. This happens because the space V_p depends only on the original basis $\{e_k\}_{k=1}^p$ and it does not depend on the covariance structure of the data. Since the subspaces V_p are nested, we can define a permutation $s : \mathbb{N} \to \mathbb{N}$, so that $\{v_{s(1)}, \ldots, v_{s(p)}\}$ provides a basis for V_p, for every p. Since the trace of an operator does not depend on the choice of the basis, we obtain

$$d_P(S_1^p, S_2^p)^2 = ||L_1^p||_{HS}^2 + ||L_2^p||_{HS}^2 - 2 \sum_{k=1}^{p} \langle \widetilde{R}^p L_2^{p*} L_1^p v_{s(k)}, v_{s(k)} \rangle$$

$$= ||L_1^p||_{HS}^2 + ||L_2^p||_{HS}^2 - 2 \sum_{k=1}^{p} \sigma_{s(k)},$$

where $\{\sigma_{s(k)}\}_{k=1}^p$ are singular values for $L_2^* L_1$. This comes from the fact that the action of the operator $L_2^{p*} L_1^p$ should be equal to the action of the operator $L_2^* L_1$ on every element belonging to the subspace V_p and $v_{s(k)} \in V_p$ for $k = 1, \ldots, p$. Finally, as $L_2^* L_1$ is trace class, the series of its singular values is absolutely con-

vergent and therefore also unconditionally convergent (convergent under any permutation). Thus,

$$\lim_{p\to+\infty} d_P(S_1^p, S_2^p)^2 = ||L_1||_{HS}^2 + ||L_2||_{HS}^2 - 2\sum_{k=1}^{+\infty} \sigma_{s(k)}$$

$$= ||L_1||_{HS}^2 + ||L_2||_{HS}^2 - 2\sum_{k=1}^{+\infty} \sigma_k = d_P(S_1, S_2)^2.$$

Chapter 2

Statistical inference for covariance operators

In this chapter we illustrate two inferential problems that can be addressed using the distances introduced in the previous chapter. First, we consider the point estimation of covariance operators. This problem have been widely considered in Functional Data Analysis but the aim was to obtain optimal estimators with respect to the Hilbert-Schmidt norm. For example, under reasonable assumptions on the random process generating functional data, the sample covariance operator is a consistent estimator, meaning that the Hilbert-Schmidt distance between the estimated operators and the true operator goes to zero. Significant contributions in this research can be found, e.g., in Besse and Ramsay (1986); Rice and Silverman (1991); Bosq (2000); Horváth and Kokoszka (2012).

However, in an object oriented approach, we would like to have estimators based on the distance that has been considered appropriate for the problem at hand. Thus, we propose here Fréchet estimators for covariance operators and different choices of metric are considered. Even if the theory of these estimators in an infinite dimensional setting is not yet fully developed, they are the natural way to take into account the specific geometrical features of covariance operators in estimation procedures.

Then, we propose a method to test the equality of the covariance operator between two groups. By means of simulation studies, it will be shown that using a distance which considers the geometry of the space, such as Square root or Procrustes distance, may result in a dramatic improve in the power of the test.

2.1 Distance-based point estimation for covariance operators

In many applications the estimation of a covariance operator is needed. In this section we illustrate an approach to estimation procedures which is coherent with the choice of the proper distance to compare covariance operators. First, the problem of averaging among covariance operators of different groups is considered as well as the estimation of the covariance operator from curves samples coming from these groups. Then, we consider the problem of interpolation and extrapolation of covariance operators.

2.1.1 Fréchet mean for a set of covariance operators

Let $S_1, \ldots, S_g$ be the covariance operators of g different groups. Then, a possible estimator of the mean covariance operator Σ is

$$\widehat{\Sigma} = \frac{1}{n_1 + \cdots + n_g}(n_1 S_1 + \cdots + n_g S_g),$$

where weights n_i, $i = 1, \ldots, g$ are the numbers of observations from which the covariance operator S_i has been obtained. As we show in the proof of Proposition 2.1.1, this formula arises from the minimisation of square Hilbert-Schmidt deviations, weighted with the number of observations, i.e.

$$\widehat{\Sigma} = \arg\min_{\Sigma} \sum_{i=1}^{g} n_i ||\Sigma - S_i||^2_{HS}.$$

If we choose a different distance to compare covariance operators, it is more coherent to use the chosen distance in this minimization problem.

More in general, the Fréchet mean of a random element S, with probability distribution μ on the space of covariance operators, can be defined as $\Sigma = \operatorname{arginf}_P \int d(S, P)^2 \mu(dS)$. If a sample S_i, $i = 1, \ldots, g$ from μ is available, a least square estimator for Σ can be defined using the weighted sample Fréchet mean:

$$\widehat{\Sigma} = \arg\inf_{S} \sum_{i=1}^{g} n_i d(S, S_i)^2.$$

The actual computation of the sample Fréchet mean $\widehat{\Sigma}$ depends on the choice of the distance $d(.,.)$. In general, it requires the solution of a high dimensional minimisation problem but some distances admit an analytic solution while for others efficient minimisation algorithms are available. Note that $\widehat{\Sigma}$ may not be unique for positively curved spaces, although it is unique for suitably concentrated data (see Kendall, 1990; Le, 2001).

Proposition 2.1.1 *For the square root distance d_S,*

$$\widehat{\Sigma} = \arg\min_S \sum_{i=1}^{g} n_i d_S(S, S_i)^2 = (\frac{1}{G}\sum_{i=1}^{g} n_i (S_i)^{\frac{1}{2}})^2, \tag{2.1}$$

where $G = n_1 + \cdots + n_g$.

Proof. We prove that in general

$$\arg\min_L \sum_{i=1}^{g} n_i ||L - L_i||_{HS}^2 = \frac{1}{G}\sum_{i=1}^{g} n_i L_i,$$

which gives the desired result for the particular case of $L_i = (S_i)^{\frac{1}{2}}$ and $L = (S)^{\frac{1}{2}}$. We have

$$\begin{aligned}
\arg\min_L \sum_{i=1}^{g} n_i ||L - L_i||_{HS}^2 &= \arg\min_L \sum_{i=1}^{g} n_i \mathrm{trace}\big((L - L_i)^*(L - L_i)\big) \\
&= \arg\min_L \sum_{i=1}^{g} n_i \{||L||_{HS}^2 + ||L_i||_{HS}^2 - 2\mathrm{trace}(L^* L_i)\} \\
&= \arg\min_L \sum_{i=1}^{g} n_i \{||L||_{HS}^2 - 2\mathrm{trace}(L^* L_i)\} \\
&= \arg\min_L \Big[G||L||_{HS}^2 - 2\mathrm{trace}\Big(L^* \sum_{i=1}^{g} n_i L_i\Big)\Big],
\end{aligned}$$

exploiting the linearity of the trace operator. It can be noticed that the second term is a scalar product in the operator space and therefore it is minimum when L is proportional to $\sum_{i=1}^{g} n_i L_i$. We thus obtain a minimisation problem in $\alpha = ||L||_{HS}$:

$$\begin{aligned}
&\arg\min_L G||L||_{HS}^2 - 2\mathrm{trace}\Big(L^* \sum_{i=1}^{g} n_i L_i\Big) \\
&\qquad = \arg\min_\alpha G\alpha^2 - 2\mathrm{trace}\Big(\frac{\alpha}{||\sum_{i=1}^{g} n_i L_i||_{HS}} \Big(\sum_{i=1}^{g} n_i L_i\Big)^* \Big(\sum_{i=1}^{g} n_i L_i\Big)\Big) \\
&\qquad = \arg\min_\alpha G\alpha^2 - 2\frac{\alpha}{||\sum_{i=1}^{g} n_i L_i||_{HS}} \underbrace{\mathrm{trace}\Big(\Big(\sum_{i=1}^{g} n_i L_i\Big)^* \Big(\sum_{i=1}^{g} n_i L_i\Big)\Big)}_{=||\sum_{i=1}^{g} n_i L_i||_{HS}^2} \\
&\qquad = \arg\min_\alpha G\alpha^2 - 2\alpha||\sum_{i=1}^{g} n_i L_i||_{HS}
\end{aligned}$$

and the minimum is reached for $\alpha = \frac{1}{G}||\sum_{i=1}^{g} n_i L_i||_{HS}$. Therefore

$$L = \frac{\alpha}{||\sum_{i=1}^{g} n_i L_i||_{HS}} \sum_{i=1}^{g} n_i L_i = \frac{1}{G} \sum_{i=1}^{g} n_i L_i.$$

□

For the Procrustes size-and-shape distance an analytic solution is not available. However, the Procrustes mean can be obtained by an adaptation of the algorithm proposed in Gower (1975). It is an iterative method that alternates a registration step and an averaging step.

1.Initialization The algorithm is initialized with $\widehat{\Sigma}^{(0)} = L^{(0)} L^{(0)*}$, where $L^{(0)} = \frac{1}{G}\sum_{i=1}^{g} n_i L_i$, L_i so that $S_i = L_i L_i^*$ and $G = n_1 + \cdots + n_g$.

2.Registration step For all the groups $i = 1, \ldots, g$, $L_i^{(k)} = L_i^{(k-1)} R_i$, where R_i is the unitary operator which minimises the Hilbert Schmidt norm of $L^{(k-1)} - L_i^{(k-1)} R_i$.

3.Averaging step The new Procrustes mean is computed: $\widehat{\Sigma}^{(k)} = L^{(k)} L^{(k)*}$, where $L^{(k)} = \frac{1}{G}\sum_{i=1}^{g} n_i L_i^{(k)}$, since this minimises $\sum_{i=1}^{g} n_i ||L - L_i^{(k)}||_{HS}^2$ as shown in the proof of Proposition 2.1.1.

Steps 2 and 3 are iterated until convergence, i.e. when the Hilbert-Schmidt norm of the difference between $L^{(k)}$ and $L^{(k-1)}$ is below a chosen tolerance. In practice, the algorithm will give a local minimum, often called Karcher mean (Karcher, 1977), in few iterations, if it is initialised with the estimate provided by (2.1).

The algorithm above is adapted from one of a number of variants of the Procrustes algorithm, all of which have been shown in the finite dimensional setting to have similar convergence properties (see Groisser, 2005). It is conjectured that analogous convergence properties are also true in the infinite dimensional setting (in particular that the finite dimensional algorithm converges to the correct infinite dimensional limit), but the geometric arguments using in the finite dimensional proof by Groisser (2005) are not immediately available for the infinite dimensional setting and we leave this for future work.

We also compared our version of the algorithm with the one proposed by Dryden and Mardia (1998), where each operator L_i is aligned to the average obtained from all the other operators, namely $\frac{1}{G-n_i}\sum_{j\neq i}^{g} n_j L_j^{(k)}$. However, this algorithm, in the examples below, provided the same result as the one above, while also having very similar convergence speed and computational burden.

2.1.2 Distance based estimation from curves sample

We have considered above the covariance operators as the starting point of the analysis, the integral kernels of these operators being estimated with traditional sample covariance function. However, a different approach is possible, using the proposed distances also in the estimation of the covariance operators. Namely, an estimate of the common covariance operator from curves samples coming from g different groups could be

$$\widehat{\Sigma} = \arg\min_{S} \sum_{i=1}^{g} \sum_{k=1}^{n_i} d(S, (f_{ik} - \bar{f}_{i\cdot}) \otimes (f_{ik} - \bar{f}_{i\cdot}))^2, \tag{2.2}$$

where $i = 1, \ldots, g$ are different groups, $k = 1, \ldots, n_i$ curves within each group, $\otimes$ indicate the tensor product $(f \otimes f)v = \langle f, v \rangle f$ and $\bar{f}_{i\cdot}$ is the sample mean in the i-th group. As in the case of sample Fréchet mean, if we choose the square root distance, we have an explicit solution for problem (2.2) and

$$\widehat{\Sigma} = (\frac{1}{\sum_{i=}^{g} n_i} \sum_{i=1}^{g} \sum_{k=1}^{n_i} (f_{ik} - \bar{f}_{i\cdot} \otimes (f_{ik} - \bar{f}_{i\cdot}))^{\frac{1}{2}})^* (\frac{1}{\sum_{i=}^{g} n_i} \sum_{i=1}^{g} \sum_{k=1}^{n_i} (f_{ik} - \bar{f}_{i\cdot} \otimes (f_{ik} - \bar{f}_{i\cdot}))^{\frac{1}{2}}).$$

For what concerns Procrustes distance, problem (2.2) can be solved with a slight modification of the algorithm illustrated above.

2.1.3 Interpolation and extrapolation

The problem of interpolation and extrapolation in non linear spaces has been studied in deep for the case of positive definite matrices, where it has been shown that using simply Euclidean metric can be very problematic (see, e.g, Dryden et al., 2009). In the infinite dimensional case, the equivalent of a Euclidean approach would be extrapolation based on kernels. Let S_1 and S_2 be two covariance operators and $s_1(s,t)$ and $s_2(s,t)$ their integral kernels. Thus, we can obtain a path passing through these kernels as

$$s(s,t)(x) = (s_1(s,t) + x(s_2(s,t) - s_1(s,t))), \tag{2.3}$$

$x \in \mathbb{R}$. However, just as in the case of positive definite matrices, extrapolation based on kernel distances does not always result in a valid kernel for a covariance operator (i.e. the associated integral operator is not non-negative definite).

Square root metric and Procrustes metric can be instead associated with the corresponding geodesic which connects the two covariance operators S_1 and S_2, being respectively

$$S_R(x) = \{(S_1)^{\frac{1}{2}} + x((S_2)^{\frac{1}{2}} - (S_1)^{\frac{1}{2}})\}^* \{(S_1)^{\frac{1}{2}} + x((S_2)^{\frac{1}{2}} - (S_1)^{\frac{1}{2}})\} \tag{2.4}$$

and

$$S_P(x) = \{(S_1)^{\frac{1}{2}} + x((S_2)^{\frac{1}{2}}\widetilde{R} - (S_1)^{\frac{1}{2}})\}\{(S_1)^{\frac{1}{2}} + x((S_2)^{\frac{1}{2}}\widetilde{R} - (S_1)^{\frac{1}{2}})\}^*, \quad (2.5)$$

with $x \in \mathbb{R}$ and $\widetilde{R}$ is an unitary operator minimising $||(S_1)^{1/2} - (S_2)^{1/2}R||^2_{HS}$. In general, analogously to the finite dimensional case, the operator $\widetilde{R}$ may be not uniquely defined if the sequence of singular values of $(S_2)^{1/2}(S_1)^{1/2}$ is degenerate (see Kent and Mardia, 2001), i.e. in our case if there is more than one zero singular value. However, any choice provides a valid geodesic with respect to the Procrustes metric.

For every x, both geodesics provide a valid covariance operator. However, in the case of extrapolation with square root geodesic, this operator can be the result of the inverse (square) operation from the space of Hilbert-Schmidt operators to the space of covariance operators. This forces negative eigenvalues, which should be set to zero, having reached the "boundary" of the space, to be positive. Extrapolating to large negative values and then squaring them would likely introduce additional variation with somewhat questionable meaning.

This effect is prevented using the Procrustes geodesic. We show this with an artificial example using covariance operators for boys and girls growth curves in the Berkeley study. Growth curves are functional data, consisting in the height of boys and girls at different ages. This is a benchmark dataset in Functional Data Analysis (see Ramsay and Silverman, 2002) and thus it provides an example of realistic covariance structures. Let S_1 and S_2 be the sample covariance operator for boys and girls curves respectively. Fig. 2.1 shows the minimum eigenvalue for the Hilbert-Schmidt operators $(S_1)^{\frac{1}{2}} + x((S_2)^{\frac{1}{2}} - (S_1)^{\frac{1}{2}})$ and $(S_1)^{\frac{1}{2}} + x((S_2)^{\frac{1}{2}}\widetilde{R} - (S_1)^{\frac{1}{2}})$, $x \in (0, 10)$. The former continuously decreases, while the latter correctly stabilizes at zero. Applying the backward map to the space of covariance operator, the square root geodesic thus presents an artifact positive eigenvalue, while the Procrustes geodesic does not suffer from this problem.

2.2 A permutation test for two - sample comparison of the covariance structure

In this section we show a second example of how the proposed distances can be used in an inferential procedure, i.e. the comparison of covariance operators of two groups. Let us consider two samples of random curves. Curves in the first sample $f_1^1(t), \ldots, f_{n_1}^1(t) \in L^2(\Omega)$ are realisations of a random process with mean $\mu(t)$ and covariance operator Σ_1. Curves in the second sample $f_1^2(t), \ldots, f_{n_2}^2(t) \in L^2(\Omega)$ are realisations of a random process with mean $\mu(t)$

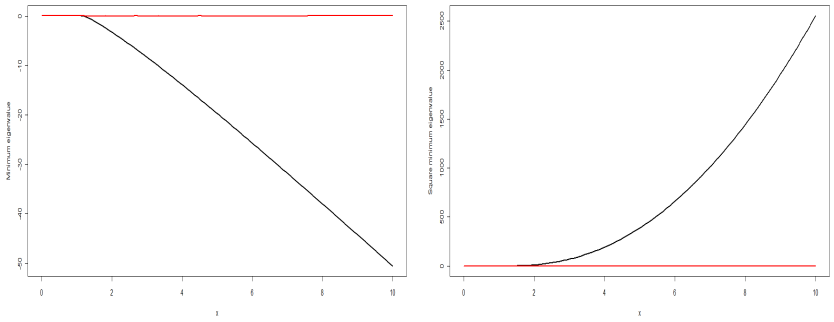

Figure 2.1: Left: Minimum eigenvalue for the extrapolated Hilbert-Schmidt operator for the square root geodesic (black) and the Procrustes geodesic(red). Right: Correspondent square eigenvalue for $S_R(x)$ (black) and $S_P(x)$ (red).

and covariance operator Σ_2. We would like to test the hypothesis

$$\mathrm{H}_0 : \Sigma_1 = \Sigma_2 \ \text{ vs } \ \mathrm{H}_1 : \Sigma_1 \neq \Sigma_2.$$

Recent works (Panaretos et al., 2010; Fremdt et al., 2012) proposed testing procedures for the equality of covariance structures by defining a test statistic through the Karhunen-Loéve expansions of the pooled covariance operator. Let S_1 and S_2 the sample covariance operators, the pooled covariance operator is

$$R = \frac{n_1}{n_1 + n_2} S_1 + \frac{n_2}{n_1 + n_2} S_2.$$

Having centered observation so that they have zero means, Panaretos et al. (2010) propose to use as test statistic

$$T = \frac{n_1 + n_2}{2} \widehat{\theta}(1 - \widehat{\theta}) \sum_{i,j=1}^{p} \frac{\langle (S_1 - S_2) v_i, v_j \rangle^2}{(\widehat{\theta}\lambda_i^1 + (1 - \widehat{\theta})\lambda_i^2)(\widehat{\theta}\lambda_j^1 + (1 - \widehat{\theta})\lambda_j^2)},$$

where $\widehat{\theta} = \frac{n_1}{n_1 + n_2}$ and $\{v_k\}_{k=1}^p$ are the first p eigenfunctions of R and

$$\lambda_l^q = \frac{1}{n_q} \sum_{k=1}^{n_q} \langle f_k^q, v_l \rangle^2,$$

for $q = 1, 2$ and $l = 1, \ldots, p$. Thus, this test statistic considers the projections of the difference operator $S_1 - S_2$ in the subspaces generated by the first p eigenfunctions of the pooled covariance operator R, normalized with respect to an estimate of the variability of the curves in those subspaces. Under gaussian assumptions, these authors show that T has an asymptotical chi-squared distribution with

$p(p+1)/2$ degrees of freedom. Fremdt et al. (2012) generalize this approach to deal also with non-gaussian process.

However, some difficulties arise in practical applications with these testing procedures. Firstly, the test statistic above depends on the chosen number of eigenfunctions to be considered and in practice this may lead to ambiguity in the decision to take, different values of p providing different results (see Fremdt et al., 2012, Section 4), even if automatic procedures can be set up for the choice of the optimal (in some sense) number of eigenfunctions (Panaretos et al., 2010). Secondly, these kinds of procedures have only asymptotic properties, while in Functional Data Analysis is often the case where the sample size is relatively small and the finite sample performance may be not sufficiently accurate in the non gaussian case (Fremdt et al., 2012).

Thus, we propose here a different approach, aiming to overcome these limitations. We would like to use the distance between two sample covariance operators to carry out inference on the difference between the true covariance operators. The complicated expression of the available distances makes it difficult to elicit their distributions, even when random curves are generated from a known parametric model. Thus, we propose to resort to a non parametric approach, namely permutation tests.

Permutation tests are non parametric tests which rely on the fact that, if there is no difference among experimental groups, the labeling of the observations is completely arbitrary. Therefore, the null hypothesis that the labels are arbitrary is tested by comparing the test statistic with its permutation distribution, i.e. the value of the test statistics for all possible permutations of labels. In practice, only a subset of permutations, chosen at random, is used to assess the distribution. A sufficient condition to apply this permutation procedure is exchangeability: under the null hypothesis, curves can be assigned indifferently to any group.

We reformulate the test using distances between covariance operators,

$$\mathrm{H}_0 : d(\Sigma_1, \Sigma_2) = 0 \ \text{ vs } \ \mathrm{H}_1 : d(\Sigma_1, \Sigma_2) \neq 0.$$

Let S_1 and S_2 be the sample covariance operators of the two groups. We use $d(S_1, S_2)$ as a test statistic, since large values of $d(S_1, S_2)$ are evidence against the null hypothesis. If we considered all the possible permutation of the assigned labels, we would obtain an exact p-value as the proportion of the distances between the sample covariance operators of the permuted curves greater than or equal to $d(S_1, S_2)$. However, the total number of permutations is usually computationally unfeasible and thus we resort to Monte Carlo method. We consider M random permutation of the labels $\{1, 2\}$ on the sample curves and we compute $d(S_1^{(m)}, S_2^{(m)})$, $m = 1, \ldots, M$, where $S_i^{(m)}$ is the sample covariance operator for the group indexed with label i in permutation m. Now the p-value of the test can

be approximated with the proportion of $d(S_1^{(m)}, S_2^{(m)})$ which are greater than or equal to $d(S_1, S_2)$,

$$\text{p} - \text{value} = \frac{\sum_{m=1}^{M} \mathbb{I}_{\{d(S_1^m, S_2^m) > d(S_1, S_2)\}}}{M} = \frac{1}{M} \sharp\{d(S_1^m, S_2^m) > d(S_1, S_2)\}.$$

For this formulation of the permutation test, equality of mean functions is essential. However, if the two groups have different (and unknown) means, an approximated permutation test can be performed, having first centered the curves using their sample means. This is a common strategy for testing scaling parameters, such as variance, for univariate real random variables (see, e.g., Good, 2005, Section 3.7.2). The test obtained is approximate in the sense that the nominal level of the test is exact only asymptotically for $n_1, n_2 \to +\infty$. This happens because the observations are only asymptotically exchangeable, due to the fact that $\hat{\mu}_i(t) = \frac{1}{n_i} \sum_{k=1}^{n_i} f_k^i(t) \to \mu_i(t)$ and therefore centered observations asymptotically do not depend on the original labels.

2.2.1 Simulation studies

We now consider simulation studies to explore the behaviour of the different distances with various modifications of the covariance structure. All the curves are simulated on $[0, 1]$ with a Gaussian process with mean $\sin(x)$ and covariance function Σ_1 and Σ_2 respectively. Observations are generated on a grid of $p = 32$ points with six different sample sizes $N = 5, 10, 20, 30, 40, 50$. Each permutation test is performed with $M = 1000$ and the test is repeated for 200 samples, so that we can evaluate the power of the test for different values of sample size and different degrees of violation of the null hypothesis. Since the aim is to investigate the performance of the different distances, we choose this simple simulation set up and two different modifications of the covariance structure. More complete simulation studies would be needed to evaluate the performance of the propose non parametric approach with respect to the existing parametric techniques but this is the scope for future works.

Fig. 2.2 shows the covariance function Σ_1 for the first group in all the simulations (where this covariance was obtained from the male curves within the Berkeley growth curve dataset (Ramsay and Silverman, 2005)).

First simulation: We consider the case in which the first two eigenvalues of Σ_2 are a convex combination of the first two eigenvalues of Σ_1, while the remaining eigenvalues and all the eigenfunctions are the same,

$$\begin{aligned} \lambda_1^2 &= \gamma \lambda_2^1 + (1 - \gamma) \lambda_1^1 \\ \lambda_2^2 &= \gamma \lambda_1^1 + (1 - \gamma) \lambda_2^1 \end{aligned} \tag{2.6}$$

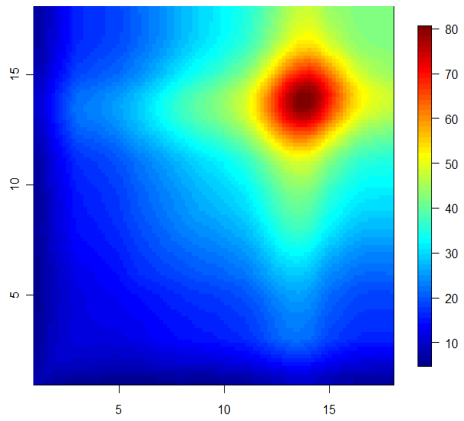

Figure 2.2: Integral kernel of the true covariance function Σ_1 for reference group.

Figure 2.3 shows the estimated power for different values of γ and N.

It is worth to mention that, despite the normality of the data, traditional parametric tests for comparison of covariances can be applied only when $N > p$, i.e. $N = 40, 50$. Indeed, the power is low also in these cases, since the sample size is small with respect to dimension $p = 32$. Note that the square root and Procrustes tests are the most powerful here, and all tests have the correct type I error probability.

Second simulation: We consider now a difference in the total variation between the covariance operators in the two groups, so that

$$\Sigma_2 = (1 + \gamma)\Sigma_1.$$

In this simulation, we can also compare the proposed method with the generalisation of the Levene test (see Anderson, 2006), since this is a procedure to test for differences in multivariate dispersion. The univariate Levene test is an analysis of variances (ANOVA) performed on the deviations from the groups means. In the multivariate case, ANOVA is performed on the set of distances between each observation and the group centroid. In our functional case, the $L^2(\mathbb{R})$ is used as distance and the sample mean as group centroid, being the minimizer of the sum of square distances within the group. Even if the simulated data are indeed Gaussian, we present results for the permutation version of this test to give a fair comparison with our method. Thus, this test uses the ANOVA F-statistics of the $L^2(\mathbb{R})$ distances between curves and group means but the p-value of the test is obtained comparing the observed value with the permutational distribution, as happens for the testing procedure based on operator distances.

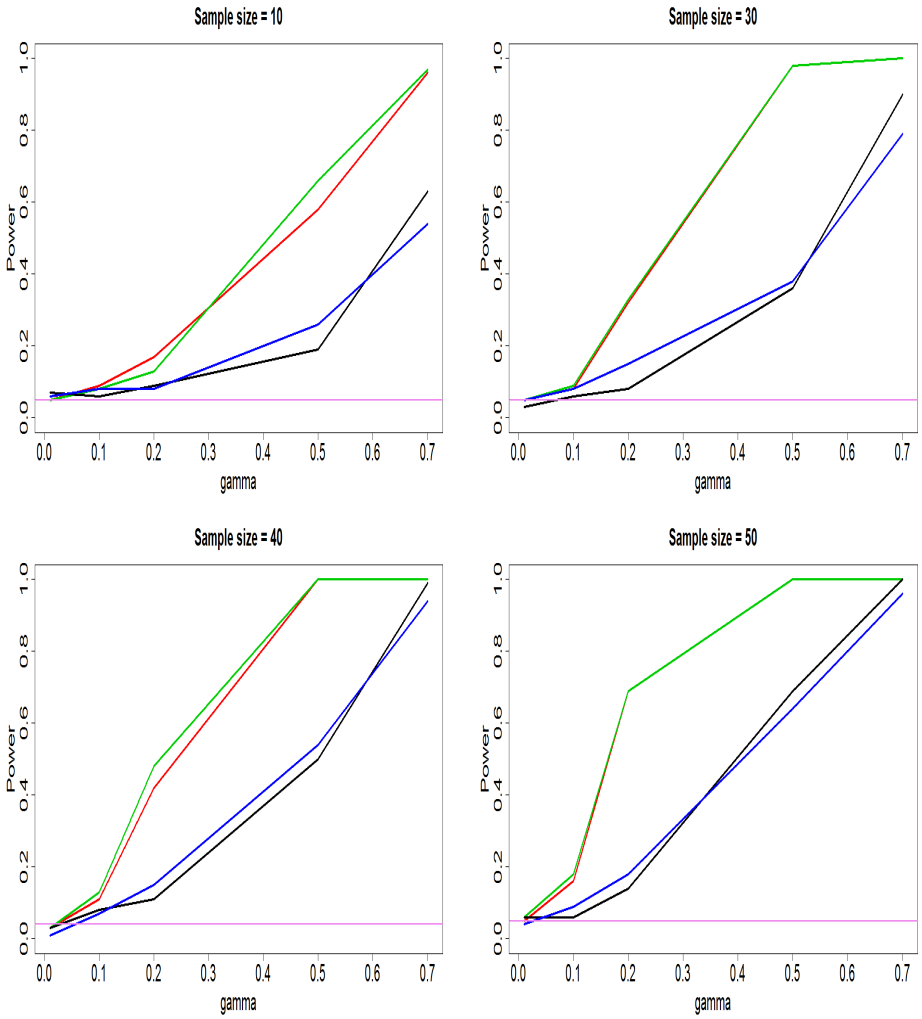

Figure 2.3: Power estimated via simulation for different values of γ and N obtained with Procrustes size-and-shape distance (red line), Square root distance (green line), Kernel $L^2(\Omega \times \Omega)$ distance (black line) and Spectral distance (blue line). The purple line shows the significance level $\alpha = 0.05$.

Fig. 2.4 shows the estimated power for different values of γ and N. For smaller N the Levene test has a high type I error. For larger N and γ, the Procrustes and square root tests are a little more powerful than those based on kernel L^2 and spectral distance, while the Levene test perform slightly worse.

We apply also the proposed permutation test to the comparison of covariance operators of the original Berkeley growth curves for male and female. The test with square root distance result in a p-value of 0.15, while the test with the Pro-

crustes distance has a p-value of 0.19. There is no strong evidence of the difference in covariance structures between the two groups, thus supporting the claim by Ramsay and Silverman (2002) that the difference between the two groups is somehow smaller than expected, being limited to phase variability.

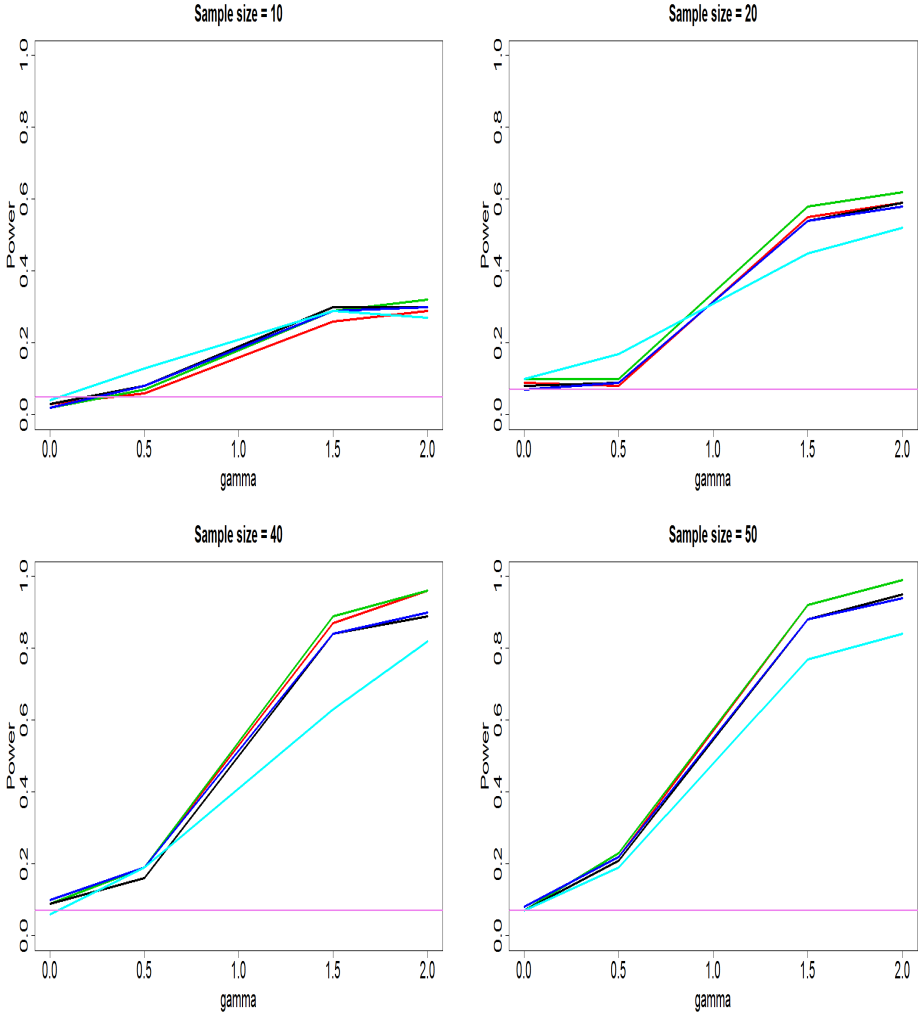

Figure 2.4: Empirical power obtained in the second simulation for different values of γ and N with Procrustes size-and-shape distance (red line), Square root distance (green line), Kernel $L^2(\Omega \times \Omega)$ distance (black line), Spectral distance (blue line) and generalised Levene test (cyan line). The purple line shows the significance level $\alpha = 0.05$.

Chapter 3

Applications

In this chapter, two applications are shown where the analysis of covariance structures may benefit from the distance-based inferential procedure illustrated in the previous chapter. The first application involves biological data, namely radius and curvature of the Inner Carotid Artery for patients suspected of being affected by cerebral aneurysm. Here the aim is to explore if different groups of patients have different covariance structures. The second application involves a linguistic problem, the analysis of relationships among Romance languages, i.e. languages with a common root in the Latin language. Data come from audio recordings of groups of speakers from different languages but the focus of the analysis is the covariance function among frequency intensities in speech, this being considered a significant feature for the language.

3.1 Data from AneuRisk Project

We illustrate here a possible application of the inferential technique described above. We consider data that have been collected within the AneuRisk Project, designed to investigate the role of vessel morphology and blood fluid dynamics on the pathogenesis of cerebral aneurysm (http://mox.polimi.it/it/progetti/aneurisk). A detailed description of the problem can be found in Sangalli et al. (2009a).

The AneuRisk data set is based on a set of three-dimensional angiographic images taken from 65 subjects, hospitalised at Niguarda Ca Granda Hospital (Milan) from September 2002 to October 2005, who were suspected of being affected by cerebral aneurysms. Out of these 65 subjects, three groups can be identified:

- 33 subjects have an aneurysm at or after the terminal bifurcation of the Internal Carotid Artery (ICA) (Upper group)
- 25 subjects have an aneurysm along the ICA (Lower group)

- 7 subjects have not had any aneurysm (No-aneurysm group).

In general, Upper group subjects are those with the most dangerous aneurysms. Qualitative considerations based on fluid dynamic and visual explorations of the data lead Sangalli et al. (2009a) to join the Lower and No-aneurysm groups in a single group, to be contrasted to the Upper group. Here, we explore possible differences between Lower and No-aneurysm groups looking at covariance operators of vessel radius and curvature.

Starting from the angiographies of the 65 patients, estimates of vessel radius and curvature are obtained with the procedure described in Sangalli et al. (2009b), resulting in a free-knots regression splines reconstruction of radius and curvatures. Each patient is therefore described by a pair of functions $R_i(s)$ and $C_i(s)$, $i = 1, \ldots, 65$, where the abscissa parameter s measures an approximate distance along the ICA, from its terminal bifurcation toward the heart (for conventional reasons, this abscissa parameter takes a negative value to highlight that the direction is opposite with respect to blood flow). These curves are defined on different intervals, thus we restrict our analysis to the region which is common to all curves (i.e., for values of abscissa between -25 and -1).

For the moment, we ignore the problem of misalignment among curves, even if we expect this to mask some of the group differences. We show later how the results change if the analysis is performed on registered curves, thus decoupling amplitude and phase variability, as suggested in the original analysis. Fig. 3.1 shows radius and curvature for the two groups, while their covariance operators can be seen in Fig. 3.2. We evaluate the kernels of the covariance operators on an equispaced grid of $p = 512$ points.

We want now to verify the equality of the two groups in terms of covariance structure, since a visual inspection of the covariance operators would seem to indicate differences. A permutation test for equality of radius covariance operators result in a p-value of 0.94 using Procrustes distance and 0.885 with Square root distance. P-values of permutation tests for equality of curvature covariance operators are 0.86 for Procrustes distance and 0.775 for Square root distance. Therefore, there is no statistical evidence for difference of covariance operators between the two groups. Thus, the decision taken in the original analysis to treat the curves as being from a single group is not rejected.

Let us now consider the two groups used in Sangalli et al. (2009a), i.e. patients with aneurysm in the upper part of the artery and patients with aneurysm in lower part of the artery or no aneurysm at all, to understand if our method can highlight differences on the covariance structures. Fig. 3.3 shows radius and curvature covariance operators. Performing the permutation test on radius covariance operators for these two groups, we find p-values less than 0.0001 both for Procrustes distance and Square root distance. For the curvature covariance operators, we

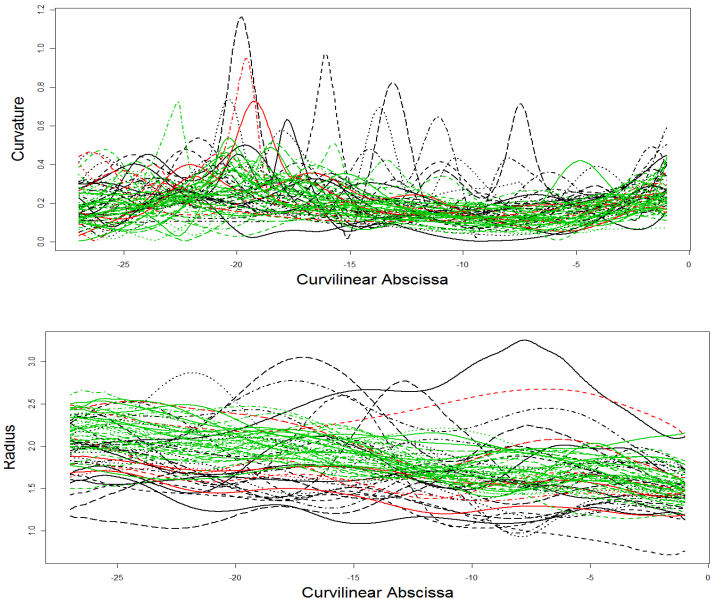

Figure 3.1: Curvature (top) and Radius(bottom) for the 65 patients in the range of abscissa common to all curves. Black coloured curves correspond to patients of the Lower group, red coloured curves to patients of the No-aneurysm group and green coloured curves to patients of the Upper group.

obtain p-values of 0.1 for Procrustes distance and 0.02 for Square root distance. Thus, both distances provide far smaller p-values than in the previous case indicating that the difference between these two groups is worth investigation. However, the evidence is somewhat weak for curvature, as the Procrustes distance is only significant at $p \leq 0.1$, with this distance being free from the arbitrary choice of decomposition.

As mentioned before, some of the difference between the two groups may also be masked by a misalignment of the curves. Sangalli et al. (2009a) pointed out that this data present a phase variability, due to biological difference among patients, that is ancillary to the analysis. For this reason, we perform the test also on the curved aligned by the method proposed in Sangalli et al. (2009a). This method focuses on the first derivative of the coordinates of the vessel centerline. It looks for the affine linear transformations on the curvilinear abscissa that maximizes an appropriate measure of similarity among patients. Thus, we obtain the warping functions h_i, $i = 1, \ldots, 65$ and a new set of radius and curvature curves $R_i(h_i(s))$ and $C_i(h_i(s))$, $i = 1, \ldots, 65$.

Fig. 3.5 shows the covariance operators of radius and curvature for the aligned curves for the group of patients with aneurysm in the lower part of the ICA and

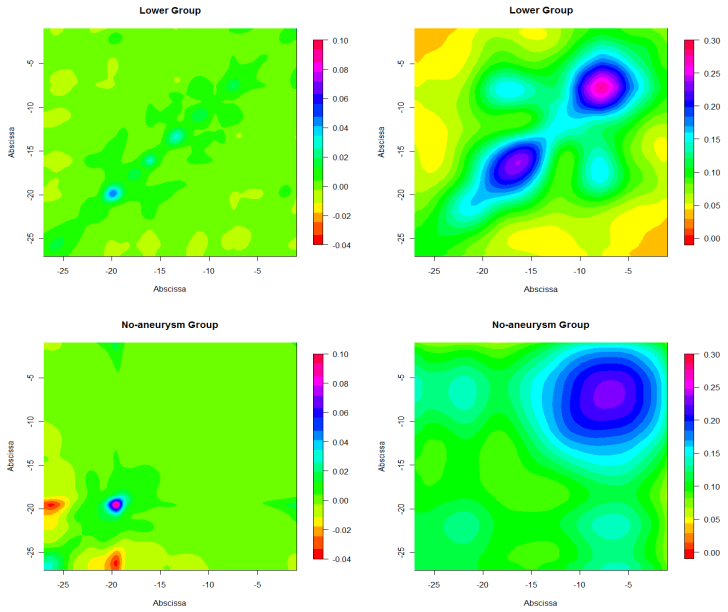

Figure 3.2: Covariance operator for curvature (left) and radius (right) for the for the Lower (first row) and No aneurysm (second row) groups.

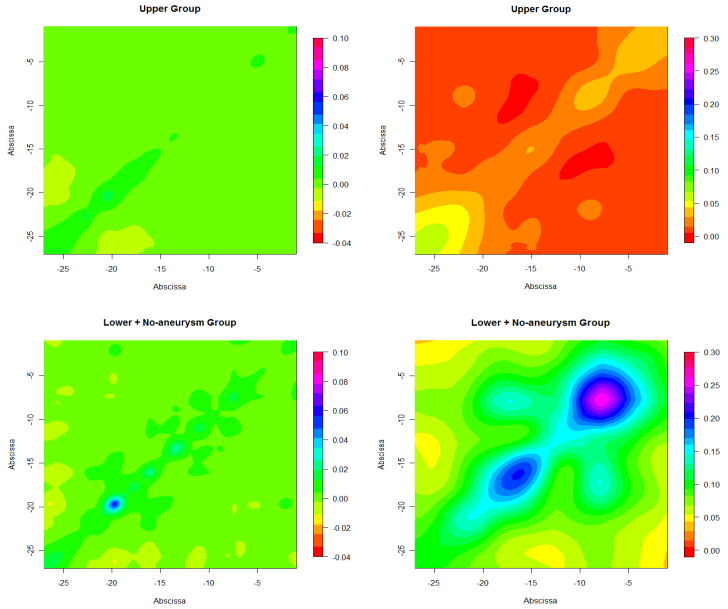

Figure 3.3: Covariance operator for curvature (left) and radius (right) for the Upper (first row) and Lower or No aneurysm (second row) groups.

for those with no aneurysm. In this case, testing the difference of radius covariance functions between these two groups results in p-values of 0.855 with Square root distance and 0.845 with Procrustes distance, while testing the difference in curvature covariance functions gives p-values of 0.61 with Square root distance and 0.54 with Procrustes distance. Fig. 3.6 shows instead the covariance operators of the aligned radius and curvature for the group of patients with aneurysm in the upper part of the ICA and for the group formed by both patients with no aneurysm and patients with aneurysm in the lower part of the ICA.

The permutation tests result in p-values lesser than 0.005 for both Square root distance and Procrustes distance and for both radius and curvature covariance functions. Indeed, having dealt with curves misalignment makes more evident the difference between these two groups (patients with aneurysm after the terminal bifurcation of the ICA and patients with aneurysm along ICA or no aneurysm) in terms of covariance structure, thus supporting results in Sangalli et al. (2009a). Moreover, this confirms also that the registration procedure is successful in separating phase and amplitude variability.

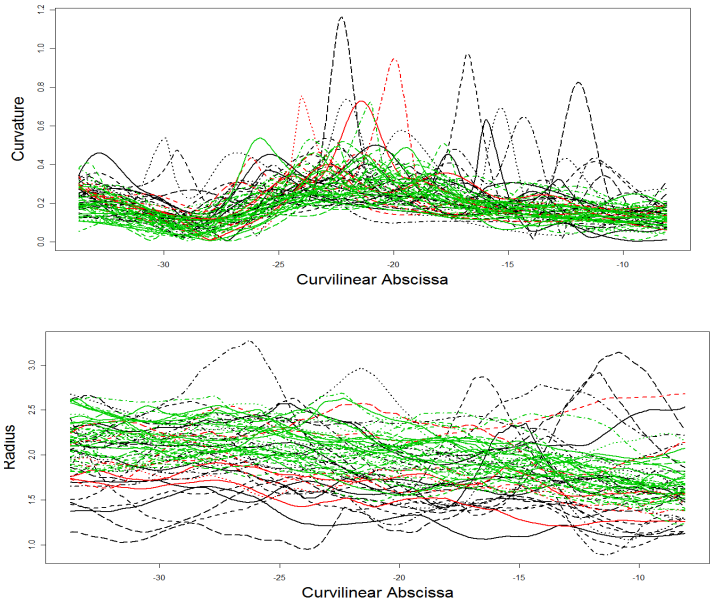

Figure 3.4: Registered Curvature (top) and Radius(bottom) for the 65 patients. Black coloured curves correspond to patients of the Lower group, red coloured curves to patients of the No-aneurysm group and green coloured curves to patients of the Upper group.

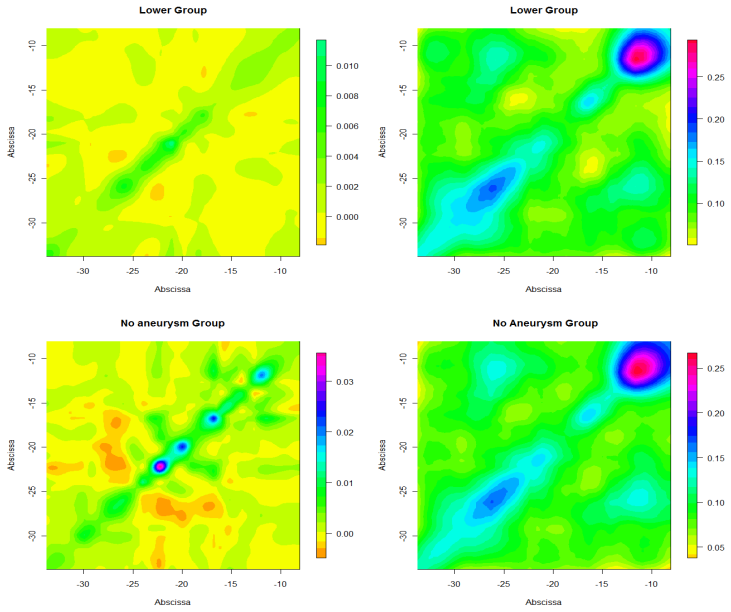

Figure 3.5: Covariance operator for curvature (left) and radius (right) for the Lower (first row) and No aneurysm (second row) groups in the case of aligned curves.

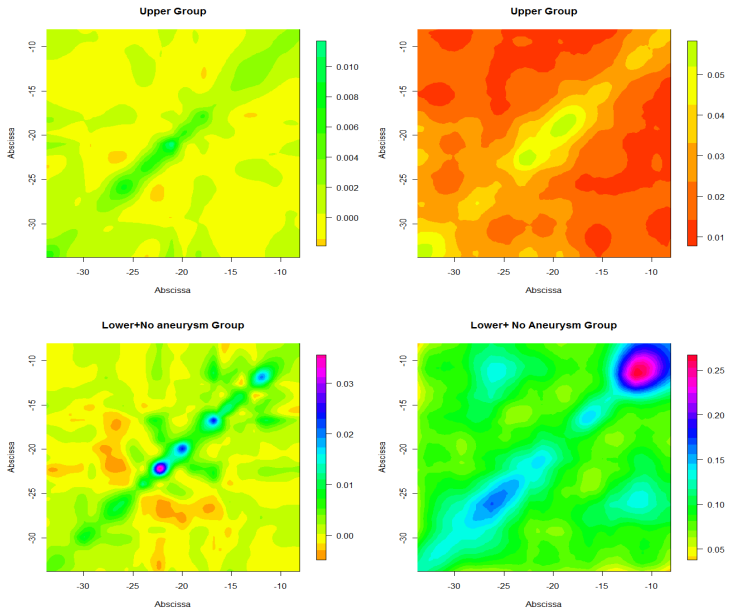

Figure 3.6: Covariance operator for curvature (left) and radius (right) for the Upper (first row) and Lower or No aneurysm (second row) groups in the case of aligned curves.

3.2 Exploring relationships among Romance languages

The traditional way of exploring relationships across languages consists of examining textual similarity. However, this neglects phonetic characteristics of the languages. Here a novel approach is proposed to compare languages on the basis of phonetic structure. People speaking different Romance languages (French, Italian, Portuguese, Iberian Spanish and American Spanish) are registered while pronouncing words in each language. The output of the registration for each word and for each speaker consists of the intensity of the sound over time and frequencies. The aim is to use this data to explore linguistic hypotheses concerning the relationships between different languages. However, while the temporal aspects of each individual word are important, we will concentrate on frequencies. Previous studies (Aston et al., 2010; Hadjipantelis et al., 2012) have indicated that covariance operators characterise languages well, and these will be the object of study. The operators summarise phonetic information for the language, while disregarding characteristics of singular speakers and words. For the scope of this work, we focus on the covariance operators among frequencies in the log-spectrogram, estimated by putting together all speakers of the language in the data set. The spectrogram is a two dimensional time-frequency image which gives localised time and frequency information across the word. We consider different time points as replicates of the same covariance operator among frequencies. It is clear that this is a significant simplification of the rich structure in the data but in itself can already lead to some interesting conclusions.

Let $f_{ijk}(t) \in L^2(\Omega)$ be a realisation of a random process, where $i = 1, \ldots, L$ are different languages, $j = 1, \ldots, n$ the groups (i.e. different time points) and $k = 1, \ldots, K$ the observations (individual speakers). As mentioned above, the working hypothesis is that the significant information of the different languages are in the language-wise covariances S_i rather than in the individual observations f_{ijk}. Here some preliminary results are reported, focusing on the covariance operator for the word "one" spoken across the different languages by a total of 23 speakers across the five languages. This word is similar in each language (coming from the common Latin root), but different enough to highlight changes across the languages (American Spanish: uno; French: un; Iberian Spanish: uno; Italian: uno; Portuguese: um). This also highlights that in this case, textual analysis is difficult as three of the languages have the same textual representation.

Fig. 3.7, 3.8 and 3.9 show the covariance operators estimated for each language with the estimator (2.2), using Square root distance Procrustes distance and Kernel distance respectively. Fig. 3.10 shows dissimilarity matrix among estimated covariance operators for each language and the corresponding dendrograms obtained with complete linkage. Indeed, it seems that focusing on the covariance operator allows the capture of some significant information about languages. Re-

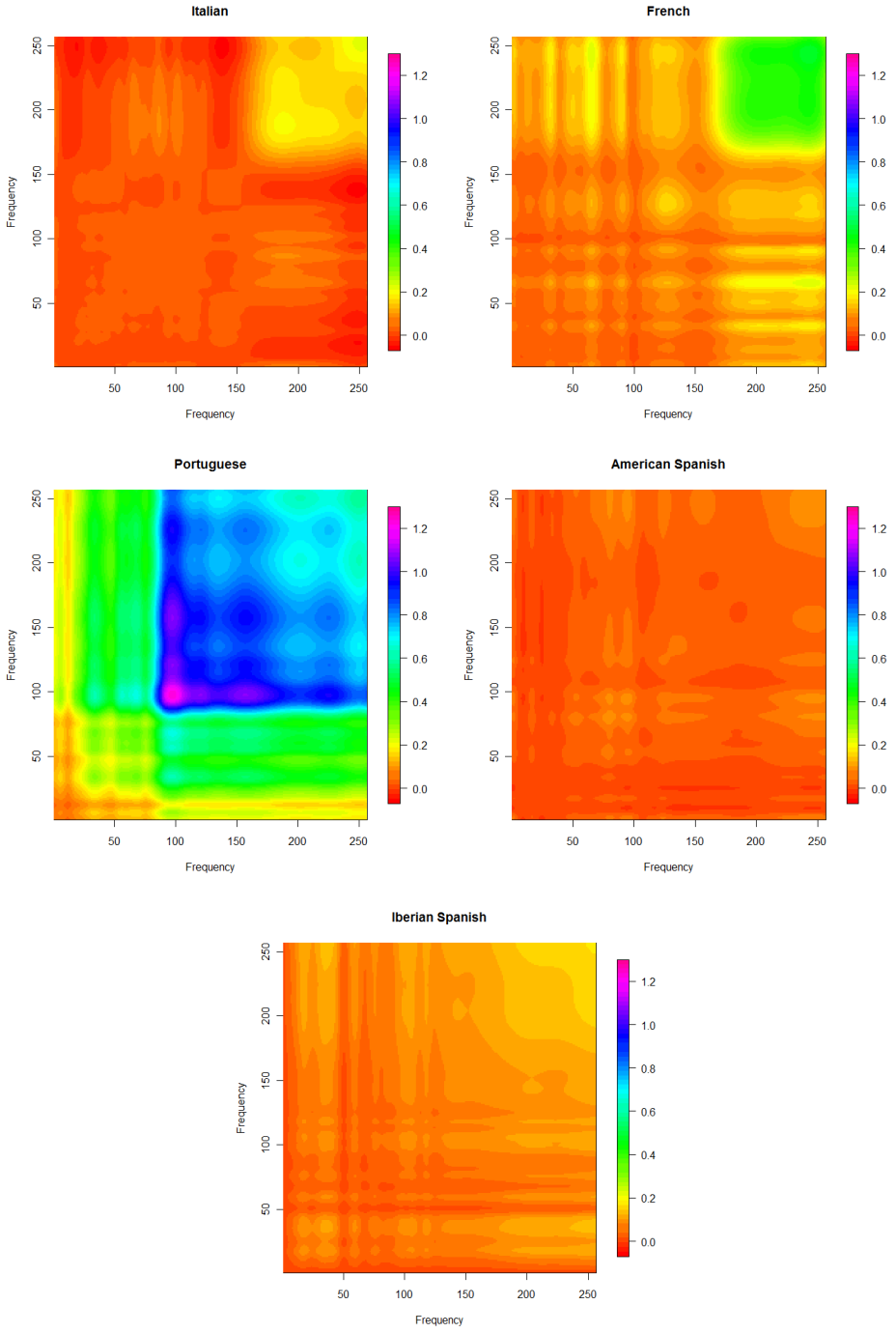

Figure 3.7: Estimates of frequency covariance operators for the five romance languages, using Square root distance.

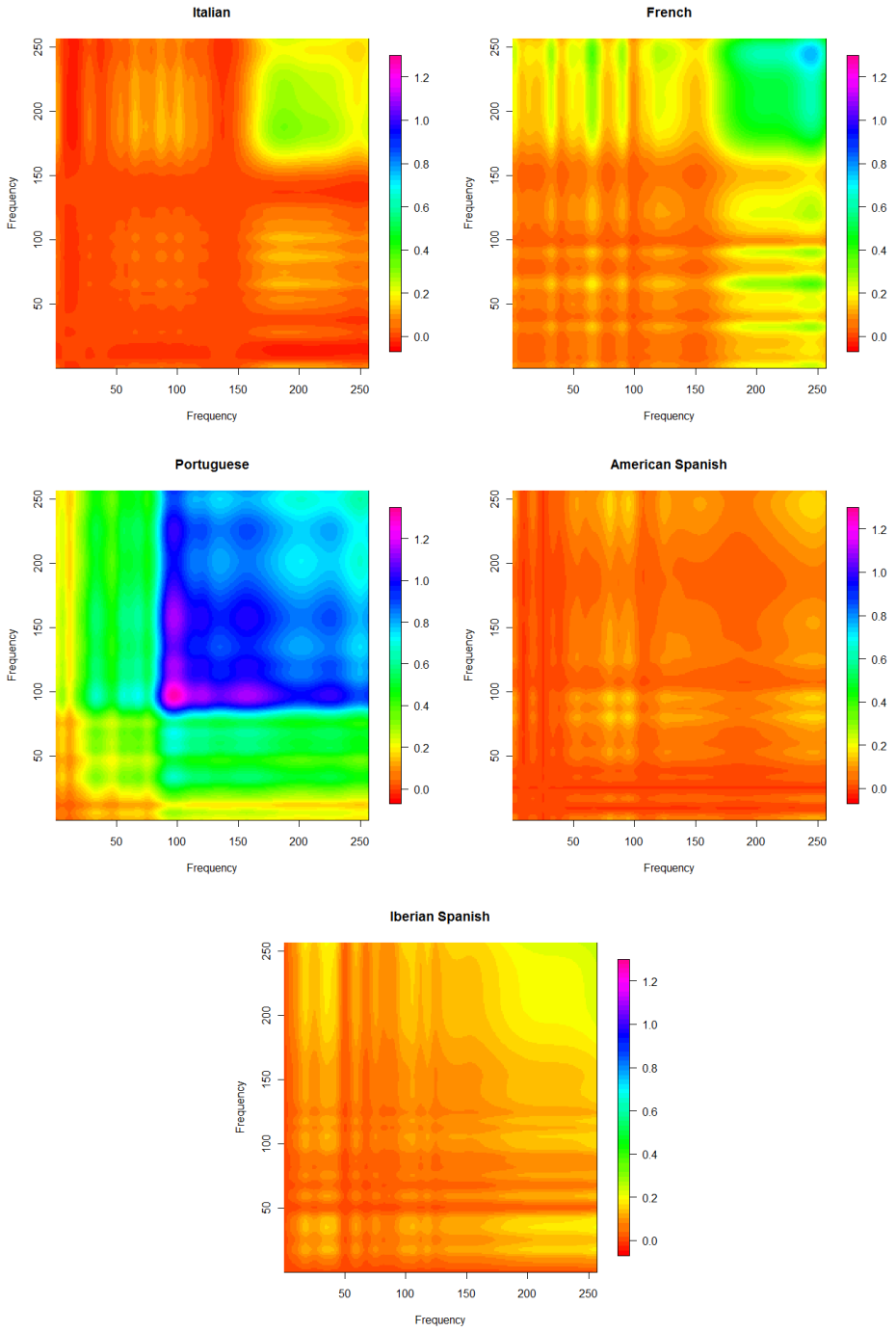

Figure 3.8: Estimates of frequency covariance operators for the five romance languages, using Procrustes distance.

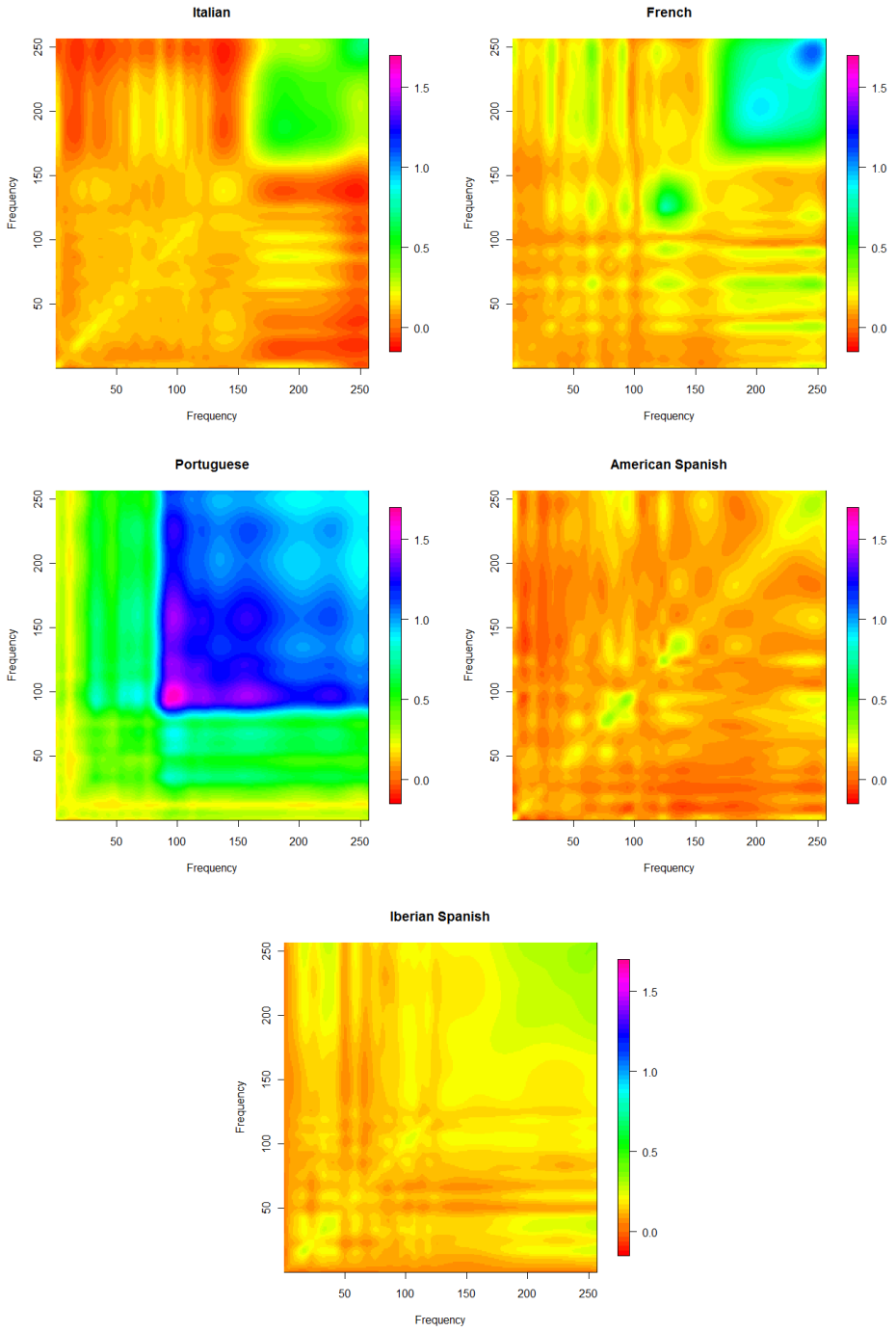

Figure 3.9: Estimates of frequency covariance operators for the five romance languages, using Kernel distance.

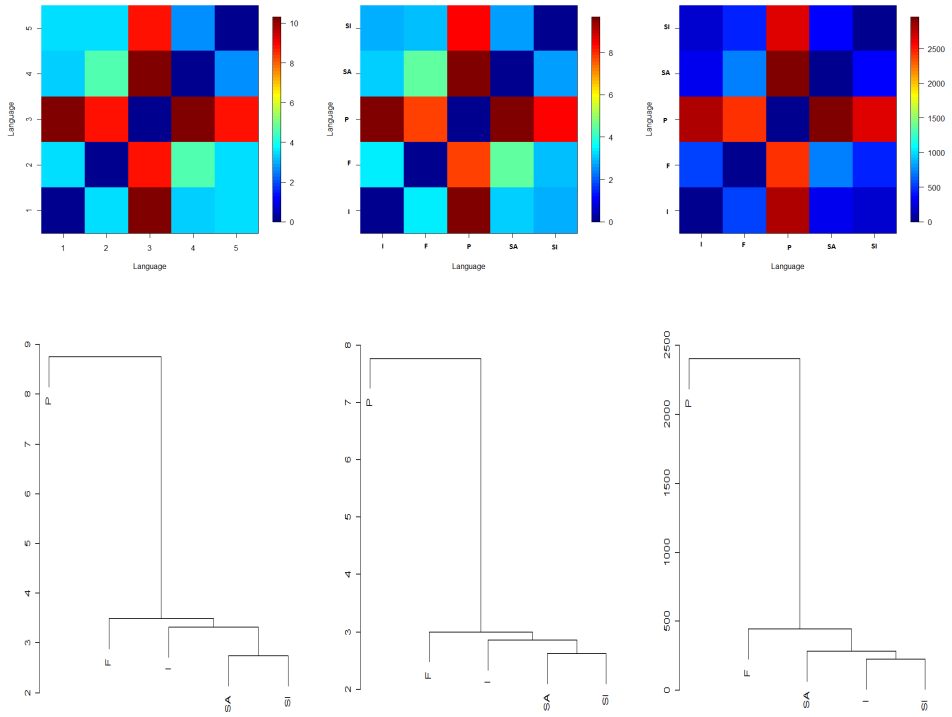

Figure 3.10: First row: Distance matrix among Fréchet estimates, obtained with Square root distance (left), Procrustes distance (center) and Kernel distance (right), where I=Italian, F=French, P=Portuguese, SA=American Spanish, SI=Iberian Spanish. Second row: Correspondent dendrograms obtained with a complete linkage.

lationships among covariance operators have features which are expected by linguistic hypotheses, such as strong similarity between the two Spanish varieties (American and Iberian) and Italian, which are correctly found using both Square root distance and Procrustes distance. These two distances provide essentially the same distance structure among languages, while Kernel distance matrix is slightly different and in worst agreement with existing linguistic knowledge. However, not all our conclusions directly support textual analysis. The distance of Portuguese from both Spanish languages is greater than expected, Moreover, for historical reasons American Spanish is expected to be nearer than Iberian Spanish to Italian, but the covariance structures indicate this is reversed. We can also compare the distance between covariance operators with a distance elicited by linguistic experts, based on geographical and historical information on languages evolution.

Fig. 3.11 reports a scatterplot of the square root distance between covariance operators and this geo-historical distance for every pairs of languages. As expected, pairs including Portuguese language behave very differently, this reflecting the exceptional difference in frequency covariance structure not unexpected by linguistic experts.

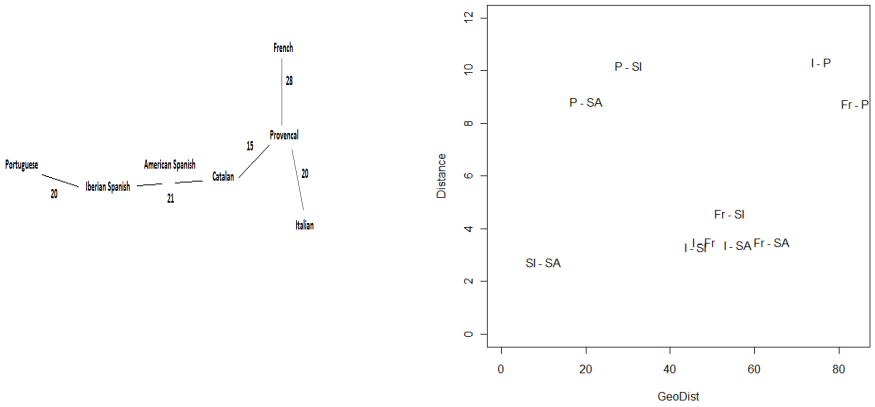

Figure 3.11: Left: Distances among languages based on historical and geographical information. Right: Scatterplot of geographical distance v.s. square root distance between covariance operators for each pair of languages.

Thus, as this analysis is currently based on the word "one", providing further assessment using a much larger corpus will be of significant interest, and is the subject of ongoing work.

A particularly interesting objective of the analysis is to provide insight into the change of the frequency structure along the path of language evolution. This would be inherently linked to extrapolation based on the distances we have proposed. The Portuguese language presents a very different covariance structure with respect to the other Romance languages. Therefore, it would be of interest to compare its frequency covariance operator with the one extrapolated from the covariances of the two Spanish languages, to see if this kind of covariance was expected (and a linear model of distance appropriate). In the opposite direction of the evolutionary path, we try also to compare Italian frequency covariance operators with the one extrapolated from the two Spanish varieties. First, we use in the extrapolation the distance-based estimates obtained with the same distance that determines the geodesic. The extrapolated covariance operator for Portuguese (P) is obtained with the method proposed in Section 2.1.3, evaluating the extrapolation line from Iberian Spanish (SI) operator to American Spanish (SA) operator at $x = d(S_{SA}, S_P)/d(S_{SA}, S_{SI})$. For Italian (I), we evaluate the line from American

Spanish to Iberian Spanish at $x = d(S_{SI}, S_I)/d(S_{SA}, S_{SI})$.

Fig. 3.12 shows the integral kernels of the extrapolated operators and the estimate obtained from Portuguese data with the same distance. Fig. 3.13 reports the same pictures for the Italian case. Table 3.1 reports the comparison between Portuguese covariance operator and the correspondent geodesic estimate obtained from Iberian and American Spanish. In this case, Procrustes extrapolation provides a better results with respect to all the considered distances. On the contrary, Square root extrapolation behaves slightly better in the extrapolation towards Italian covariance operator (see Table 3.3). This is not surprising, since the extrapolation for Italian covariance operator is a "short distance" extrapolation, where the square root mapping does not introduce large artificial effects.

We consider also the application of the three extrapolation methods to the same covariance estimates to decouple the effect of the estimation and those of the extrapolation. Table 3.2 and 3.4 show the comparison between the extrapolation starting from kernel-based estimates of the covariance operators of the two Spanish varieties and the kernel-based estimates of Portuguese and Italian respectively. Here the Procrustes method is shown to be better both for Portuguese and for Italian, even if the advantage is far greater for the former. Thus, we can conclude that the extrapolation based on Procrustes geodesic is to be preferred, as expected from the theory.

Thus, we can conclude that some features of the most "extreme" language in the family can be expected, such as a higher variability in the high frequency. On the other hand, unexpected features are also present, for example a higher variability also in the middle range frequencies for Portuguese, and these are worth of deeper linguistic explorations.

Table 3.1: Comparison between estimates Portuguese covariance operator and the extrapolation from the two Spanish varieties.

	Square root geodesic	Procrustes geodesic	Kernel extrapolation
Square root estimate	10.57292	8.022398	NaN
Procrustes estimate	11.03729	8.186329	NaN
Kernel estimate	12.15648	8.95952	NaN

Square root distance

	Square root geodesic	Procrustes geodesic	Kernel extrapolation
Square root estimate	10.28457	7.45902	NaN
Procrustes estimate	10.77405	7.631172	NaN
Kernel estimate	11.78495	8.33287	NaN

Procrustes distance

	Square root geodesic	Procrustes geodesic	Kernel extrapolation
Square root estimate	2161.635	725.2036	2187.593
Procrustes estimate	2349.609	870.1691	2375.824
Kernel estimate	3004.293	1448.528	4471.765

Kernel distance

Table 3.2: Comparison between kernel-based estimate of Portuguese covariance operator and the extrapolation from kernel-based estimates of the two Spanish varieties.

	Square root geodesic	Procrustes geodesic	Kernel extrapolation
Square root distance	14.5469	13.42664	NaN
Procrustes distance	13.83603	12.77588	NaN
Kernel distance	3107.697	2524.992	5372.538

Table 3.3: Comparison between estimates of Italian covariance operator and the extrapolation from the two Spanish varieties.

	Square root geodesic	Procrustes geodesic	Kernel extrapolation
Square root estimate	3.552607	4.192906	NaN
Procrustes estimate	4.093186	4.179287	NaN
Kernel estimate	6.799505	6.370495	NaN

Square root distance

	Square root geodesic	Procrustes geodesic	Kernel extrapolation
Square root estimate	3.132686	3.674005	NaN
Procrustes estimate	3.613229	3.319093	NaN
Kernel estimate	6.494973	6.002709	NaN

Procrustes distance

	Square root geodesic	Procrustes geodesic	Kernel extrapolation
Square root estimate	111.4949	275.9129	282.5581
Procrustes estimate	228.2282	233.6398	239.9034
Kernel estimate	479.4734	250.2278	424.3027

Kernel distance

Table 3.4: Comparison between kernel-based estimate of Italian covariance operator and the extrapolation from kernel-based estimates of the two Spanish varieties.

	Square root geodesic	Procrustes geodesic	Kernel extrapolation
Square root distance	6.282191	6.228176	NaN
Procrustes distance	5.755854	5.69689	NaN
Kernel distance	329.8802	258.4509	232.9852

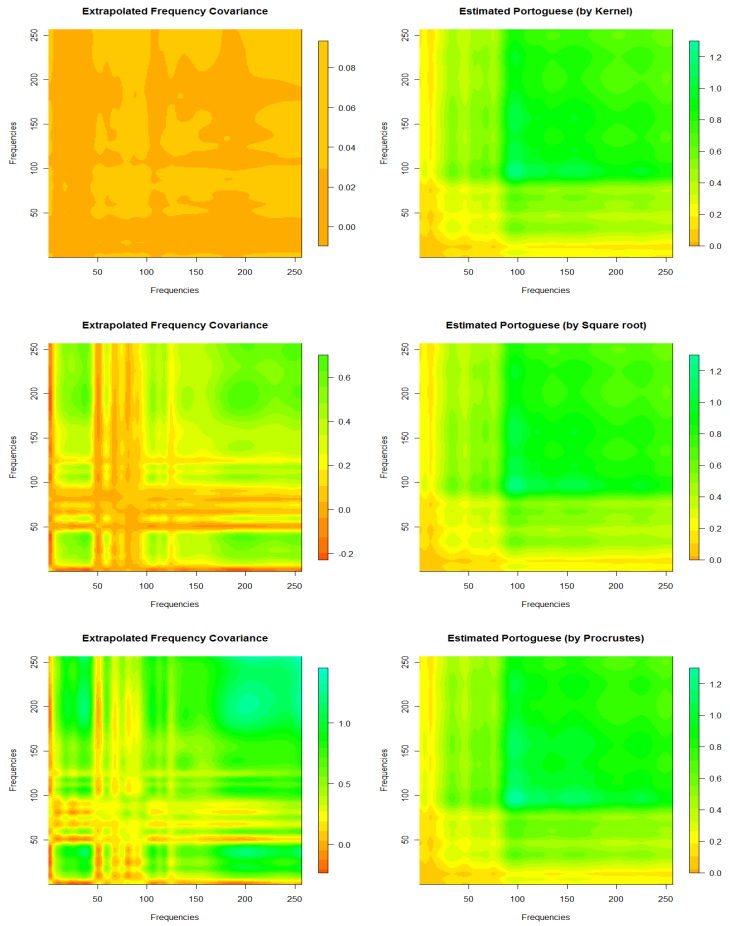

Figure 3.12: First row: Kernel extrapolated for Portuguese from the two Spanish languages using equation (2.3) (left). It is not a valid kernel for covariance operators, since the associated integral operator is not non negative definite. Sample covariance function for Portuguese speakers (right). Second Row: Covariance operator extrapolated for Portuguese from the two Spanish languages with the square root mapping of equation (2.4) (left) and covariance operator estimated from Portuguese speakers using equation (2.2) and square root distance (right). Third row: Covariance operator extrapolated for Portuguese from the two Spanish languages with the Procrustes alignment in equation (2.5) (left) and covariance operator estimated from Portuguese speakers using equation (2.2) and Procrustes distance(right).

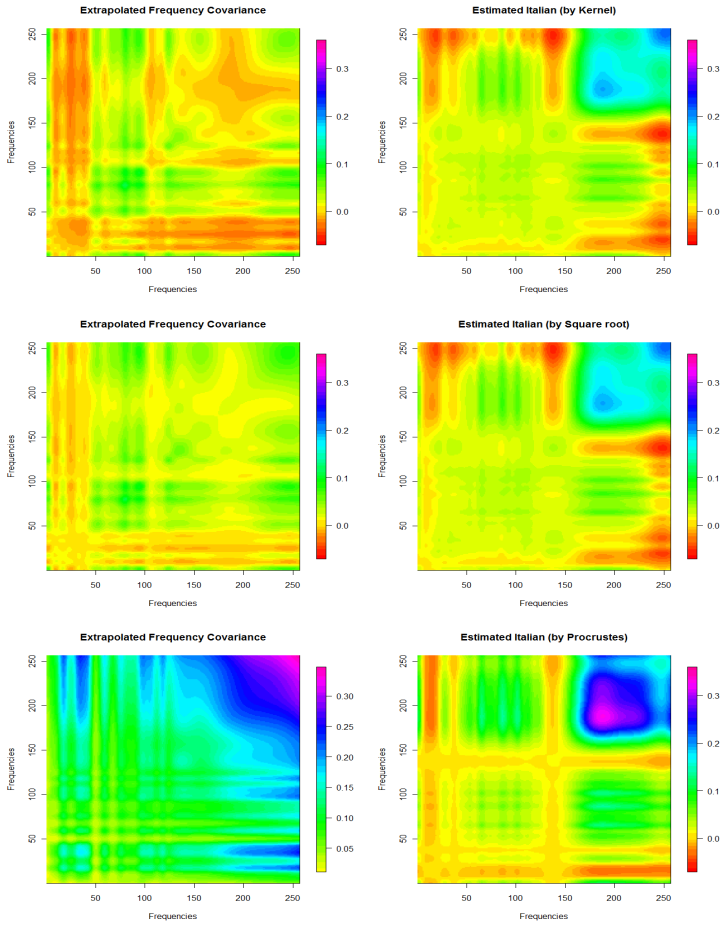

Figure 3.13: First row: Kernel extrapolated for Italian from the two Spanish languages using equation (2.3) (left). It is not a valid kernel for covariance operators, since the associated integral operator is not non negative definite. Sample covariance function for Italian speakers (right). Second Row: Covariance operator extrapolated for Italian from the two Spanish languages with the square root mapping of equation (2.4) (left) and covariance operator estimated from Italian speakers using equation (2.2) and square root distance (right). Third row: Covariance operator extrapolated for Italian from the two Spanish languages with the Procrustes alignment in equation (2.5) (left) and covariance operator estimated from Italian speakers using equation (2.2) and Procrustes distance(right).

Part II

Spatial statistics for covariance matrices

This part of the dissertation is focused on the development of spatial statistical methods for Riemannian data, with a particular focus on the case of covariance matrices. Little attention has been paid to this problem, while in many applications data are spatially distributed. In the general context of complex data, this issue has recently received much attention within the field of functional data analysis (see Delicado et al., 2012; Gromenko et al., 2012; Menafoglio et al., 2012) but the extension to non Euclidean data is even a greater challenge because they do not belong to a vector space.

We move here the first steps in the direction of developing a statistics theory for data belonging to a Riemannian manifold. Tools for the description of spatial dependence are proposed and the problem of estimation of the mean in the presence of spatial dependence is addressed. Chapter 5 introduces a semivariogram for covariance matrix data and an estimator for the mean from a sample of spatially correlated covariance matrices. A model for generating samples from a random field of spatially correlated positive definite matrices is illustrated and simulated data are used to evaluate the performance of the proposed estimator of the mean. If observations are spatially located on an irregular grid, this method provides better estimates than those obtained ignoring spatial dependence. We apply this method to obtain a better estimate of the covariance matrix between temperature and precipitations in the province of Quebec, Canada.

Then, in Chapter 6, a local tangent space approximation is used for describing a non stationary random field taking value in a Riemannian manifold. The choice of a local linear approximation is needed because the definition of a deterministic drift and a correlated error term directly on the manifold is not straightforward. The methods here introduced rely only on the definition of a distance among data and on the locally Euclidean structure of the manifold. Thus, applications to any Riemannian manifold is possible, once the geometry of the manifold has been carefully considered.

Chapter 4

Statistical analysis of positive definite symmetric matrices

Positive definite symmetric matrices are an important instance of data belonging to a Riemannian manifold. In this section, we introduce notation and a few metrics, together with their properties, that we deem useful when dealing with data that are positive definite symmetric matrices. A broad introduction to the statistical analysis of this kind of data can be found, e.g., in Pennec et al. (2006) or Dryden et al. (2009).

Let $PD(p)$ indicate the Riemannian manifold of positive definite symmetric matrices of dimension p. It is a convex subset of $\mathbb{R}^{p(p+1)/2}$ but it is not a linear space: in general, a linear combination of elements of $PD(p)$ does not belong to $PD(p)$. Moreover, the Euclidean distance in $\mathbb{R}^{p(p+1)/2}$ is not suitable to compare positive definite symmetric matrices (see Moakher, 2005, for details). Thus, more appropriate metrics need to be used for statistical analysis. A good choice could be the Riemannian distance: the shortest path between two points on the manifold. A description of the properties of Riemannian manifolds in general, and of $PD(p)$ in particular, can be found in Moakher and Zéraï (2011) and references therein.

Let $Sym(p)$ be the space of symmetric matrices of dimension p. The tangent space to the manifold of positive definite symmetric matrices of dimension p in the point $S \in PD(p)$ is $T_S PD(p) = Sym(p)$. It is worth to recall that for every pair $(S, A) \in PD(p) \times Sym(p)$, it exists an unique geodesic curve $\gamma(t)$ such that

$$\begin{aligned} \gamma(0) &= S \\ \gamma'(0) &= A \end{aligned}$$

and this curve has the expression

$$\gamma(t) = S^{\frac{1}{2}} \exp(tS^{-\frac{1}{2}} A S^{-\frac{1}{2}}) S^{\frac{1}{2}}$$

where $\exp(C)$ indicates the exponential matrix of $C \in Sym(p)$. The exponential map of $PD(p)$ in S is defined as the point at $t = 1$ of this geodesic, thus

$$\exp_S(A) = S^{\frac{1}{2}} \exp(S^{-\frac{1}{2}} A S^{-\frac{1}{2}}) S^{\frac{1}{2}}.$$

Thus, the exponential map takes the geodesic passing through S with "direction" A and follows it until $t = 1$. The exponential map has an inverse, the logarithmic map, defined as

$$\log_S(P) = S^{\frac{1}{2}} \log(S^{-\frac{1}{2}} P S^{-\frac{1}{2}}) S^{\frac{1}{2}},$$

where $\log(D)$ is the logarithmic matrix of $D \in PD(p)$. The logarithmic map returns the tangent element A that allows the correspondent geodesic to reach P in $t = 1$.

The *Riemannian distance* between elements $P_1, P_2 \in PD(p)$ is the length of the geodesic connecting P_1 and P_2, i.e.

$$d_R(P_1, P_2) = || \log(P_1^{-1/2} P_2 P_1^{-1/2}) ||_F = \sqrt{\sum_{i=1}^{n} (\log \sigma_i)^2},$$

where the σ_i are the eigenvalues of the matrix $P_1^{-1} P_2$ and $||.||_F$ is the Froebenius norm for matrices, defined as

$$||A||_F = \sqrt{\text{trace}(A^T A)}.$$

This distance is also called *trace metric*, for instance in Yuan et al. (2012).

Once a metric has been introduced in $PD(p)$, we can address the problem of estimating the mean given a sample of positive definite symmetric matrices. In recent years, many authors (Fletcher et al., 2004; Pennec et al., 2006; Dryden et al., 2009) proposed to use the Fréchet mean for a more coherent approach in dealing with data belonging to a Riemannian manifold. The Fréchet mean of a random element S, with probability distribution μ on a Riemannian manifold, is defined as

$$\Sigma_R = \text{arginf}_P \int d_R(S, P)^2 \mu(dS)$$

and it can be estimated with the sample Fréchet mean

$$\widehat{\Sigma}_R = \text{arginf}_P \sum_{i=1}^{n} d_R(S_i, P)^2, \tag{4.1}$$

where S_i, $i = 1, \ldots, n$ is a sample from μ. For the $PD(p)$ case, both the Fréchet mean and the sample Fréchet mean exist and are unique (see, e.g, Moakher and Zéraï, 2011). By means of extensive comparisons, Dryden et al. (2009) show that

using estimators based on the Riemannian distance, or its approximation, gives better results than the estimator based on Euclidean metric.

Analogously, the variance of S can be defined as

$$\sigma^2 = \mathrm{Var}(S) = \mathbb{E}[d_R(S, \Sigma_R)^2]$$

and estimated with the sample variance

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} d_R(S_i, \widehat{\Sigma}_R)^2.$$

In practical applications, using the Riemannian distance could be computationally expensive. For this reason, other distances have been proposed to compare two positive definite symmetric matrices. For example, we may consider the Cholesky decomposition of the positive definite symmetric matrix P, i.e. the lower triangular matrix with positive entries $L = \mathrm{chol}(P)$ such that $P = LL^T$. Then, Wang et al. (2004) defined a *Cholesky distance* between two positive definite symmetric matrices as

$$d_C(P_1, P_2) = ||\mathrm{chol}(P_1) - \mathrm{chol}(P_2)||_F.$$

Using the Cholesky distance, the sample Fréchet mean for a sample S_i, $i = 1, \ldots, n$, is easily computed:

$$\widehat{\Sigma}_C = \widehat{\Delta}_C \widehat{\Delta}_C^T, \quad \text{where } \widehat{\Delta}_C = \frac{1}{n}\sum_{i=1}^{n} \mathrm{chol}(S_i).$$

Another possibility is to resort to the *square root distance* (Dryden et al., 2009):

$$d_S(P_1, P_2) = ||P_1^{\frac{1}{2}} - P_2^{\frac{1}{2}}||_F,$$

where $P^{\frac{1}{2}}$ is the matrix square root of P. Also for this case a simple formula exists for the sample Fréchet mean which minimizes square root distances from a sample $S_1, \ldots, S_n$ of positive definite symmetric matrices:

$$\widehat{\Sigma}_S = \widehat{\Delta}_S \widehat{\Delta}_S^T, \quad \text{where } \widehat{\Delta}_S = \frac{1}{n}\sum_{i=1}^{n} S_i^{\frac{1}{2}}.$$

It is worth noticing that the square root distance is also defined for non negative definite matrices. Thus, it is to be preferred in applications where matrix data may have zero eigenvalues, or very small eigenvalues which lead to instability in the computation of the Riemannian distance or the Cholesky decomposition.

In the following, we propose methods that are based on a general distance $d(.,.)$ on the manifold. In practice, the appropriate distance has to be chosen by looking at the problem at hand while weighing computational efficiency.

We will need also to exploit the identification between $Sym(p)$ and $\mathbb{R}^{p(p+1)}$, defining a map which associates each symmetric matrix with a vector of $\mathbb{R}^{p(p+1)}$. Many choices are of course possible, we indicate with $vec(A)$ the row-wise vectorization of the non redundant elements of a symmetric matrix A and with $vec^{-1}(a)$ the inverse operation, defined for $a \in \mathbb{R}^{p(p+1)}$. For example, for $p = 2$, we have

$$vec(A) = \begin{pmatrix} A_{11} \\ A_{12} \\ A_{22} \end{pmatrix}$$

and

$$vec^{-1}(a) = \begin{pmatrix} a_1 & a_2 \\ a_2 & a_3 \end{pmatrix}.$$

Chapter 5

Estimation of the mean for spatially dependent non Euclidean data

In this chapter we address the problem of estimating the mean from a spatially dependent sample of positive definite symmetric matrices, following the exposition in Pigoli and Secchi (2012). However, the proposed technique relies only on the definition of distance and thus it can be easily applied to any metric space, even if for the interpretation of the method we sometimes refer to the Riemannian geometry of $PD(p)$.

A generalization of the Fréchet mean (4.1) is proposed, which takes into account spatial dependence among Riemannian observation. To do this, we also define a semivariogram for Riemannian data.

5.1 Semivariogram for positive definite symmetric matrices

Let us consider the random field

$$\{S(\mathbf{s}) \in PD(p) : \mathbf{s} \in D\} \tag{5.1}$$

where D is a subset of R^d, $E[S(\mathbf{s})] = \Sigma \in PD(p)$ for every $\mathbf{s} \in D$. Since our aim is to perform the statistical analysis from a single incomplete realization of the random field, we ask the spatial dependence between $S(\mathbf{s}_1)$ and $S(\mathbf{s}_2)$ to be a function only of $\mathbf{h} = \mathbf{s}_1 - \mathbf{s}_2$, for $\mathbf{s}_1, \mathbf{s}_2 \in D$. This can be formally stated using the notion of joint probability measure on the manifold (see Pennec, 2006, for more details about probability measures on manifolds). For $\mathbf{s}_1, \ldots, \mathbf{s}_n \in D$, consider the finite-dimensional measure

$$\mu_{\mathbf{s}_1,\ldots,\mathbf{s}_n}(\Gamma_1, \ldots, \Gamma_n) = P(S(\mathbf{s}_1) \in \Gamma_1, \ldots, S(\mathbf{s}_1) \in \Gamma_n),$$

for all possible $\Gamma_1, \ldots, \Gamma_n$ in the Borelian σ-field of $PD(p)$. We require the random field to be strictly stationary, i.e. for every finite set $\mathbf{s}_1, \ldots, \mathbf{s}_n \in D$,

$$\mu_{\mathbf{s}_1,\ldots,\mathbf{s}_n}(\Gamma_1, \ldots, \Gamma_n) = \mu_{\mathbf{s}_1+\mathbf{h},\ldots,\mathbf{s}_n+\mathbf{h}}(\Gamma_1, \ldots, \Gamma_n)$$

for all possible $\Gamma_1, \ldots, \Gamma_n$ in the Borelian σ-field of $PD(p)$ and for all $\mathbf{h} \in R^d$ such that $\mathbf{s}_1 + \mathbf{h}, \ldots, \mathbf{s}_n + \mathbf{h} \in D$.

In general, the definition of a covariance between two random elements on a Riemannian manifold is not straightforward, but in this particular setting a natural extension of the variogram seems to be available. Indeed, in the one dimensional Euclidean setting the variogram is defined as

$$2\widetilde{\gamma}_E(\mathbf{h}) = Var(x(\mathbf{s}+\mathbf{h}) - x(\mathbf{s})) = E[(x(\mathbf{s}+\mathbf{h}) - x(\mathbf{s}))^2] - E[x(\mathbf{s}+\mathbf{h}) - x(\mathbf{s})]^2 =$$

$$= E[(x(\mathbf{s}+\mathbf{h}) - x(\mathbf{s}))^2] - (E[x(\mathbf{s}+\mathbf{h})] - E[x(\mathbf{s})])^2$$

i.e, the expected value of the squared Euclidean distance between the random variables minus the square Euclidean distance between their expected values. Hence, we may generalize the notion of variogram by substituting the Euclidean distance with a more appropriate distance, based on the geometry of the Riemannian manifold. By analogy with its definition in spatial statistics for Euclidean data (see, e.g, Cressie, 1993), we define the variogram for a positive definite matrix field as

$$2\widetilde{\gamma}(\mathbf{h}) \doteq E[d(S(\mathbf{s}+\mathbf{h}), S(\mathbf{s}))^2] - d(E[S(\mathbf{s}+\mathbf{h})], E[S(\mathbf{s})])^2 \qquad (5.2)$$

and consequently,

$$\mathrm{Var}(S(\mathbf{s})) = \lim_{||\mathbf{h}||\to 0} \widetilde{\gamma}(\mathbf{h}), \quad \mathrm{Cov}(S(\mathbf{s}), S(\mathbf{s}+\mathbf{h})) = \mathrm{Var}(S(\mathbf{s})) - \widetilde{\gamma}(\mathbf{h}) \quad (5.3)$$

when the limit exists. Since we assume $E[S(\mathbf{s})] = \Sigma$ for every $\mathbf{s} \in D$, the Riemannian semivariogram simply becomes

$$\widetilde{\gamma}(\mathbf{h}) = \frac{1}{2} E[d(S(\mathbf{s}+\mathbf{h}), S(\mathbf{s}))^2].$$

In practice, we require that spatial correlation depends only on the length of the distance between two points $\mathbf{s}_1$ and $\mathbf{s}_2$, thus restricting to the case of an isotropic semivariogram, where $\widetilde{\gamma}(\mathbf{h}) = \gamma(||\mathbf{h}||)$. This assumption is useful for estimation, but it can be removed in applications when information about the anisotropic structure of the field generating the data is available. Thus, in the presence of a sample $(S(\mathbf{s}_1), ..., S(\mathbf{s}_n))$ generated by the random field (5.1), the isotropic semivariogram γ can be estimated from the empirical Riemannian distances, for instance by means of the classical estimator illustrated in Cressie (1993):

$$\widehat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{(\mathbf{s}_i,\mathbf{s}_j)\in N(h)} d(S(\mathbf{s}_i), S(\mathbf{s}_j))^2,$$

where $N(h) = \{(\mathbf{s}_i, \mathbf{s}_j) \in D : h-\Delta < ||\mathbf{s}_i - \mathbf{s}_j|| < h+\Delta; i, j = 1, \ldots, n\}$, Δ is a positive (small) quantity acting as a smoothing parameter, $h = ||\mathbf{h}||$ and $|N(h)|$ is the number of couples $(\mathbf{s}_i, \mathbf{s}_j)$ belonging to $N(h)$. Finally, a model semivariogram can be fitted to the empirical semivariogram, via least squares. As it happens in the Euclidean setting, an accurate estimation of the semivariogram is crucial for subsequent analysis. All the guidelines and methods developed for vector data can also be easily applied to the estimation of $\widetilde{\gamma}$.

5.1.1 Stochastic dependence in non Euclidean spaces

The development of statistical methods for the analysis of samples generated by random fields of positive definite matrices asks for a definition of stochastic dependence between two random elements taking values on a Riemannian manifold. In Euclidean spaces, the covariance is a measure of linear dependence between two random variables. However, in a non Euclidean framework linear dependence cannot be properly captured. While in many applications geodesics can be used as a surrogate of linear subspaces, this is not possible here, since an element of the manifold does not generate a geodesic in the way a random vector in a Euclidean space generates a linear subspace, thus allowing for measuring linear dependence.

The definition proposed in the previous section is based on the difference between the common variance of the random elements and half the expected value of their square distance. In this section, we explore properties and limits of this definition to fully understand the peculiarity of a non linear space for what concerns stochastic dependence.

Let (A, B) be a random vector whose components are positive definite matrices. The spatial model proposed in the previous section leads to a covariance between the random matrices A and B defined as

$$\mathrm{Cov}(A, B) := \frac{1}{2}\{\sigma_A^2 + \sigma_B^2 - (\mathbb{E}[d(A, B)^2] - d(\mathbb{E}[A], \mathbb{E}[B])^2)\} \tag{5.4}$$

where, for $S = A, B$, we set $\mathbb{E}[S] = \mathrm{arginf}_{\Sigma}\mathbb{E}[d(S, \Sigma)^2]$ and $\sigma_S^2 = \mathrm{Var}(S) = \mathbb{E}[d(S, \mathbb{E}[S])^2]$.

In the Euclidean setting, covariance is a measure of how close to a linear subspace observations are expected to lie, e.g. a straight line in $\mathbb{R}^2$. No linear subspaces exist on a Riemannian manifold, unless locally. In this framework the covariance measures how "near" observations are expected to be, with respect to their variability (i.e., the variance of the individual random elements A and B). A negative covariance indicates that A and B are expected to be farther apart than what it is to be expected by looking only at their marginal means and variances.

To better understand (5.4), we may focus on the special case $\mathbb{E}[A] = \mathbb{E}[B] = \Sigma$

and $Var[A] = Var[B] = \sigma^2$, which is of interest in spatial models. Then

$$\mathrm{Cov}(A, B) \leq \sigma^2,$$

since $\mathbb{E}[d(A, B)^2] \geq 0$. The maximum value is taken for $A = B$ and $Cov(A, A) = \sigma^2$. Therefore the covariance defined in (5.4) has an upper limit that is reached when the random elements are the same.

5.2 Estimation of the mean from a spatially correlated sample on a Riemannian manifold

This section addresses the problem of estimating the mean given a sample of spatially correlated positive definite symmetric matrices. The influence of spatial correlation on estimation and prediction is well known in the traditional Euclidean setting (see, e.g., Cressie, 1993) and it has been recently highlighted also for the case of functional data (Gromenko and Kokoszka, 2011). In particular, in the presence of strong spatial correlation, the sample mean can be inefficient as estimator for the mean of the population, having larger variance than estimators that take into account spatial dependence, see Cressie (1993, Section 1.3) for a proof in the case of real valued random variables and Gromenko and Kokoszka (2011), for extensive simulation studies on functional data. Indeed, in the presence of highly irregular spatial designs, the sample may contain a great amount of data coming from close by locations, together with a few isolated and distant observations. If spatial correlation is strong, data from close by locations are expected to provide similar information; their influence on the estimate should be mitigated, with respect to the few data coming from distant locations.

We propose an estimator for the mean Σ of a random field $S \in PD(p)$ which generalizes the estimator proposed by Gromenko and Kokoszka (2011) for a linear space. It is defined as a weighted sample Fréchet mean:

$$W = \mathrm{arginf}_P \sum_{i=1}^{n} \lambda_i d(S(\mathbf{s}_i), P)^2, \tag{5.5}$$

where $S(\mathbf{s}_i)$ is the observation of the random field S at location $\mathbf{s}_i \in D$. Weights λ_i have to be chosen taking into account the spatial dependence among observations. Analogously to Section 4, we add a subscript to indicate the distance that has been used in the estimation procedure: W_R is the weighted sample Fréchet mean using the Riemannian distance and W_S is the weighted sample Fréchet mean using the square root distance. Minimization of the weighted sum of square distance to estimate Fréchet mean on the manifold has been first proposed in Dryden

et al. (2009) in the context of smoothing for Diffusion Tensor Imaging fields. Their aim is to estimate the diffusion tensor for each point of the domain, starting from noisy and discrete observations. Thus, weights are chosen as function of the distance of each observations from the point where the estimate is needed. This approach has been recently developed in Yuan et al. (2012), where a local polynomial regression estimator for the conditional mean $\mathbb{E}[S(\mathbf{s})|\mathbf{s} = s_0]$ is introduced.

Differently from these previous works, we here want to estimate the unconditional mean of the random field (5.1), starting from a spatially correlated sample of data belonging to the manifold. This leads to a different choice of weights λ_i, that should now take into account the dependence among the random elements the data are realizations of. Following the analogy with the Euclidean setting, we ask the weights λ_i to solve the quadratic constrained minimization problem:

$$\min \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j Cov(S(\mathbf{s}_i), S(\mathbf{s}_j)), \qquad \sum_{i=1}^{n} \lambda_i = 1, \lambda_i \geq 0 \text{ for } i = 1, ..., n. \quad (5.6)$$

In the Euclidean case, (5.6) is equivalent to the minimization of the mean square error, but this need not be true for a general Riemannian manifold. However, choosing the weights λ_i as the solution of problem (5.6) meets the qualitative request to attribute less influence to subsets of data which are strongly correlated. We also ask the weights to be non negative to avoid instability in the minimization on the manifold, since, in any case, the solution of the minimization problem would not result in a linear combination of the data. Many numerical methods exist to solve the quadratic programming problem set in (5.6). We resort to that proposed in Goldfarb and Idnani (1983). The covariance structure $Cov(S(\mathbf{s}_i), S(\mathbf{s}_j))$ is obtained from the model semivariogram estimated with the procedure illustrated in the previous section.

5.3 Simulation studies

In this section we present a simulation study to test the performance of the proposed mean estimator. To do this, we introduce a simple method for simulating a random field of positive definite matrices with spatial correlation. Then, we use the simulated field to compare the weighted sample Fréchet mean W_S with the usual sample Fréchet mean $\widehat{\Sigma}_S$, for different experimental designs. Here, we choose the square root distance to compare two positive definite matrices for computational savings and to avoid problems with nearly singular matrices.

5.3.1 Simulation of a random field in $PD(2)$

We want to simulate a positive definite symmetric matrix field $S(\mathbf{s}) \in PD(2)$ with mean Σ and a spatial correlation structure. This is obtained through the sample covariance matrices of the realizations of a gaussian random vector field $\mathbf{v}$. This is a relatively easy way to address the problem of simulating from a multivariate random variable where each entry is a positive definite matrix and a dependence is present among different entries.

Let $\mathbf{s} \in D \subset \mathrm{R}^2$ indicate the spatial coordinates of two independent gaussian random field $x(\mathbf{s})$, $y(\mathbf{s})$, with $\mathbf{0}$ mean and spatial covariance

$$\mathrm{Cov}(x(\mathbf{s}_i), x(\mathbf{s}_j)) = \mathrm{Cov}(y(\mathbf{s}_i), y(\mathbf{s}_j)) = \begin{cases} \exp(-q\|\mathbf{s}_i - \mathbf{s}_j\|^2) & \|\mathbf{s}_i - \mathbf{s}_j\|^2 > 0; \\ 1 & \|\mathbf{s}_i - \mathbf{s}_j\|^2 = 0, \end{cases}$$

for $\mathbf{s}_i, \mathbf{s}_j \in D$.

Given a 2×2 matrix A, say $A = (1, 1; 0, 1)$, the random vector field $\mathbf{v}(\mathbf{s}) = A(x(\mathbf{s}), y(\mathbf{s}))^T$ has covariance matrix $\Sigma = AA^T = (2, 1; 1, 1)$. We generate N realizations of the random vector field $\mathbf{v}(\mathbf{s})$ and compute the sample covariance matrix

$$S(\mathbf{s}) = \frac{1}{N-1} \sum_{k=1}^{N} (\mathbf{v}_k(\mathbf{s}) - \bar{\mathbf{v}}(s))(\mathbf{v}_k(\mathbf{s}) - \bar{\mathbf{v}}(s))^T \sim \mathrm{Wishart}(\Sigma, N-1).$$

The positive definite symmetric matrix field $S(\mathbf{s})$ has thus mean Σ and it has a spatial dependence structure inherited by the spatial correlation of the underlying vector field $\mathbf{v}(\mathrm{s})$. The law of the random field $S(\mathbf{s})$ depends on the parameters q and N, which determine respectively the spatial dependence and the variability. In Fig. 5.1 some realizations of the matrix random field are reported for different values of q and N, using ellipses to represent 2×2 positive definite symmetric matrices. It can be seen that larger values of q correspond to lower spatial dependence and larger values of N to lower variability. By inspecting the realizations of the field, we can guess that for q and N both small, taking into account spatial dependence improves the estimate of the unconditional mean Σ. Indeed, for large values of N the variability of the field is so small that every single observation is a good representative of the mean and so every estimation techniques is adequate. Of course, when q is large, no spatial dependence is present, observations are independent and thus the sample Fréchet mean is the proper estimator. Hereafter we focus on the case when q and N are small.

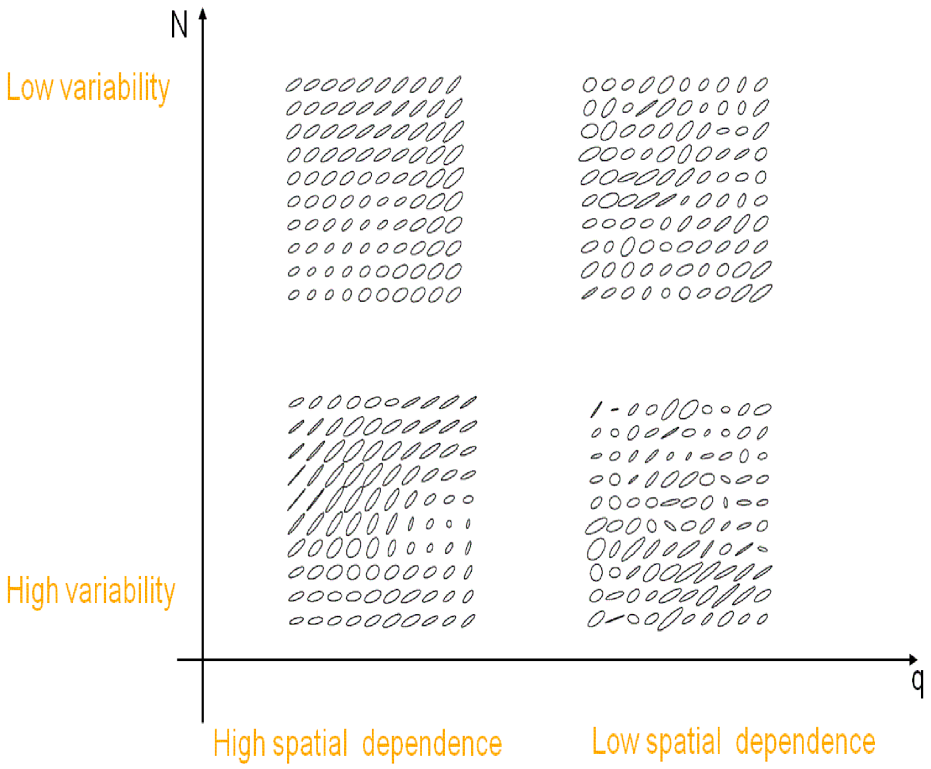

Figure 5.1: Simulation of the positive definite random fields, for different values of q and N. Each statistical unit $S(\mathbf{s_i})$ (a 2×2 positive definite symmetric matrix) is represented as an ellipse that is centered in $\mathbf{s}_i$ and has axis $\sqrt{\sigma_j}\mathbf{e}_j$, where $S(\mathbf{s}_i)\mathbf{e}_j = \sigma_j \mathbf{e}_j$ for $j = 1, 2$.

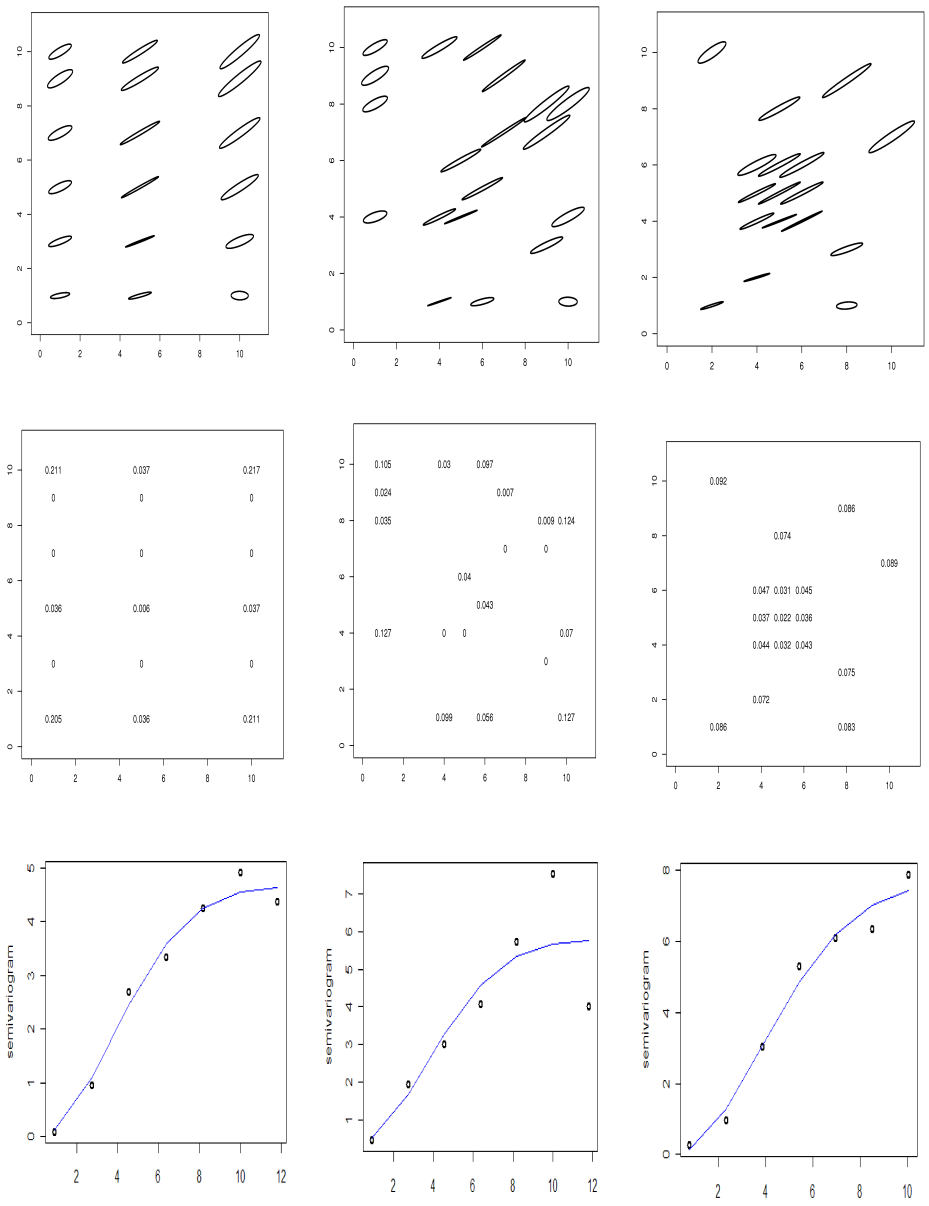

Figure 5.2: First row: three datasets obtained in the first simulation for the three experimental designs: regular grid (left), irregular (middle) and clustered (right). Second row: weights λ_i assigned to each location, rounded down to the third decimal digit, for the first simulated field and the three experimental designs. Third row: empirical semivariograms obtained from the three experimental designs in the first simulation. A fitted gaussian model is superimposed to the empirical semivariogram (solid line).

5.3.2 Estimation of the mean Σ of the simulated field

We now compare the proposed estimator with the sample Fréchet mean for three different experimental designs. We simulate 20 realizations of the random field $S(\mathbf{s})$ on a rectangular grid, setting $q = 0.01$ and $N = 4$, which is a case of high variability and high spatial dependence. We then subsample each realization in different points $\mathbf{s}_i$, obtaining different sets of observations for each experimental design. Fig. 5.2 shows the datasets for the first simulation: each statistical unit $S(\mathbf{s_i})$ (a 2×2 positive definite symmetric matrix) is represented as an ellipse that is centered in $\mathbf{s}_i$ and has axis $\sqrt{\sigma_j}\mathbf{e}_j$, where $S(\mathbf{s}_i)\mathbf{e}_j = \sigma_j \mathbf{e}_j$ for $j = 1, 2$. The first experimental design corresponds to a regular grid, the second to an irregular grid, while the third grid presents a cluster of spatial locations. The same picture shows the empirical semivariogram obtained for each dataset with a superimposed gaussian semivariogram fitted via least squares. For each realization of the random field S, we estimated the mean for the three experimental designs both with the sample Fréchet mean $\widehat{\Sigma}_S$ and the weighted Fréchet mean W_S. Fig. 5.3 shows the boxplots of the distances $d_S(\widehat{\Sigma}_S, \Sigma)$ and $d_S(W_S, \Sigma)$ for the three experimental designs. The weighted estimator behaves better, especially in the case of clustered data, where it is able to disregard some of the redundant information coming from points in the cluster. Fig. 5.2 shows also the weights λ_i obtained in the first simulation for the three experimental designs.

5.4 Applications to the estimation of mean covariance structure for meteorological variables

The simulation studies of the previous section support the tenet that the estimate of the mean covariance could be improved by taking into account data spatial dependence. As an illustrative application, we consider the problem of estimation of the mean covariance between different meteorological variables, say temperature and precipitation. Temperature and precipitation are two very important climatic variables. Their co-variability is also of interest: a better understanding of their relationship can provide insights on the precipitation-forming process or improve weather forecasting methods. Moreover, relative behavior of temperature and precipitation affects agricultural production (Lobell and Burke, 2008, see[). For a broader introduction to the importance of the temperature - precipitations relationship and its estimate see, e.g., Trenberth and Shea (2005) and references therein.

We focus on the Quebec province, Canada. Data from Canadian meteorological stations are made available by Environment Canada on the website http://climate.weatheroffice.gc.ca. Indeed different measurement stations provide

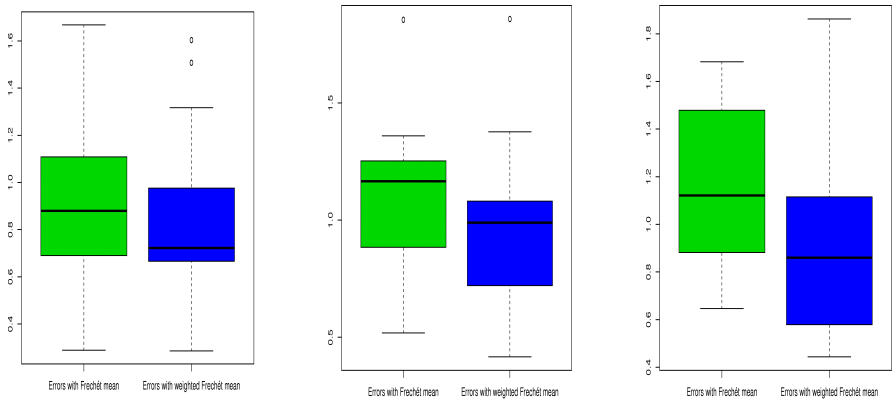

Figure 5.3: Boxplot of $d_S(\widehat{\Sigma}_S, \Sigma)$ (left) and $d_S(W_S, \Sigma)$ (right) for the three experimental designs: regular pattern (left), irregular (center) and clustered (right).

meteorological data along time and a first idea could be to bundle all data together in order to estimate the covariance between meteorological variables. This procedure is questionable since it does not take into proper account the spatial distribution of the measurement stations, which can be far from a regular grid on the region of interest. Analyzing similar data, coming from Canadian meteorological stations, Gromenko and Kokoszka (2011) point out the relevance of spatial dependence between measurement stations when estimating the monthly mean temperature function.

Fig. 5.4 shows the map of Quebec and the meteorological stations for which monthly data for temperature and precipitation are available, from 1983 to 1992. We assume that the monthly variation of the mean covariance between temperature and precipitation stays unchanged along the years of this short time period. The goal is to estimate the mean covariance between temperature and precipitation for each month of the year. Thus, for each meteorological station, we use

the 10-year measures of temperature and precipitation to estimate a 2×2 sample covariance matrix for every month from January to December. Fig. 5.4 shows the ellipse representation of these covariance matrices for January.

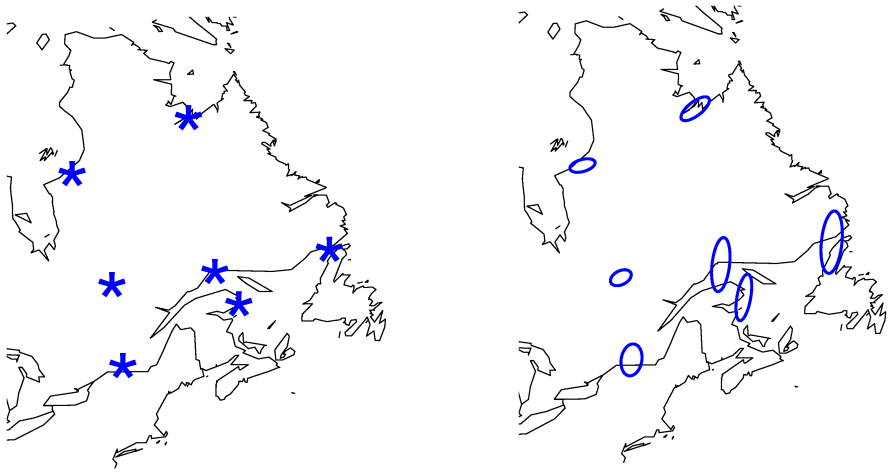

Figure 5.4: Left: Map of Quebec. Blue stars indicate positions of meteorological stations. Right: For each meteorological station an ellipse is plotted, representing 2×2 covariance matrix between temperature and precipitations in January.

Locations of the meteorological stations form an irregular pattern within Quebec. Thus, we expect that taking into account spatial dependence leads to a more accurate estimate of the mean covariance between temperature and precipitation in Quebec. We also assume that spatial dependence is constant along time. This allows to have more data for variogram estimation, which is a crucial point in the analysis. Fig. 5.5 shows the empirical semivariogram estimated with the method proposed in Section 5.1, with a superimposed fitted gaussian variogram, and the weights for each station obtained by solving (5.6). It is interesting to notice that three stations are associated with almost zero weights: this means that they are bringing redundant information for the estimation of the mean covariance structure.

An ellipse representation of the estimates obtained with sample Fréchet mean $\widehat{\Sigma}_S$ and weighted sample Fréchet mean W_S for the 12 months of the year appears in Fig. 5.6. The two estimators provide similar estimates for the winter period, from October to February, where a positive correlation exists between temperature and precipitation in the coldest months of the year. This is in agreement with Isaac

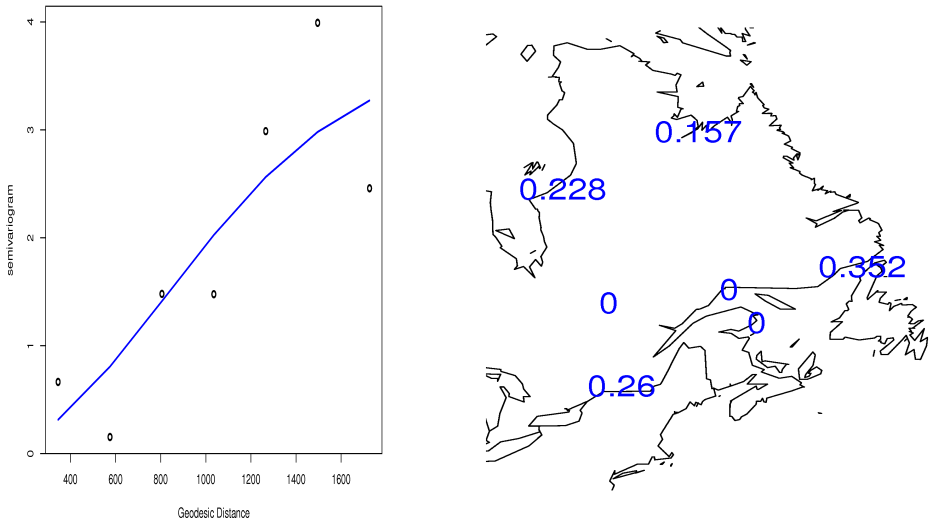

Figure 5.5: Left: Empirical semivariogram for the covariance matrix between temperature and precipitation (black points) and least squares fitting of a gaussian semivariogram. Right: Weights given to every station for the estimation of the average covariance matrix. Weights are rounded down to the third decimal digit.

and Stuart (1991), where correlation between daily temperature and precipitation is considered for the whole Canada by looking at the temperature precipitation index, i.e. the percentage of precipitation occurring at temperatures colder than the median daily temperature. They found that in January more precipitation is observed in relatively warm days.

The weighted sample Fréchet mean W_S provides a quite different estimate for the beginning of spring (March and April), where no correlation seems to be present, while the sample Fréchet mean $\widehat{\Sigma}_S$ would suggest a positive correlation. For April, Isaac and Stuart (1991) found great variability of the temperature precipitation index in the different Canadian provinces. For Quebec, however, it is around 50%, thus suggesting no correlation. The two estimates agree again for May and June (negative correlation), while for summer months estimates provided by the weighted sample Fréchet mean W_S suggest a different total variation for these covariance matrices (lower in July and September, greater in August) but both of them indicate that there is no correlation between temperature and precipitation, thus agreeing with Isaac and Stuart (1991) who report a temperature precipitation index around 50% for Quebec and Ontario in July, contrary to the trend of all the other Canadian provinces.

In conclusion, estimates provided by the proposed estimator W_S are in full

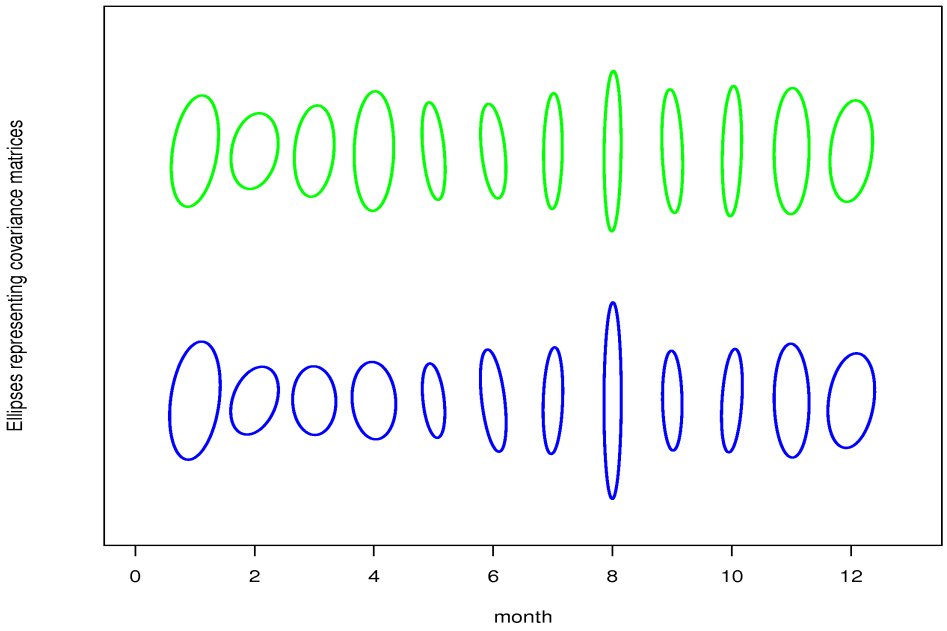

Figure 5.6: Ellipses representing the estimated covariance matrix between temperature and precipitations in Quebec, for the twelve months of the year. First Row: Sample Fréchet mean $\widehat{\Sigma}_S$. Second row: Weighted Fréchet mean W_S.

agreement with previous analysis of Canadian climate, while ignoring spatial dependence among measurement stations leads to anomalous results for March and April. Moreover, dealing with the covariance matrix, rather than with the temperature precipitation index, supplies also information about temperature and precipitation variability. Differences between the estimates of total variability provided by W_S and $\widehat{\Sigma}_S$ are concentrated in summer months. In August, our method estimates a greater total variability, while in July and September a lower one.

5.4.1 Choice of a different design for meteorological stations

As shown in Section 5.4, the spatial correlation among the meteorological stations of Quebec implies that three of them bring no significant information for the estimation of the mean covariance between temperature and precipitation. We now imagine to have the possibility to add a new meteorological station. We assume that the spatial dependence between data generated by the meteorological stations is described by the gaussian semivariogram represented in Fig. 5.5, estimated via

least square from the empirical semivariogram. We aim at finding a site for the new station that makes the weights λ_i, given to the stations for mean covariance estimation, as close as possible to $1/n$, being n the total number of meteorological stations, the new one included. This would provide an estimator (5.5) with a smaller variance.

Let us superimpose a fine grid of points on the region of interest and indicate with x and y the latitude and the longitude of a point on the grid. For each grid point (x, y), we solve problem (5.6) pretending that the new station is located in (x, y). We thus obtain a new weight $\lambda_i(x, y)$ for each of the n meteorological stations. The utility of positioning the new station in (x, y) is defined as

$$U(x, y) = 1 - \sum_{i=1}^{n} (\lambda_i(x, y) - \frac{1}{n})^2.$$

We now look for the site (x, y) where the utility $U(x, y)$ is maximized. Fig. 5.7 shows the utility function on the Quebec province and the site for the new station maximizing it. Of course, the exercise considers only the problem of estimation of the mean covariance between temperature and precipitation, disregarding other quantities of interest for meteorological analysis.

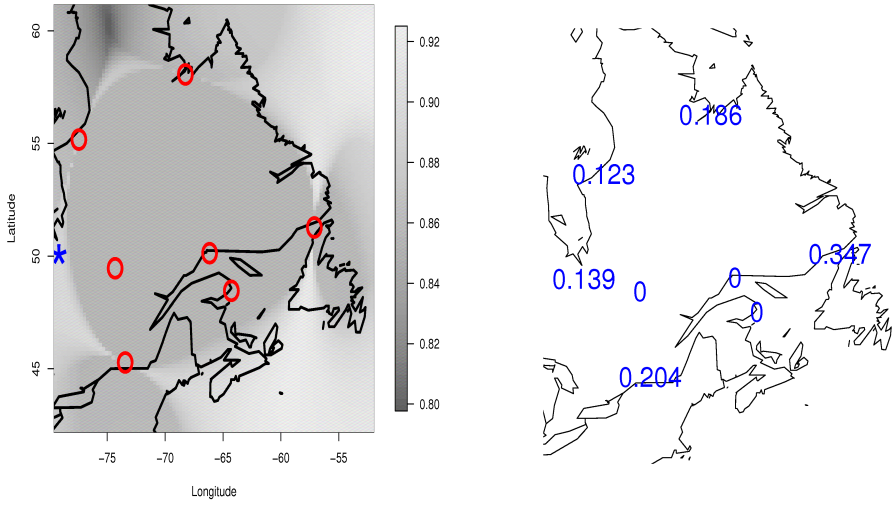

Figure 5.7: Left: Utility function $U(x, y)$ evaluated on the Quebec province. Locations of the already existing station are indicated by red circles, while the blue star corresponds to the maximum of the utility on the Quebec province. Right: Weights assigned to the new set of measurements stations.

Chapter 6

Kriging interpolation for Riemannian data

In this chapter a novel proposal for kriging of Riemannian data is described. This method is illustrated for the special case of positive definite symmetric matrices, but it can be easily applied every time that a tangent space and correspondent logarithmic and exponential map are available.

Let $S(\mathbf{s}_1), \ldots, S(\mathbf{s}_n) \in PD(p)$ be observations from a realization of a random field and $\mathbf{s}_1, \ldots, \mathbf{s}_n \in \Omega \subseteq \mathbb{R}^2$. The aim is to predict the value of the same realization in an unobserved location $\mathbf{s}_0$.

Many works have recently considered the problem of dealing with manifold-valued response variables. Some of them propose non parametric (see Yuan et al., 2012, and referece therein) or semi-parametric (see Shi et al., 2012) approaches but this implies a lack of interpretability or the reduction of multivariate predictors to univariate features. In particular, these approaches do not allow to introduce the spatial information in the prediction procedure.

In this dissertation, a different line of research is followed, along the lines of those who try to extend to manifold-valued data parametric (generalized) linear models (see, e.g. Fletcher, 2012). We propose here a novel linear regression model for Riemannian data based on a tangent space approximation. This model has been developed with the aim to take into account a drift effect in the universal kriging for manifold data. However, it could be useful in general to address multivariate regression in the context of Riemannian data.

6.1 A tangent space model for Riemannian data

Under the hypothesis that the dispersion of the observations on the manifold is not too large, a tangent space can be used to approximate data in a linear space,

where an additive model can be used to describe the relationship between response variable and covariates. This allows to extend well established statistical tools for regression models to the context of manifold valued response variable. Let us consider the model

$$S(\mathbf{x};\boldsymbol{\beta},\Sigma) = \exp_{\Sigma}(A(\mathbf{x};\boldsymbol{\beta}) + \Delta) \tag{6.1}$$

where $\Sigma \in PD(p)$ and $A(\mathbf{x};\boldsymbol{\beta}) \in Sym(p)$ depends on the matrix $\boldsymbol{\beta} \in \mathbb{R}^{p(p+1)/2\times(r+1)}$ of unknown deterministic parameters, r being the number of predictors and $\mathbf{x}$ the covariates vector. $\Delta \in Sym(p)$ is a random matrix such as $\mathbb{E}[vec(\Delta)] = 0$ and $\mathrm{Cov}(vec(\Delta)) = \sigma^2\mathrm{I}$. Thus, this manifold valued random variable is generated following the geodesic passing through Σ with tangent vector $A(\mathbf{x};\boldsymbol{\beta}) + \Delta$: the geodesic to be followed to obtain the realization of the variable is controlled by the covariates vector $\mathbf{x}$ but an additive error is also present. The expression of the symmetric matrix A has to be specified and in this work we deal with the linear model

$$vec_k(A) = \boldsymbol{\beta}_{k\cdot}^T\mathbf{x}, \tag{6.2}$$

for $1 \leq k \leq p(p+1)/2$, where $\boldsymbol{\beta}_{k\cdot} \in \mathbb{R}^{(r+1)}$ is the kth-row of the parameter matrix. This strategy, involving a tangent space approximation, can be followed also to generalize to manifold valued response variables more complex model, such as non linear or generalized regression.

Let us consider a sample $(\mathbf{x}_1, S_1), \ldots, (\mathbf{x}_n, S_n)$. The goal is to fit the tangent plane approximation which best models the relationship between $\mathbf{x}$ and S, i.e. to find

$$(\widehat{\Sigma}, \widehat{\boldsymbol{\beta}}) = \arg\min_{\Sigma,\boldsymbol{\beta}} \frac{1}{2}\sum_{i=1}^{n} ||\boldsymbol{\beta}\mathbf{x}_i - vec(\log_{\Sigma}(S_i))||^2_{\mathbb{R}^{p(p+1)/2}}, \tag{6.3}$$

where $\log_{\Sigma}$ indicates the logarithmic map that projects each element of $PD(p)$ on the tangent space of $PD(p)$ in Σ. Thus, our estimator looks for the linear model in the tangent space which minimizes the error sum of squares. It has to be noticed that the solution of problem (6.3) is a maximum likelihood estimator if $vec(\Delta)$ has a multivariate normal distribution.

Indeed, given a sample $(\mathbf{x}_1, S_1), \ldots, (\mathbf{x}_n, S_n)$, the likelihood function for the multivariate gaussian model on the tangent space is

$$L(\Sigma, \boldsymbol{\beta}, \sigma^2 | S_i, \mathbf{x}_i, i = 1, \ldots, n) = \prod_{i=1}^{n} \frac{1}{(2\pi\sigma^2)^{p(p+1)/2}} e^{-\frac{1}{2\sigma^2}\mathbf{y}_i^T\mathbf{y}_i},$$

where $\mathbf{y}_i = vec(\log_{\Sigma}(S_i)) - \boldsymbol{\beta}\mathbf{x}_i)$. The correspondent log-likelihood is therefore

$$\begin{aligned} l(\Sigma, \boldsymbol{\beta}, \sigma^2 | S_i, \mathbf{x}_i, i = 1, \ldots, n) \;=\; & -\tfrac{np(p+1)}{4}\log(2\pi\sigma^2) + \\ & -\tfrac{1}{2\sigma^2}\textstyle\sum_{i=1}^{n}\sum_{k=1}^{p(p+1)/2}(vec(\log_{\Sigma}(S_i))_k - \boldsymbol{\beta}_{k\cdot}^T\mathbf{x}_i)^2 \end{aligned}$$

and its maximization is equivalent to the minimization of problem (6.3).

For a given known Σ, minimization of (6.3) is an ordinary least square problem on the tangent space, since

$$\sum_{i=1}^{n} ||\boldsymbol{\beta}\mathbf{x}_i - \log_\Sigma(S_i)||^2_{\mathbb{R}^{p(p+1)/2}} = \sum_{i=1}^{n} \sum_{k=1}^{p(p+1)/2} (\boldsymbol{\beta}_{k.}^T \mathbf{x}_i - vec(\log_\Sigma(S_i))_k)^2 =$$

$$= \sum_{k=1}^{p(p+1)/2} ||\boldsymbol{\beta}_{k.}^T \mathbf{X} - \mathbf{Y}_{k.}(\Sigma)||^2_{\mathbb{R}^n},$$

where $\mathbf{X}$ is the $(r+1) \times n$ matrix whose columns are the vectors of predictors and $\mathbf{Y}(\Sigma)i.$ is the i-th row of the matrix such that $Y_{ki}(\Sigma) = vec(\log_\Sigma(S_i))_k$. Thus, each term of the previous sum is minimized with respect to β by the ordinary least square solution

$$\widehat{\boldsymbol{\beta}}_{k.}(\Sigma) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}_{k.}(\Sigma) \tag{6.4}$$

and the global solution should be in the form $(\Sigma, \widehat{\boldsymbol{\beta}}(\Sigma))$. Finally, the problem

$$\min_\Sigma \frac{1}{2} \sum_{i=1}^{n} ||\widehat{\boldsymbol{\beta}}(\Sigma)\mathbf{x}_i - vec(\log_\Sigma(S_i))||^2_{\mathbb{R}^{p(p+1)/2}} \tag{6.5}$$

has to be numerically solved. Here we resort to Nelder-Mead algorithm implemented in the `optim()` function in the R software (R Development Core Team, 2009), with constrains to ensure the matrix to be positive definite. This works well for two or three dimensional matrices, while in case of higher dimensional matrices more efficient optimization tools would be needed, for example gradient descend or Newton methods on the manifold (see, e.g, Dedieu et al., 2003). In this way, we obtain an estimate for the parameters:

$$\begin{cases} \widehat{\Sigma} = \arg\min_\Sigma \frac{1}{2} \sum_{i=1}^{n} ||\widehat{\boldsymbol{\beta}}(\Sigma)\mathbf{x}_i - vec(\log_\Sigma(S_i))||^2_{\mathbb{R}^{p(p+1)/2}}, \\ \\ \widehat{\boldsymbol{\beta}}_{k.} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}_{k.}(\widehat{\Sigma}) \quad \text{for} \quad 1 \leq k \leq p(p+1)/2 \end{cases}$$

This method asks to specify the model (6.2) for the matrix A. Since this model is defined on the tangent space, well established techniques can be used for model selection, for example cross-validation.

6.2 Universal and ordinary kriging

In this section we apply the tangent space approximation above to non-stationary manifold-valued random field. The main idea is to use a tangent space model to

approximate the geometry of the manifold and to refer to the tangent space to deal with spatial dependence. Thus, the proposed model is

$$S(\mathbf{s};\boldsymbol{\beta},\Sigma) = \exp_{\Sigma}(A(\mathbf{x}(\mathbf{s});\boldsymbol{\beta}) + \Delta(\mathbf{s})) \tag{6.6}$$

where $\Delta(\mathbf{s})$ is a zero-mean second-order stationary random field taking values in the Euclidean space of symmetric matrices.

The estimation procedure follows the idea illustrated, e.g., in Cressie (1995) and it is based on an iterative algorithm. First, the estimate of the tangent space model is initialized with the estimator proposed in the previous section for the case of independent errors. Then, the spatial dependence of the random field Δ on the tangent space is estimated with traditional methods for Euclidean data, considering the residuals $\Delta_i^{(0)} = A(\mathbf{x}(\mathbf{s}_i);\widehat{\boldsymbol{\beta}}) - \log_{\Sigma}(S_i)$ as an incomplete realization of the random field Δ. The space of symmetric matrices is a linear space, endowed with the scalar product $\langle A, B\rangle_F = \mathrm{trace}(AB)$ and the distance $||A-B||_F^2 = \mathrm{trace}((A-B)(A-B))$. Thus, all the well known tools from spatial statistics can be used.

Once the spatial covariance structure has been estimated, a new estimate for the tangent space model is provided substituting the least square estimators for the model coefficients with the generalized least square estimators

$$\widehat{\boldsymbol{\beta}}_{GLS}(\Sigma) = \arg\min_{\boldsymbol{\beta}} \sum_{k=1}^{p(p+1)/2} (\boldsymbol{\beta}_{k.}^T\mathbf{X} - \mathbf{Y}_{k.}(\Sigma))^T\Gamma^{-1}(\boldsymbol{\beta}_{k.}^T\mathbf{X} - \mathbf{Y}_{k.}(\Sigma)) =$$

$$= \begin{pmatrix} (\mathbf{X}^T\Gamma^{-1}\mathbf{X})^{-1}\mathbf{X}^T\Gamma^{-1}\mathbf{Y}_{1.}(\Sigma) \\ \vdots \\ (\mathbf{X}^T\Gamma^{-1}\mathbf{X})^{-1}\mathbf{X}^T\Gamma^{-1}\mathbf{Y}_{p(p+1)/2.}(\Sigma) \end{pmatrix}$$

where Γ_{sr} is the covariance between Δ_s and Δ_r. Thus, the new minimization problem would be

$$\widehat{\Sigma} = \arg\min_{\Sigma} \frac{1}{2}\sum_{i=1}^{n} ||\widehat{\boldsymbol{\beta}}_{GLS}(\Sigma)^T\mathbf{x}(\mathbf{s}_i) - vec(\log_{\Sigma}(S_i))||^2_{\mathbb{R}^{p(p+1)/2}}.$$

This problem can be solved with the Nelder-Mead algorithm, even if at each step some subiterations are needed to estimate the spatial covariance structure and thus the model coefficients. The model being estimated, a kriging interpolation on the residuals provides an estimate for the field S in the unobserved location $\mathbf{s}_0$,

$$\widehat{S}(\mathbf{x}(\mathbf{s}_0)|S(\mathbf{x}(\mathbf{s}_1)),\ldots,S(\mathbf{x}(\mathbf{s}_n))) = \exp_{\widehat{\Sigma}}(vec^{-1}(\widehat{\boldsymbol{\beta}}_{GLS}(\widehat{\Sigma})^T\mathbf{x}(\mathbf{s}_0)) + \sum_{i=1}^{n}\lambda_i^0\boldsymbol{\Delta}_i),$$

where $\boldsymbol{\Delta}_i$ are the residuals in the known locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ and $\boldsymbol{\lambda}^0 = (\lambda_1^0, \ldots, \lambda_n^0)^T$ is the vector of weights from the ordinary kriging estimation. Let

$$\mathbf{c} = (\mathrm{cov}(\Delta(\mathbf{s}_1), \Delta(\mathbf{s}_0)), \ldots, \mathrm{cov}(\Delta(\mathbf{s}_n), \Delta(\mathbf{s}_0)))^T$$

and $\mathbf{1} = (1, \ldots, 1)^T$, then

$$\boldsymbol{\lambda_0} = \Gamma^{-1}\mathbf{c} + \Gamma^{-1}\mathbf{1}\frac{1 - (\mathbf{1}^T\Gamma^{-1}c)}{\mathbf{1}^T\Gamma^{-1}\mathbf{1}}.$$

This model includes as a special case the ordinary kriging, where the deterministic part of the model does not depend on spatial location.

6.3 Kriging for Quebec meteorological data

In this section the proposed kriging prediction is applied to the covariance matrix between temperature and precipitations in Quebec. Data are the same which have been considered in Section 5.4 but the aim of the analysis is now different, since we want to predict the covariance matrix in unobserved locations, starting from the incomplete realization of the field. At first, we still make the hypothesis that the matrix random field has a constant mean and thus only a constant term is included in the tangent space model, obtaining an ordinary kriging of the random field. Looking at the empirical semivariogram in the tangent space, we can choose a exponential model for spatial dependence. However, since we are working on different tangent spaces for the different months, we have to estimate separate semivariogram for each month, unlike what we have done in Section 5.4. This means that fewer data are available for the estimation of the spatial structure. Fig. 6.1 shows the estimated matrix field for the months of January and July, as well as the correspondent semivariogram on the tangent space.

Focusing separately on each month for the estimation of the spatial structure allows to notice that the hypothesis of stationarity may not be adequate for all months. In January, for example, the semivariogram suggests to introduce a non stationary drift term. Here the choice of an appropriate model is needed. We investigated linear and quadratic models with respect to longitude and latitude, including an interaction term. We found that the only choice which allows to have a stationary residual term is a linear term on the longitude, the final model being

$$A(\mathrm{Longitude}_i, \mathrm{Latitude}_i) = \beta_0 + \beta_{Long}\mathrm{Longitude}_i.$$

A possible meteorological interpretation is that more negative values for the longitude mean a greater distance from the Atlantic Ocean.

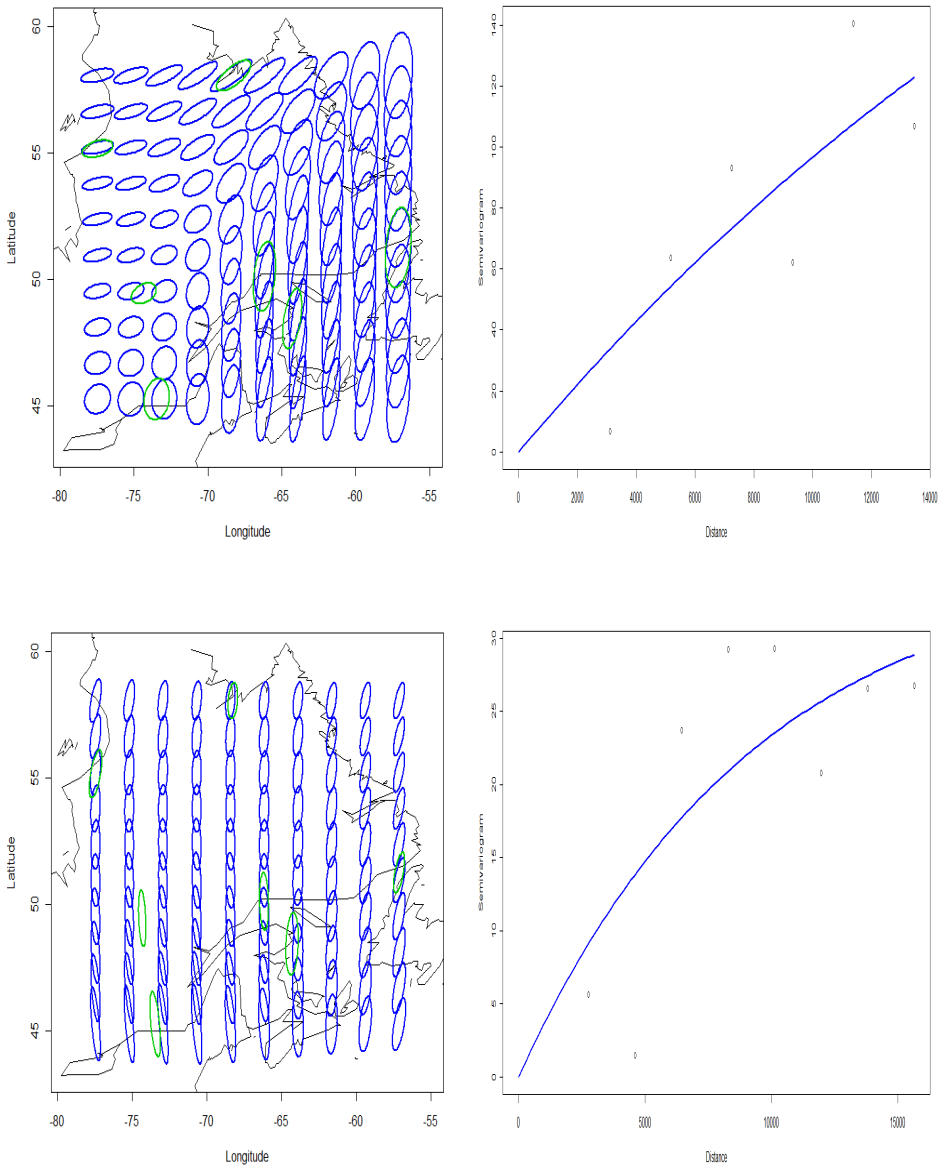

Figure 6.1: Ordinary kriging for covariance matrix between temperature and precipitations, green ellipses indicating original data (left) and empirical and model exponential semivariogram (right) for the months of January (first row) and July (second row).

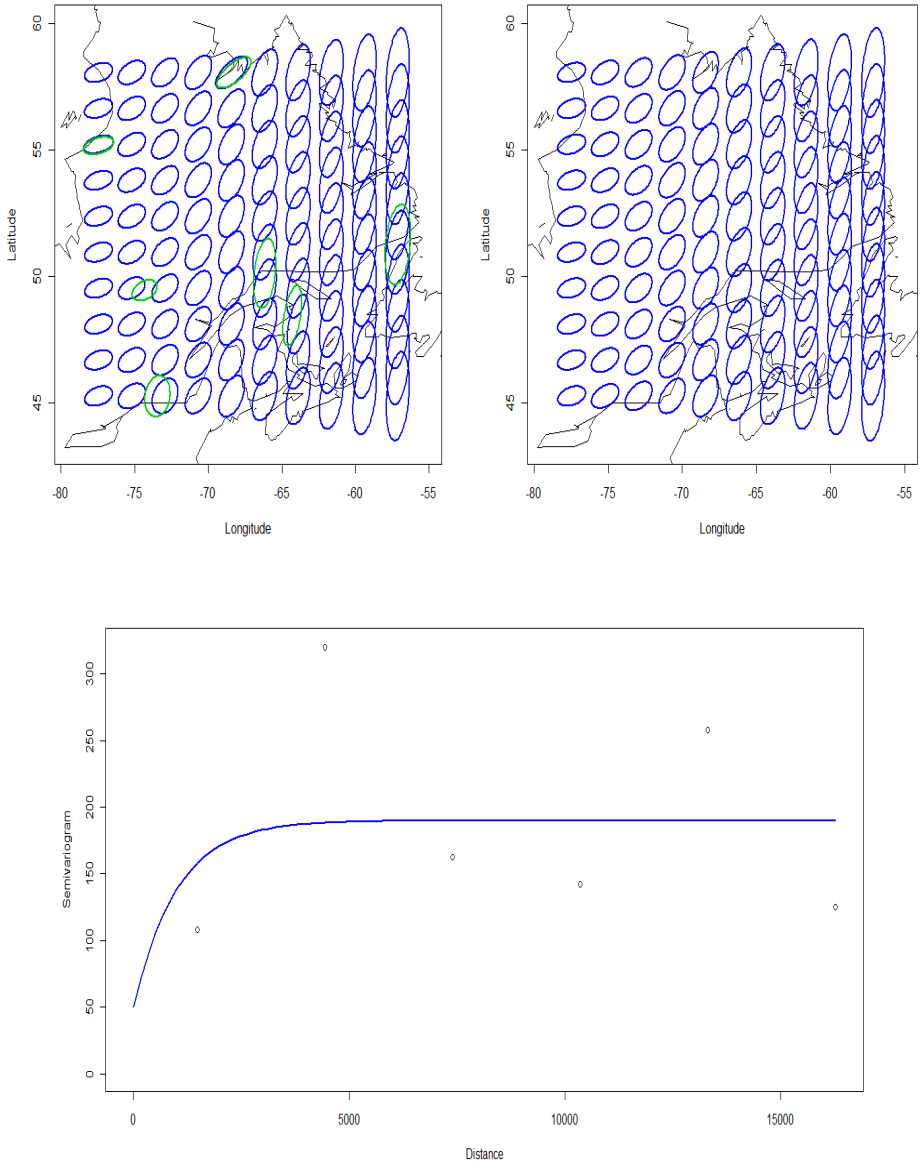

Figure 6.2: Universal kriging for covariance matrix between temperature and precipitations, green ellipses indicating original data (top left), drift term depending on longitude (top right) and empirical and model exponential semivariogram (bottom) of the residuals for the month of January.

The estimate of the coefficient vector for longitude is

$$\widehat{\boldsymbol{\beta}}_{Long} = \begin{pmatrix} 0.04956154 \\ -0.3550782 \\ 2.543298 \end{pmatrix}$$

and its effect on the covariance matrix field can be appreciated in Fig. 6.2, which shows the semivariogram for the residuals and a representation both of the drift term and of the kriging estimation of the matrix field. It is worth to notice that spatial dependence seems to be very low, once the drift has been removed, the deterministic trend explaining most of the variability. This is not completely unexpected, since Menafoglio et al. (2012) report similar results for temperature curves when using average meteorological data. This is implicit in our data, since we use yearly replicates for the estimation of covariance and thus intra-year dependence is not visible.

Conclusions and further development

This dissertation is in the framework of Object Oriented Data Analysis, focusing in particular on covariance operators as elements of manifolds. The overall aim is to find out how taking into account the non Euclidean structure of these mathematical objects can benefit the statistical analysis.

In Part I the problem of dealing with covariance operators on $L^2(\mathbb{R})$ has been investigated. The choice of the appropriate metric is crucial in the analysis of covariance operators: some suitable metrics have been proposed and their properties have been highlighted. In particular two metrics have been successfully applied to infinite dimensional covariance operators: the square root operator distance and the Procrustes size-and-shape distance. Both these metrics rely on the mapping of the covariance operators to a suitable space of Hilbert-Schmidt operators where a linear distance, i.e. the Hilbert-Schmidt norm of the difference of the operators, is appropriate. The square root operator distance uses the square root map defined in (1.1), while the Procrustes distance allows for unitary transformations in the space of Hilbert-Schmidt operators, thus taking into account the arbitrariness of the representation given by the chosen mapping transformation.

On the basis of an appropriate metric, statistical methods can be developed to deal with covariance operators in a functional data analysis framework. Here the notable cases of point estimation of covariance operators and two groups hypothesis testing are illustrated. The latter technique has proved useful for the analysis of the AneuRisk data, where investigating the covariance structures of different groups supports the results of previously published analysis. Moreover, in some applications the covariance operator itself is the object of interest, as shown with the linguistic data of Section 3.2. Here, linguistic scholars suppose that the covariance operator of frequency intensities catches significant phonetic features of the language. Using the square root and Procrustes distances to analyse these covariance operators for some Romance languages, preliminary results support this hypothesis, while also detecting some phonetic structures which deserve deeper linguistic explorations.

Many other developments are of course possible, both in the theoretical aspects of the proposed statistical methods and in data analysis. In particular, there is considerable scope for development of the consistency properties for estimators proposed in Section 2 and of necessary conditions for (local) convergence of the Procrustes algorithm in infinite dimensions. The proposed extrapolation procedure is a seminal idea for the development of linear models for covariance operators, this being a promising line of research for future works. The analysis of linguistic data is only preliminary and more significant results are expected when many words are taken simultaneously into account. Moreover, interactions with linguistic scholars can suggest other possible investigations of the data.

In Part II, we introduced spatial statistics methods which take into account the specific nature of the non Euclidean data at hand. We introduced a semivariogram whose definition consistently relies only on the notion of distance between two elements of the space to which data belong. This allows to tackle the problem of estimation of the mean from a spatially correlated sample of non Euclidean data. The proposed method relies only on the notion of distance between non Euclidean data and therefore it can be applied to any Riemannian manifold. Here we have focused on the notable case of positive definite matrices to show the effectiveness of our approach, both with simulations and with a significant real data application. However, our work can be easily adapted to other kinds of non Euclidean data. An example could be the covariance operator considered in Part I of this thesis: taking into account spatial dependence could generate interesting analysis for that kinds of problems. Indeed, Horváth and Kokoszka (2012) show that taking into account spatial dependence provides better estimates for functional principal component, i.e. for the correspondent covariance operator. However, they deal with covariance operators as Euclidean objects, looking for estimators that minimize Hilbert-Schmidt norm. Our approach can be easily generalized to this infinite dimensional setting and it would allow to consider other possible metrics for covariance operators.

In Section 5.4 we apply our method to a meteorological problem, the estimation of the covariance matrix between temperature and precipitation in the province of Quebec (Canada). We show that taking into account spatial dependence provides estimates that are in a better agreement with known meteorological information.

Finally, we illustrate a possible way to deal with the non stationarity of the covariance matrix field: ordinary or universal kriging based on a tangent space approximation of the manifold. This allows for the consideration of spatial dependance in smoothing procedures, such as those proposed in Dryden et al. (2009) or Yuan et al. (2012). The statistical analysis of a non stationary Riemannian field is still an ongoing research. Deeper investigation are needed for a better understanding of the relationship between the spatial dependence on the tangent space and

those measured directly on the manifold. Moreover, theoretical properties of the estimators for parameters of the tangent space model still have to be investigated, as well as the feasibility of solving the least squares problem in more general situation.

In conclusion, the statistical analysis of non Euclidean data is a very promising research field which will become more and more important with the diffusion of complex measurement systems. In particular, dealing properly with covariance operators bring new insight to the statistical analysis, as shown here in some practical applications.

Bibliography

Anderson, M.J. (2006) "Distance-Based Tests for Homogeneity of Multivariate Dispersions", *Biometrics*, **62**:245-253.

Aston, J. A. D., Chiou, J.-M. and Evans, J. P. (2010) "Linguistic Pitch Analysis using Functional Principal Component Mixed Effect Models". *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **59**:297–317.

Baker, C. R. (1970) "Mutual Information for Gaussian Processes", *SIAM Journal on Applied Mathematics*, **19**:451–458.

Besse, P. and Ramsay, J. O. (1986) "Principal components analysis of sampled functions", *Psychometrika*, **51**:285-311.

Bosq, D. (2000) Linear processes in function spaces, Springer, New York.

Cressie, N.A.C. (1993) Statistics for spatial data, Revised edition. Wiley, New York.

Dedieu, J., Priouret, P. and Malajovich, G. (2003) "Newtons method on Riemannian manifolds: convariant alpha theory", *IMA Journal of Numerical Analysis*, **23**:395–419.

Delicado, P., Giraldo, R., Comas, C. and Mateu, J. (2010) "Statistics for spatial functional data: some recent contributions", *Environmetrics*, **21**:224–239.

Dryden, I. L. and Mardia, K. V. (1998), Statistical Shape Analysis. Wiley, Chichester.

Dryden, I.L., Koloydenko, A., Zhou, D. (2009), "Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging", *The Annals of Applied Statistics*, **3**:1102-1123.

El Karoui, N. (2008) "Operator norm consistent estimation of large-dimensional sparse covariance matrices", *The Annals of Statistics*, **36**:2717–2756.

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

Fletcher, P.T. (2012) "Geodesic Regression and the Theory of Least Squares on Riemannian Manifolds", *International Journal of Computer Vision*, in press.

Fletcher, P.T., Conglin Lu, Pizer, S.M., Sarang Joshi (2004) Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions in Medical Imaging*, **23**:995–1005.

Fremdt, S., Steinebach, J.G., Horváth, L. and Kokoszka, P. (2012), "Testing the Equality of Covariance Operators in Functional Samples", *Scandinavian Journal of Statistics*, in press.

Goldfarb, D. And Idnani, A. (1983) " A numerically stable dual method for solving strictly convex quadratic programs". *Mathematical Programming*, **27**:1-33.

Good, P.I. (2005) Permutation, Parametric and Bootstrap Tests of Hypotheses, Third Edition. Springer, New York.

Gromenko, O., Kokoszka, P. (2011) " Estimation and testing for geostatistical functional data". In: Ferraty, F. (Ed.). Recent Advances in Functional Data Analysis and Related Topics. Springer-Verlag, Berlin, 155–160.

Gromenko, O., Kokoszka, P., Zhu, L., Sojka, J. (2012) "Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends", *The Annals of Applied Statistics*, **6**:669–696.

Groisser, D. (2005) "On the convergence of some Procrustean averaging algorithms", *Stochastics: An International Journal of Probability and Stochastic Processes*, **77**:31-60.

Gower, J. C. (1975) "Generalized Procrustes analysis", *Psychometrika*, **40**:33–50.

Hadjipantelis, P. Z., Aston, J. A. D., and Evans, J. P. (2012) "Characterizing fundamental frequency in Mandarin: A functional principal component approach utilizing mixed effect models", *Journal of the Acoustical Society of America*, **131**:4651–4664.

Horváth, L. and Kokoszka, P. (2012), Inference for Functional Data with Applications, Springer, New York.

Isaac, G.A., Stuart R.A. (1991) "Temperature-Precipitation Relationships for Canadian Stations", *Journal of Climate*, **5**:822-830.

Jung, S., Foskey, M., Marron J.S. (2011) "Principal Arc Analysis on direct product manifolds", *The Annals of Applied Statistics*, **5**:578–603.

Karcher, H. (1977) "Riemannian center of mass and mollifier smoothing", *Communications on Pure and Applied Mathematics*, **30**:509–541.

Kendall, W.S. (1990) "Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence", *Proceedings of the London Mathematical Society*, **61**:371–406.

Kent, J.T and Mardia, K.V. (2001) "Shape, Procrustes tangent projections and bilateral symmetry", *Biometrika*, **88**, 469-485.

Le, H. (2001) "Locating Fréchet means with application to shape spaces", *Advances in Applied Probability*, **33**:324-338.

Lobell, D.B., Burke, M.B. (2008) "Why are agricultural impacts of climate so uncertain? The importance of temperature relative to precipitation", *Environmental Research Letters*, **3**, 034007.

Menafoglio, A., Dalla Rosa, M. and Secchi, P. (2012) "A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space", Tech. rep. MOX 34/2012 http://mox.polimi.it/it/progetti/pubblicazioni/quaderni/34-2012.pdf

Moakher, M. (2005) "On the Averaging of Symmetric Positive-Definite Tensors". *Journal of Elasticity*, **82**:273–296.

Moakher, M., Zéraï, M. (2011) "The Riemannian Geometry of the Space of Positive-Definite Matrices and Its Application to the Regularization of Positive-Definite Matrix-Valued Data", *Journal of Mathematical Imaging and Vision*, **40**:171–187.

Panaretos, V. M., Kraus, D. and Maddocks, J.H. (2010) "Second-Order Comparison of Gaussian Random Functions and the Geometry of DNA Minicircles", *Journal of the American Statistical Association (Theory and Methods)*, **105**:670-682.

Pennec, X. (2006) "Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements". *Journal of Mathematical Imaging and Vision*, **25**:127-154.

Pigoli, D. and Secchi P. (2012) "Estimation of the mean for spatially dependent data belonging to a Riemannian manifold", *Electronic Journal of Statistics*, **6**:1926–1942.

Pennec, X., Fillard, P., Ayache, N. (2006) "A Riemannian framework for tensor computing", *International Journal of Computer Vision*, **6**:41-66.

R Development Core Team (2009) "R: A language and environment for statistical computing", R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org.

Rice, J.A. and Silverman, B.W. (1991) " Estimating the Mean and Covariance Structure Nonparametrically When the Data are Curves", *Journal of the Royal Statistical Society Ser. B, Methodological*, **53**:233–243.

Ramsay, J.O. and Silverman, B.W. (2002), Applied functional data analysis. Methods and case studies. Springer-Verlag, New York.

Ramsay, J.O. and Silverman, B.W. (2005) Functional data analysis (2nd ed.), Springer, New York.

Sangalli, L. M., Secchi, P., Vantini, S., and Veneziani, A. (2009a) "A Case Study in Exploratory Functional Data Analysis: Geometrical Features of the Internal Carotid Artery", *Journal of American Statistical Association*, **104**:37-48.

Sangalli, L. M., Secchi, P., Vantini, S., and Veneziani, A. (2009b) "Efficient estimation of three-dimensional curves and their derivatives by free-knot regression splines, applied to the analysis of inner carotid artery centrelines", *Journal of the Royal Statistical Society Ser. C, Applied Statistics*, **58**:285-306.

Shi, X., Zhu, H., Ibrahim, J.G., Liang, F., Lieberman, J. and Styner, M. (2012) "Intrinsic Regression Models for Medial Representation of Subcortical Structures", *Journal of American Statistical Association*, **107**:12–23.

Schwartzman, A., Dougherty, R.F., Taylor, J.E. (2010) "Group Comparison of Eigenvalues and Eigenvectors of Diffusion Tensors", *Journal of American Statistical Association*, **105**:588–599.

Ten Berge, J. M. F. (1977) "Orthogonal Procrustes rotation for two or more matrices",*Psychometrika*, **42**:267–276.

Trenberth, K.E. and Shea, D.J. (2005) "Relationship between precipitation and surface temperature", *Geophysical Research Letters*, **32**, L14703.

Wang, H. and Marron, J.S. (2007) "Object oriented data analysis: Sets of trees", *The Annals of Statistics*, **35**:1849–1873.

Wang, Z., Vemuri, B., Chen, Y. And Mareci, T. (2004) "A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex DWI", *IEEE Transactions in Medical Imaging*, **23**:930-939.

Yuan, Y., Zhu, H., Lin, W. and Marron, J.S. (2012) "Local polynomial regression for symmetric positive definite matrices", *Journal of the Royal Statistical Society Ser. B, Methodological*, **74**:697-719.

Zhu, K. (2007) Operator theory in function spaces (2nd ed.), American Mathematical Society, New York.