Politecnico di Milano

Facoltà di Ingegneria dei Sistemi
Corso di Laurea in Ingegneria Matematica

Tesi di Laurea Specialistica

Functional Data Analysis of CFD Simulations: the Systolic Wall Shear Stress Map of the Internal Carotid Artery

Candidato:
Andrea Boneschi
Matricola 735665

Relatore:
Dr. Simone Vantini

Anno Accademico 2009–2010

"Innovazione è fare qualcosa che gli altri non fanno"
Andrea Pontremoli
(Personal communication, 2010)

Contents

List of Figures

Sommario

In questa tesi di laurea, situata all'interno del progetto Aneurisk, ci proponiamo di indagare possibili relazioni tra il Wall Shear Stress sulle pareti della Arteria Carotide Interna e l'insorgenza di aneurismi cerebrali. I dati analizzati provengono da simulazioni numeriche di emodinamica, implementate presso il Mathematics and Computer Science Department della *Emory University*.
Abbiamo scelto di analizzare le simulazioni attraverso l'approccio molto innovativo di *Functional Data Analysis*. A quanto ci risulti, questa è la prima volta che un simile tipo di dato venga analizzato statisticamente. I modelli numerici simulano il valore di Wall Shear Stress per ogni punto della ICA, che è rappresentata da una superficie bidimensionale.
In questo lavoro, presentiamo le tecniche ed i risultati ottenuti dalle analisi statistiche condotte secondo l'approccio funzionale.

Nel **Capitolo 1** vengono delineati gli obiettivi del presente studio in rapporto alla letteratura scientifica disponibile sul soggetto e viene descritto il processo di raccolta dei dati provenienti da immagini medicali.
Nella **Sezione 2.1** viene presentata una breve introduzione alla *Functional Data Analysis*, sottolineandone le peculiarità rispetto all'approccio classico.
Nel **Capitolo 2** viene descritta la procedura di smoothing dei dati, ovvero, in questa fase i dati discreti vengono trasformati in dati continui, dedicando ampia trattazione alla scelta del bandwidth.
Nel **Capitolo 3** viene trattato il problema della registrazione, mostrando la trasformazione a dati continui allineati.
Nel **Capitolo 4** vengono descritti tecniche e risultati dell'analisi delle componenti principali funzionali, elemento caratterizzante di questo lavoro.
Infine, nel **Capitolo 5** vengono esposte le conclusioni di questa tesi di laurea, in relazione ai risultati ottenuti precedentemente nel progetto AneuRisk, suggerendo, inoltre, nuove direzioni di ricerca per possibili sviluppi futuri del progetto.

Abstract

In this work, situated in the context of the AneuRisk project, we investigate possible relationships between Wall Shear Stress on Internal Carotid Artery wall and the origin of cerebral aneurysms. Our data come from fluidynamics numerical simulations performed by the Mathematics and Computer Science Department of Emory University.
We decided to analyze CFD simulations by means of a very innovative approach that is Functional Data Analysis.
At our knowledge, this the very first time that this kind of data is statistically analyzed. Simulations provide us Wall Shear Stress values for every point of ICA wall, which is represented as a two dimensional surface.
In this work the techniques and results obtained by using Functional Data Approach on Wall Shear Stress surfaces are presented.

In **Chapter 1**, purposes of the present work, with respect to the scientific literature, are outlined and the process of collection and elicitation of raw data from medical imagery is described. A first exploratory analysis about the composition of patients' set is presented.
In **Section 2.1**, a brief introduction to Functional Data Analysis is presented, underlying its peculiarities with rapport to the classical approach.
In **Chapter 2**, we describe the smoothing procedure, i.e. in this step discrete data are transformed into continuous ones, focusing on the choice of the bandwidth.
In **Chapter 3**, the registration problem is treated, showing the transition from continuous to continuous and aligned data, ready to be analyzed.
In **Chapter 4**, the statistical analysis on continuous, aligned, reduced data is performed by means of the Functional Principal Component Analysis.
Finally, in **Chapter 5**, conclusions of this work are exposed, keeping into considerations also all the results previously found in the AneuRisk project, suggesting other research directions suitable for possible future developments of the program.

Chapter 1

Problem setting and data collection

1.1 The cerebral aneurysmal pathology

Cerebral aneurysms are deformations of cerebral vessels characterized by a bulge of the vessel wall. This pathology is common in the adult population, and usually it is asymptomatic and not disrupting. Epidemiological studies show that between 1% and 6% of adults develop a cerebral aneurysm during their lives ([6]). Rupture of a cerebral aneurysm is usually a tragic event, even if it is a very rare event (about one event in every 10,000 adults per year). To have a rough idea of the damages caused by an aneurysm rupture, one can consider that out of nine people with rupture, three are expected to die before reaching any emergency room, two to die after being hospitalized, two to survive with permanent cerebral lesions and only two to come back to former living habits (Boccardi E., personal communication, 2005).
When a cerebral artery blows out, the skull rigidness and the heart beats make pressure on the brain suddenly and rhythmically increase at each systolic peak. A very high pressure over the brain tissues is reached in just few seconds.

Rupture-preventing therapies are available but they are not without risk. Therapies may be endovascular or surgical; they are shown in Figure 1.1. The surgical treatment - introduced by W. Dandy in 1937 - avoids the blood to enter the aneurysm by clipping the aneurysm neck by means of external platinum clips. The endovascular treatment - introduced by G. Guglielmi in 1991 - consists in filling the aneurysm by means of platinum coils from inside the artery. In both cases the idea is to limit blood pressure in the aneurysm by mechanically forcing blood not to flow into the aneurysmal bulge. Both approaches are quite expensive and risky for patient's life, so any tool providing useful indications about aneurysm rupture probability

would be extremely helpful to address the physician choice of treating or not treating.

A useful classification of cerebral aneurysms is related to their position along the vascular tree (as suggested by [7]). Most cerebral aneurysms appear along the arteries constituting the circle of Willis or closely related to it. Circle of Willis is a complex loop of cerebral arteries at the base of the brain whose arrangement creates redundancies in the cerebral circulation. It is thought to be an emergency mechanism for supplying blood to the brain. Indeed, whether one of the inflow arteries is narrowed or occluded, the blood can flow from the other vessels preserving sufficient cerebral perfusion to guarantee oxygenation of cerebral cells.
A diagram of Willis circle is given in Figure 1.2.
From the front to the back, the arteries constituting the circle of Willis are: the Arterial Communicating Artery (ACoA), the proximal parts of the two Anterior Cerebral Arteries (ACA), the distal parts of the two Internal Carotid Arteries (ICA), the two Posterior Communicating Arteries (PCoA) and the proximal parts of the two Posterior Cerebral Arteries (PCA). The arteries providing blood to the circle are the two Internal Carotid Arteries (ICA) and the Basilar Artery (BA) that is originated by the merging of the two Vertebral Arteries (VA). The main outflow arteries are the two Anterior Cerebral Arteries (ACA), the two Median Cerebral Arteries (MCA) and the two Posterior Cerebral Arteries (PCA).

In order to perform statistical analysis, the medical classification of patients used by the physicians of Niguarda Ca'Granda Hospital Milano has been adopted in all the AneuRisk project to identify homogeneous groups of patients. Patients are divided into two groups:

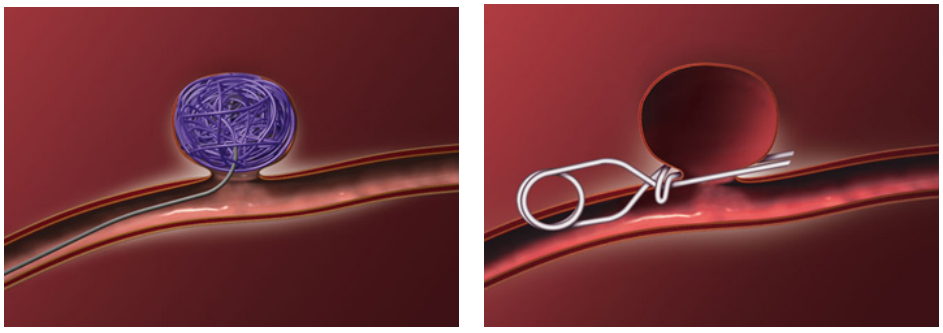


Figure 1.1: On the left, aneurysm endovascular treatment by means of coils. On the right, aneurysm surgical treatment by means of clips (courtesy of http://www.med.umich.edu).

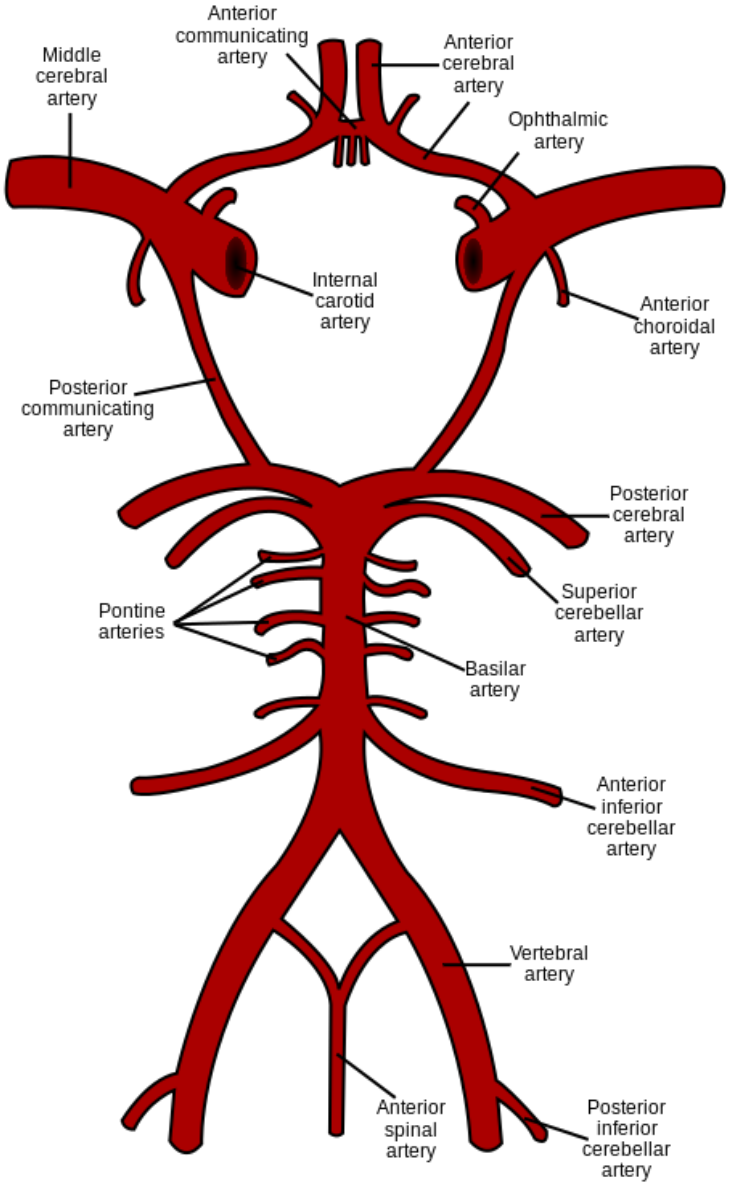


Figure 1.2: View of a complete circle of Willis and the related arteries.

Upper Group. These are the patients with an aneurysm downstream of the ICA, i.e. at the final bifurcation of the ICA or along the MCA, the ACA or the ACoA;

Lower Group. These are the remaining patients, i.e. those with an aneurysm along the ICA or healthy.

Note that no patients with an aneurysm along PCA's or BA are present in the dataset. Indeed, vessels supplied by the left ICA, the right ICA and the BA, cannot be simultaneously detected during the medical exam. Since in these phases of the project we are mainly interested in the analysis of the ICA, all medical examinations not providing data for the ICA have been ignored.

In the scientific community there is not a commonly shared theory about origin and reasons of aneurysms. Several possible explanations are provided by the medical literature. A brief summary is reported in Section 1.1 of [1]. In medical literature, biomechanical properties of artery walls and hemodynamic factors, such as Wall Shear Stress and pressure, are often considered as possible explanatory factors. Hemodynamics is strictly dependent on vascular geometry as reported in [8], [9] and [10]. The role played by morphology on the hemodynamics has been deeply analyzed by [11].
An adimensional index, called "Dean number $\mathcal{D}$", has been proposed in [12] to describe different possible flow situations.
This index depends on blood viscosity and density (quite easy to measure), mean velocity (to be computed by numerical simulations) and two geometrical quantities: vessel radius and curvature.

1.2 The present work and the AneuRisk project

The challenge of this master thesis is to focus the statistical analysis on the fluidynamics factors suggested by the medical literature, such as Wall Shear Stress on the Internal Carotid Artery walls.
Obviously Wall Shear Stress can not be measured directly on the patients or obtained by three dimensional angiographies, because it is a dynamic feature, varying in time, with rapport to heart beating.
So, the mathematical models for blood circulation in the cerebral district studied in [2] have been exploited. These models have been implemented numerically by Emory University producing simulations of the Wall Shear Stress on the artery walls of all the patients previously analyzed in [1] and [3] with rapport to vessel radius and curvature.
In the present work we propose to apply Function Data Analysis techniques to the data obtained by these numerical simulations.

In Chapter 1.3 a first exploratory analysis about the composition of patients' set is presented.
In Chapter 2 we describe the smoothing procedure, i.e. in this step discrete data are transformed into continuous ones, focusing on the choice of the bandwidth.
In Chapter 3 the registration problem is treated, showing the transition from not-aligned to aligned data, ready to be analyzed.
In Chapter 4 the statistical analysis on continuous, aligned data is implemented by means of the Functional Principal Component Analysis, which is the essential statistical tool of this work. All steps of the method are described. Analysis of the scores and interpretation of Principal Component function are performed.
Finally, in Chapter 5 conclusions of this work are exposed, keeping into considerations also all the results previously found in the AneuRisk project, suggesting other research directions suitable for future possible developments of the program.

The context of this work is the AneuRisk Project, a scientific research program that aims at evaluating the role played by different factors in the pathogenesis of cerebral aneurysms and their eventual rupture, with the final purpose to estimate probabilities for aneurysms origin and rupture, in order to provide information which may support physicians' choices. As explained in [1] two different kinds of outcomes were expected to be provided by the project: theoretical results pertaining the development of new mathematical, statistical and engineering methods and practical results consisting in the realization of an automatic procedure providing, for a generic hospitalized patient, a realtime support to medical decisions.

Many fields of sciences are implicated in the project, ranging from medicine to statistics passing by neuroradiology, image reconstruction, bioengineering and finally computational fluidynamics.
Many actors has participated to the project activities: three universities (Politecnico di Milano, Università degli Studi di Milano and Emory University), research groups of public institutes, such as Mario Negri Institute for Pharmacological Researches and the Niguarda Ca'Granda Hospital. Fondazione Politecnico di Milano and Siemens Medical Solutions Italy have funded a PhD position and three research fellowships.
It is evident that collaborative interaction between the different actors has been fundamental for the outcomes of the project and this has been a very challenging aspect.

The main production in the AneuRisk project about statistical analysis corresponds to [1] and [3], where the geometric aspects of arteries dealing with Dean number $\mathcal{D}$, vessel radius and curvature, have been properly inves-

tigated with the tools of Functional Data Analysis, whose details are given in Section 2.1.

1.3 Description, Exploration and Preprocessing of data

In this Section the dataset analyzed in this work is introduced, giving details about all the information contained and features of the sample.
AneuRisk dataset is a very particular dataset in terms of its origin and the different techniques which have made possible the data collection, to which many actors have given their active contributions. In fact, as well outlined in Section a1.2, interaction between professionals coming from very different scientific disciplines has been one of most challenging aspects of the program.
The dataset analyzed in the present study has been built starting from three-dimensional angiographies of $n = 51$ patients hospitalized at the *Neuroradiology Department of Niguarda Ca' Granda Hospital*, Milan, from September 2002 to October 2005.
So the sample is not independent because a sort of "selection" of the statistical units has been performed, by the fact that only people reporting illnesses which possible involvement of cerebral arteries have been submitted to the angiographic investigation. In fact, the three-dimensional angiography is usually conducted by means of contrast liquid and it results in a quite expensive and invasive exam, so it can not be conducted as routine check.
Some of the analyzed patients are affected by an aneurysm along one of the left or right Internal Carotid Artery, others have an aneurysm at the terminal bifurcation of the ICA or after it and, finally, a little part of them are healthy. None of the patients has other severe diseases affecting the cerebral vascular system, apart aneurysms.
People affected by aneurysmal pathology are further divided up into two groups: patients with broken aneurysms or with not broken ones. This dataset contains male and female patients and patients'ages range from 32 to 105 years.

Because of the non-independence of the sample, our work is limited at an *observational study*, without any pretension to conclude absolute inferential results.
Indeed, the exploratory Dataset Analysis reported in Figure 1.3 shows us that the sample is quite homogeneous in respect to the analyzed factors. In fact, percentages of the position of aneurysms along ICA don't differ significantly, (the p-value of the test for equal proportions is 9,6%) and the percentages of broken or not broken are not far from 0.5 at all (the p-value

of the same test as before is 100%). Neither gender percentages differ significantly from 0.5 (the p-value of the same test as before is 65,2%).
Finally, patients' ages appear normally distributed (the p-value of the Shapiro-Wilk test is 5%).

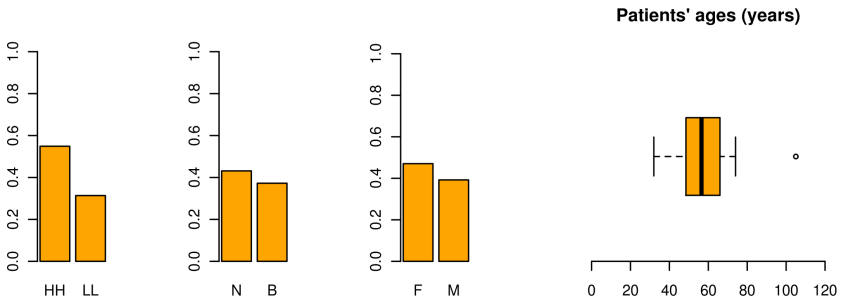


Figure 1.3: Descriptive Statistics of analyzed dataset

Geometrical and morphological features of patient's carotid are reconstructed from a 3D angiography by means of a specific algorithm of image reconstruction implemented by Istituto Mario Negri and coded in the Vascular Modeling Toolkit, available at *http://vmtk.sourceforge.net*, (details are available in Chapter 2 of [1]).
Then, hemodynamics of blood flow inside ICA have been computed starting from the geometrical reconstruction of the Internal Carotid Artery, by using fluidynamics models studied in [2].
For every patient, Wall Shear Stress vector has been mathematically modeled by the function $\mathbf{WSS}_i$ defined by:

$$\mathbf{WSS}_i : \mathbb{R}^3 \longmapsto \mathbb{R}^{3^+} \tag{1.1}$$

$$\mathbf{WSS}_i \;=\; (\mathbf{WSS_x}(x,y,z),\ \mathbf{WSS_y}(x,y,z),\ \mathbf{WSS_z}(x,y,z)) \tag{1.2}$$

$\mathbf{WSS}_i$ maps every point $(x,\ y,\ z)$ of the Internal Carotid Artery to the vector $(\mathbf{WSS_x},\ \mathbf{WSS_y},\ \mathbf{WSS_z})$ evaluated in that point.
Fludynamics models have been implemented numerically by the Department of Mathematics and Computer Science of Emory University, Georgia, USA, after having discretized the n reconstructions of ICAs. These numerical simulations resulted in the computing of the Wall Shear Stress vector values of all the nodes $(x_j,\ y_j,\ z_j)$, $j \;=\; 1, \ldots,\ m_i$ defined by the discretization of patient i.
The discretization in 3D space (x, y, z) chosen by the numerical team is different for every patient and the number of nodes is in the order of 10^5.

In the present work all the analysis is focused on Wall Shear Stress values at the instant of the systolic peak $t = 2.24s$, so our functional data do not depend on time.

A reference system is necessary both for the Internal Carotid Artery mathematical and numerical models in order to define the position of points. So, a parametrization was to be introduced.
The three coordinates $(x,\ y,\ z)$ have respectively the meaning of left–right, up–down, and front–back. Note that these coordinates are not absolute, but are relative to the cubic volume analyzed during 3D angiography. A parametrization more useful is the cylindrical one defined by (S, ρ, Θ), where S is the curvilinear abscissa along ICA, Θ is the angle in radiants and ρ the local radius. More in detail, curvilinear abscissa S is computed along ICA *centerline*, which is an imaginary line situated in the center of ICA lumen.

Before any statistical analysis, a preprocessing of data has been necessary to produce discrete data. This phase of the study has been realized in strict interaction with the team of the Department of Mathematics and Computer Science of Emory University, who performed the numerical simulations. Firstly, the straight pipes added to ICA extremities in order to produce a numerically stable hemodynamics simulation have been eliminated. Two different approaches have been applied to achieve the purpose on both sides of ICA: from the positive-side and the negative-side of S. In the first case, the first node of the centerline with a positive value of S is identified, then the whole part of ICA with values of S greater then the node identified has been removed. In the second one, it has been suggested to cut away a terminal part of ICA having length equal to ten times the value of local radius at the last node of the centerline having negative value of S.

After having completed this preprocessing on data, the first exploratory data analysis conducted has been the graphical analysis of patients' centerlines.
For each patient $i = 1, \ldots,\ n$ the plot of centerlines nodes in the 3D (x, y, z) space has been realized. In Figure 1.4 an example of this plot is reported.
Secondly, the three centerlines projections on the three planes defined respectively by Equations $z = 0$, $x = 0$, $y = 0$ have been computed, and in Figure 1.5 the respective example of these three plots is reported.
As already outlined, one of main purposes and innovations of the present work compared to the previous ones, which are presented in detail in Section 1.2, is to find out correlations between aneurysmal pathology and values of Wall Shear Stress. In this early phase of exploration of Wall Shear Stress effects, we have chosen to analyze the information carried by the Euclidean norm in $\mathbb{R}^3$, $|\mathbf{WSS}|$, of the **WSS** vector only:

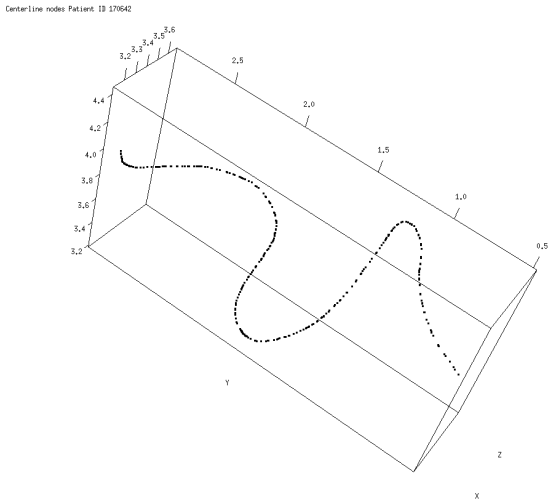


Figure 1.4: Centerline nodes of patient ID=170642

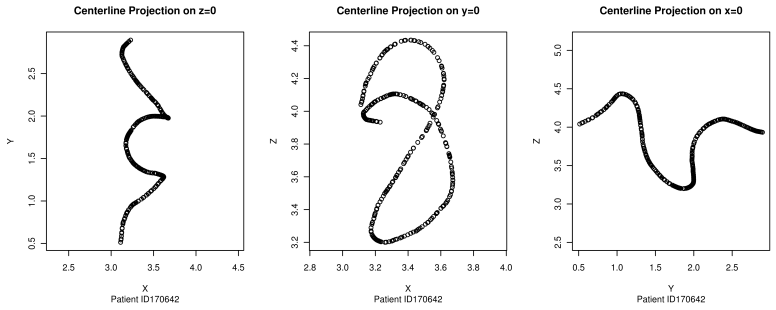


Figure 1.5: The three projections of patient ID=170642 Centerline nodes

$$|\mathbf{WSS}| = \sqrt{\mathbf{WSS_x}^2 \ + \ \mathbf{WSS_y}^2 \ + \ \mathbf{WSS_z}^2} \tag{1.3}$$

Further and deeper analysis on the role played by the three components of Wall Shear Stress vector along the three coordinates x, y, z, respectively $\mathbf{WSS_x}$, $\mathbf{WSS_y}$, $\mathbf{WSS_z}$, may be explored in the future as possible perspectives.
So, for every patient $i = 1, \ldots, n$, for every node $j = 1, \ldots, m_i$ of ICA walls surface the value of $|\mathbf{WSS}|$ is available. From this moment on, ICA surfaces will be represented by cutting the carotids along S, obtaining a two-dimensional surface parametrized by two coordinates: curvilinear abscissa S and angle Θ. This representation will be named *open carotid representation* for the sequel.
An example of *open carotid representation* is given in Figure 1.6.
The functions object of the statistical analysis of this work are the functions f_i 2π-periodic on Θ, defined for every patient $i = 1, \ldots, n$ as follows:

$$f_i : [a, \ b] \times [-\pi, \pi] \longmapsto \mathbb{R}^+ \tag{1.4}$$

$$f_i(S, \Theta) = |\mathbf{WSS}|(S, \Theta) \tag{1.5}$$

where a and b delimit the interval of values of curvilinear abscissa of the stretch of ICA analyzed.

In order to maintain natural proportions and measure units between two coordinates of the ICA in the *open carotid representation*, the coordinate Θ originally expressed in radiants has been changed into a new one, $\tilde{\Theta} = \Theta \times \tilde{R}$ expressed in centimeters, obtained by multiplying the old value for the constant $\tilde{R}$, the same for all patients, defined as the arithmetic mean of all local radius of ICAs, $\tilde{R} = 0.1894875$ *cm*.

Another important peculiarity of the present dataset is the *2π-periodicity* on Θ (*$2\pi \times \tilde{R}$-periodicity* on the new coordinate $\tilde{\Theta}$), and it is immediate to realize that this feature comes directly from the cutting made in the *open carotid representation* along ICA: every point placed on the cutting line is splitted into two nodes placed at the extremities of $\tilde{\Theta}$ and they must have the same $|\mathbf{WSS}|$ value, being $|\mathbf{WSS}|$ a continue function. This has to be strongly kept into consideration when conceiving the smoothing of data, treated in Section 2.

Before computing the smoothed data a grid of nodes common to all the $n = 51$ patients has been defined. This grid $\mathbf{G}$ has been built on the intersection of all the $n = 51$ patients' grids.
Practically, we have placed $n_{\tilde{\Theta}} = 70$ nodes on the $\tilde{\Theta}$ common range [-0.5952925; 0.5952925], the mesh of grid is squared and so the step is 0.01700836 *cm* on both coordinates. So, on the S common range, [-3.386841;

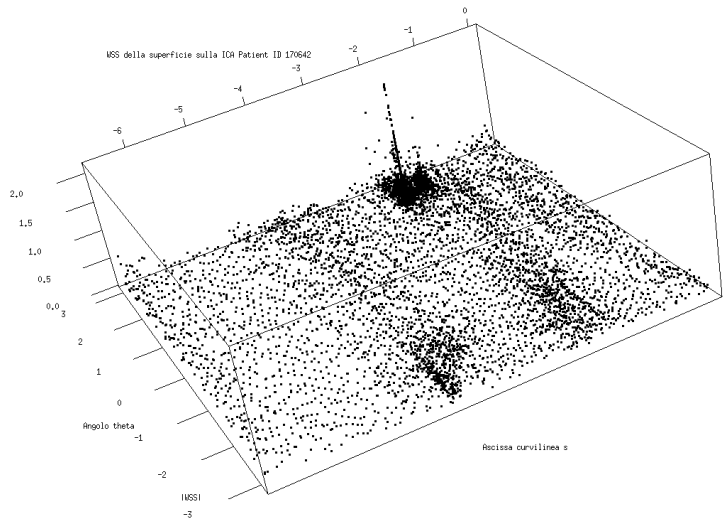


Figure 1.6: Open carotid representation of $|\mathbf{WSS}|$ of patient ID=170642

-0.508594] cm, $n_S = 171$ equally spaced nodes have been fixed.

The grid $\mathbf{G}$ is common to all patients by construction, and it contains $n_{\mathbf{G}} = n_{\tilde{\Theta}} \times n_S = 11970$ nodes and so $\mathbf{G}$ will be the domain used to evaluate the data in the continuous form.

All computations presented in the present work have been implemented and performed in `R 2.12.0`.

Chapter 2

Smoothing of dataset

2.1 The Functional Data Analysis approach

Functional Data Analysis, in the sequel FDA, is one of the most recent branches of statistics and it deals with data which are *functions*, in the mathematical sense of the term.
It is very far from the classical univariate or multivariate statistics, where the statistical units are scalar values or vectors, because all usual concepts, defined on vectorial spaces are not projected into *functional* spaces.
The goals of functional data analysis are essentially the same those of other branch of statistics, according to [5]:

- to represent the data in ways that aid further analysis;
- to display the data so as to highlight various characteristics;
- to study important sources of pattern and variation among the data;
- to explain variation in an outcome or dependent variable by using input or independent variable information;
- to compare two or more sets of data with respect to certain types of variation, where two sets of data can contain different sets of replicates of the same functions, or different functions for a common set of replicates.

On the contrary, Functional Data need preliminary operations before any statistical study. In fact, the first task is to convert the discrete values of observations into a real function $x_i(s)$. This can be done by following different approaches, such as *smoothing* or *interpolation*.
Secondly, to perform every analysis we would be interested in having all data expressed according to the same reference system. This is not obvious and this purpose is reached by means of *registration*, also known as *alignment*.

Finally, the statistical analysis is performed. Many techniques are available in rapport to the aim of the study. One of the most important techniques is the Functional Principal Component Analysis (FPCA), which allows to identify the dominant or substantial *modes of variation*, reducing significantly the dimension of dataset.
A peculiarity of FDA is the importance of *derivatives* of data, which are contained in data and can bring relevant information at our analysis. Further details on Functional Data Analysis are available in [5] or on the website:

`http://www.functionaldata.org`

2.2 The Kernel Smoothing Methods

An unavoidable feature a mathematical model representing a real phenomenon has to respect necessarily is the high fidelity between the quantitative variables of the model and the behavior of quantitative aspects of the phenomenon.
It is quite evident that in real arteries Wall Shear Stress on the walls varies in a *smooth* way, because like in all real systems there are specific dynamics which impose a minimum transition space, making immediate changes impossible to happen.
At this step we have discrete data, but we are interested in having a continuous representation of the Wall Shear Stress on the Internal Carotid Artery walls.
The main purpose of this phase of the work is to obtain the data in a continuous and regularized form starting from discrete data. The obtained functions have to be necessarily continuous but they may be also C^k, for any $k > 1$ hopefully. This procedure is very important and has to be performed very carefully in order to transfer the maximum possible amount of relevant information available from discrete data into the continuous ones, without losing important parts of it. Several techniques are available and they differ from each other according to complexity and features of the smoothed functions produced. The simplest method is the local weighted average but it has been discarded because of the important bias produced by the model near the boundaries of the domain and this is not acceptable at all; this effect is particularly bad on S coordinate, where the absence of periodicity does not help in limiting the bias.
So in the present work we have chosen a locally weighted linear regression. The different behaviors of these two models at the boundaries are shown in Figure 2.1.

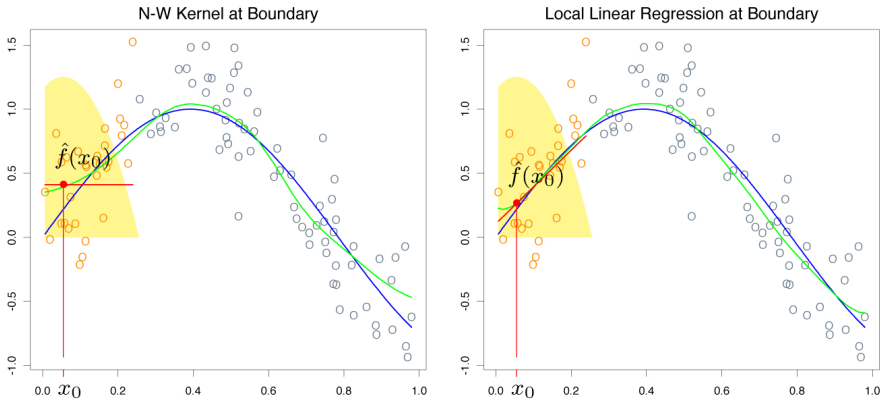


Figure 2.1: Behavior of locally weighted average and locally weighted linear regression at the boundaries (courtesy of [4])

Given a patient i, its target nodes $\mathbf{x_0}$ are the nodes of $\mathbf{G}$, computed in Section 1.3 common to all patients, on the other hand, the training nodes $\mathbf{x_j}$ $j = 1, \ldots, m_i$ are the nodes of ICA walls provided by the numerical simulation of $|\mathbf{WSS}|$ for patient i, finally, y_j, $j = 1, \ldots, m_i$ is the value of $|\mathbf{WSS}|$ for patient i observed at training node $\mathbf{x_j}$. Vector $\mathbf{y}$ is the vector containing all the m_i observations y_j.
In this practice the aim is to estimate the $n = 51$ regression functions $f_i(S, \tilde{\Theta})$ over the domain by fitting a different model separately at each target node $\mathbf{x_0}$ and this is done by using only closest observations to the target node $\mathbf{x_0}$, which are multiplied by weights decreasing as the distance from the target node increases.

The localization is defined by the weighting function $K_\lambda(\mathbf{x_0}, \mathbf{x_j})$, $j = 1, \ldots, m_i$, called *kernel* centred in $\mathbf{x_0}$ and evaluated in $\mathbf{x_j}$, and it corresponds to the weights decreasing as the distance from the target node increases.
Different kernel functions will be characterized in the sequel. So this regression model estimates the value of function $\mathbf{f}_i$ as:

$$\hat{f}(s, \tilde{\theta})_i = \mathbb{E}[\, |\mathbf{WSS}| \mid S = s, \tilde{\Theta} = \tilde{\theta}] \tag{2.1}$$

At each target node $\mathbf{x_0} = (s_0, \tilde{\theta}_0)$, a local linear model of the following form is computed:

$$|\mathbf{WSS}|(S, \tilde{\Theta}) = \beta_0(S_0, \tilde{\Theta}_0) + \beta_1(S_0, \tilde{\Theta}_0)\,(S - S_0) + \beta_2(S_0, \tilde{\Theta}_0)\,(\tilde{\Theta} - \tilde{\Theta}_0) + \varepsilon \tag{2.2}$$

where

$$\boldsymbol{\beta}(\mathbf{x_0}) \;=\; [\beta_0(\mathbf{x_0}),\ \beta_1(\mathbf{x_0}),\ \beta_2(\mathbf{x_0})]^T \tag{2.3}$$

For each linear model, the vector of parameters $\boldsymbol{\beta}(\mathbf{x_0})$ is computed by solving a weighted least squares problem:

$$min_{\beta_0(\mathbf{x_0}),\beta_1(\mathbf{x_0}),\beta_2(\mathbf{x_0})} \;\sum_{j=1}^{m} K_\lambda(\mathbf{x_0},\mathbf{x_j})[|\mathbf{WSS}_j|-\beta_0(\mathbf{x_0})-\beta_1(\mathbf{x_0})\, s_j - \beta_2(\mathbf{x_0})\tilde{\theta}_j]^2 \tag{2.4}$$

The model has been fitted using all, but not only, the m_i training nodes of patient i, but finally the regression function $\hat{f}_i$ is evaluated over $\mathbf{G}$ nodes only.
The estimate function is $\hat{f}_i(\mathbf{x_0}) \;=\; \hat{\beta}_0(\mathbf{x_0})$.

The solution of problem in Equation (2.4) is given in Equation (2.5):

$$\hat{f}(\mathbf{x_0}) \;=\; \mathbf{b}(\mathbf{x_0})^T\ (\mathbf{B}^T\mathbf{W}(\mathbf{x_0})\ \mathbf{B})^{-1}\ \mathbf{B}^T\ \mathbf{W}(\mathbf{x_0})\ \mathbf{y} \tag{2.5}$$

where $\mathbf{b}(\mathbf{x}) \;=\; [1, s, \tilde{\theta}]^T$ and $\mathbf{B}$ is a $(m_i \;\times\; 3)$ matrix defined as

$$\mathbf{B} \;=\; \begin{pmatrix} \mathbf{b}(\mathbf{x_1})^T \\ \dots \\ \dots \\ \mathbf{b}(\mathbf{x_j})^T \\ \dots \\ \dots \\ \mathbf{b}(\mathbf{x_{m_i}})^T \end{pmatrix}$$

$\mathbf{W}(\mathbf{x_0})$ is a diagonal $(m_i \;\times\; m_i)$ matrix containing all the evaluations of the kernel function centred in $\mathbf{x_0}$, as it is shown below:

$$\mathbf{W}(\mathbf{x_0}) \;=\; \begin{pmatrix} K_\lambda(\mathbf{x_0},\mathbf{x_1}) & 0 & \dots & \dots & 0 \\ 0 & K_\lambda(\mathbf{x_0},\mathbf{x_2}) & 0 & \dots & 0 \\ 0 & \dots & K_\lambda(\mathbf{x_0},\mathbf{x_j}) & \dots & 0 \\ 0 & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 & K_\lambda(\mathbf{x_0},\mathbf{x_{m_i}}) \end{pmatrix}$$

An equivalent formulation of (2.5) is

$$\hat{f}(\mathbf{x_0}) \;=\; \sum_{j=1}^{m_j} l_i(\mathbf{x_0})\ y_i \;=\; \mathbf{l}(\mathbf{x_0})^{\mathbf{T}}\ \mathbf{y} \tag{2.6}$$

Equation (2.6) is very important because it highlights that the estimate of f given by model (2.5) is linear in the observations y_j, in fact $\mathbf{l}(\mathbf{x_0})$ does not involve $\mathbf{y}$. The vector of these weights $\mathbf{l}(\mathbf{x_0})$ combines the weighting kernels $K_\lambda(\mathbf{x_0}, \mathbf{x_j})$ and the least squares operations, components of $\mathbf{l}(\mathbf{x_0})$ are often referred to as *equivalent kernels*.

One of main goals of Smoothing methods is that the result of this process, $\hat{\mathbf{f}}_i$, can be a C^∞ function and so its derivatives of all orders can be easily estimated. Continuity features of the final smoothed function depend directly on continuity features of the Kernel function $K_\lambda(\mathbf{x_0}, \mathbf{x_j})$ uniquely.

In the statistical scientific literature different shapes of kernel are available, the most frequently used are Gaussian Kernel, Rectangular Kernel and Epanechnikov quadratic Kernel. The Gaussian Kernel has the great advantage to be C^∞, but its support is infinite and this can lead significantly to longer computing times. On the other hand, Rectangular and Epanechnikov ones has finite support but they are not even C^1 in all the domain. In Figure 2.2 a plot of the three most frequently applied Kernel functions is reported. Further details about shapes of Kernels functions are available in Chapter 6 of [4].

More in detail, for the purposes of the present work the Kernel Smoothing method chosen has to maintain absolutely the periodicity on $\tilde{\Theta}$ coordinate, in fact, the ICA *open carotid representation* is a fictitious representation generated artificially by cutting the carotid along S, so, in this practice a point of ICA wall situated on the cutting line is splitted into two different points placed at the extremities of the 2D surface representing ICA.
We have implemented a bivariate Gaussain Kernel centred on the target node $\mathbf{x_0}$ and having the following matrix as variance-covariance matrix:

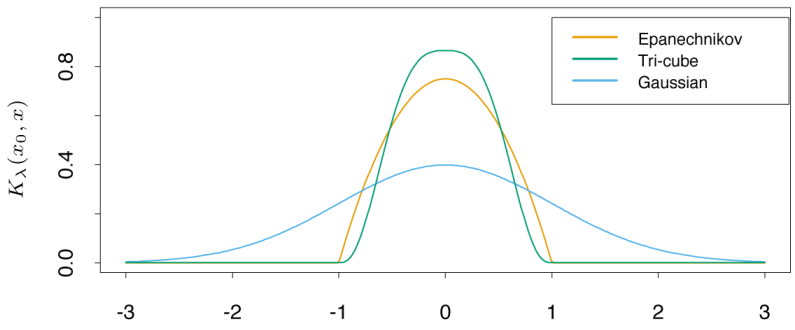


Figure 2.2: Most frequently applied Kernel functions (courtesy of [4])

$$\mathbf{\Sigma} = \begin{bmatrix} \lambda^2 & 0 \\ 0 & \lambda^2 \end{bmatrix}$$

Bandwidth value is the parameter which filters the amount of information from discrete data that will be maintained in the smoothed function, discriminating between primary information and noise along the direction considered.
We have chosen the same bandwidth values for the two directions defined by S and $\tilde{\Theta}$, assuming the signal-to-noise ratio to be isotropic on the two directions.
The bivariate Gaussian kernel $K_\lambda(\mathbf{x_0}, \mathbf{x_j})$ is defined in Equation (2.7).

$$K_\lambda(\mathbf{x_0},\ \mathbf{x_j}) \ = \ \frac{1}{2\ \pi\ |\mathbf{\Sigma}|^{\frac{1}{2}}}\ e^{-\ \frac{1}{2}\ (\mathbf{x_j}\ -\ \mathbf{x_0})^T\ \mathbf{\Sigma}^{-1}\ (\mathbf{x_j}\ -\ \mathbf{x_0})} \tag{2.7}$$

In order to obtain $2\pi \times \tilde{R}$ periodic estimates of Wall Shear Stress functions as outcome of smoothing process, theoretically we should replicate infinite Kernel functions along the $\tilde{\Theta}$ direction. This is not very easy in the practice, so we have augmented data artificially, in fact the fitting of the model has been implemented not only on the original training nodes m_i, but on the training set $\mathbf{T} = \{\mathbf{O} \cup \mathbf{A} \cup \mathbf{S}\}$ where $\mathbf{O}$ is the set of original m_i training nodes obtained at the end of processes described in Section 1.3, $\mathbf{A}$ is a translated repetition of $\mathbf{O}$ where each value $\tilde{\theta}_j$ of $\tilde{\Theta}$ has been changed with $\tilde{\theta}_j + 2\pi \times \tilde{R}$, and finally $\mathbf{S}$ is an analogous repetition translated on the other side of the original dataset.
This approach is fully equivalent to the replication of kernel functions.
Being the support of Gaussian Kernel infinite, in order to maintain the desired periodicity infinite replicates of data for each side should be theoretically included in $\mathbf{T}$, but in the practice, in order to obtain periodic estimates of the Wall Shear Stress functions, it is enough to replicate data one time only by side.

2.3 Bandwidth selection

The second relevant factor defining a Kernel Function is the width λ of the local neighborhood, which will be referred to as *bandwidth*. Large values for λ imply lower variance (averages over more observations) but higher bias. Smoothing process plays a fundamental role in the outcomes of statistical analysis because it filters all the original data and it may lead to restrain all the significant information. On the other hand, it is necessary to discriminate between important information and noise, otherwise the statistical models and the conclusions they lead to are likely to be affected by *overfitting* phenomenon.

So, it is necessary to fix an optimum value to obtain the best possible balance between loss of primary information and overfitting.

In literature the approaches usually chosen to select the bandwidth value are the expected Kullback-Leibler cross-validation, or least-squares cross validation using the method of Racine and Li (2004) and Li and Racine (2004). All these methods split the sample into *training set* and *validation set*. The first is used to estimate the model and the second to test its performances. The procedure is repeated several times choosing different splitting of the sample and the validation results are averaged over all the results.
Both methods are implemented in the function `npregbw` of the cited `R` package but unfortunately they are too expensive in term of computational costs for our purposes, so we have chosen a heuristic approach. In spite of this choice, the crossavalidation remains a very interesting approach and it might be explored as future perspective of the present work.

We have based the choice of λ parameter analyzing extremal cases. To achieve this, the $|\mathbf{WSS}|$ data of all patients have been graphically inspected in order to choose the three cases with the most irregular patterns. The three patients have the following ID: 149198I, 199926, 213558 and their Wall Shear Stress are given in Figures 2.3 and 2.4.

These three test cases have been chosen because they all are far in different directions from the typical map, in fact in the first one there are three important peaks in the left side, in correspondence with lower values pf S. The second map is more regular and higher values of $|\mathbf{WSS}|$ are reached for higher values of S. Lastly, in the third one there is a symmetrical structure

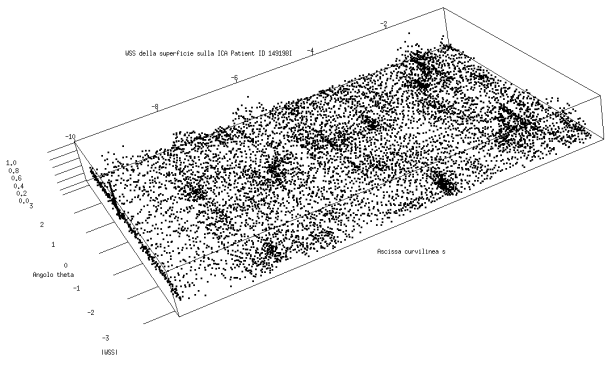


Figure 2.3: Test case patient ID 149198I Wall Shear Stress map

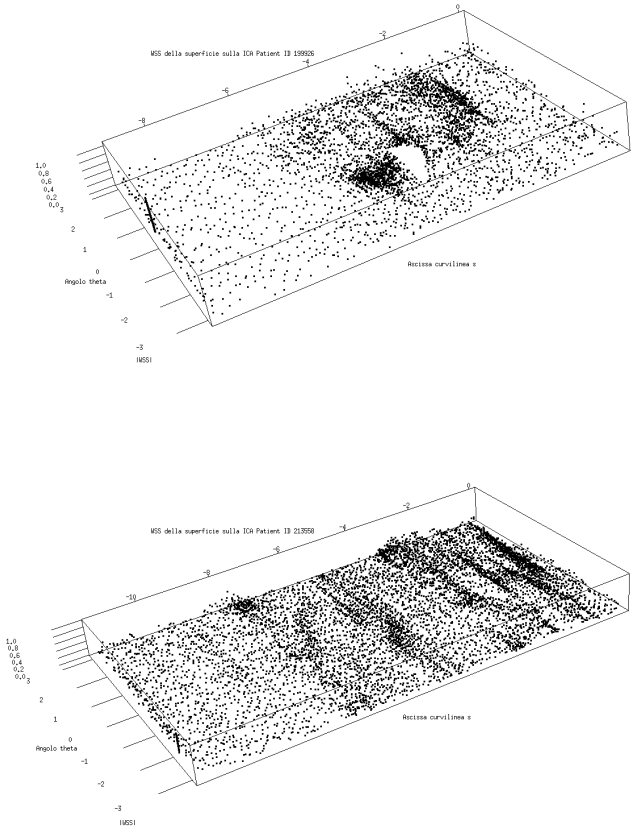


Figure 2.4: Test cases patients ID 199926 (top) and ID 213558 (bottom) Wall Shear Stress maps

which has its maximum values at the extremities.
At this point, for each of this three test cases, seven different smoothed dataset have been computed using the values for λ given in Equation (2.8). The computing of smoothing of functions f_i has been entirely performed by using package `np` of `R 2.12.0`.

$$\boldsymbol{\lambda}_{test} \;=\; [0.01, 0.02, 0.03, 0.04, 0.05, 0.10, 0.15]^T \tag{2.8}$$

In Figures from 2.5 to 2.15 the smoothed data of three test patients are given for the all the four values of bandwidth 0.03, 0.05, 0.10 and 0.15.
A preliminary inspection of the smoothing for bandwidth values 0.03, 0.05, 0.10 and 0.15 has been conducted.
It is evident that it is not possible to choose 0.03 or 0.05 as λ value because the model is led to follow the training data in a too strict way. So it is necessary to select between 0.10 and 0.15.

From a further exploration, $\lambda \;=\; 0.10\ cm$ has been selected, and the plot of smoothed data of the three patients taken as test cases are given in Figure 2.17 where also discrete initial data have been reported.

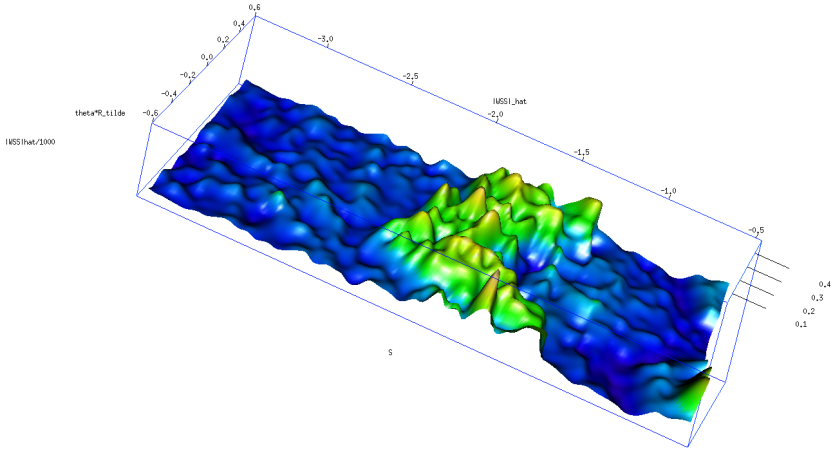


Figure 2.5: Smoothing test case patient ID 149198I - $\lambda = 0.3$

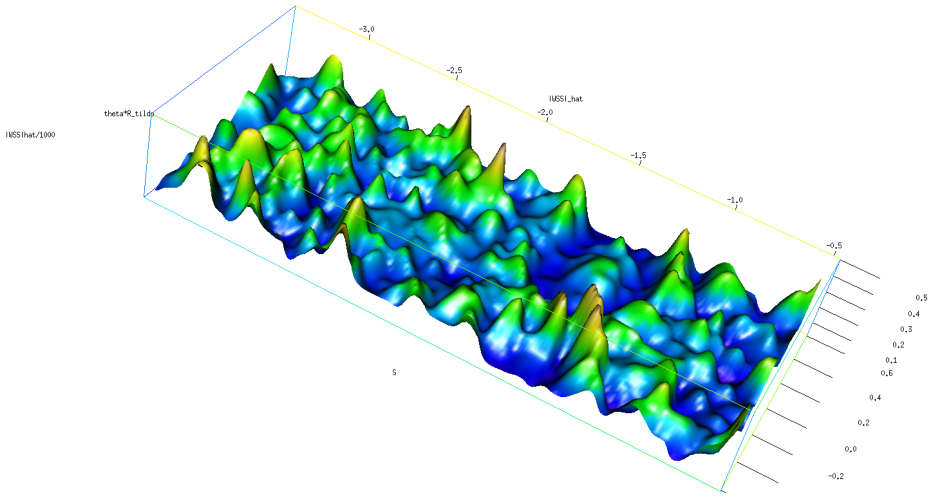


Figure 2.6: Smoothing test case patient ID 199926 - $\lambda = 0.3$

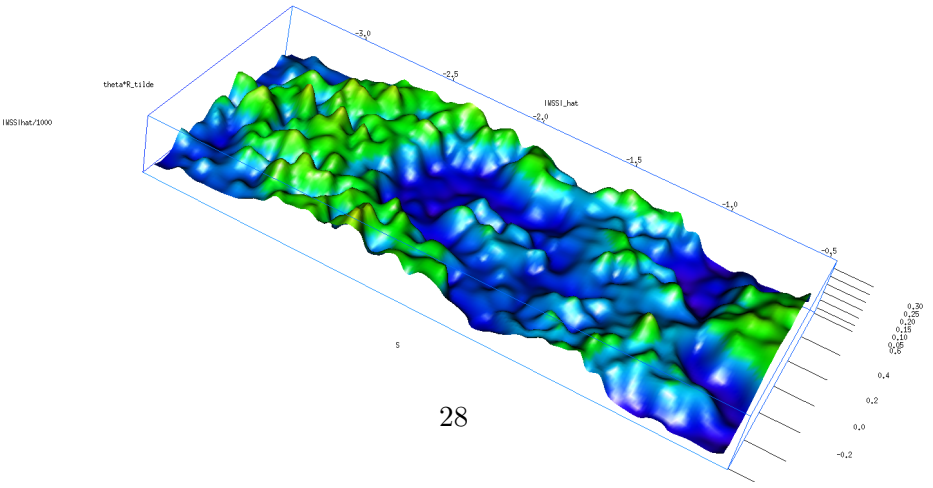


Figure 2.7: Smoothing test case patient ID 213558 - $\lambda = 0.3$

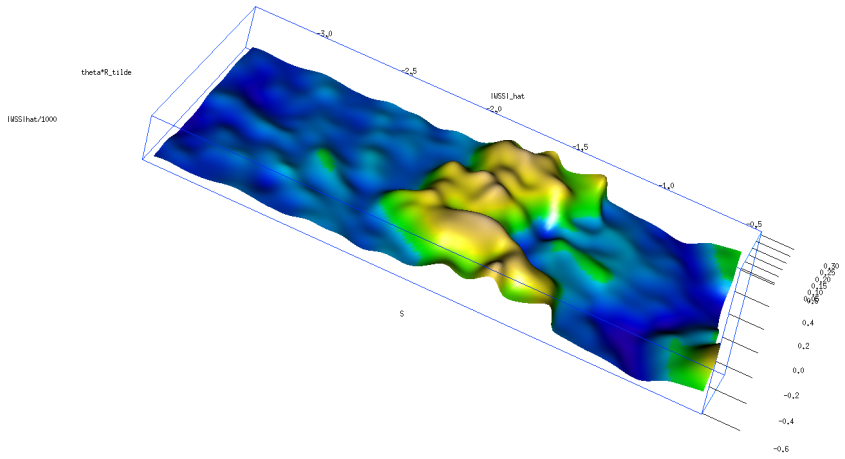


Figure 2.8: Smoothing test case patient ID 149198I - $\lambda = 0.5$

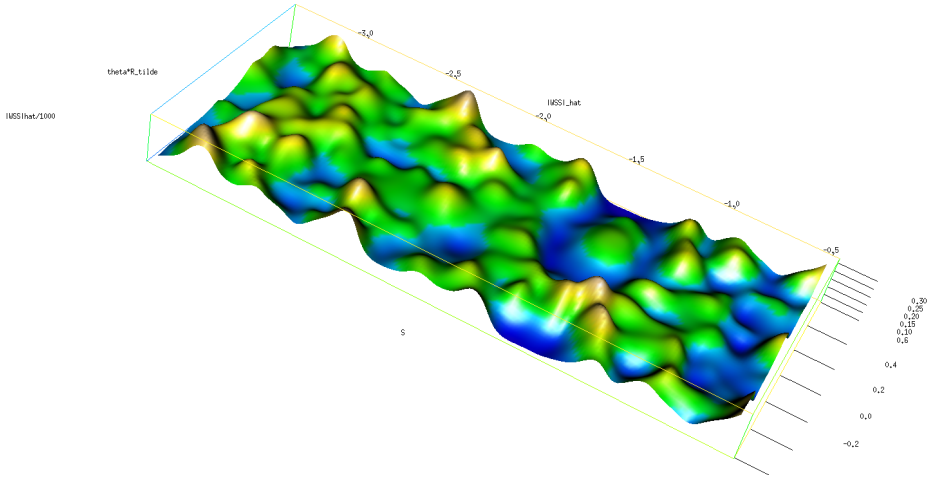


Figure 2.9: Smoothing test case patient ID 199926 - $\lambda = 0.5$

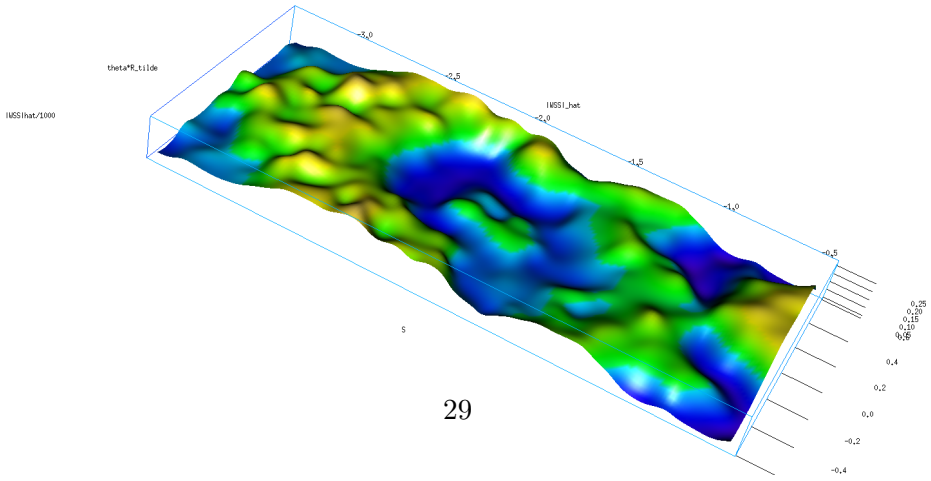


Figure 2.10: Smoothing test case patient ID 213558 - $\lambda = 0.5$

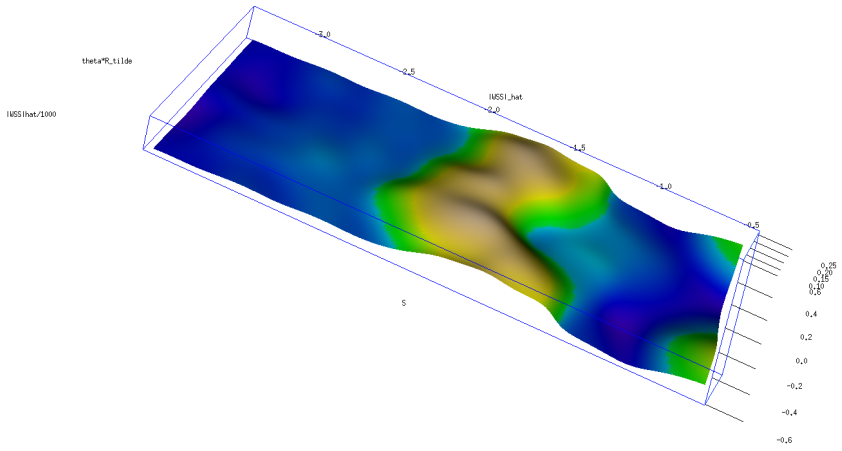


Figure 2.11: Smoothing test case patient ID 149198I - $\lambda = 0.10$

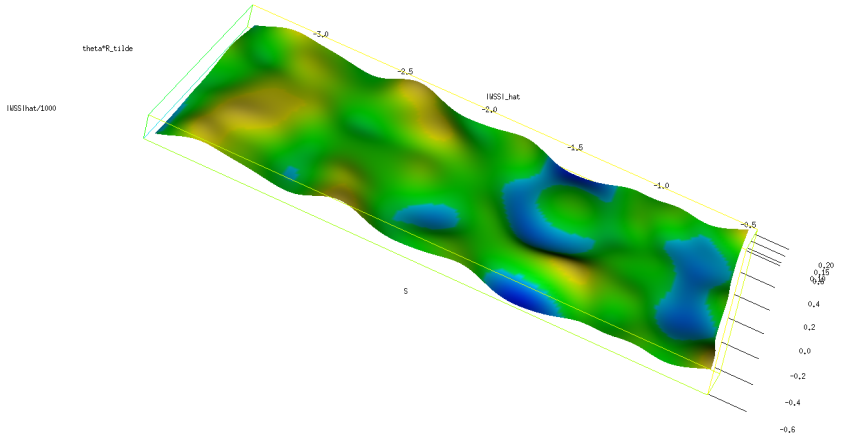


Figure 2.12: Smoothing test case patient ID 199926 - $\lambda = 0.10$

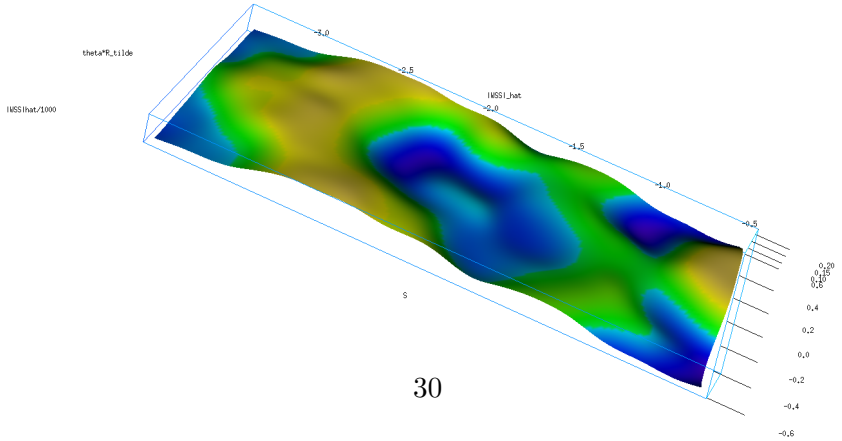


Figure 2.13: Smoothing test case patient ID 213558 - $\lambda = 0.10$

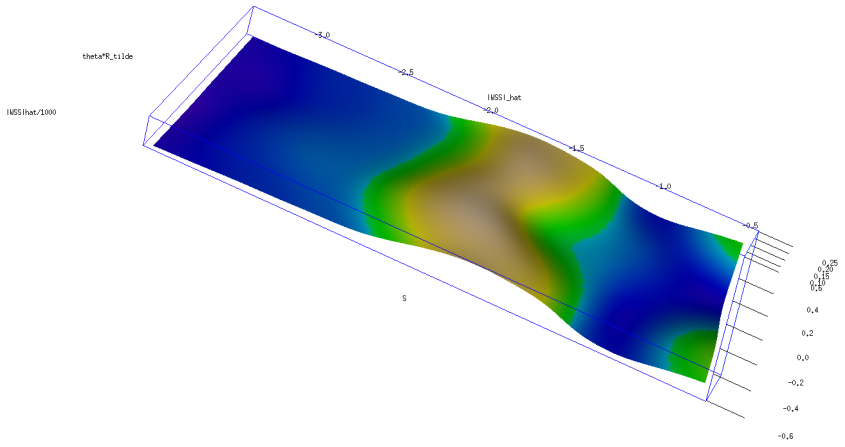


Figure 2.14: Smoothing test case patient ID 149198I - $\lambda = 0.15$

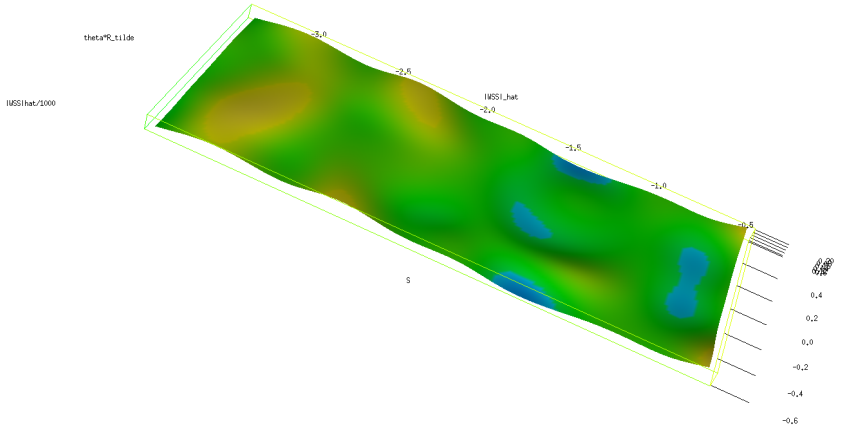


Figure 2.15: Smoothing test case patient ID 199926 - $\lambda = 0.15$

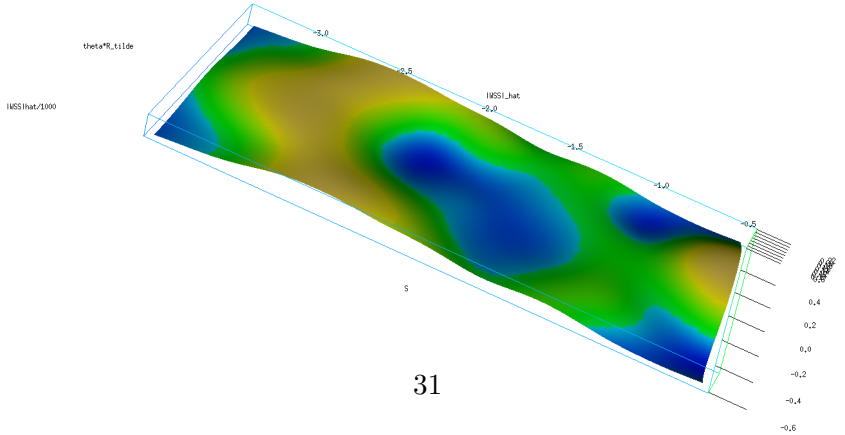
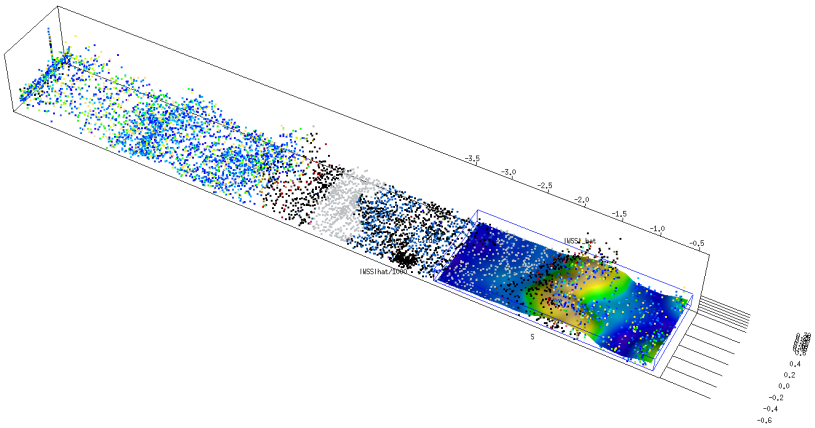
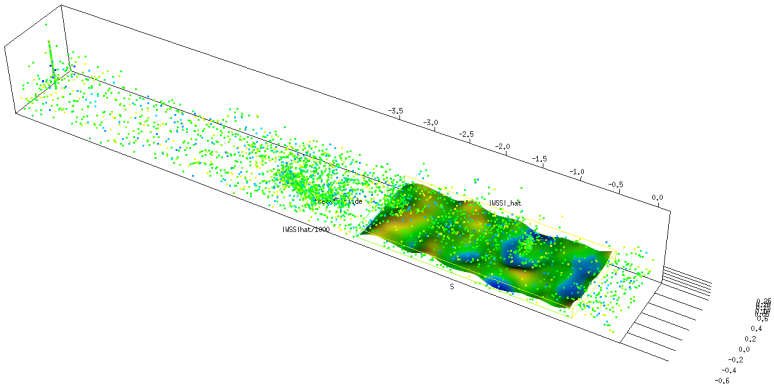


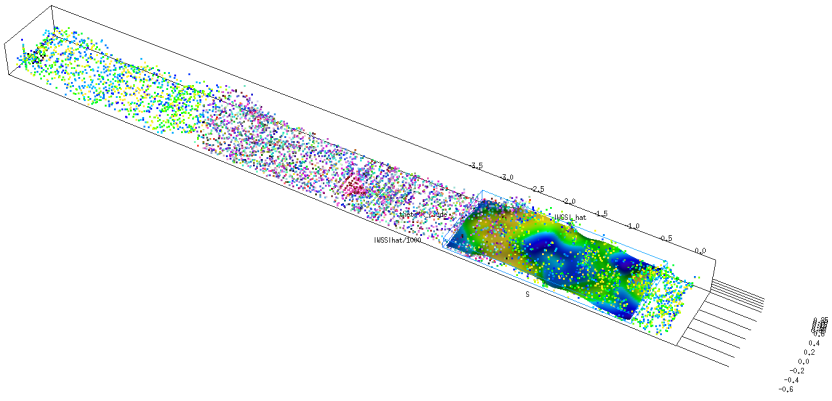
Figure 2.16: Smoothing test case patient ID 213558 - $\lambda = 0.15$



(a) Patient ID 149198I



(b) Patient ID 199916



(c) Patient ID 213558

Figure 2.17: Final Smoothed data for three test cases - $\lambda = 0.10$

Chapter 3

Registration of data

3.1 Amplitude variability and phase variability

At this point of the study, we have obtained the smoothed data, they are in the continuous form but they may not be ready to perform the statistical analysis yet, because as shown in Chapter 7 of [5], stochastic phenomena may contain inherently two different forms of variability: *amplitude variability* and *phase variability*.
The first one pertains to the values assumed by the $|\mathbf{WSS}(S,\ \tilde{\Theta})|$ functions, while, the second one is the variation in the spatial placements of the feature considered by the first one, disregarding their size.
In order to give an illustrative idea of data, average value over $\tilde{\Theta}$ for fixed S, $|\mathbf{WSS}|_{\tilde{\Theta}}(\tilde{\Theta})$, and average value over S for fixed $\tilde{\Theta}$ $|\mathbf{WSS}|_S(S)$, for all the $n = 51$ patients, which are reported respectively in Figures 3.1 and 3.6.

In each plot the sample mean is reported in a bold continuous black line. These functions are defined in Equations (3.1) and (3.2):

$$|\mathbf{WSS}|_S(S) = \frac{\sum_{j=1}^{n_{\tilde{\Theta}}} |\mathbf{WSS}|(S,\ \tilde{\Theta} = \tilde{\theta}_j)}{n_{\tilde{\Theta}}}, \tag{3.1}$$

$$|\mathbf{WSS}|_{\tilde{\Theta}}(\tilde{\Theta}) = \frac{\sum_{j=1}^{n_S} |\mathbf{WSS}|(S = s_j,\ \tilde{\Theta})}{n_S}. \tag{3.2}$$

Looking at Figure 3.1, it is immediate to see that data are *not* aligned, in fact for some groups of patients there are similar structures in the distributions of $|\mathbf{WSS}|$ peaks, for example, the highest pink and purple curves have two peaks, one at each extremity of the domain of S, but they are not placed at the same point. This might be due to phase variability. These peaks do not reach the same values: maximum value of pink curve is greater than purple one, and this aspect has to do with amplitude variability.

Data have been captured in respect to the *rigid* reference system defined

by the angiographic instruments and not the natural biologic one, which is different from patient to patient.
The rigid reference system may not be directly relevant to the biologic dynamics of human body, in fact, it makes no sense to pretend to have a unique artificial reference system to compare ICAs from different patients. So, it might be useful to warp the $n = 51$ scales of the curves obtained at the end of the smoothing phase, by transforming their arguments according to an appropriate procedure which is referred to as *registration* or *alignment*, which has the purpose of finding a good matching of homologous points of ICA between patients, in order to identify the natural reference system.

This work deals with, in particular, the phase variability and, as outlined before, $\tilde{\Theta}$ is a $2\pi \times \tilde{R}$ periodic coordinate, so it is immediately clear how important a good matching across patients is. So, we have registered data on $\tilde{\Theta}$ and all details of registration algorithm are given in the sequel.

Another very important reason reinforcing the choice of registering data on $\tilde{\Theta}$ coordinate is that, as explained before, also position of the zero point for $\tilde{\Theta}$ has been fixed on the mathematical model of every patient's ICA according to numerical algorithms only, such as the parallel transport, and so another arbitrary element is introduced in the original data.

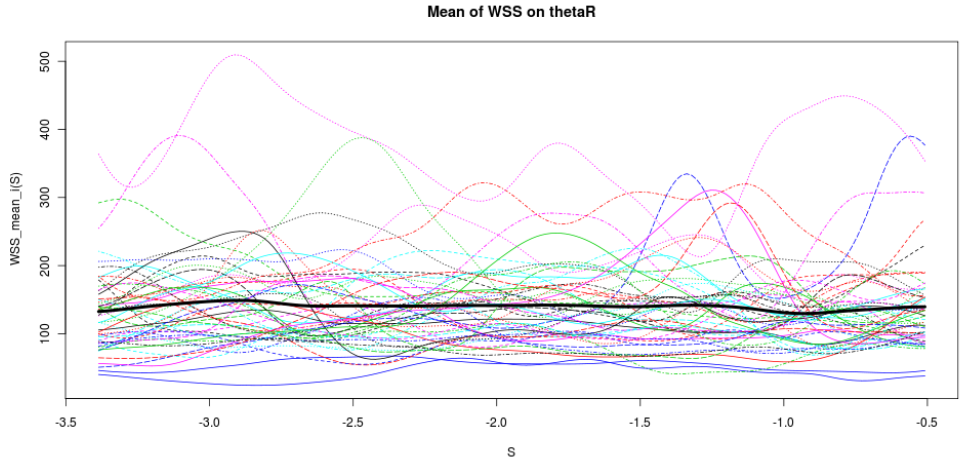


Figure 3.1: $|\mathbf{WSS}|_S(S)_i \; i = 1, \ldots, 51$

3.2 Choice of registration criterion

The final purpose of registration is to obtain all transformed data modified according to a common reference system so as to perform all the analysis. $\tilde{\Theta}_{range}$ is defined as the range $[\tilde{\Theta}_{min} = -\pi \times \tilde{R}; \tilde{\Theta}_{max} = \pi \times \tilde{R}]$ of $\tilde{\Theta}$ values of the grid $\mathbf{G}$, and it is the interval [-0.5952925; 0.5952925] cm; similarly S_{range} is $[S_{min}; S_{max}]$ = [-3.386841; -0.508594]. For the present problem of registration along $\tilde{\Theta}$ the functional space $\mathbf{W}$ chosen is the set of all $2\pi \times \tilde{R}$ one-to-one periodic maps g_α defined as:

$$g_\alpha : \tilde{\Theta}_{range} \rightarrow \tilde{\Theta}_{range} \tag{3.3}$$

$$g_\alpha(\tilde{\Theta}) = \tilde{\Theta} + \alpha, \quad \alpha \in \tilde{\Theta}_{range} \tag{3.4}$$

in order to affect minimally the data, without any further dilatation, or other non-linear operations.
Equations (3.4) specify that translated data must be contained in the same departure set, so as to preserve the periodicity previously described.

$\mathbf{W}$ is the algebraic group of $2\pi \times \tilde{R}$ periodic translations over $\tilde{\Theta}$ in respect to composition of functions, in fact:

- it is closed with respect to composition;
- it has a neutral element ($g_0(\tilde{\Theta}) = \tilde{\Theta}$);
- every $g_\alpha(\tilde{\Theta})$ admits an inverse element with respect to composition, which is $h_\alpha(\tilde{\Theta}) = \tilde{\Theta} - \alpha$.

The group structure of $\mathbf{W}$, in particular the fact that it is closed with respect to composition, supports the iterative procedure presented in the sequel.

For every function f defined on $\tilde{\Theta}$, its periodic L^2_* norm, $\|f\|_{L^2_*}$ is given by:

$$\|f\|^2_{L^2_*} = \int_{\tilde{\Theta}_{range}} f^2 d\tilde{\Theta}. \tag{3.5}$$

For every couple of functions $f(\tilde{\Theta}), h(\tilde{\Theta})$, for every element $g_\alpha \in \mathbf{W}$, it holds the (3.6):

$$\|f(\tilde{\Theta}) - h(\tilde{\Theta})\|_{L^2_*} = \|f(g_\alpha(\tilde{\Theta})) - h(g_\alpha(\tilde{\Theta}))\|_{L^2_*} \tag{3.6}$$

So, the L^2_* periodic norm is invariant with rapport to g_α angular translations. This a very important aspect, because distance between functions is *not* modified by registration and this will result particularly desirable in the

Functional Data Analysis.

Different approaches are available to fix the mechanics of estimating the *shifts* α_i of patient i, as shown in Chapter 7 of [5].
In the present work a global registration criterion is defined. The iterative procedure is conducted as follow:

1. Definition of all possible values for α_k admitted to be the final shift values α_i for all the patients

$$\alpha_k \;=\; \frac{2\pi\tilde{R}}{n_{\tilde{\Theta}}}k \;=\; 0.1894875\; k\; cm \qquad k \;=\; -\frac{n_{\tilde{\Theta}}}{2},\ldots,\frac{n_{\tilde{\Theta}}}{2} \quad i \;=\; 1,\ldots,n \tag{3.7}$$

2. Computing of the estimate $\hat{\mu}_0(S,\tilde{\Theta})$ of the mean of not registered data, in this case data are smooth enough to take the sample average directly;

3. Computing of the global registration criterion $\texttt{REGSSE}_{i,k,0}$ for the patient i, for every shift α_k at iteration 0, i.e. the sum of squares of vertical discrepancies between the shifted curve and the sample mean function:

$$\texttt{REGSSE}_{i,k,0} \;=\; \int_{\tilde{\Theta}}\int_{S}[|\mathbf{WSS}(S,g_{\alpha_k}(\tilde{\Theta}))| \;-\; \hat{\mu}_0(S,\tilde{\Theta})]^2\; dS\; d\tilde{\Theta} \tag{3.8}$$

4. Identification of the value $\tilde{k}_{i,0}$ which minimizes (3.8). Every patient's data are translated according to the $g_{\alpha_{\tilde{k}_{i,0}}}$ found;

5. A new updated estimate $\hat{\mu}_l(S,\tilde{\Theta})$ is computed starting from new data;

6. Steps from 3 to 5 are repeated with the difference that at iteration l the $\texttt{REGSSE}_{i,k,l}$ measures the difference between data of patient i, aligned by shift $g_{\alpha_{k_{i,l}}}$ given by

$$g_{\alpha_{k_{i,l}}} \;=\; g_{\alpha_k} \circ g_{\alpha_{\tilde{k}_{i,l-1}}} \circ \;\ldots\; \circ\; g_{\alpha_{\tilde{k}_{i,0}}} \tag{3.9}$$

   and the estimate of their average $\hat{\mu}_l(S,\tilde{\Theta})$ with rapport to they are being registered.
   The procedure is repeated until an interaction q such that $g_{\alpha_{\tilde{k}_{i,t}}} = g_{\alpha_{\tilde{k}_{i,t-1}}}$ for every $t \;>\; q$. Final α_i is given by

$$\alpha_i \;=\; g_{\alpha_{\tilde{k}_{i,\tilde{q}}}} \circ g_{\alpha_{\tilde{k}_{i,\tilde{q}-1}}} \circ \;\ldots\; \circ\; g_{\alpha_{\tilde{k}_{i,0}}} \tag{3.10}$$

   In any case, in our process a maximum iteration number I_{max} has been fixed equal to 20, according to our expectations based upon previous studies.

This approach is commonly known in the statistical literature as the *Procrustes method*.
For every patient i, at every iteration l $\Delta^1_{i,l}$ and $\Delta^2_{i,l}$ are computed so as to analyze the convergence to a final α_i value.

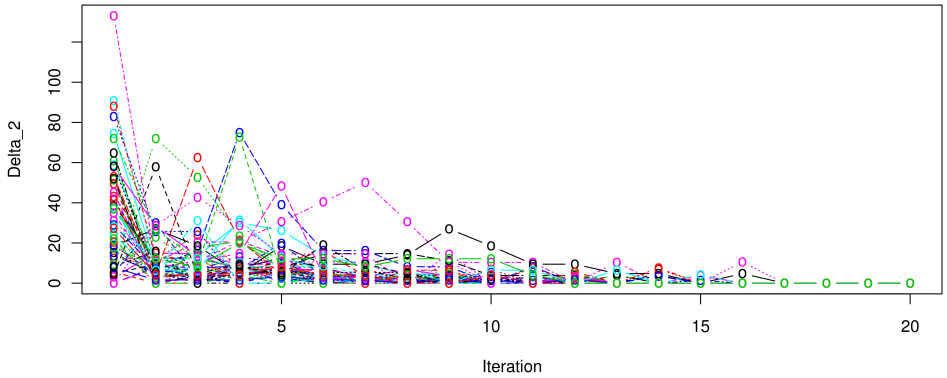


Figure 3.2: Matplot of $\Delta^2_{i,l}$, $i = 1, \ldots, n$, $l = 1, \ldots, I_{max}$

$$\Delta^1_{i,l} = \| |\mathbf{WSS}_i(S, g_{\alpha_{\tilde{k}_{i,l}}})| - \hat{\mu}_l(S, \tilde{\Theta})| \|_{L^2} \tag{3.11}$$

$$\Delta^2_{i,l} = \| |\mathbf{WSS}_i(S,\ g_{\alpha_{\tilde{k}_{i,l}}})| - |\mathbf{WSS}_i(S,\ g_{\alpha_{\tilde{k}_{i,l-1}}})| \|_{L^2} \tag{3.12}$$

From Figure 3.2, where Δ^2_i is reported for every patient, it is clear that from iteration 17 the algorithm does not change shift $\alpha_{\tilde{k}_{i,l}}$ for any patient.
In Figure 3.3 $\Delta^1_{i=28}$ for patient 195206 is plotted in order to show how L^2 norm of the distance between registered data and the mean with respect to data have been aligned decreases. In any case, for some patients, like the one given in Figure 3.4, at the first iterations phenomena of settlement may occur before convergence.
From all these graphical considerations, we are allowed to conclude that Procrustes alignment algorithm converges to final values of shift α_i which allow data registration according to a unique metric.

$|\mathbf{WSS}|_{\tilde{\Theta}}(\tilde{\Theta})_i$ of the final registered and not registered data are reported in Figures 3.5 and 3.6 respectively.
In the first of the two pictures, it is shown that the effect of registration is very nonessential, but it can be appreciated in regard to the plots of smoothed data not aligned.

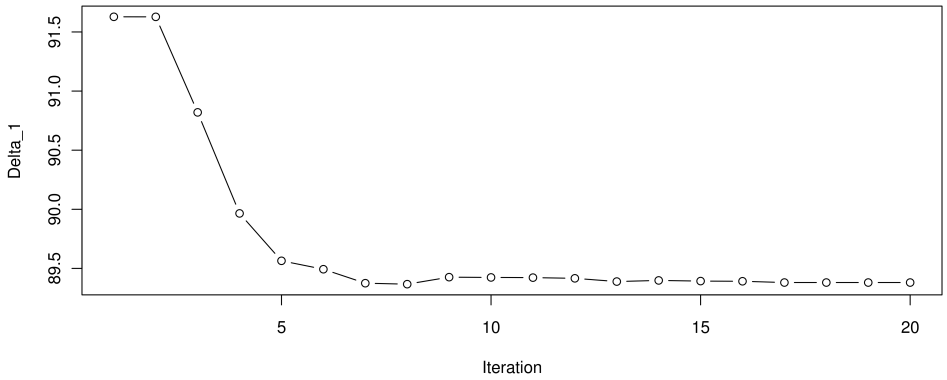


Figure 3.3: Plot of $\Delta^1_{i=28,l}$ for patient 195206, $l = 1, \ldots, I_{max}$

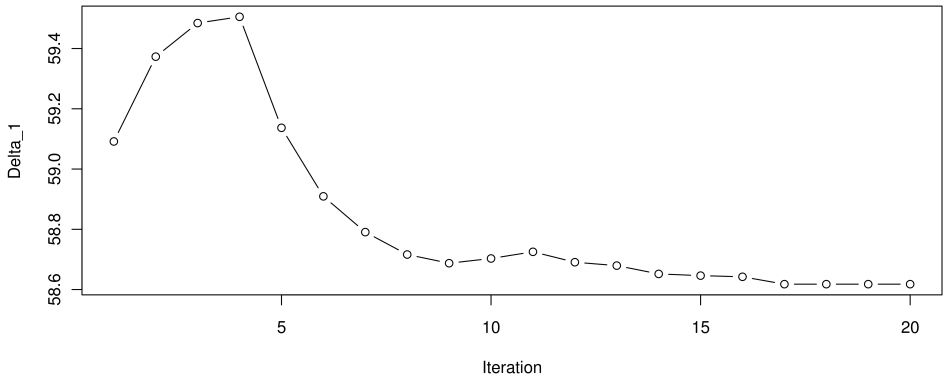


Figure 3.4: Plot of $\Delta^1_{i=12,l}$ for patient 147589, $l = 1, \ldots, I_{max}$

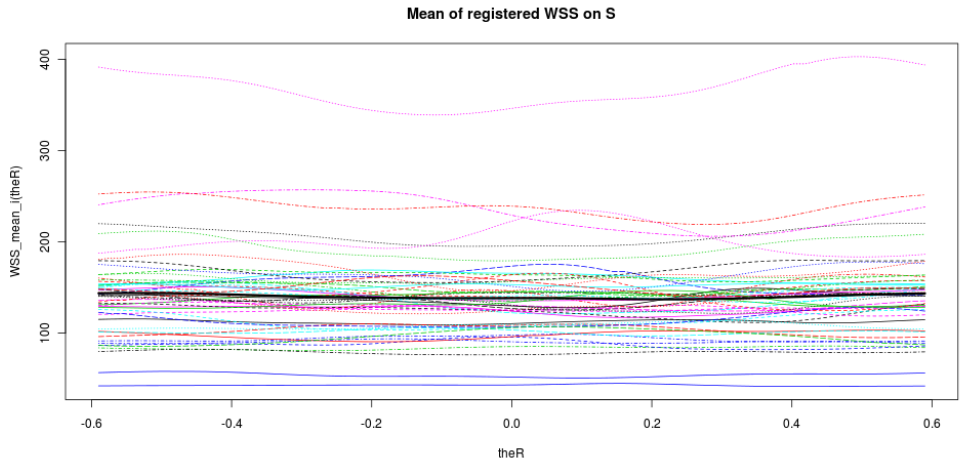


Figure 3.5: $|\mathbf{WSS}|_{\tilde{\Theta}}(\tilde{\Theta})_i$ $i = 1, \ldots, 51$ - Registered data

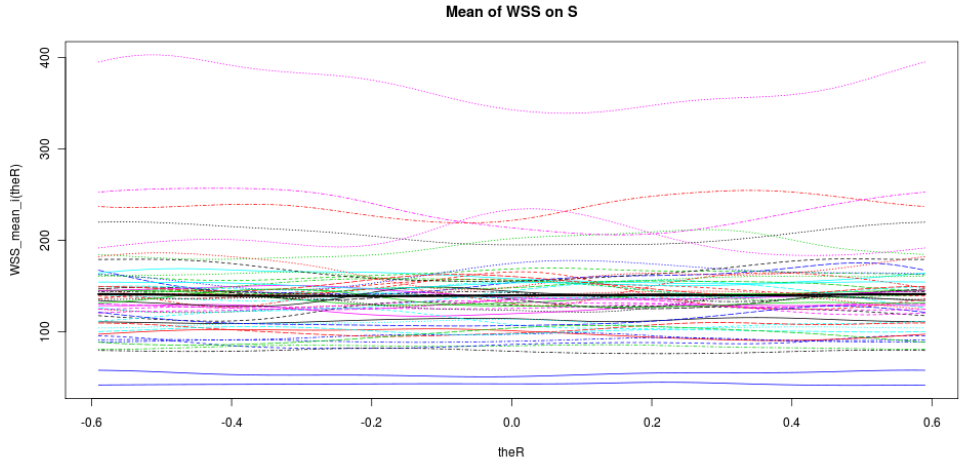


Figure 3.6: $|\mathbf{WSS}|_{\tilde{\Theta}}(\tilde{\Theta})_i$ $i = 1, \ldots, 51$ - Not registered data

3.3 Analysis of the shifts α_i

As already described before, registration has the principal aim of transforming all data to the same reference system, different from artificial one, given by the angiographic exam. At this step a statistical analysis of shifts α_i with rapport to information about aneurysms may be interesting in order to see if the matching of homologous points between patients changes for different typologies of patients.
The analysis proposed is just an observational study, because α_i are periodic data on a circumference, which require particular analysis techniques and a precise inferential analysis lies outside the purposes of the present work.

Initially, an exploratory analysis is conducted to investigate significant differences between the two groups that every biological feature contained in the AneuRisk dataset defines:

- Position of the aneurysm according to lower-upper groups (ICA - Willis);
- Position of the aneurysm on left-right carotid (L - R);
- Aneurysm broken or not (B - N);
- Gender of the patient (F - M);

Definition of Lower and Upper groups is the same as in [1].
Analysis have been performed by means of boxplots in Figure 3.7, where also p-values of T-test for equal means are provided; hypothesis of normality of data in all groups has been verified by Shapiro-Wilk test, and the lowest p-value of these tests is 0.06.

Regarding this early study we conclude that shift parameters α_i are uniformly spread on $\tilde{\Theta}_{range}$ in the groups.
From this and from the inspection of Figures 3.5 and 3.6, phase variability seems to be a nuisance variability to our aim, artificially introduced from the numerical algorithm of image reconstruction and origin point positioning.
The only significant variability is along the S coordinate. In fact, a future perspective may be the exploration of these data taking into consideration variability along S only, representing the Wall Shear Stress with a simpler model which requires less computational costs and which provides more accurate estimates.

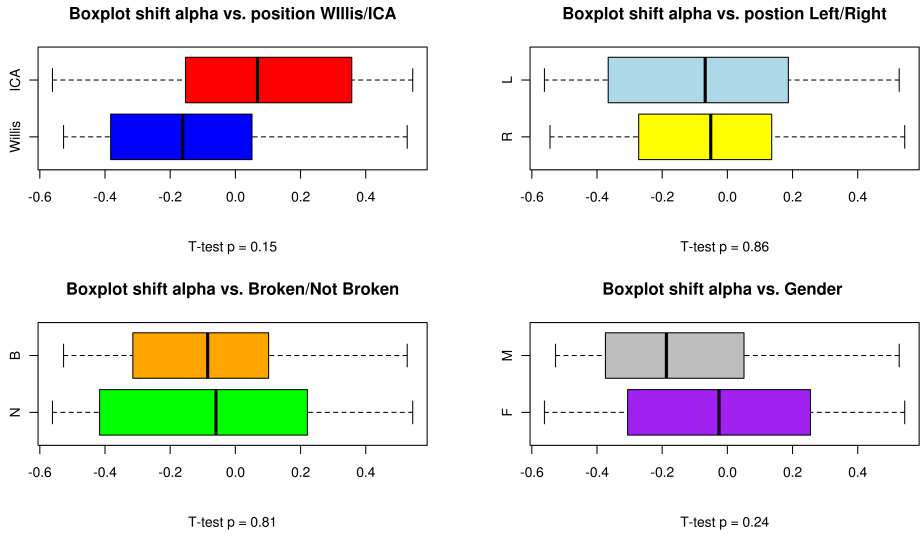


Figure 3.7: Boxplots of shift α

Chapter 4

Functional Principal Component Analysis

At this moment, after all processes described in previous chapters of the present work, we have smoothed and aligned data, which means that discrete initial data have been transformed into n continuous functions $|\mathbf{WSS}|_i$ of S and $\tilde{\Theta}$ according to the open carotid representation. These functions have also the great advantage to be aligned, i.e. all the n surfaces modeling mathematically the ICA walls are represented with rapport to a natural reference system, which is not the artificial one introduced by the angiographic instruments.

The center of this work consists in the analysis of data obtained from the above procedures with Functional Data approach, whose peculiarities and advantages compared to other more classical approaches have been described in Section 2.1.

The Functional Principal Component Analysis (FPCA) is a very powerful statistical tool which allow data dimensional reduction, by projecting data in a lower dimensional space whose generators can be chosen in order to explain important amounts of variability by means of a very low dimension space.
In fact, following this approach, we are going to compute the Principal Component functions, (PC), and select between them, by a very smart and simple procedure, the functions which best explain the variability contained in data. The analysis of selected PC turns out to be extremely useful as first attempt to understand or interpret variability of phenomena of which no *law-driven* model is available, especially for applicative problems.

Practically, the PFCA performed in the present work has been conducted following the paradigm brilliantly adopted in [1] and [3], which can be sum-

marized in these macro-steps:

- Computation of Functional Principal Components;
- Data dimensional reduction;
- Scores computation;
- Data projections computation;
- Interpretation of the results.

In order to investigate the effect of data registration on the final results, we have conducted the Functional Principal Component Analysis, in the sequel FPCA, separately on registered data and not registered ones.
In Sections 4.1, 4.2, 4.4, 4.3, and 4.5, all details of FPCA steps are given and results about registered data commented.

4.1 Computation of Functional Principal Components

The first step, according to FPCA approach, is the computation of the autocovariance function:

$$\Sigma_{\mathbf{WSS}}(t,u) \;=\; \mathbb{E}[(|\mathbf{WSS}(t)| \;-\; \mathbb{E}[|\mathbf{WSS}(t)|])\;\; (|\mathbf{WSS}(u)| \;-\; \mathbb{E}[|\mathbf{WSS}(u)|])] \tag{4.1}$$

In order to have a simpler notation, t and u indicate two points of the domain $S_{range} \times \tilde{\Theta}_{range}$, each of them corresponds to a couple $(s, \tilde{\theta})$.
Numerically, *sample autocovariance* $\hat{v}(t,u)$ function has been evaluated in all the $n_{\mathbf{G}} \times n_{\mathbf{G}} \;=\; 11970 \times 11970 \;=\; 143280900$ possible pairs of $\mathbf{G}$ nodes.

Referring to the same notation adopted in Section 2.1, results set out in Appendix A.5.2 of [5] can be applied in order to obtain the estimates $\hat{\xi}_p(u)$ of the principal component weight functions $\xi_p(u)$. Each of these functions satisies

$$\int_{S_{range} \times \tilde{\Theta}_{range}} \hat{v}(u,t)\hat{\xi}(t)\; dt \;=\; \hat{\rho}\hat{\xi}(u) \tag{4.2}$$

for an appropriate eigenvalue $\hat{\rho}$. The left side of (4.2) is an *integral transform* $\hat{V}$ of the weight function $\hat{\xi}$. This integral transform is called *covariance sample operator* $\hat{V}$ and it is defined by

$$\hat{V}\hat{\xi} \;=\; \int_{S_{range}\times\tilde{\Theta}_{range}} \hat{v}(\cdot,t)\hat{\xi}(t)\; dt \tag{4.3}$$

Therefore, with the new elements introduced above, the *eigenequation* may be expressed directly as

$$\hat{V}\hat{\xi} \;=\; \hat{\rho}\hat{\xi} \tag{4.4}$$

So, our statistical problem corresponds to the mathematical problem of finding eigenvalues and eigenfunctions of the sample covariance operator $\hat{V}$, which is a self-adjoint linear operator.
Spectral theorem for compact self-adjoint operators guarantees that $\hat{V}$ admits an orthonormal basis consisting of eigenfunctions and eigenvalues, which are all real and form a sequence converging to zero.
So, we have

$$\|\hat{\xi}_p(u)\|_{L^2} \;=\; 1, \quad \forall p \;\in \mathbb{N} \tag{4.5}$$

$$<\hat{\xi}_p,\; \hat{\xi}_q> \;=\; 0, \quad \forall p \;\neq\; q \tag{4.6}$$

where $<\cdot,\cdot>$ denotes the usual inner product in $L^2(S_{range}\times\tilde{\Theta}_{range})$ defined by

$$<f,g> \;=\; \int_{\tilde{\Theta}} \int_{S} f\cdot g\; dS\; d\tilde{\Theta} \tag{4.7}$$

Moreover, from the nature of operator $\hat{V}$, all its eigenvalues are non negative. Summing up all this, we are able to obtain a particular decomposition of operator $\hat{V}$ where the $\hat{\rho}_p$ are in decreasing order and all eigenfunctions are orthonormal:

$$\hat{V} = \sum_{p=1}^{\infty} \hat{\rho}_p \;<\cdot,\hat{\xi}_p>\; \hat{\xi}_p \tag{4.8}$$

The rank of operator $\hat{V}$ is $n \;-\; 1 = \;50$, so $\hat{V}$ admits 50 nonzero eigenvalues only.

Numerically, we have computed eigenvalues and eigenfunctions by implementing the *power method* because this method allows us to compute only the first k eigenfunctions with maximum modulus and the importance of this will be shown in the sequel.

4.2 Reduction of dimensionality of data

The statistical meaning of $\hat{\rho}_p$ is fundamental for Functional Principal Component Analysis, because it corresponds to the variability explained by the corresponding eigenfunction $\hat{\xi}_p$.
From what has been described in Section 4.1, the idea of FPCA is to project the data into a lower-dimension new space, generated by the orthogonal eigenfunctions of operator $\hat{V}$.

The main challenge of FPCA is to represent data in a lower-dimension space, that is however able to explain an important amount of data variability. This trade-off is controlled by the opportune choice of the parameter k, corresponding to the number of the first greatest eigenvalues $\hat{\rho}_p$ to be considered.

At this point, we have computed the first 50 pairs $(\hat{\rho}_p, \hat{\xi}_p)$ $p = 1, \dots, 50$ eigenvalues, in fact, being zero all the others $\hat{\rho}_p$, $p = 51, \dots, 11970$, it is no use computing them.
The value for k must be great enough to contain the eigenvalues contributing significantly, excluding, on the contrary, all the remaining eigenvalues. In order to choose k properly, we have investigated plots reported in Figure 4.1 and 4.2, which represent the fraction of explained total variance explained by each eigenvalue $\hat{\rho}_p$, $\frac{\hat{\rho}_p}{\sum_{p=1}^{50} \hat{\rho}_p}$, and the cumulated fraction of explained total variance explained by the first k eigenvalues $\frac{\sum_{p=1}^{k} \hat{\rho}_p}{\sum_{p=1}^{50} \hat{\rho}_p}$.

This exploration suggest $k = 6$. Indeed, we see that in correspondence of

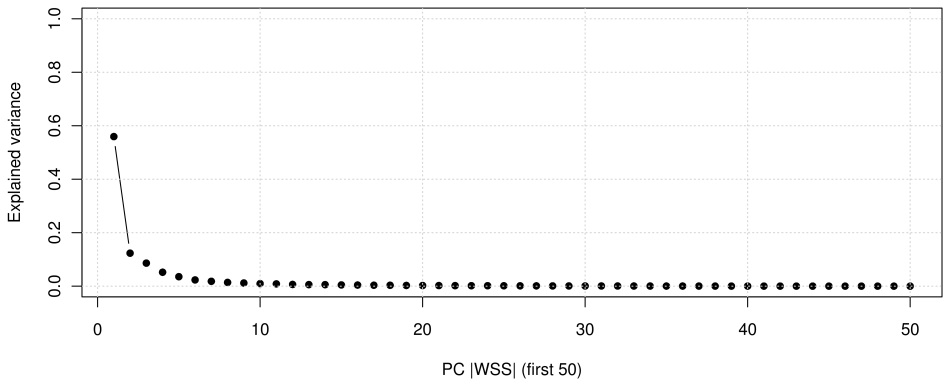


Figure 4.1: Variance explained by eigenvalue $\hat{\rho}_p$

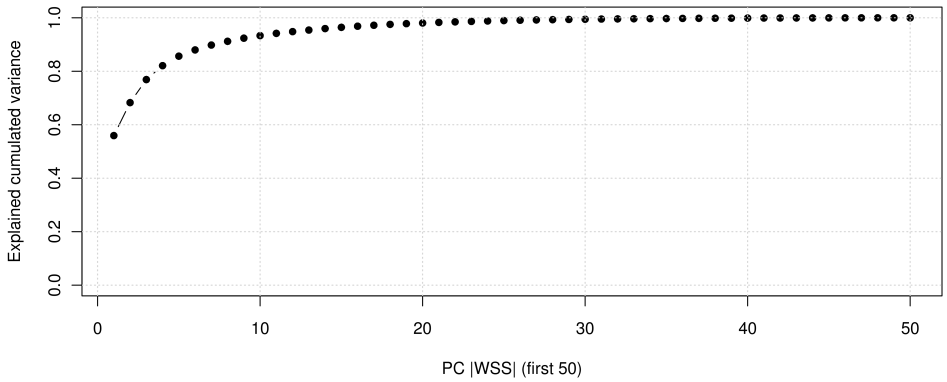


Figure 4.2: Cumulated variance explained by first k eigenvalues

that value, in Figure 4.1 there is a leap, and in Figure 4.2 an elbow.
So, by means of the application of the result (4.8) of the spectral theorem, we have a restricted decomposition $\hat{V}_*$ of operator $\hat{V}$, limited to the first k eigenfunctions, which takes the following form:

$$\hat{V}_* = \sum_{p=1}^{k=6} \hat{\rho}_p \ < \cdot, \hat{\xi}_p > \ \hat{\xi}_p \tag{4.9}$$

At this point we have computed the first $k = 6$ sample eigenfunctions which generate a functional space with lower dimension, on which we are going to project our original data. This new k coordinates are orthogonal and allow us to identify easily how the variability is distributed.

4.3 Analysis of Principal Components

As already seen, we have reduced the dimension of the functional space to a new space, generated by the first $k = 6$ eigenfunctions of operator $\hat{V}$. This set of eigenfunction explains the $89,29\%$ variability of data. So the present step has the aim of interpreting this six eigenfunctions with rapport to the initial problem, in fact, as explained previously, these eigenfunctions has no "a-priori" obvious meaning.

To analyze Principal Components, in the sequel PC, we have visualized each PC function as perturbation of the mean of $|\mathbf{WSS}|$, i.e. always adopting the *open carotid representation*, the overall mean is traced (grey surface) and the functions are obtained by adding (red surface) and subtracting (blue surface), on the same plot, a multiple of the principal component examined.

The multiple is defined by a multiplicative constant equal to the sample standard deviation of the correspondent scores.
This approach allows us to evaluate properly information coming from the PC function with respect to mean function $|\overline{\mathbf{WSS}(S, \widetilde{\Theta})}|$.
The Principal Component functions are traced in Figures 4.3, 4.4, 4.5, 4.6, 4.7, and 4.8.

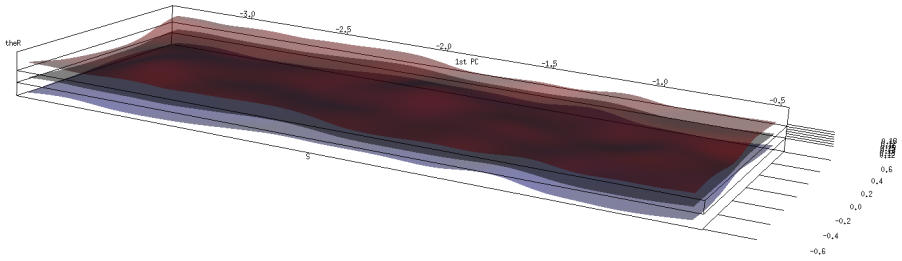


Figure 4.3: 1st PC (55,93%)

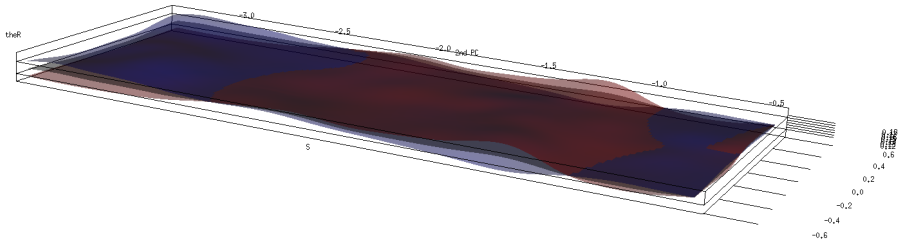


Figure 4.4: 2nd PC (12,34%)

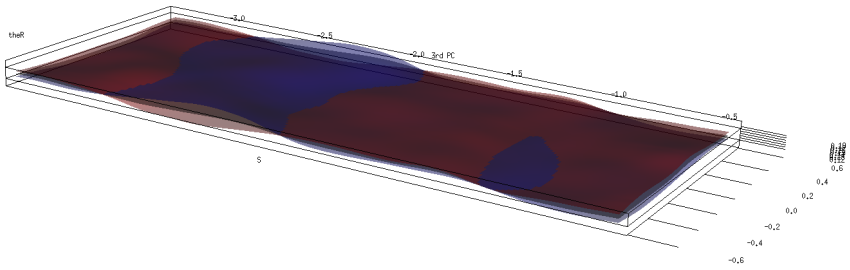


Figure 4.5: 3rd PC (8.62%)

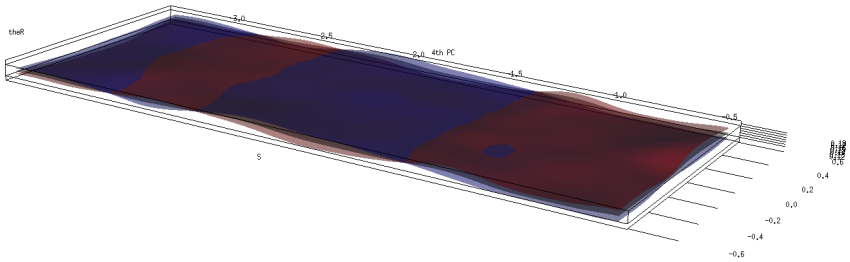


Figure 4.6: 4th PC (5,21%)

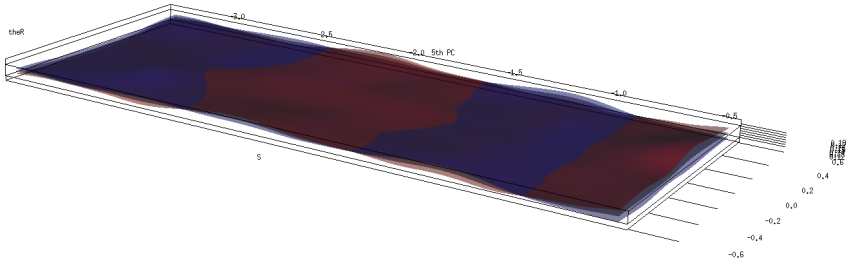


Figure 4.7: 5th PC (3,55%)

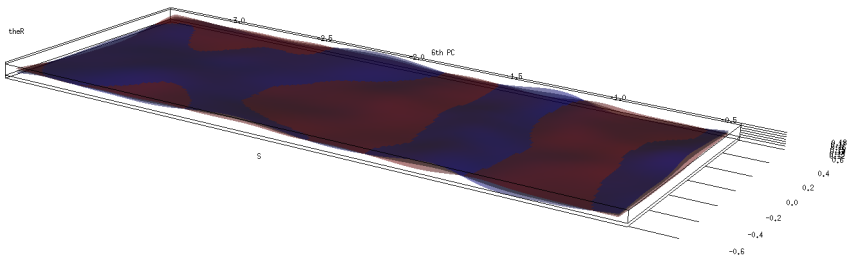


Figure 4.8: 6th PC (2,33%)

4.4 Analysis of the scores

According to the results found in Section 4.2, from this point on, we will limit the analysis to the first $k = 6$ eigenfunctions.
The *score* $\gamma_{i,p}$ corresponding to the ith obserbeed curve $|\mathbf{WSS}|_i$ and the pth estimated eigenfunction $\hat{\xi}_p$ is defined as the component along $\hat{\xi}_p$ of the ith observed curve $|\mathbf{WSS}|_i$ centred around the sample mean $\overline{|\mathbf{WSS}|}$:

$$\gamma_{i,p} = \int_{\tilde{\Theta}} \int_{S} (|\mathbf{WSS}(S, \tilde{\Theta})|_i - \overline{|\mathbf{WSS}(S, \tilde{\Theta})|}) \, \hat{\xi}_p(S, \tilde{\Theta}) \, dS \, d\tilde{\Theta} \tag{4.10}$$

where $\overline{|\mathbf{WSS}|}$ is defined by:

$$\overline{|\mathbf{WSS}|} = \sum_{i=1}^{n=51} |\mathbf{WSS}(S, \tilde{\Theta})|_i \tag{4.11}$$

So, at each patient are associated $k = 6$ scores, one for each eigenfunction.

The statistical interpretation of $\gamma_{i,p}$ is a fundamental part of the present work, in fact score value $\gamma_{i,p}$ quantifies the influence of the p^{th} eigenfunction on the deviation of i^{th} patient from the average behavior, i.e. $\overline{|\mathbf{WSS}|}$.

In order to investigate Wall Shear Stress maps on the ICA with respect to aneurysmal pathology features, the analysis of the scores plays a fundamental role.
So we have analyzed the scores corresponding to the $k = 6$ selected eigenfunctions selected, investigating scores regarding groups Lower/Upper, Broken/Not Broken, Left/Right, Male/Female gender. We have conducted an exploration by means of separate boxplots according to different groups, and hypothesis tests for equal central values; in fact scores in the groups are not always normally distributed, so T-test for equal means cannot be applied. As in all the analysis performed in the present work normality hypothesis has been checked through the Shapiro-Wilk test.

To test the equality of central values, in our case the median, the Wilcoxon test has been implemented. It is a non parametric permutation test which operates on the ranks. Further details about non parametric hypothesis testing are available in [13].

In Figures 4.9, 4.10, 4.11 and 4.12 are reported all boxplots introduced for the analysis of scores described in the present Section.

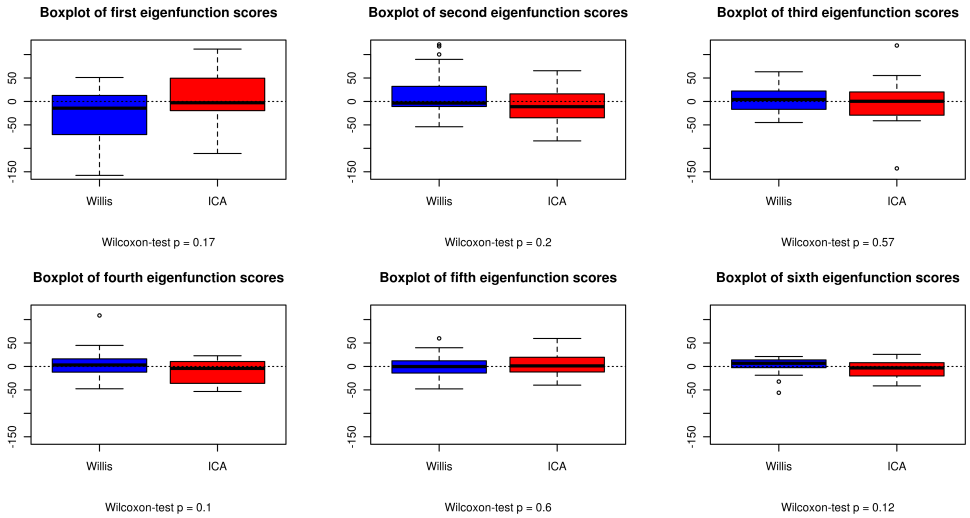


Figure 4.9: Boxplots of $\gamma_{i,p}$ for Lower(ICA)/Upper (Willis)

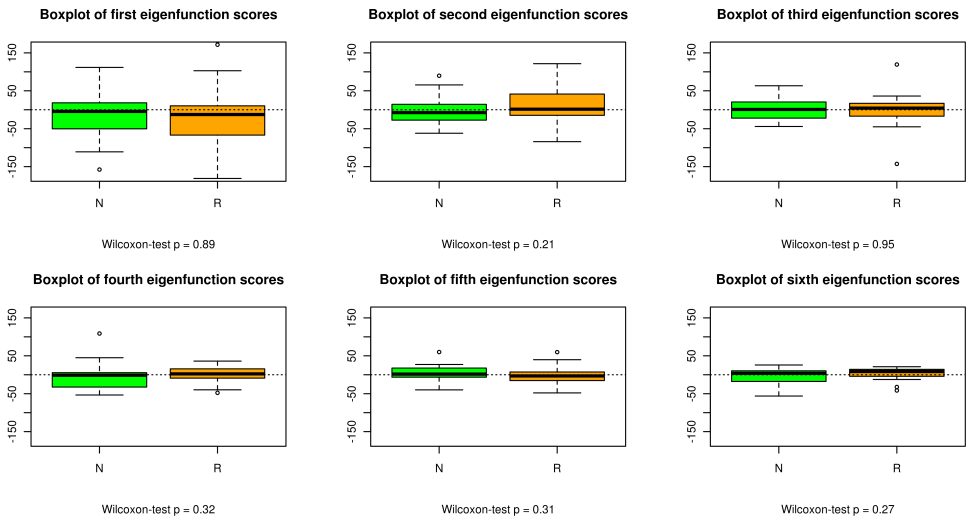


Figure 4.10: Boxplots of $\gamma_{i,p}$ for Broken (R)/Not Broken (N)

By means of a MANOVA global testing on the scores of the first $k = 6$ PC functions, we have looked for the presence of significative differences between groups, but no significative result was found.

All results of the FPCA presented in this Chapter are discussed in Section 4.6.

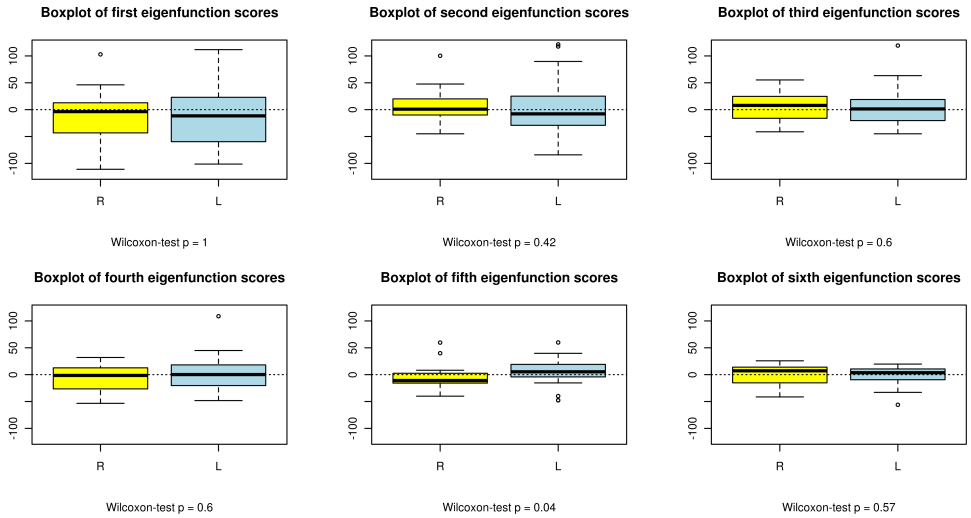


Figure 4.11: Boxplots of $\gamma_{i,p}$ for Left/Right

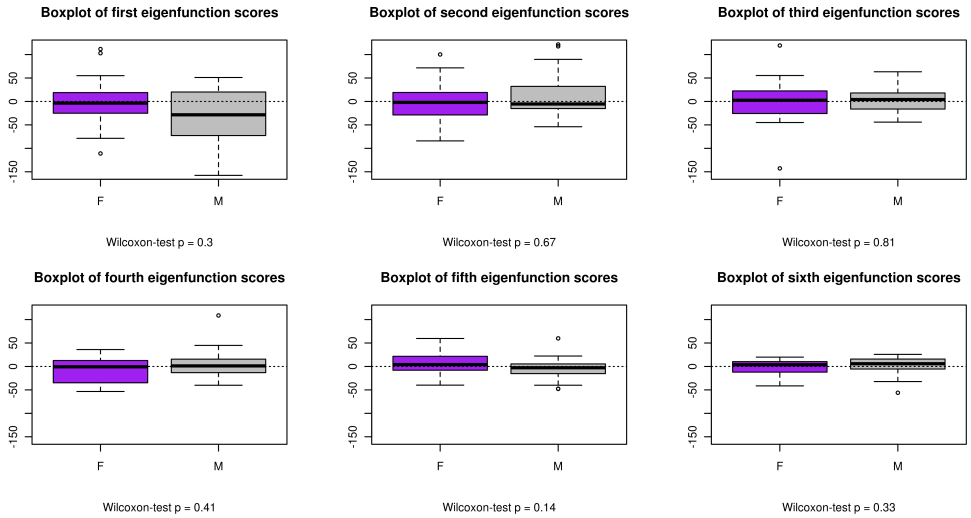


Figure 4.12: Boxplots of $\gamma_{i,p}$ for Male/Female

4.5 Analysis of projections of data

The last step of the FPCA consists in the computation of projections of all the $|\mathbf{WSS}|_i$ functions on the $k = 6$ PC functions selected.
Projection of $|\mathbf{WSS}|_i$ on Principal Component p is given by:

$$|\mathbf{WSS}|_{i,p} = \overline{|\mathbf{WSS}|} + \gamma_{i,p} \times \hat{\xi}_p \qquad i = 1, \ldots, 51 \quad p = 1, \ldots, 6 \qquad (4.12)$$

Each projection contains the information carried by the overall mean and by the p^{th} Principal Component function, i.e. we are focusing on the variability explained by the involved PC uniquely.
These projections have the purpose to make clear how each Principal Component function has impact on every patient.

In each of Figures 4.13, 4.14, 4.15, 4.16, 4.17 and 4.18 are reported projections the $n = 51$ projections $|\mathbf{WSS}|_{i,p}$ on the considered PC, coloured according their values.

In each of Figures 4.19, 4.20, 4.21, 4.22, 4.23 and 4.24 are reported the $n = 51$ projections $|\mathbf{WSS}|_{i,p}$ on the considered PC, coloured according Upper(blue)/Lower(red) group. It is quite evident from these plots that eigenfunctions do not allow to discriminate between patients of the Upper or Lower group.

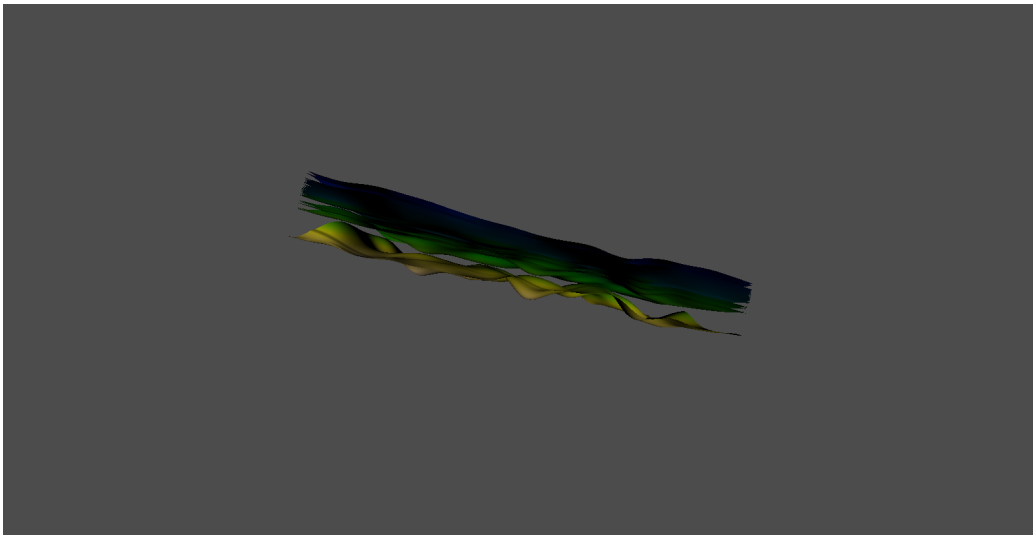


Figure 4.13: Projections on 1st PC

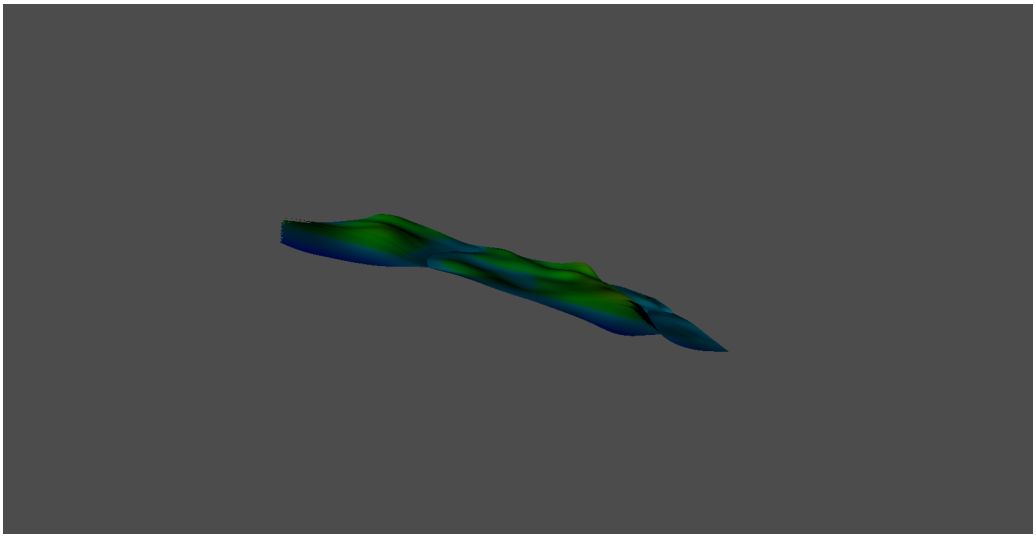


Figure 4.14: Projections on 2nd PC

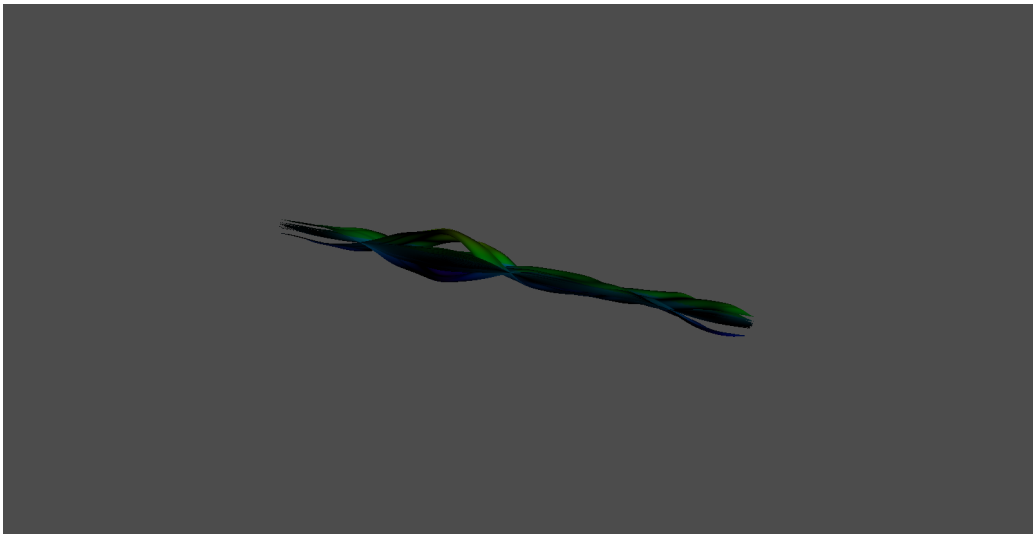


Figure 4.15: Projections on 3rd PC

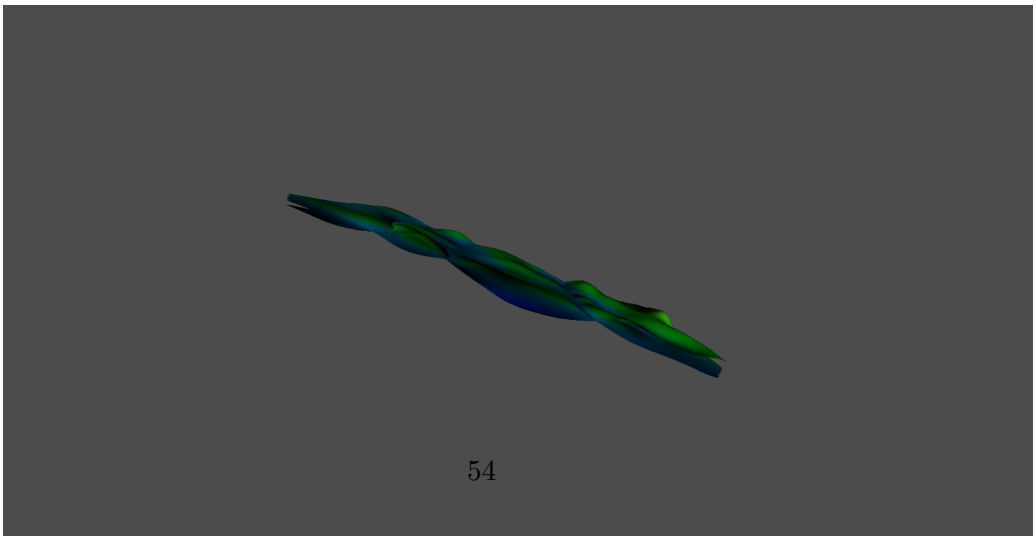


Figure 4.16: Projections on 4th PC

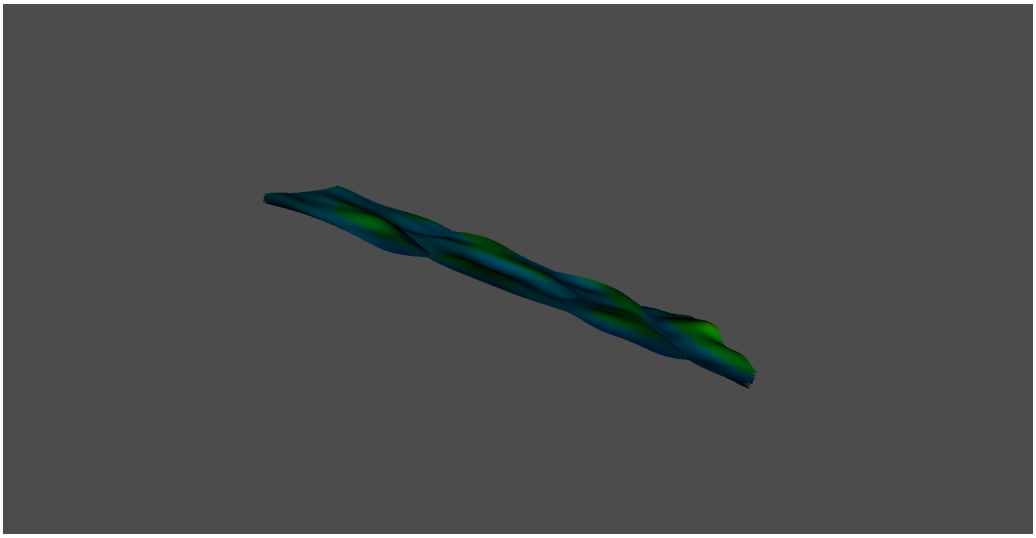


Figure 4.17: Projections on 5th PC

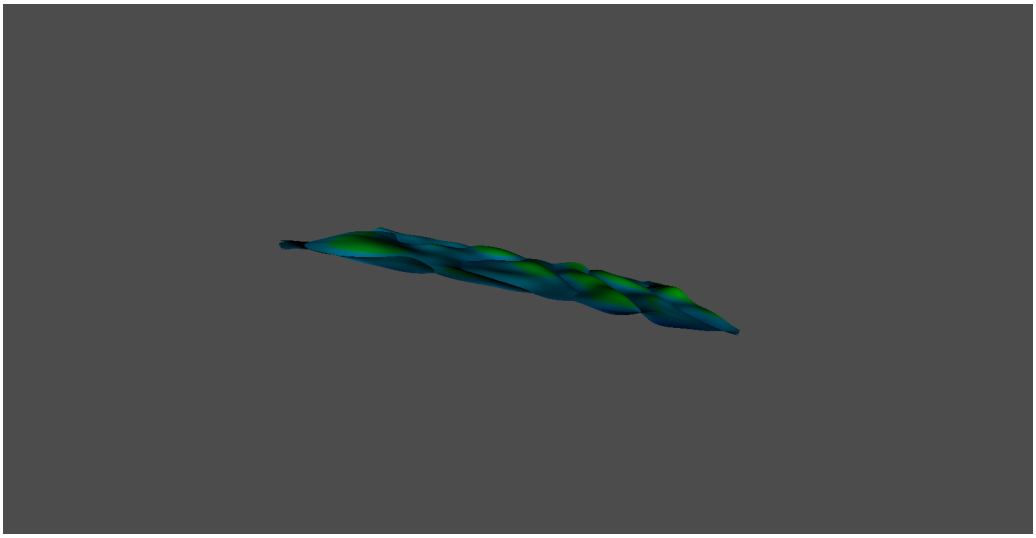


Figure 4.18: Projections on 6th PC

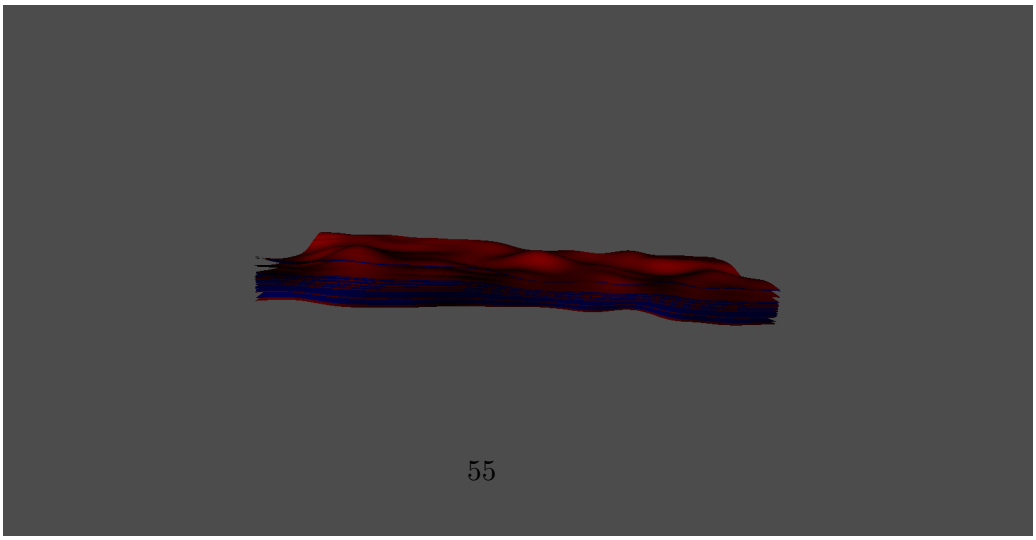


Figure 4.19: Projections on 1st PC - Upper/Lower groups

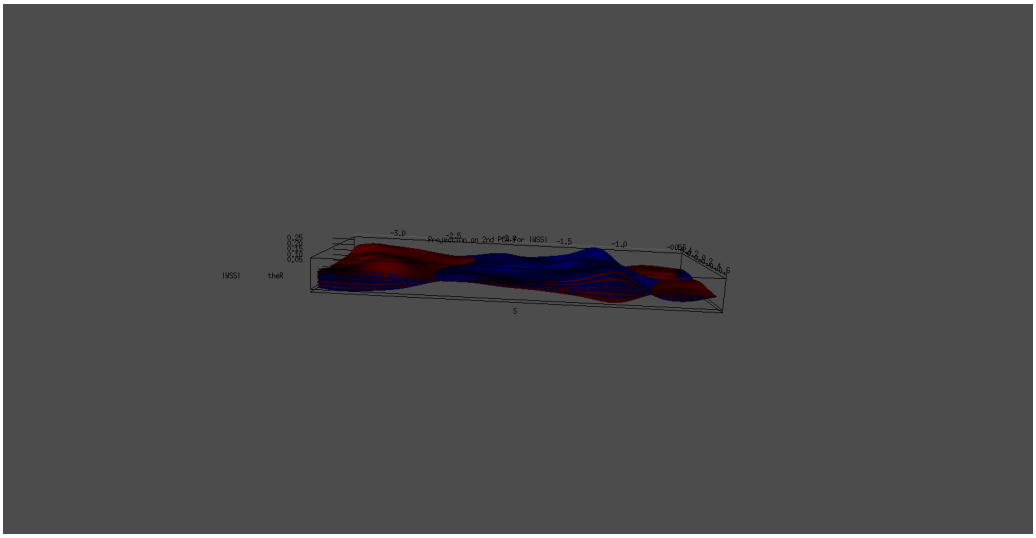


Figure 4.20: Projections on 2nd PC - Upper/Lower groups

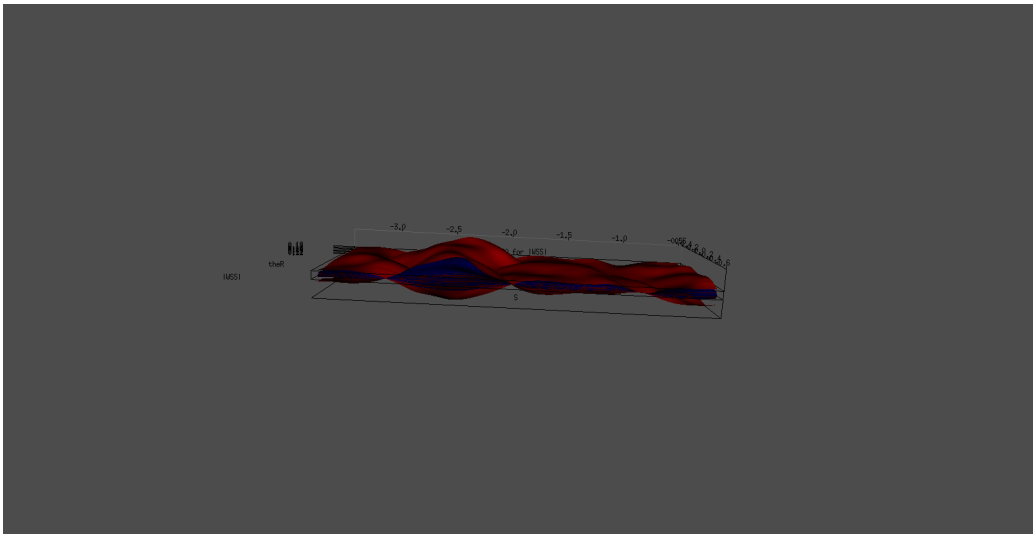


Figure 4.21: Projections on 3rd PC - Upper/Lower groups

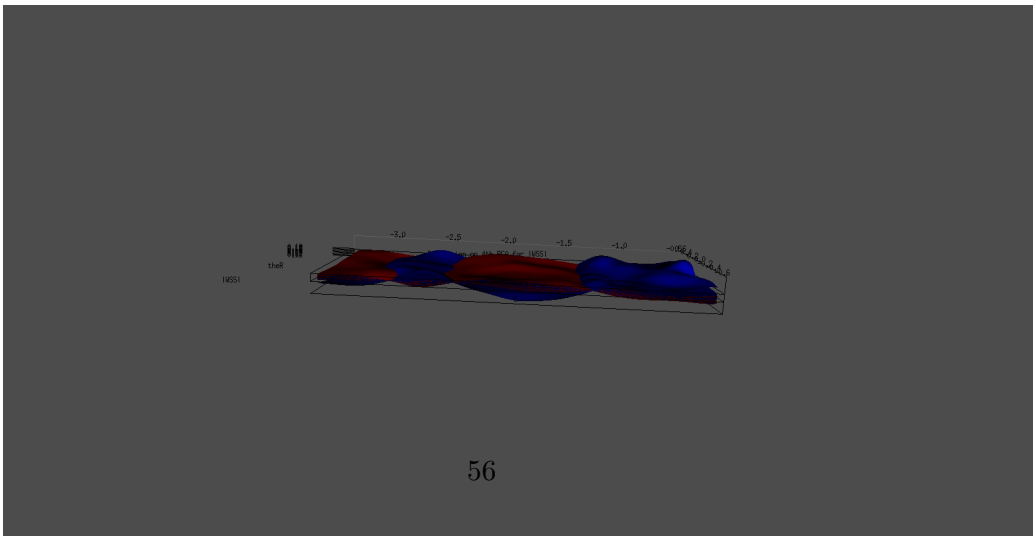


Figure 4.22: Projections on 4th PC - Upper/Lower groups

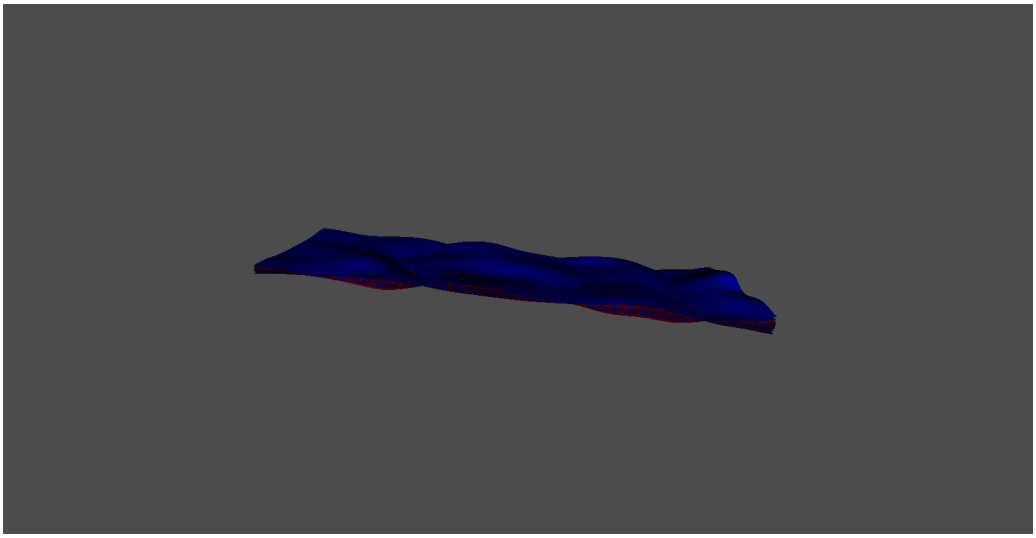


Figure 4.23: Projections on 5th PC - Upper/Lower groups

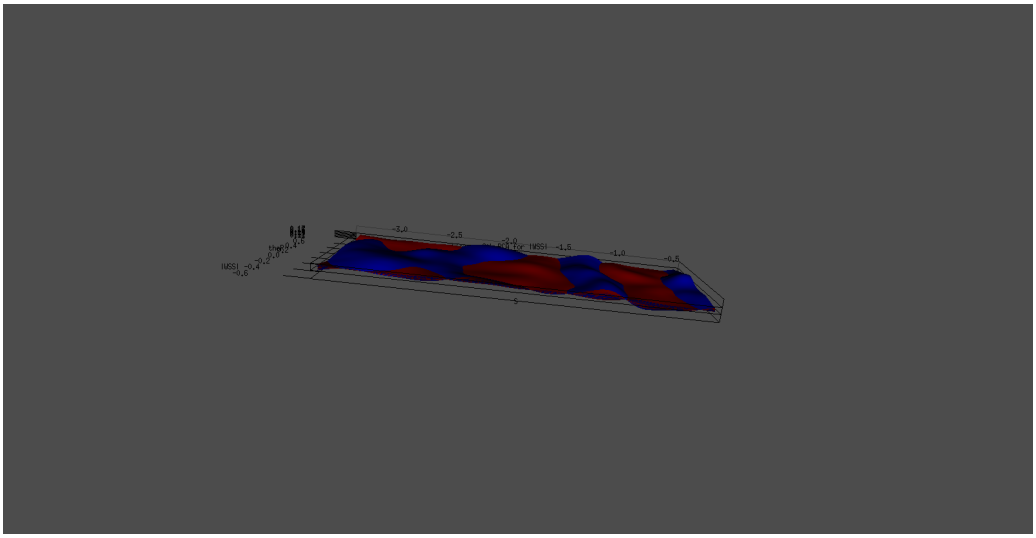


Figure 4.24: Projections on 6th PC - Upper/Lower groups

4.6 Interpretation of the results

Inspecting the plots of PC functions as perturbation of the mean, we remark that first PC function might provide an early important information, discriminating between patients with a high value of the average function $|\overline{\mathbf{WSS}}|$ from patients with a low one.

Analyzing second PC function, it is possible to remark that this function is almost constant on $\tilde{\Theta}$, varying on S only. Moreover, looking at the red surface, it is immediate to see that the 2^{nd} PC function is monotone on almost the whose reconstruction of ICA surface, increasing in the direction towards the zero, until a peak is reached, after which the function becomes decreasing.
The function has two peaks, a negative one about at $S = -2,75\ cm$, and a positive one at $S = -1,25\ cm$. It is interesting to note that 2^{nd} PC function has a symmetrical behavior with respect to the axis $S = -2,25\ cm$.
This PC function seems to suggest having a periodic sinusoidal behavior. Blood circulation is certainly a period phenomenon on time, and the frequency is given by the heart. So this corresponds to periodic phenomena in the points of ICA, such as the intensity of $|\mathbf{WSS}|$.
A possible interpretation of this PC function might be that this function corresponds to the spatial periodic variations in $|\mathbf{WSS}|$ with respect to the mean function.
These results are the same for registered data and not registered.

Analyzing third PC function we remark again a periodic behavior even if the location of the peaks is not the same as the one measured for second PC. A possible interpretation of this PC function might be similar to the one of the second PC, but on this $S-$periodic function has a different spatial frequency.
Similar interpretations might be attributed to fourth and fifth PC functions.

No evident interpretation can be provided to sixth PC function.
The analysis of the scores shows that for each eigenfunction there is no statistical evidence that scores are different with respect to groups defined by aneurysms' features. This is true for all features provided by the AneuRisk dataset.
If the macro-groups Upper/Lower are considered separately things do not change significantly.
The absence of any significative difference between groups of scores, considering different combinations of groups and models, investigated by the MANOVA global testing turns out to be surprising compared to literature and physicians' expectations.
Looking at PC functions also, this makes us suppose that *time* has an im-

portant role in the considered phenomenon.
Analysis of time-dependent Wall Shear Stress models should be performed as futures perspective.

These results are the same for registered data and not registered.
In conclusion, we may say the registration of data do not play any relevant role in the analysis of $|\mathbf{WSS}|$.

Chapter 5

Conclusions and future perspectives of AneuRisk program

This work had the aim of investigating the presence of possible relationships between the Wall Shear Stress values of the distal part of the Internal Carotid Artery obtained by CFD simulations and the origin of cerebral aneurysms in arterial district downstream of the ICA. This work is situated in the context of AneuRisk project, a research program set up in 2005 in order to look for effects of some factors in the onset of aneurysms. These factors, such as geometry and hemodynamics, were suspected by medical literature to have some active role in this pathology. This project was meant to provide information which could support physicians in their decisions about the therapy to adopt for the therapy of aneurysmal pathology.
The values of Wall Shear Stress have been simulated numerically from mathematical models of blood circulation in the cerebral district by Emory University.
In order to make results of this work easy to interpret, ICA has been modeled according to the *open carotid representation*, i.e. carotid is represented by a two dimensional surface, just as the artery were cut longitudinally and plied till it were completely plain. To parametrize the surface, we have chosen the curvilinear abscissa and an angular coordinate.

This is an observational study, so raw data have been explored to see how patients analyzed were distributed with respect to gender, age, position of the aneurysm and its eventual rupture, in order to verify the unbiasedness of the dataset in terms of these features.
The next step of the work has been focused on the estimates of n Wall Shear Stress continuous functions of the two coordinates of the ICA surface. Kernel Regression Method has been implemented after an empirical analysis

supporting the choice of bandwidth.
The need for data aligned to a natural system of coordinates, not defined arbitrarily by the angiography technology, has resulted in the registration process. This process has been conducted on the angular coordinate only.
After these preliminary phases, the central core of the present thesis has consisted in the Functional Principal Component Analysis, i.e. the challenge to succeed in describing an important amount of variability, by reducing as much as possible the dimension of the functional space where Wall Shear Stress functions are defined. This objective is reached by solving the problem of the search of the first greater eigenvalues of the sample autocovariance operator and the corresponding eigenfunctions. Principal Component functions (PC) have been analyzed in order to interpret the information they contain. According to the formulation of Functional Principal Component Analysis adopted in the work, they are related to the variability shown by patients that can not be explained by the mean function. After a brief analysis, we have chosen to limit our study to the first six Principal Component functions.
The first one seems to discriminate between patients having an high value of the average function $|\overline{\mathbf{WSS}}|$ from the patients with a low one. Other interesting information come from the second, third, fourth ad fifth PC, which show two important features: the first one is that, as suspected from the earliest exploratory analysis, all the within-patient variability is along curvilinear abscissa direction, suggesting that no information would be lost performing analysis of these data by means of simpler carotid models, averaging over angular coordinate and keeping the curvilinear coordinate only.
The second relevant aspect outlined is the behavior along curvilinear abscissa of these three functions, which seems to be sinusoidal and periodic. Blood circulation in the arteries is a periodic phenomenon both in time and space, in fact, Wall Shear Stress changes periodically on all the points of the arterial walls. So, these Principal Components may correspond to periodic variations with respect to an overall mean computed averaging for every point of the carotid on the time period (the first PC function). These plots suggest that each of these functions may correspond to a different spatial frequency.
Moreover, an exploratory analysis of the scores corresponding to first six PC functions has been performed with rapport to the groups Upper/Lower, Rupted/Not ruptured, Left/Right, Male/Female looking for relationships between the PC scores and medical information about the location of the aneurysms, but no statistical evidence was found.
Neither a MANOVA global test performed on the scores of first six PC functions reveals significative differences between groups.
In addition to the results coming from PC functions analysis, this makes us suppose a *time* effect in the studied phenomenon, which has not been captured by the present study, because all the analysis have been performed

at the systolic peak instant.

As already outlined, this work has to be positioned in the context of AneuRisk project, considering also all the previous studies which delineated the directions of research we have followed. So, in spite of some interesting results found and the implementation of techniques never applied before in this project, we consider this only an initial effort to investigate ICA hemodynamics in rapport to aneurysms onset.
This work, in turn, has outlined possible future perspectives for the future developments of the project.
In the immediate future, efforts will be related to the repetition of similar analysis on data registered also on the curvilinear coordinate. Potentially, simpler one dimensional models of ICA might be taken into consideration to analyze Wall Shear Stress, such as a model obtained by computing the mean on the angular coordinate. A simpler model would have many advantages both in terms of computational costs and application easiness. In fact, visualization of functions $\mathbb{R} \longrightarrow \mathbb{R}$ would be more immediate for physicians than functions $\mathbb{R}^2 \longrightarrow \mathbb{R}$.
A further research direction would be the analysis of spatial and temporal derivatives of the $|\mathbf{WSS}|$ function.
In addition, a Fourier analysis of Functional Principal Component Analysis would be interesting to test our suspects about their different periodicities. Moreover, a new model approach including also the time dependency in the Wall Shear Stress value should be evaluated.
An innovative approach might also schedule analysis of Wall Shear Stress models considering curvilinear coordinate and time only.
In a longer term perspective, it would be desirable to enlarge the dataset in order to make possible more robust results.

Bibliography

[1] Vantini, S. (2008): *"Functional Data Analysis of the Geometrical Features of the Internal Carotid Artery"*, Ph. D. course in Mathematical Engineering thesis, Milan.

[2] Passerini, T., (2009): *"Computational hemodynamics of the cerebral circulation: multiscale modeling from the circle of Willis to cerebral aneurysms"*, Ph. D. course in Mathematical Engineering thesis, Milan.

[3] Sangalli, L.M., Secchi, P., Vantini, S., Veneziani, A. (2009): "A Case Study in Exploratory Functional Data Analysis: Geometrical Features of the Internal Carotid Artery", *Journal of the American Statistical Association*, Vol. 104, No.485, 37-48.

[4] Friedman, J., Hastie, T., Tibshirani, R., (2009): *"The Elements of Statistical Learning"*, Springer series in statistics, New York.

[5] Ramsay, J.O., Silverman, B.W., (2005): *"Functional Data Analysis"*, Springer series in statistics, New York.

[6] Rinkel, G.J., Djibuti, M., Algra, A., and Van Gijn, J., (1998), "Prevalence and Risk of Rupture of Intracranial Aneurysms: A Systematic Review," *Stroke*, 29, 251-256.

[7] Kayembe, K., Sasahara, M., Hazam, F. (1984), "Cerebral Aneurysms and Variations in the Circle of Willis," *Stroke*, 15, 846–850.

[8] Hoi, Y., Meng, H., Woodward, S. H., Bendok, B. R., Hanel, R. A., Guterman, L.R., and Hopkins, L. N. (2004), "Effects of Arterial Geometry on Aneurysm Growth: Three-Dimensional Computational Fluid Dynamics Study," *Journal of Neurosurgery*, 101, 676–681.

[9] Hassan, T., Timofeev, E. V., Saito, T., Shimizu, H., Ezura, M., Matsumoto, Y., Takayama, K., Tominaga, T., and Takahashi, A. (2005), "A Proposed Parent Vessel Geometry-Based Categorization of Saccular Intracranial Aneurysms: Computational Flow Dynamics Analysis of the Risk Factors for Lesion Rupture," *Journal of Neurosurgery*, 103, 662–680.

[10] Castro, M. A., Putman, C. M., and Cebral, J. R. (2006), "Computational Fluid Dynamics Modeling of Intracranial Aneurysms: Effect of Parent Artery Segmentation on Intra-Aneurysmal Hemodynamics," *American Journal of Neuroradiology*, 27, 1703–1709.

[11] Berger, S. A., Talbot, L., and Yao, L. S. (1983), "Flow in Curved Pipes", *Annual Review of Fluid Mechanics*, 15, 461–512.

[12] Jitchote, W., and Robertson, A. M. (2000), "Flow of Second Order Fluids in Curved Pipes", *Journal of Non-Newtonian Fluid Mechanics*, 90, 91–116.

[13] Wilcoxon, F., (1945), "Individual comparisons by ranking methods", *Biometrics*, 1, 80-83.