POLITECNICO DI MILANO

School of Industrial and Information Engineering
Master of Science degree in Mathematical Engineering

Functional statistical analysis of fnirs data applied to cerebral hemodynamics

Master Thesis

Giovanni DI FABIO, 804369

Supervisor: Prof. Anna Maria Paganoni
Co-supervisor: Prof. Alessandro Torricelli

Accademic year 2015-2016

La statistica

Sai ched'è la statistica? È 'na cosa
che serve pe' fa' un conto in generale
de la gente che nasce, che sta male,
che more, che va in carcere e che sposa.

Ma pe' me la statistica curiosa
è dove c'entra la percentuale,
pe' via che, lì, la media è sempre eguale
puro co' la persona bisognosa.

Me spiego: da li conti che se fanno
seconno le statistiche d'adesso
risurta che te tocca un pollo all' anno:

e, se nun entra ne le spese tue,
t'entra ne la statistica lo stesso
perché c'è un antro che ne magna due.

Trilussa

Ai miei nonni.

Contents

List of Figures

List of Tables

Introduction

The present work concerns the analysis of data deriving from the application on the brain of Near Infrared Spectroscopy (NIRS).

Our study is based on the results obtained by Bonomini et. al (2015) who adopt a statistical method on NIRS data for the detection of neural activation.

This thesis differs from the previous work, because we use a functional statistical method to approach the same problem. In fact, we are interested in clustering NIRS functional data in activated and in not-activated areas of the brain and in confirming if these results are in according with the previous statistical study.

This work is composed by the following parts.

Chapter 1 is an introduction to Near Infrared Spectroscopy (NIRS) with a brief description of its techniques in Section 1.1. In Sections 1.2, 1.3, 1.4, 1.5 is exposed a review of the present literature. Section 1.6 describes briefly softwares for NIRS data analysis.

Chapter 2 provides a description of dataset: characteristics and gathering of data.

Chapter 3 describes methods to acquire functional data. We propose two different basis to represent the data: the Fourier basis in Section 3.2 and the Bspline basis in Section 3.3.

Chapter 4 reports procedure to obtain a functional K-means clustering (Section 4.1). Section 4.2 explain how to get the best number of clusters to fit the data. The results of K-means algorithm for the concentrations of O_2Hb and HHb are provided in Section 4.3, while those of tHb(total hemoglobin concentration) and StO_2 (tissue oxygen saturation) are in Section 4.4.

In Chapter 5 we analyze the K-means algorithm applied to the function median of the data. Sections 5.2 and 5.3 are used to introduce the concept of band depth and modified band depth which are essential to define the functional median, developed in Section 5.4. Results of K-means algorithm, using the functional median of the data, are reported in Section 5.5.

Chapter 6 is about covariance-based cluster algorithm. Section 6.2 proposes the definition of covariance operators for functional data. In Section 6.4 we discuss the max-swap algorithm to cluster data. Two different variations of the algorithm are proposed to estimate the covariance of data: the covariance operators of the groups (Section 6.3) and the Shrinkage estimator of covariance (Section 6.5). In Section 6.6 are shown the results of these analysis.

This work was conducted in collaboration with the Physics Department of the Politecnico of Milan.

Chapter 1

Introduction to NIRS

This thesis is a continuation of the article of Bonomini et al. (2015) using a novel statistical approach: *functional statistics*.

In this first chapter we want to summarize the history and the most useful informations about Near Infrared Spectroscopy. A large part of the following informations are obtained from Bonomini et al. (2015).

1.1 An optical method with several applications

Near Infrared Spectroscopy (NIRS) is a spectroscopic method that uses the near-infrared region of the electromagnetic spectrum (i.e radiations with wavelengths from 650 nm to 950 nm) to inspect the composition of materials.

The discovery of near-infrared energy is ascribed to William Herschel in 19 th century and the first industrial application began in the 1950s in agriculture, food science and animal feeds. At the beginning it was used in a few context because it was an expensive technology and there were difficulties in handling and evaluating results. In 1990s with the electronics and computer development NIRS became a more influential tool for scientific research. Nowadays, NIRS has many new applications in pharmaceutical, combustion products and cosmetics.

Recently, NIRS has began to be employed also in medical diagnostic and medicine research, with application in ergonomics, neonatal research and childhood brain development analysis, urology, neurology and functional neuroimaging.

The use of NIRS with the purpose of functional neuroimaging is best known as functional Near-Infrared Spectroscopy (fNIRS). The functional activation of human cerebral cortex can be explored by fNIRS and thanks to this the human functional brain mapping research has gained a new dimension. These studies are feasible because NIRS allows to monitor changes in oxygenated hemoglobin (O_2Hb) and deoxygenated hemoglobin (HHb) caused by neural activity.

fNIRS has become a complementary and alternative technique to functional Magnetic Resonance Imaging (fMRI), with the advantages of higher temporal resolution, excellent sensitivity to hemoglobin (fMRI can only monitor concentration changes in HHb), fewer motion artefacts and a cheaper and more portable technology.

The basic and main characteristics of NIRS are summarized as follows:

1. Human tissues are relatively transparent to light in the NIR spectral windows (650 : 900 nm);

2. NIR light is either absorbed by pigments compounds (chromophores) or scattered in tissues;

3. NIR light is able to penetrate human tissues, since the dominant factor in its tissue transport is scattering, which is typically about 100 times more probable than absorption;

4. The relatively high attenuation of NIR light in tissue is due to the main chromophore hemoglobin located in small vessels ($<$ 1 mm in diameter) of the microcirculation, such as capillary, arteriolar and venular beds. NIR is weakly sensitive to blood vessels $>$ 1 mm because they completely absorb the light. Given the fact that arterial blood volume fraction is approximately 30% in human brains, the NIR technique offers the possibility to obtain information mainly concerning oxygenation changes occurring within the venous compartment. The absorption spectrum of hemoglobin depends on its level of oxygenation. Another important aspect of NIR is that it is a non-invasive technique of analysis because it can be performed in natural environments without the need for restrain or sedation.

During the experiment, the patient wears a helmet with a variable number of sources and detectors. Photons emitted in the NIR range by sources penetrate in the head, being absorbed, transmitted or reflected depending on the optical proprieties of the medium they pass through. Reflected photons are then absorbed by detectors, placed some centimeters apart from the emission point. Highly sensitive detectors are required, since the intensity of the reflected light is very low.

Adequat depth of NIR light penetration (almost one half of the source-detector distance) can be achieved using a source-detector distance around 3 cm. The selection of the optimal source-detector distance depends on NIR light intensity and wavelength, as well as the age of the patient and the head region measured.

While the spatial resolution is moderate and it degrades rapidly with increasing depth in the brain, the temporal resolution of hemoglobin detection with NIRS is not acquisition limited and can be up to milliseconds, much faster then the hemodynamic response itself.

NIRS permits a quantitative monitoring of important physiological parameters as oxygenated hemoglobin (O_2Hb) and deoxygenated hemoglobin (HHb). From these two quantities is possible to evaluate another two important quantities:

$$\begin{cases} tHb = O_2Hb + HHb \\ StO_2 = 100 \cdot \frac{O_2Hb}{O_2Hb+HHb} \end{cases}$$

where tHb is the total hemoglobin concentration and StO_2 is the tissue oxygen saturation.

Three different NIRS techniques are used, each based on a specific type of light.

The first one that is also the simplest and the cheapest is "continuous wave" (CW) fNIRS. This NIRS system uses light sources which emit light at a constant frequency and amplitude, continuously in time. Absorption changes are determined by measuring the attenuation of incident light. The main drawback of this technology is that it does not have any knowledge of photon path-length and so changes in concentration are relative to an unknown path-length.

A more complex and more precise system is "time domain" (TD) fNIRS. In this type of spectroscopy photons are emitted through pulses with a temporal width of some tens of picoseconds. The photon paths can be obtained evaluating delay between emission and detection, thus it is possible to reconstruct the depth reached by photons and to determine the absolute absorption coefficients of the crossed media. In this way the absolute concentration values of O_2Hb and HHb can be retrieved.

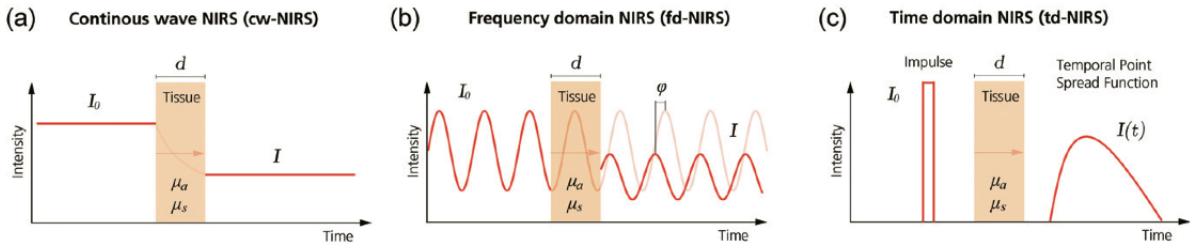

Figure 1.1: The emission/detection of NIR waves for the three types of NIRS systems.

This method is more expansive and technically complex than CW fNIRS, because it requires high-speed source and detectors. Nevertheless it provides the most information, as well as a good spatial resolution . It also enables the correction of the noise of the superficial layers.

The datasets we use are obtained with this technique.

The last NIRS system is "frequency domain" (FD) . In frequency domain systems, NIR laser sources provide an amplitude modulated sinusoid at frequencies near one hundred megahertz (100 MHz). Changes in the back-scattered signal's amplitude and phase provide information on changes in the concentration of hemoglobin and provide direct measurement of absorption coefficients obviating the need for information about photon path-length. Because of the need for modulated lasers as well as phasic measurements, frequency domain systems are more technically complex than continuous wave systems. However, these systems are capable of providing absolute concentrations of O_2Hb and HHb.

The behavior of the three NIRS systems in time is shown in Figure 1.1.

1.2 The State of the Art

As explained before, using NIRS to measure hemoglobin concentration has a lot of advantages, anyway, NIRS techniques presents also some disadvantages as a modest spatial resolution (1-3 cm), a limited penetration depth and the presence of noises.

fNIRS data are usually corrupted by three different noise: physiological noise (such as interferences provoked by heartbeat, breathing, blood pressure), instrument noise (such as shot noise, with a Poisson distribution, and measurement noise, which is assumed to be a Gaussian white noise) and experiment errors (motion artefacts, potential sensitivity to hair absorption, interferences from external sources). Thus it is advantageous to pre-process data, in order to remove some interferences and to improve the spatial sensitivity of registered measurements. After subtracting noise, data are usually analyzed to detect active areas of the brain.

The following subsections are a brief summary about preprocessing data and active channel detection.

1.3 Systemic interference correction

Are called *systemic physiological interference* the noises caused by heart beat, respiration, blood pressure variation and other slow variation. *Systemic physiological interference* can arise from at

two spatial sources: in the superficial layers (such as skull and scalp) and inside brain, due to factors such as heart activity, respiration and spontaneous low frequency oscillations.

In empirical studies of brain function using NIRS, the amount of global interference varies from subject to subject and from time to time. In some cases, the amount of interference is small, other times the amount of interference is too large for the evoked brain activity to be detected without signal processing.

The contribution of this interference in NIRS signal is amplified because the light is both introduced and collected at the surface of the scalp. This back-reflection geometry makes NIRS very sensitive to the superficial layers of the head which contain no brain signal but exhibit strong systemic fluctuations. For this reason systemic interference from superficial layers often results as the dominating noise component.

Several methods have been explored for the removal of global interference and improvement of evoked brain activity measurements.

One of the most common technique is low-pass filtering (e.g. Franceshini and Boas (2004)). This approach suits perfectly the suppression of high frequency interferences, such as heart beat, but it does not remove noise from respiration and other low frequency phenomena. If the noise has the same frequency of the hemodynamic response, a frequency based filtering is also inappropriate, because it may misrepresent the information on cerebral activity.

Other methods for reduction of systemic physiological interference include wavelet filtering (e.g. Lina et al. (2010)) and Principal Component Analysis (e.g. Franceschini et al. (2006)).

Another way to isolate low frequency noise in superficial layer is "multi-distance" approach. This approach is based on the photon transport theory. Photons propagating through a highly scattering tissue travel along a zig-zig path before they are detected. The collective photon propagation follows a roughly banana-shaped patter when reflection geometry is used.

With an appropriate source and detector placement, we can use the signal from close source-detector (S-D1 in Figure 1.2) to detect superficial hemodynamic changes and signal from the most distant source-detector (S-D2 in Figure 1.2) to detect hemodynamical changes in both superficial and deep layer. In this way the superficial layer signals acquired from S-D1 are used to estimate systemic interference in S-D2. This approach is applied in Gagnon et al. (2012) and Saager and Berger (2005).

Several algorithms have been developed to perform the regression of the small separation measurements. These include linear minimum mean square estimation (LMMSE) (Saager and Berger (2005), (2008); Gregg et al. (2010); Saager et al. (2011)), adaptive filtering (Zhang et al. (2007)) and state-space modeling with Kalman filter estimation (Gagnon et al. (2011)).

The most critical point in multi-distance approach is the difficulty in determining photons path length in each head layer.A partial pathlength of each layer can be predicted computationally by Monte Carlo simulation (Unemaya and Yamada (2009)).

As said in the previous section, we use TD fNIRS also because data obtained with this technology have a systemic noise much lower than CW fNIRS. With TD fNIRS is possible to estimate the superficial hemodynamics and to distinguish between photons from deep and shallow layers.

1.4 Motion artifacts correction

Due to its many attractive qualities, fNIR is an ideal candidate for monitoring cortical function in the brain while subjects are engaged in various real life or experimental tasks. However, another

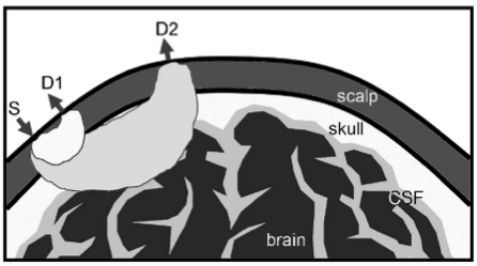

Figure 1.2: Multi-distance source-detector approach. Superficial layer interference acquired from S-D1 are used to estimate systemic interference in S-D2 (Zhang et al. (2007)).

important limitation of optical data in these applications is the artifact caused by motion of the head, better called *motion artifacts* (MA). Head movement can cause the NIR detectors to shift and lose contact with the skin, exposing them to either ambient light or to light emitted directly from the NIR sources or reflected from the skin, rather than being reflected from tissue in regions of interest. These effects cause sudden increases in the NIR data. Another consequence of head movement is that can cause the blood to move toward (or away from) the area that is being monitored, increasing (or decreasing) the amount of oxygen, hence result in an increase (or decrease) in the measured data. Since the dynamics of this type of motion artifact are slow, they can easily be confused with the actual hemodynamic response due to brain activation. The motion artifacts in NIR studies are a serious problem for real life applications where head immobility is undesirable or untenable. Hence, cleaning the NIR data from motion artifacts is an important and necessary task in order to deploy NIRS as a brain monitoring technology in its full potential to many real life application areas.

Several different methods were suggested to solve this problem. These methods can be categorized as follows: (i) methods based on the temporal characteristics of the signal, (ii) methods based on the spatial characteristics of the signal and (iii) measuring noise independently and subtracting it from the signal.

The first category includes methods such as: Wiener filtering (Izzetoglu et al. (2005)), wavelet minimum description length (Wavelet-MDL) detrending (Jang et al. (2009)) and a correlation-based signal improvement (CBSI) method (Cui et al. (2010)). A Wiener filter does not need an external input signal because the filter coefficients are only estimated based on the statistics of the measured signal. This requires two calibration measurements before the actual measurement: one measurement without MA and other with MA. The disadvantage is that the only types of MA eliminated are those present in the calibration measurement. Additionally, in animals or neonates, it is difficult to avoid MA during the calibration measurement.

Algorithms in the second category, for example, eigenvector based algorithms, usually assume that signal due to noise is broadly spatially distributed compared to signal due to neural activity. However, with eigenvector based methods one often needs to make subjective decisions about the number of components to keep, and they are difficult to apply to real-time processing.

In the third category, noise is measured independently by additional hardware. For example, head motion can be measured using an accelerometer, or noise can be measured using a channel with a very short distance between the emitter and detector so that the infrared light does not pass through the brain tissue. The resultant signal must be noise, which can be subtracted from the signal.

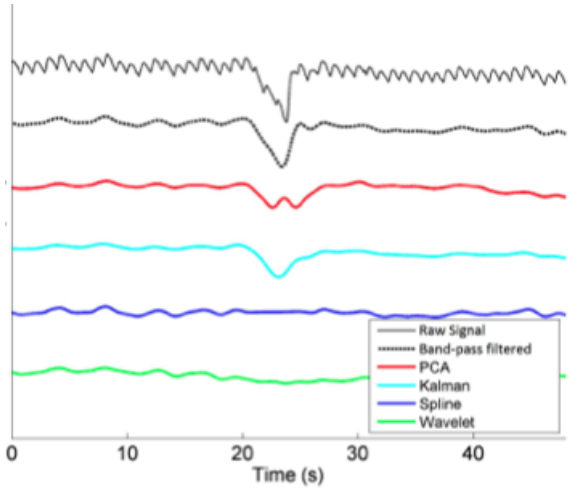

Figure 1.3: Motion Artefact corrected trough different techniques (Cooper et al. (2012)).

Figure 1.3 shows how motion artifact can be corrected trough different techniques.

1.5 Statistical Inference for brain activation detection

In the early ages of fNIRS studies, brain signal detection was usually attempted through visual inspection (Murata et al. (2002)) or simple thresholding with some preprocessing steps. For instance, Benaron et al. (2000) control if the hemoglobine concentration increase: if signal exceed 2 standard deviation from mean, the activation was assumed. However, such heuristic approaches are prone to error, especially when the noise and interference levels increase, so more rigorous statistical analyses were required. Hence, various statistical analysis methods have been proposed to detect activation.

Hoshi et al. (2003) proposed t-test for the difference between the mean of two groups: the hemoglobin concentration during task period and the hemoglobin concentration during rest period. In these analysis one hypothesis test was implemented for each channel and P-values were adjusted with Bonferroni correction.

Also one-way or multi-way ANOVA analysis was proposed to compare different groups concentrations. This approach take average values during the task period as data to avoid any assumption of the exact shape or timing of the time course of changes in O_2Hb and HHb in response to stimuli. This is the limit of this type of analysis because it does not use the time course of data, which is quite important in fNIRS data.

Many researchers are interested in understanding the fNIRS time course. Then the Generalized Linear Model (GLM), that was used in fMRI analysis because of the similarities of the two technologies, was proposed for fNIRS. GLM assumes that data can be represented as a linear combination of some predictors plus an error term. Schroeter et al. (2004) were the first to apply the GLM to analyze fNIRS data, then numerous authors have employed GLM to study fNIRS data.

Bonomini et al. (2015) use as predictors convolutions between a boxcar function and the hemodynamic response function (HRF). These convolutions represent a sort of temporal evolution of

the "ideal" hemoglobin concentration in an active channel. These convolutions represent a sort of temporal evolution of the "ideal" hemoglobin concentration in an active channel. In fact neuronal activity consumes oxygen carried by hemoglobin, provoking an increase of the blood flux. This blood flux variation is called hemodynamic response and it reaches a peak after 4-5 s, then it decreases to the initial level. Figure 1.4 shows the hemodynamic response function, that describes this phenomenon. The authors through a linear regression model and a K-means method estimated the hemodynamic activation in fNIRS data sets.

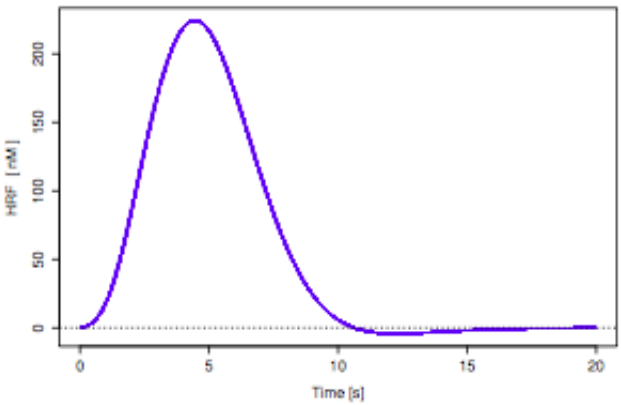

Figure 1.4: The hemodynamic response function.

Pourshoghi et al. (2016) proposed the application of functional data analysis (FDA) (for more details, see chapter 3) on fNIRS signals for the development a method of pain perception. The FDA was applied on the collected fNIRS data to convert discrete samples into continuous curves. This method allows to represent the curves as a linear combination of basis functions. Pourshoghiet al. (2016) utilized bases coefficients as features that represent the shape of the signals and used them to train a Support Vector Machine (SVM) to classify, using hierarchical algorithm, the signals based on the level of induced pain.

In our analysis we also use FDA applied to fNIRS but with a different application. Our aim is to classify active channels using a functional k means evaluating the distance between functional data. In this work we utilized L^2 and H^1 distance (for more details see Section 4.1).

1.6 Softwares to process functional optical data

Over past decades, with the development of NIRS also methods to process functional optical data were needed.

The first public domain software package for fNIRS data analysis was HomER[1] (acronym for Hemodynamic Evoked Response). The software provides a graphical user interface and MATLAB[2] scripts for both the preprocessing and the standard statistics on fNIRS data. HomER has been

[1]http://www.nmr.mgh.harvard.edu/PMI/, distributed by Massachusetts General Hospital.
[2]MATLAB and Statistics Toolbox Release 2015, The MathWorks, Inc., Natick, Massachusetts, USA.

upgraded and the new release HomER2 more easily supports group analyses and re-configuration of the processing stream, and it integrates users algorithms into the processing stream.

Another free software is Functional Optical Signal Analysis (fOSA[3]), which offers MATLAB based functions for a basic analysis of fNIRS data, incorporating several filters for signal denoising and providing also the Statistical Parametric Mapping (SPM) methodology for statistical analysis based on the GLM approach.

More focused on the development of SPM routines is the non-commercial MATLAB-based software NIRS-SPM[4]. A novelty introduced by this program is represented by a voxel based alignment between interpolated maps instead of an inter-subjects realignment of optodes, in order to facilitate the group analysis.

Another software is NIRS analysis package (NAP[5]) which allows noise removal and GLM analysis, as well as anatomical registration of the measurements. fNIRSOFT [6] is a stand-alone software to process, analyse and visualize fNIRS signals through a graphical user interface and/or scripting distributed by BIOPAC Systems, Inc.

Finally POTATo [7] (Platform for Optical Topography Analysis Tools) is a software package for fNIRS signal processing and analysis, developed by Hitachi, Ltd.

To perform our functional statistical analysis on fNIRS we use R[8] software.

[3]Koh P. H., Glaser D. E., Flandin G., Butterworth B., Maki A., Delpy D., Elwell C. E. (2007) Functional Optical Signal Analysis (fOSA): A Software Tool for NIRS Data Processing Incorporating Statistical Parametric Mapping (SPM). Journal of Biomedical Optics, vol. 12, 6: 064010.

[4]http://bisp.kaist.ac.kr/NIRS-SPM.html

[5]T. Fekete, D. Rubin, J. M. Carlson, and L. R. Mujica-Parodi, "The NIRS Analysis Package: noise reduction and statistical inference," PLoS One, vol. 6, p. e24322, 2011.

[6]http://www.biopac.com/fNIR-Software-Professional-Edition

[7]http://www.hitachi.co.jp/products/ot/analyze/kaiseki_en.html

[8]R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Chapter 2

The datasets

We will use two types of data: *in vivo* data and *synthetic* data.

In vivo data are collected through a real fNIRS experiment on volunteers, while *Synthetic* data gathering derive from simulations.

Both dataset register the concentration in superficial and lower layer of the brain of O_2Hb and HHb in 30 fixed positions, called *channels*, in a healthy adult during a motor task. For the sake of simplicity, we call as UP and DOWN the data registered respectively from the superficial and lower layer of the brain.

The hemoglobin measurements are recorded with a TD fNIRS system. The disposition of the channels on the brain is reported in the Figure below.

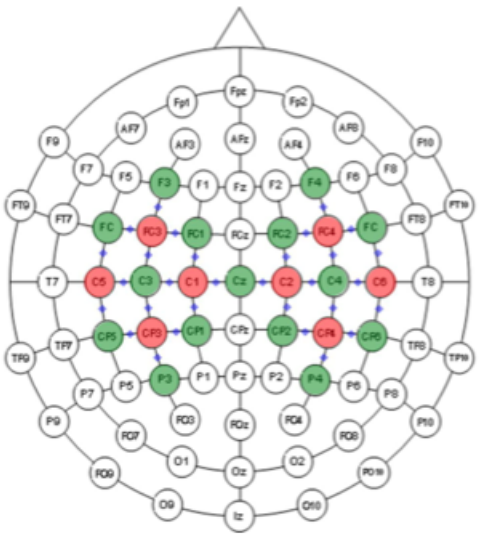

Figure 2.1: Disposition of the channels on the brain

Hemoglobin concentrations of in vivo and virtual subjects are registered during a *right hand grip* experiment. During the experiment subjects alternate periods of rest and task and, after an initial resting period of 50s, 10 trials are performed. Each trial has the following structure: 10s of rest; 20s of right hand gripping (task); 10s of rest. At the end of the last trial subjects have a resting

period of 40 s.

The experiment lasts 490 s but for our study we have considered only the 10 trials period (400s).

The fNIRS system registers the concentration of O_2Hb and HHb on each second. The complete dataset is composed by 30 vectors of length 400s for O_2Hb and HHb measures.

Figure 2.2: Structure of the experiment

2.1 In vivo data

We have O_2Hb and HHb measures from 12 right-handed healthy volunteers. Data were acquired by a multi-channel dual-wavelength TD fNIRS medical device developed by the Physics Department of the Politecnico di Milano (Contini et al. (2009)).

Subjects are 7 males and 5 females, of age 32.2±10 years.

During task periods they squeezed a soft ball in the right hand at a rate of 2 Hz, guided by a metronome. Instructions about the movement and rest periods were given by presenting a picture on a screen, which always had a fixation cross in the center.

Hemodynamic parameters were estimated by the Physics Department of the Politecnico di Milano trough the following steps:

1. calculating the baseline optical proprieties and the absorption changes in the upper and lower layer;

2. gathering the hemoglobin concentrations from the absorption coefficients.

About the physic experiment it is interesting to understand how to estimate the photon path length in the superficial layers and deep layers.

The simplest way to estimate the photon path length, usually, consists in using the formula $l = vt$, where v is the speed of light in tissue and t is a characteristics time, set to $500 \cdot 10^{-12}$ s for "early" photons (the ones that pass through superficial layers), set to $1500 \cdot 10^{-12}$ s for photons from deep layers (they are called "late"ß photons). The risk of this method consists in overestimating the path in deep regions and underestimating hemoglobin concentration.

In our work we do not use this method to estimate the photon path length because the Physics Department of Politecnico di Milano developed a more precise one (Zucchelli et al. (2013)).

In this work the photon propagation model is refined trough a more precise computation of the path length travelled by photons within each layer the tissue is composed of. The non-idealities of the fNIRS system set-up and the heterogeneous structure of the human head are also considered. This method leads to a reduction of photon path in deep regions, and to an increase in estimated hemoglobin concentration. In this way we find higher hemoglobin amplitudes and a more precise hemoglobin estimation. The signals are more corrupted by noise because this method is more sensitive to few photons.

All the statistical analysis will be conducted separately and independently on each subject.

Subjects are identified trough a number. They are 0173, 0176, 0179, 0182, 0185, 0188, 0191, 0197, 0200, 0203, 0206, 0209.

Some subjects present several channels with very high noise. These channels could create some problems in activation detection.

During the 40s trials, activated channels should have an increase in O_2Hb and a decrease in HHb .

2.2 Synthetic data

Simulated data mimic real multichannel fNIRS measurements on a healthy adult during the *right hand grip* experiment before described. The procedure to create synthetic data is developed by Physics Department of Politecnico di Milano. It consits in the following steps.

Defining the geometry of the head

The head is modelled as a bilayered medium, where the upper layer is 1 cm thick, and the lower layer is ideally a semi-infinite medium. To a first approximation in fact this geometry can be used to simulate fNIRS measurements on the head of an adult, where an extra-cerebral layer (composed by scalp, skull and cerebrospinal fluid) overlays the intra-cerebral one (gray and white matter).

Specification of the hemodynamic parameters and the HRF

Hemoglobin concentrations are calculated for both layers of the head. They are simulated by considering reference values of 12 μM for the O_2Hb and 7 μM for the HHb in the superficial layer, and reference values of 30 μM for the O_2Hb and 20 μM for the HHb in the lower layer.

The experiment considers an ideal situation where a neuronal activation is generated by a motor task, and no physiological oscillations occur in the superficial layer. The O_2Hb and HHb concentrations in the upper layer are simulated to be constant at reference values during the whole experiment, while the concentrations in the lower layer are perturbed in some channels so as to mimic a hemodynamic response in correspondence to the task periods.

This superimposed response profile is calculated as a convolution of a boxcar function,representing the task and the rest alternation, with the Hemodynamic Response Function (HRF) evoked by a single stimulus. By following the method proposed by Scarpa et al. (2013) the HRF is modelled as a linear combination of two different gamma-variant time-dependent functions Γ_n:

$$HRF(t) = \alpha \cdot (\Gamma_n(t, \tau_1, \rho_1) + \beta \cdot \Gamma_n(t, \tau_2, \rho_2))$$

with

$$\Gamma_n(t, \tau_j, \rho_j) = \frac{1}{\rho! \tau_j} \Big(\frac{t - \rho_j}{\tau_j} \Big)^p e^{-\frac{t-\rho_j}{\tau_j}} \delta(t - \rho_j)$$

$$\delta(t - \rho_j) = \begin{cases} 1 & t - \rho_j \geq 0 \\ 0 & otherwise \end{cases}$$

where *alpha* tunes the amplitude, τ_j and ρ_j tune the response width and the onset time respectively. Variability in amplitude of about 5% is considered among the different trials to account for possible differences in the execution of the task and/ or in the functional response.

The peak of the HRF for the O_2Hb is chosen to be around 1.555 $\pm$ 0.075 μM; the HRF for the HHb is inverted and with a maximum set at -1/3 with respect to the O_2Hb response. The

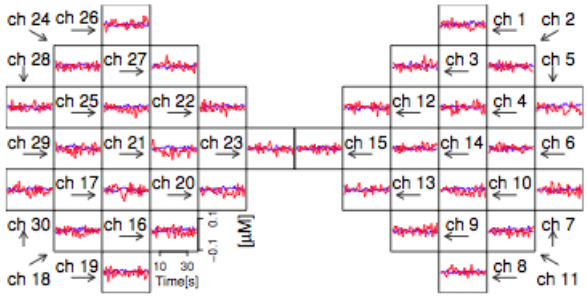

Figure 2.3: Folded data for the superficial layer.

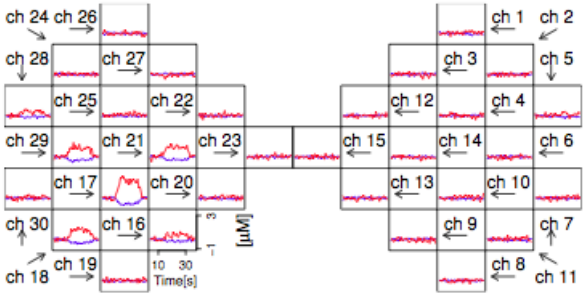

Figure 2.4: Folded data for the deep layer.

free parameters are chosen so as to create a HRF similar to the one expected for the motor task of interest ($\alpha = 1282, \beta = 0.17, \tau_1 = 1, \tau_2 = 1, \rho_1 = -0.5, \rho_2 = 3.5$).

To simulate an actual neuronal activation localized around the central positions of the hemisphere contralateral to the movement, channels considered as activated are: channel 16 (with an intensity of 25% HRF), 17 (100% HRF), 18 (50% HRF), 21 (50% HRF), 28 (25% HRF), 29 (50% HRF).

Obtaining the absorption coefficients

The absorption coefficients at two wavelength (690 and 820 nm) for both layers are computed from these hemoglobin concentration changes by exploiting the Lambert Beer law and the a priori knowledge of the specific absorption of O_2Hb and HHb.

Generate distributions of photon time-of-flight

A forward model for photon diffusion in a bilayered geometry is used to generate synthetic time-resolved reflectance curves for each channel by using as input parameters the optical proprieties and the source detector distance (fixed at 3 cm). A count rate of $5 \cdot 10^5$ ph/s is considered, the integration time is at 1 s, and Poisson noise is added to the simulated curves to mimic real measurements.

After this procedure, as for in vivo data, hemodynamic parameters are obtained estimating the baseline proprieties and the absorption changes in the upper and deep layer and calculating the hemodynamic parameters from the absorption coefficients, following the method proposed by Zucchelli et al. (2013). The resulting hemoglobin concentrations for synthetic data are reported in Figures 2.3 and 2.4

Chapter 3

Functional data

In order to analyze these data, we decided to adopt a functional approach.

The basic philosophy of functional data analysis is to think about observed data functions as single entities, rather than as a sequence of individual observations. The term *functional* in reference to observed data refers to the intrinsic structure of the data rather than to their explicit form.

Functional data are usually observed and recorded discretely as n pairs $(t_j; y_j)$ where y_j is a snapshot of the function at time t_j. Time is the continuum over which functional data are recorded, we refer to t_j as such. We call x the functional form of the observations and we assume the existence of a function x giving rise to the observed data.We want to declare that the underlying function x is *smooth*, so that a pair of adjacent data values y_j and y_{j+1} are necessarily linked together to some extent and unlikely to be too different from each other.

By smooth, we usually mean that function x possesses one or more derivatives. We will usually want to use the discrete data y_j, $j = 1, ..., n$ to estimate the function x and at the same time a certain number of its derivatives.

The actual observed data, however, may not be at all smooth due to the presence of measurement errors.

Smoothness, in the sense of possessing a certain number of derivates, is a propriety of the function x, and may not be at all obvious in the raw data vector $y = (y_1, ..., y_n)$ owning the presence of the measurement error. We express this in notation as

$$y_j = x(t_j) + \epsilon_j$$

where the noise ϵ_j contributes a roughness to the raw data. One of the tasks in representing the raw data as functions may be to attempt to filter out this noise as efficiently as possible.

3.1 Representing functions by basis functions

To represent functions by basis function we choose the $H^1(\mathbb{R})$ functional space defined as:

$$H^1(\mathbb{R}) = \{f : \mathbb{R} \longrightarrow \mathbb{R} \mid \quad f \in L^2(\mathbb{R}), f' \in L^2(\mathbb{R})\}$$

A basis function system is a set of known functions ϕ_k that are mathematically independent of each other and that have the property that we can approximate arbitrarily well any function by taking a weighted sum or linear combination of a sufficiently large number K of these functions.

Basis function procedures represent a function x by a linear expansion

$$x(t) = \sum_{k=1}^{K} c_k \phi_k(t)$$

in terms of K known basis function ϕ_k.

We can express the previous formula in matrix notation as

$$x(t) = \mathbf{c}'\boldsymbol{\phi} = \boldsymbol{\phi}'\mathbf{c}.$$

The *dimension* of the expansion is therefore K.

An exact *interpolation* is achieved when $K = n$ in the sense that we can choose the coefficients c_k to yield $x(t_j) = y_j \ \forall j$. Therefore the degree to which the data y_j are *smoothed* is determined by the number K of basis functions. Consequently, we do not view a basis system as defined by a fixed number K of parameters, but rather we see K itself as a parameter that we choose according to the characteristics of the data.

The choice of basis is particularly important also for a derivative estimate

$$D\widehat{x}(t) = \sum_{k=1}^{K} \widehat{c}_k D\phi_k(t) = \widehat{\mathbf{c}}' D\boldsymbol{\phi}.$$

The notation $D\widehat{x}(t)$ is used to represent the derivative of the function $\widehat{x}$ and $D\phi_k(t)$ is used to represent the derivative of ϕ_k. This notation produces cleaner formulas than dx/dt.

Bases that work well for functions estimation may give poor derivative estimates.

In our work we used *B-spline basis* and *Fourier basis*.

3.2 Fourier basis

Fourier basis expansion is provided by the Fourier series:

$$\widehat{x}(t) = c_0 + c_1 sin(\omega t) + c_2 cos(\omega t) + c_3 sin(2\omega t) + c_4 cos(2\omega t) + ...$$

defined by the basis $\phi_0(t) = 1$, $\phi_{2r-1}(t) = sin(r\omega t)$ and $\phi_{2r}(t) = cos(r\omega t)$. This basis is periodic and the parameter ω determines the period $2\pi/\omega$.

Called n the number of values of $x(t)$ that we need, the Fast Fourier Transform (FTT) makes it possible to find all the coefficients extremely efficiently when n is a power of 2 and the arguments are equally spaced, in this case we can find both the coefficients c_k and all n smooth values at $x(t_j)$ in $\mathcal{O}$(n logn) operations.

Derivate estimation in a Fourier basis is simple since

$$Dsin(r\omega t) = r\omega cos(r\omega t)$$

$$Dcos(r\omega t) = -r\omega sin(r\omega t)$$

This implies that the Fourier expansion of Dx has coefficients

$$(0, c_1, -\omega c_2, 2\omega c_3, -2\omega c_4, ...)$$

and the D^2x has coefficients

$$(0, -\omega^2 c_1, -\omega^2 c_2, -4\omega^2 c_3, -4\omega^2 c_4, ...)$$

A Fourier series is especially useful for describing the periodic data. Fourier series generally yield expansions which are uniformly smooth. But they are inappropriate for data known or suspected to reflect discontinuities in the function itself or in low order derivatives.

We used Fourier basis thinking that our data is periodic. We did not set a period because R[1] is able to choose the best period that fits the data. We chose the number of basis for each channel through the generalized cross-validation (GCV) criterion. We run a R code varying n in a range of values from 6 to 30 and set the number of basis equal to value that minimizes the GCV.

In the Figure 3.1 are shown the data functions obtained through the Fourier basis for O_2Hb Down measurements, in the plot each function represents a different channel. Plots for HHb Down data, O_2Hb Up Data and HHb Up data are given respectively in Figures 3.3, 3.5 and 3.7.

Figures 3.2, 3.4, 3.6 and 3.8 are the first derivatives of the previous functional data.

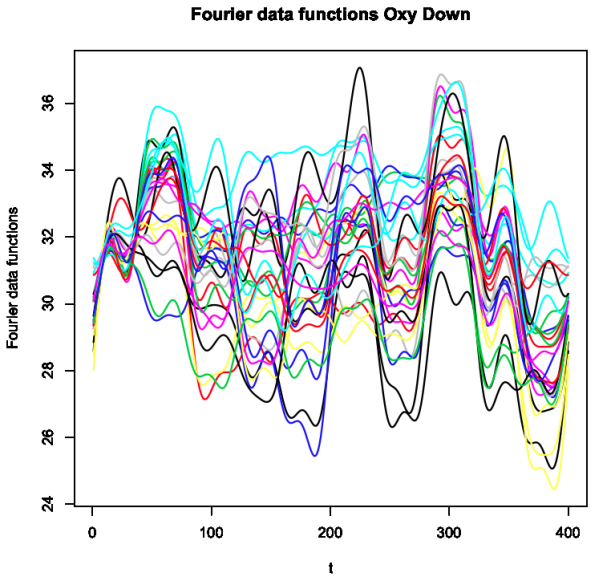

Figure 3.1: Fourier data functions for O_2Hb Down measurements

3.3 B-spline basis

Spline functions are the most common choice of approximation system for non-periodic functional data or parameters.

Splines are polynomial segments joined end to end. Splines combine the fast computation of polynomials with substantially greater flexibility. Moreover, basis system have been developed for spline functions that require an amount of computation that is proportional to n ($\mathcal{O}$(n)).

[1]R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

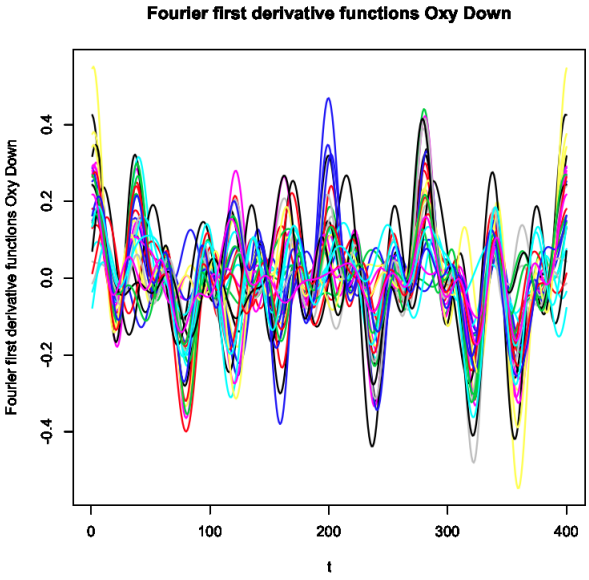

Figure 3.2: Fourier data first derivative functions for O_2Hb Down measurements

The first step in defining a spline is to divide the interval over which a function is to be approximated into L subintervals separated by values τ_l, $l = 1, ..., L-1$ that are called *breakpoints* or *knots*.

Over each interval, a spline is a polynomial of specified order m. The *order* of a polynomial is the number of constants required to define it, and is one more than its *degree*.

Adjacent polynomials join smoothly at the breakpoint, so the function values are constrained to be equal at their junction. Moreover, derivatives up to order $m-2$ must also match up at these junctions.

In our work we used equally spaced knots and cubic splines in order to have two continuous derivatives.

Figure 3.9 on page 20 is an example of Bspline basis for the $30th$ channel, the spline is cubic with 24 interior knots, shown as vertical dashed lines.

As did for Fourier basis, we chose the number of basis through the GCV criterion. The number of knots depends on n and is equal to:

$$nknots = nbasis - norder + 2$$

In the Figure 3.10 on page 20 are shown the data functions obtained through the Bspline basis for O_2Hb Down measurements, in the plot each function represents a different channel; in the same way we drew the plots 3.12 for HHb Down data, 3.14 for O_2Hb Up Data and 3.16 for HHb Up data.

Figures 3.11, 3.13, 3.15 and 3.17 are the first derivatives of the previous functional data.

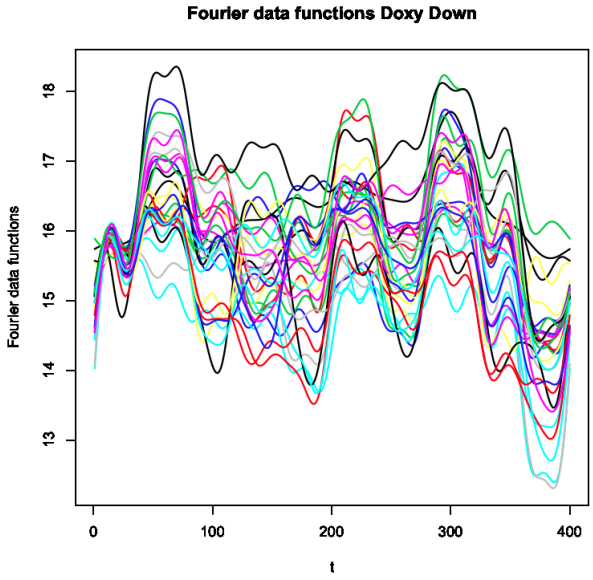

Figure 3.3: Fourier data functions for HHb Down measurements

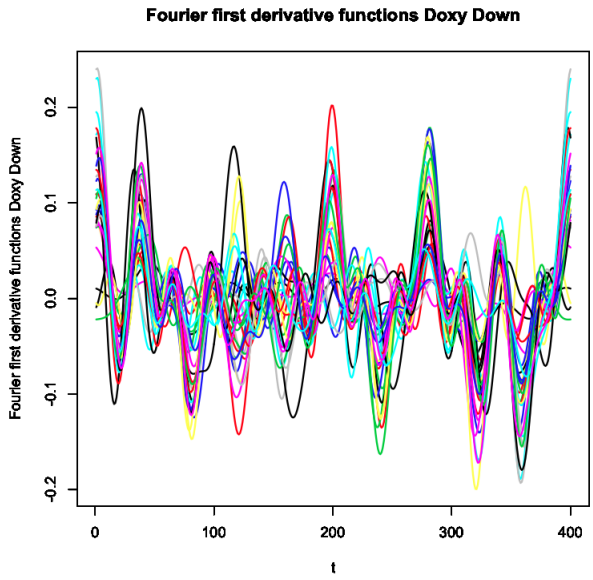

Figure 3.4: Fourier data first derivative functions for HHb Down measurements

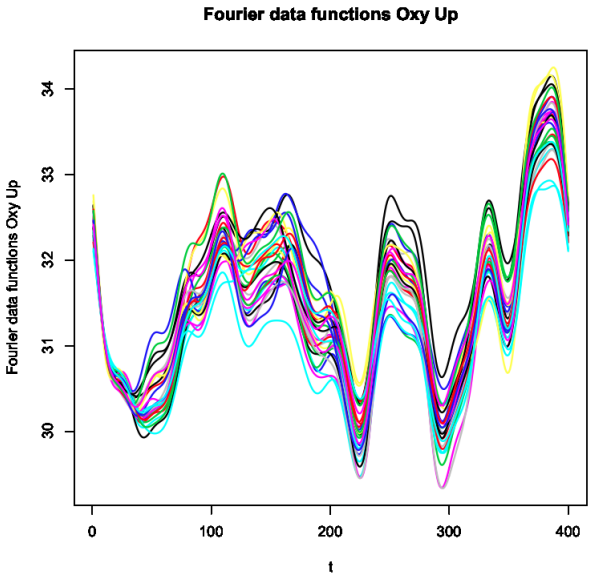

Figure 3.5: Fourier data functions for O_2Hb Up measurements

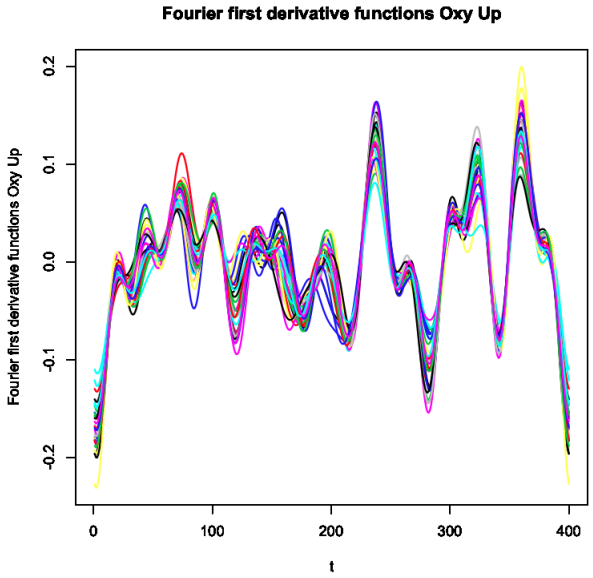

Figure 3.6: Fourier data first derivative functions for O_2Hb Up measurements

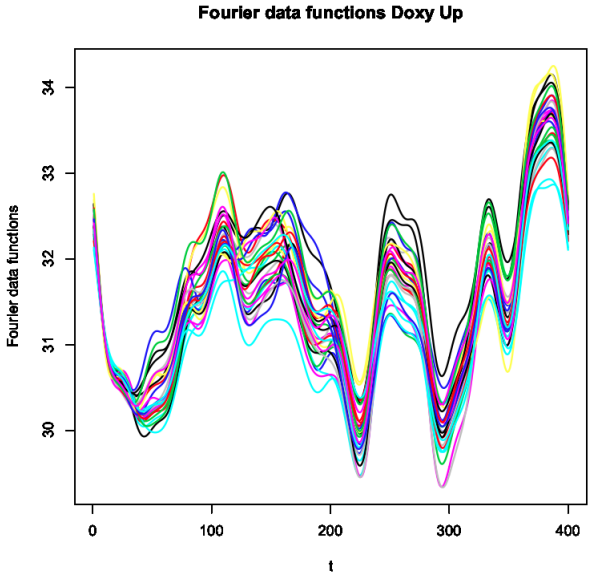

Figure 3.7: Fourier data functions for HHb Up measurements

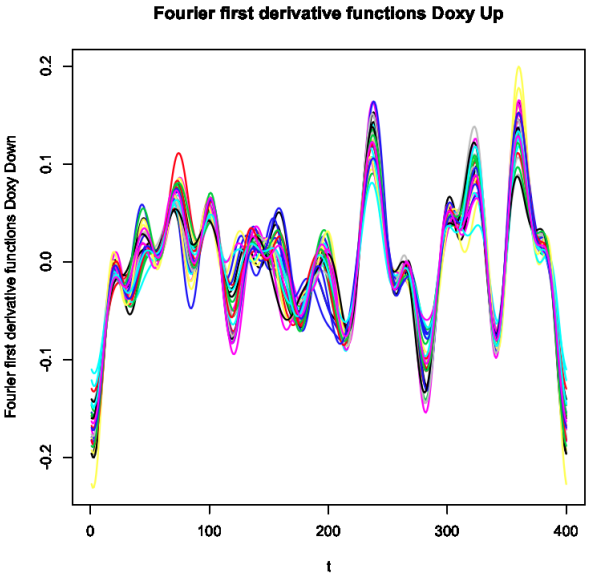

Figure 3.8: Fourier data first derivative functions for HHb Up measurements

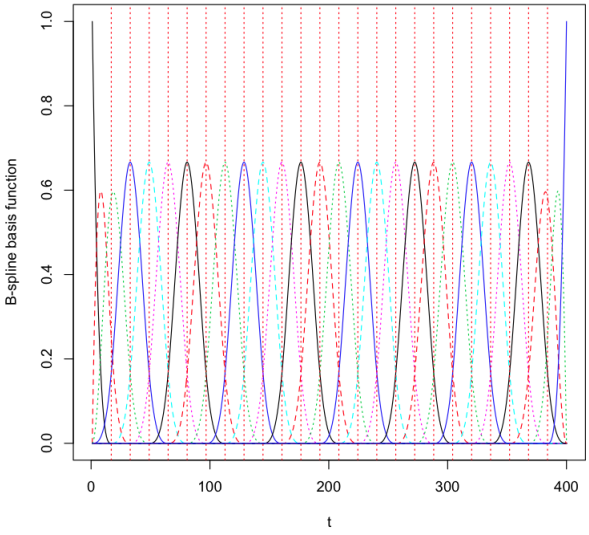

Figure 3.9: B spline basis for the 30^{th} channel. n = 28, m = 4, knots =26

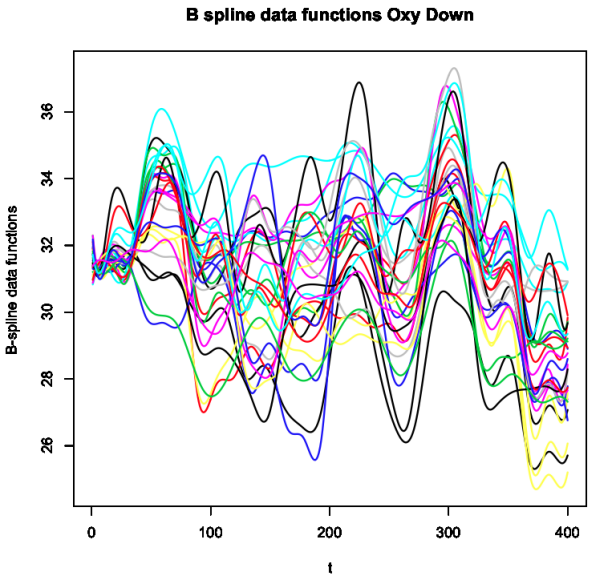

Figure 3.10: B spline data functions for O_2Hb Down measurements

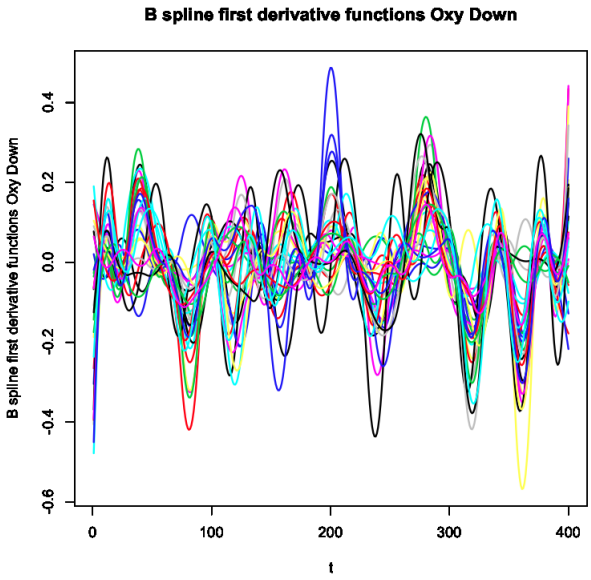

Figure 3.11: B spline data first derivative functions for O_2Hb Down measurements

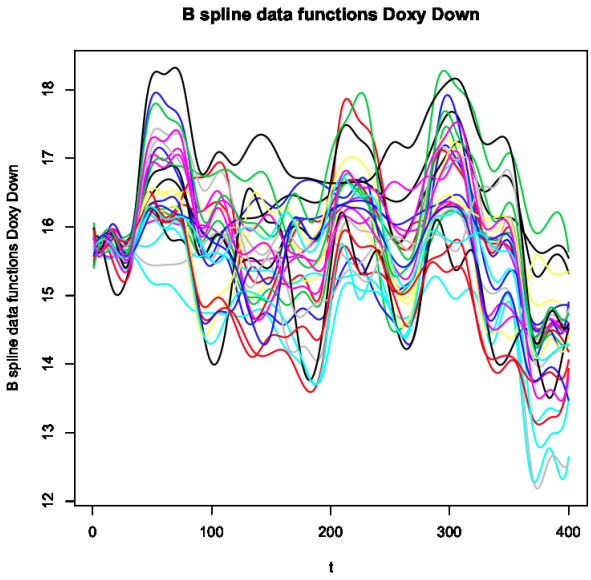

Figure 3.12: B spline data functions for HHb Down measurements

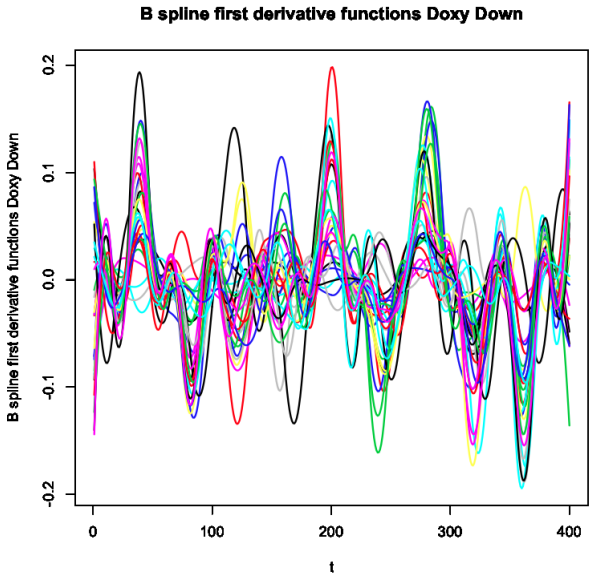

Figure 3.13: B spline data first derivative functions for HHb Down measurements

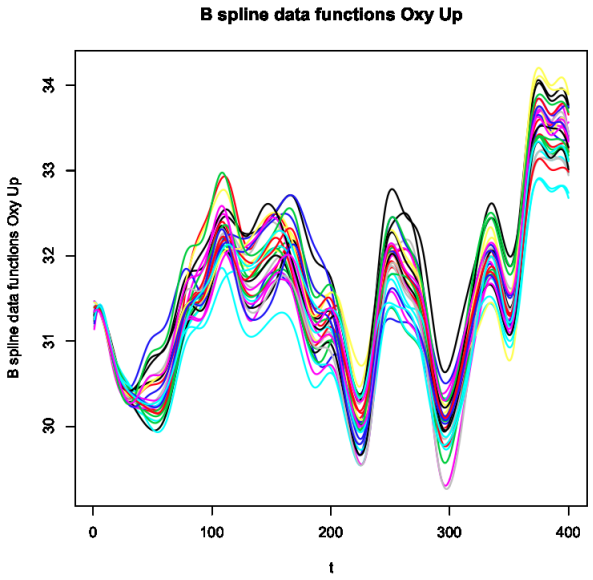

Figure 3.14: B spline data functions for O_2Hb Up measurements

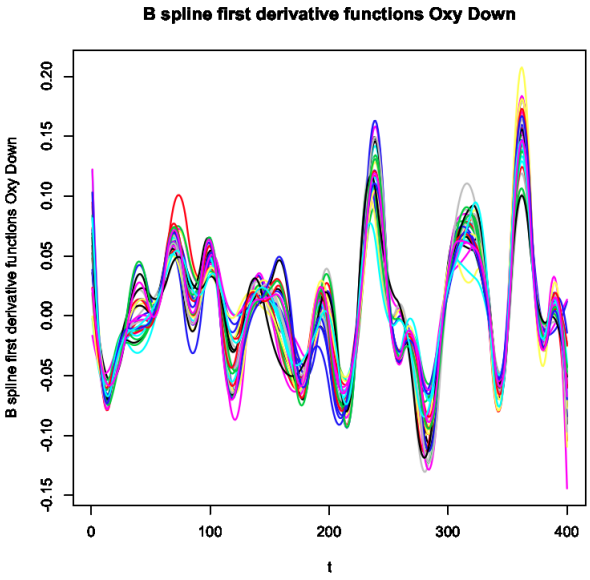

Figure 3.15: B spline data first derivative functions for O_2Hb Up measurements

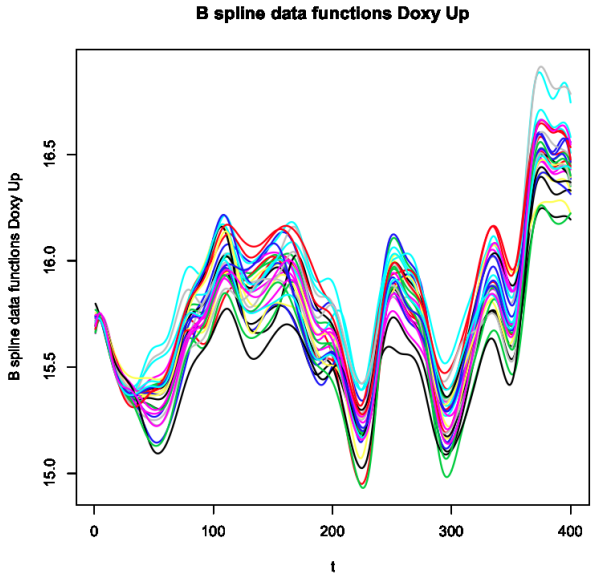

Figure 3.16: B spline data functions for HHb Up measurements

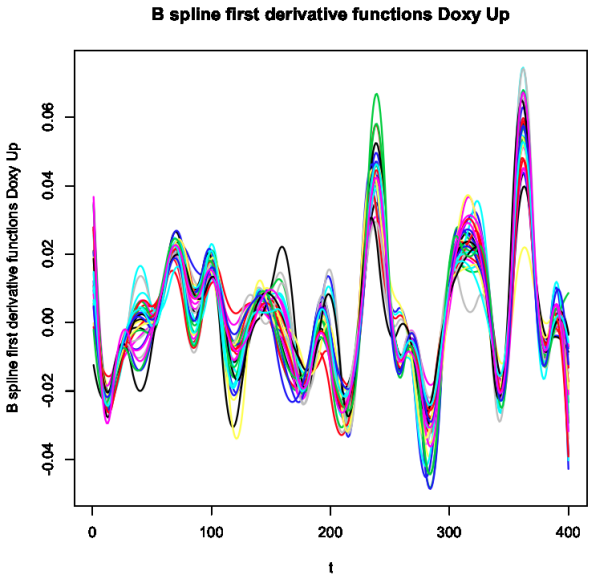

Figure 3.17: B spline data first derivative functions for HHb Up measurements

Chapter 4

K-means algorithm

The main objective of our studies is the detection of neural activation in the brain. For this reason we are interested in clustering channels in *activated* and in *not-activated*.

A main point is that we do not have a certain way to classify channels between *activated* and *not-activated* but we can assess our results by comparison with Bonomini et al. (2015). In the following schedule are summarized activeted channels resulting from Bonomini et al. (2015).

	O_2Hb	HHb
Down data	16-17-18-21-28-29	16-17-18-21-28-29
Up data	$\emptyset$	$\emptyset$

Table 4.1: *Activated* channels resulting in Bonomini et al. (2015) for Down and Up data.

In our work will be called *activated* channels the channels that are gathered with those in according to the previous schedule.

In this chapter we propose the use of a functional data analysis technique to perform a clustering of smoothed hemoglobin concentrations. We ran the K-means algorithm for O_2Hb and HHb concentrations for both UP and DOWN data. Moreover, to confirm these results, we propose the same kind of analysis on two new data: the total hemoglobin concentration (tHb) and the tissue oxygen saturation (StO_2). These new data are given by:

$$\begin{cases} tHb = O_2Hb + HHb \\ StO_2 = 100 \cdot \frac{O_2Hb}{tHb} \end{cases}$$

Figure 4.1 is an example of an acquiring procedure for tHb data. The tHb and StO_2 data are obtained through the functional data of O_2Hb and HHb evaluated for $t = 1, ..., 400$ for UP and DOWN data. The plots 4.2 4.3 show the functional tHb for DOWN and UP data. Figures 4.4 4.5 are the tissue oxygen saturation respectively for DOWN and UP data.

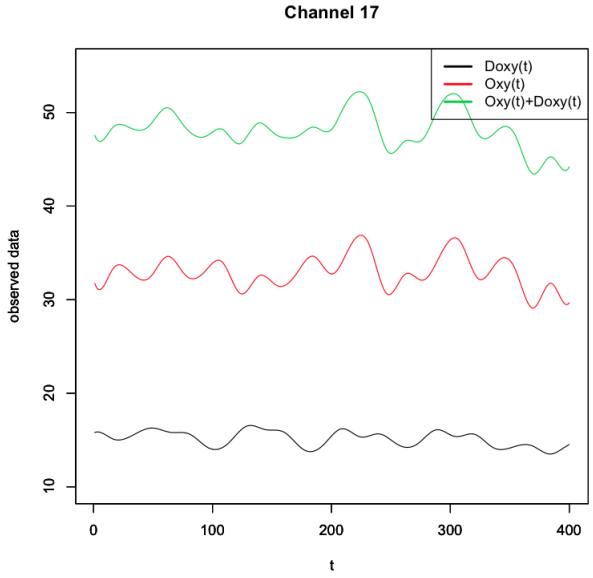

Figure 4.1: Example of tHb measurements for channel 17

4.1 Functional classification

We analyze the 30 channels according to a functional K-means clustering procedure.

$F_i(t) : T \to \mathbb{R}$ is the functional data for $i = 1, .., 30$ channels where $T = (0, 400)$ represents the time domain. To develop the clustering procedure we assume that $F_i(t) \in H^1(T, \mathbb{R})$ and we take for guaranteed the following distance to evaluate the distance between functions:

$$d_1\{F_i(t), F_j(t)\} = \sqrt{\int_T \{F_i(t) - F_j(t)\}^2 dt + \int_T \{DF_i(t) - DF_j(t)\}^2 dt}$$

$$d_2\{F_i(t), F_j(t)\} = \sqrt{\int_T \{F_i(t) - F_j(t)\}^2 dt}$$

In the previous formula $DF_i(t)$ is the first derivative of the functional data of the channel i.

The distance d_1 is the natural distance in the space $H^1(T, \mathbb{R})$, instead the distance d_2 is the natural distance in $L^2(T, \mathbb{R})$. To compare the differences between the final channel clustering we ran the k means algorithm using both distance defined.

The functional K-means clustering algorithm is an iterative procedure, alternating a step of *clustering assignment*, where each curve is assigned to a cluster, and a step of *centroid calculation* where the centroid functional for each cluster is defined.

K-means algorithm is a *clustering partitioning method*: the algorithm divides the dataset into k clusters, where the integer k needs to be specified by the user. Typically, the user runs the algorithm for a range of k values. For each k, the algorithm carries out the clustering and also yields a *quality index*, which allows the user to select a value of k.

For a fixed number of cluster k, the algorithm select random between the curves in the dataset a set of initial centroids $\{\phi_1^{(0)}(t), .., \phi_k^{(0)}(t)\}$. After the initializing, the algorithm iteratively repeats two basic steps. At a generic iteration $m > 0$ the two steps are performed as follow:

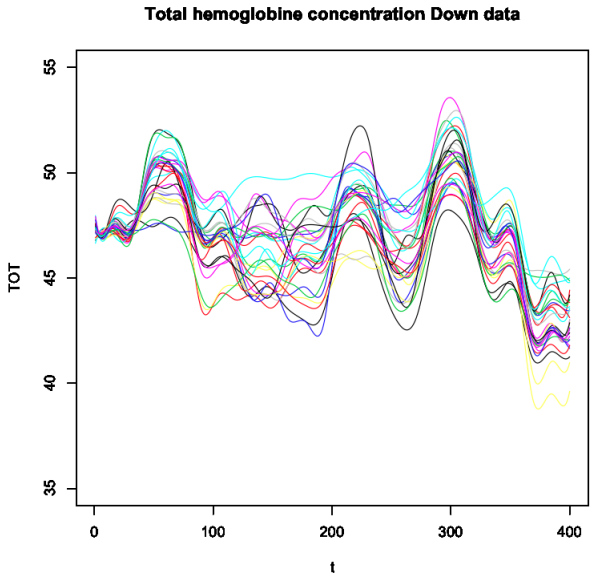

Figure 4.2: tHb data functions Down Data

Step 1 (clustering assignment step) : each curve is assigned to the cluster whose centroid at the $(m-1)^{th}$ iteration is the nearest according to the distance d. This means that the choice of the mth cluster assignment of the ith channel, $i = 1, .., 30$ is

$$C_i^m = \underset{l=1,..,k}{\operatorname{argmin}}\, d\{F_i(t), \phi_l^{m-1}(t)\}$$

Step 2 (centroid calculation step) : the identification of centroids $\{\phi_1^{(m)}(t), .., \phi_k^{(m)}(t)\}$ is performed by solving the optimization problem

$$\phi_l^{(m)}(t) = \underset{\phi \in \Omega_d}{\operatorname{argmin}} \sum_{i:C_i^m=l} d\{F_i(t), \phi(t)\}^2$$

where $C_i^{(m)}$ is the cluster assignment of the ith channel at the current iteration, d is a distance and Ω_d is the Hilbert space where the chosen distance d is natural.

The algorithm is stopped when the same cluster assignment is obtained at two subsequent iterations.

Using this stop criterion, the convergence of the algorithm is ensured in 8-10 iterations.

4.2 Silhouette method

The K-means algorithm, obviously, depends on the number of the clusters k. In this section we want to explain the *silhouette method*. The *silhouette method* through the *silhouette value* help us to select the best number of clusters for K-means.

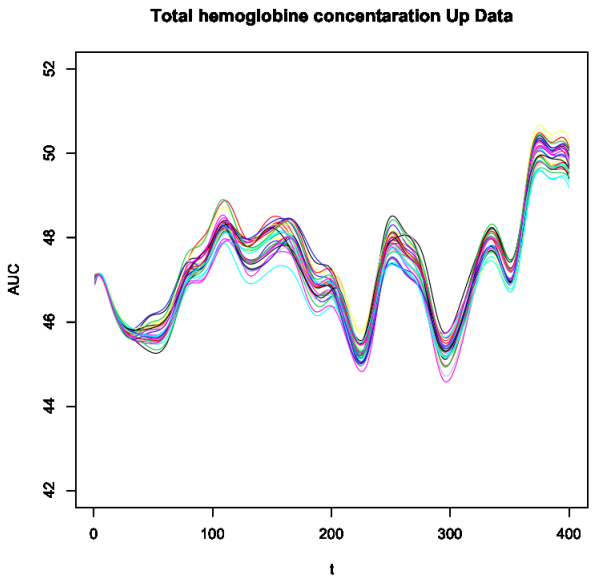

Figure 4.3: tHb data functions Up Data

This part is intended to be an explanation of the *silhouette method*.
As said before, the K-means is a partitioning method. We ran K-means for $k = 2, 3, 4, 5, 6$.
For each channel i we denote by $\widetilde{C}$ the cluster which i belongs and compute

$$a(i) = \frac{1}{\widetilde{C} - 1} \sum_{\substack{j \in \widetilde{C} \\ j \neq i}} d\{\boldsymbol{F}_i, \boldsymbol{F}_j\}$$

$|\widetilde{C}|$ is the cardinality of the cluster $\widetilde{C}$ and d is one of the distance defined before.
$a(i)$ is the average dissimilarity of i to all other objects of $\widetilde{C}$.

Now consider any cluster C different from $\widetilde{C}$ and compute

$$d(i, C) = \frac{1}{|C|} \sum_{j \in C} d\{\boldsymbol{F}_i, \boldsymbol{F}_j\}$$

$d(i, C)$ is the average dissimilarity of the channel i to all the objects of C.

After computing $d(i, C)$ for all clusters $C \neq \widetilde{C}$ we take the smallest of those:

$$b(i) = \min_{C \neq \widetilde{C}} d(i, C)$$

The *silhouette value* $s(i)$ of the object i is defined as:

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}}$$

$s(i)$ lies between -1 and 1 and can be interpreted as follows:

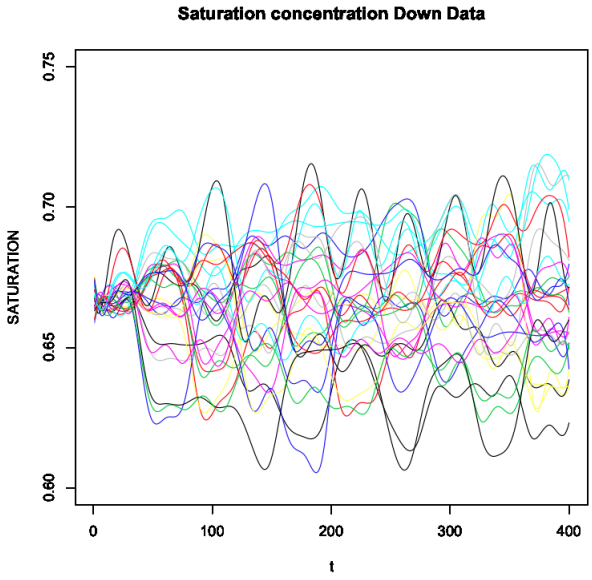

Figure 4.4: StO_2 data functions Down Data

$s(i) \approx 1 \Rightarrow$ channel i is well classified
$s(i) \approx 0 \Rightarrow$ channel i lies intermediate between two clusters
$s(i) \approx 1 \Rightarrow$ channel i is badly classified

For instance, Figure 4.9 shows the silhouette plot for the clustering of HHb Down data using two clusters. In the plot each line is representative of a channel and the colour line reveal the cluster which they belong. The meaning of Figures 4.12, 4.15, 4.18, 4.21, 4.24, 4.27 and 4.30 are the same of previous one.

The silhouette of cluster $\widetilde{C}$ is a plot of all its $s(i)$, ranked in a decreasing order. The entire silhouette plot shows the silhouettes of all clusters below each other, so the quality of clusters can be compared.

The *average silhouette width*, or *silhouette value*, is the mean of all silhouette width. The best k number of clusters for the dataset is that with the higher *silhouette value*.

In the Figure 4.8 the dots are the silhouette value derived from the K-means algorithm for different number of clusters k. In this case and in all others that we studied we chose $k = 2$ because is the highest silhouette value.

In the same way it is possible to explain the plots 4.11, 4.14, 4.17, 4.20, 4.23, 4.26 and 4.29.

4.3 Results

The aim of this analysis is to detect what areas of the brain are *activated*.

We perform a K-means clustering algorithm for both Up and Down data for the O_2Hb and HHb measurements, using the two different distances described before and functional data obtained

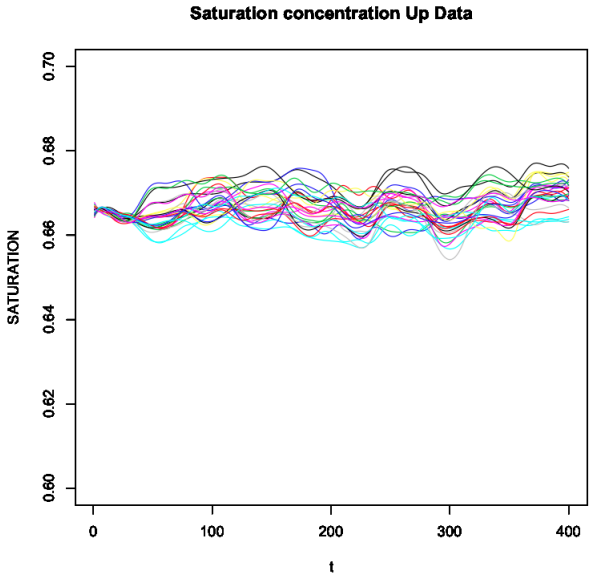

Figure 4.5: StO_2 data functions Up Data

through the Fourier and Bspline basis.

The final silhouette plots obtained by clustering the 30 channels according to a K-means procedure using the distance d_2, Bspline basis and setting $k = 2, 3, 4, 5, 6$ are shown in the Figure (4.11, 4.14, 4.17, 4.20, 4.23, 4.26 and 4.29). As it is evident from the plots, the best grouping structure is obtained setting $k = 2$. We repeat the same analysis using the d_1 distance and Fourier basis but the results are similar to the previous one, therefore we decide to continue our study using the d_2 distance and Bspline basis.

The final classification that is obtained choosing $k = 2$ is shown in the Figures 4.10 4.13, 4.16, 4.19, 4.22, 4.25, 4.28 and 4.31. In the plots each channel is coloured according to the cluster to which it belongs to. The marked lines are the medoids functions, one for each cluster.

Figures 4.6 4.7 show p-values resulting from Bonomini et al. (2015).

Reminding that our work is based on Bonomini et al. (2015), where we found activated channels only for down data and that these are 16-17-18-21-28-29, the most interesting thing that we can notice analyzing our clusters is that channels 16-17-18-19-21-28-29 are gather together both for Up and Down data.

Anyway, in the same cluster we found other channels 2-3-20-27 that registered low p-values in study Bonomini et al. (2015), therefore we could conclude that this cluster assignment is coherent with the previous study. In the Figures are shown the p-values of study Bonomini et al. (2015) for each channel for O_2Hb and HHb for DOWN and UP data.

We found the same cluster assignment for both Down and Up data but, as we can see from the plot, the clustering assignment for Up data seems to be artificial because the distance between medoid functions is small.

We investigate this issue in the next chapter computing the functional median for each cluster in the ten subintervals.

	O_2Hb	HHb
Down data	2-3-15-16-17-18-20-21-27-28-29	2-7-10-15-16-17-18-21-23-26-28-29

Table 4.2: *Activated* channels resulting in our study for Down data.

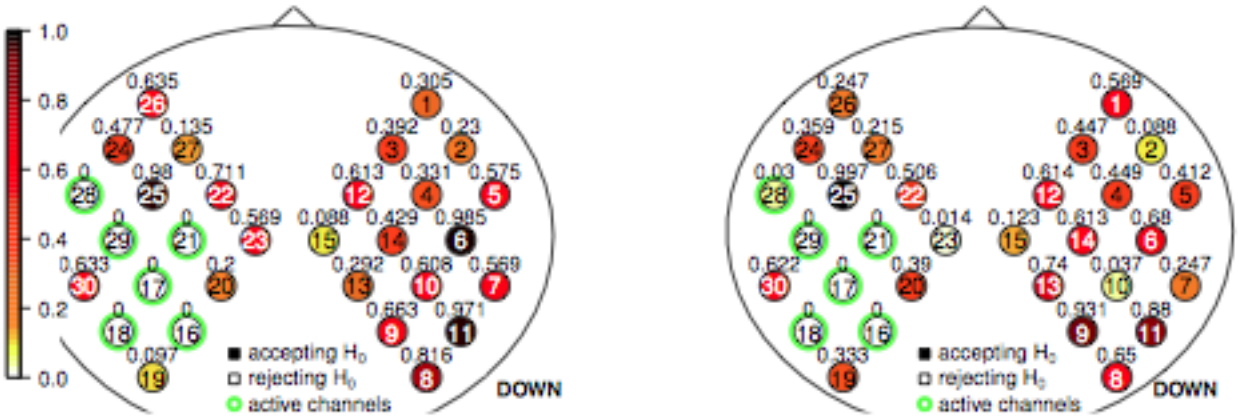

Figure 4.6: P-values of channels resulting from Bonomini et al. (2015) in Down O_2Hb (left) and HHb (right) data.

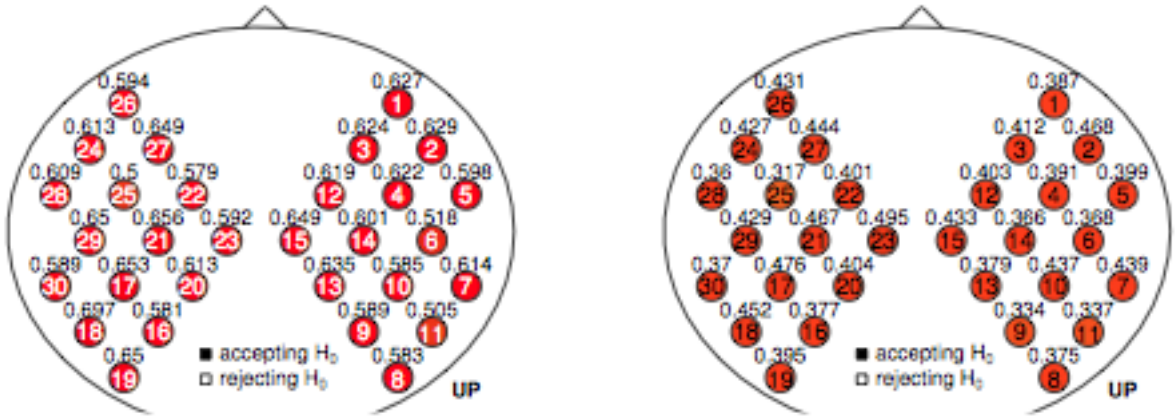

Figure 4.7: P-values of channels resulting from Bonomini et al. (2015) in Up O_2Hb(left) and HHb(right) data.

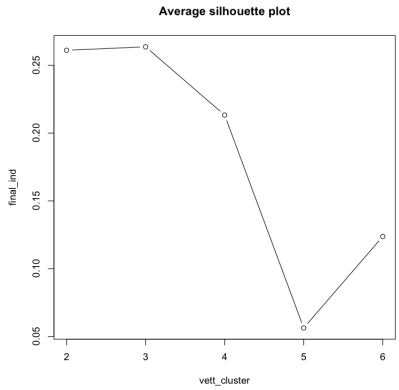

Figure 4.8: Silhouette Average plot of HHb Down data obtained through Bspline basis and L_2 distance

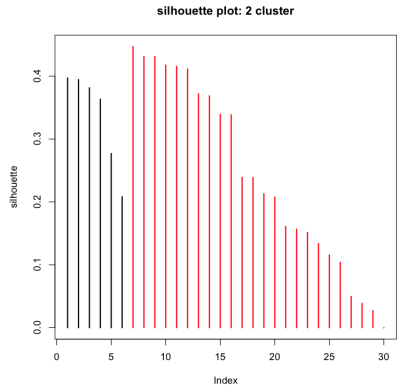

Figure 4.9: Silhouette plot for 2 clusters of HHb Down data obtained through Bspline basis and L_2 distance

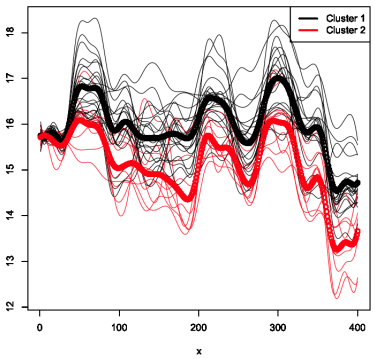

Figure 4.10: Cluster plot for 2 clusters of HHb Down data obtained through Bspline basis and L_2 distance

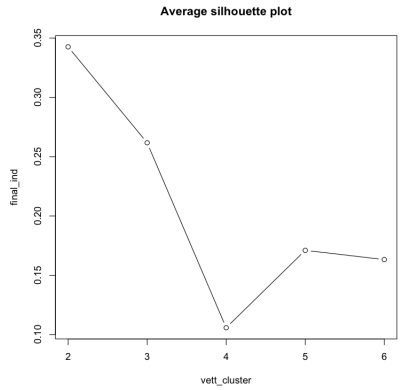

Figure 4.11: Silhouette Average plot of HHb Down data obtained through Fourier basis and L_2 distance

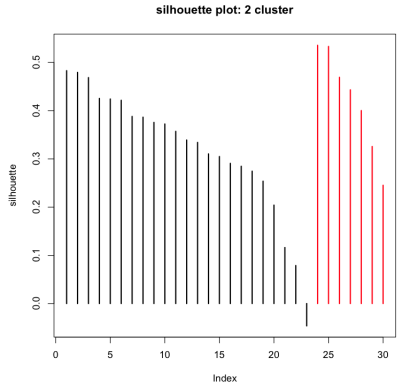

Figure 4.12: Silhouette plot for 2 clusters of HHb Down data obtained through Fourier basis and L_2 distance

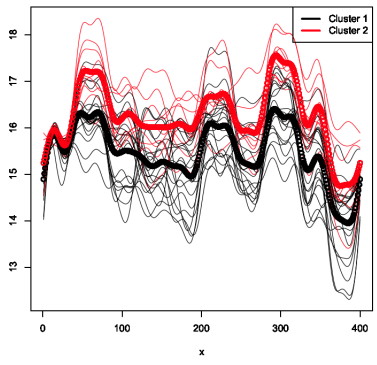

Figure 4.13: Cluster Plot for 2 cluster of HHb Down data obtained through Fourier basis and L_2 distance

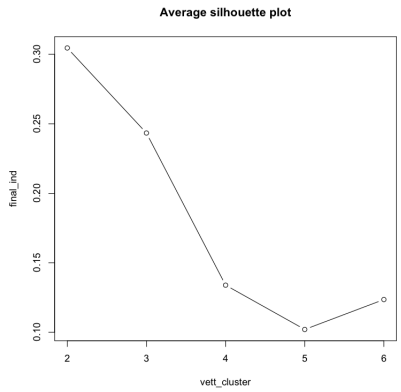

Figure 4.14: Silhouette Average plot for HHb Up data obtained through Bspline basis and L_2 distance

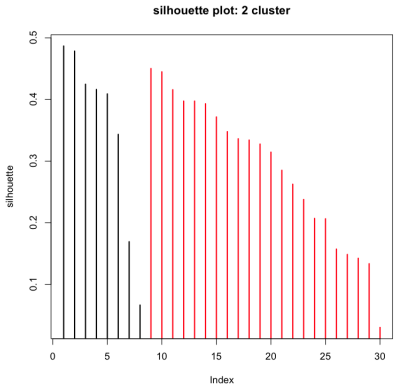

Figure 4.15: Silhouette plot for 2 clusters of HHb Up data obtained through Bspline basis and L_2 distance

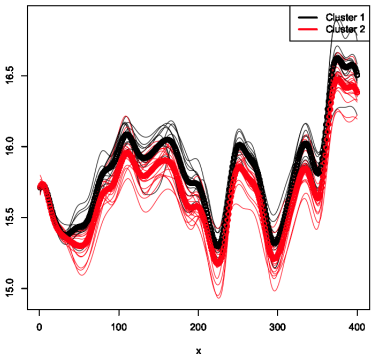

Figure 4.16: Cluster plot for 2 clusters of HHb Up data obtained through Bspline basis and L_2 distance

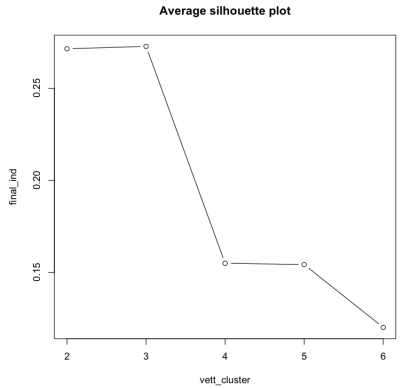

Figure 4.17: Silhouette Average plot for HHb Up data obtained through Fourier basis and L_2 distance

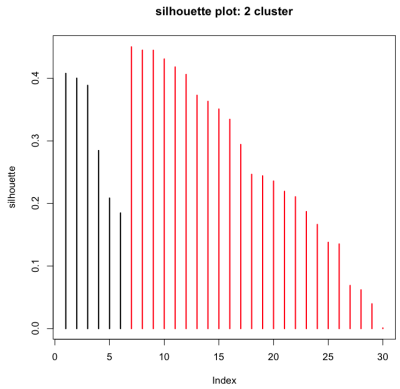

Figure 4.18: Silhouette plot for 2 clusters of HHb Up data obtained through Fourier basis and L_2 distance

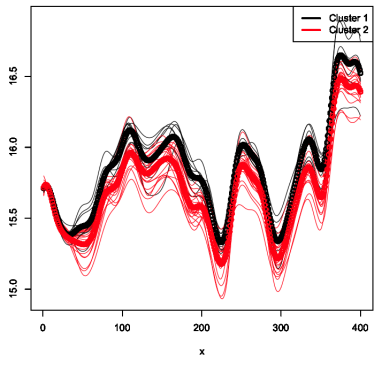

Figure 4.19: Cluster Plot for 2 clusters of HHb Up data obtained through Fourier basis and L_2 distance

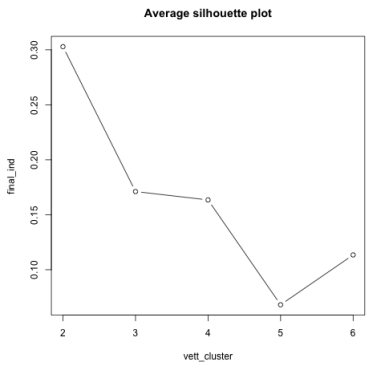

Figure 4.20: Silhouette Average plot for O_2Hb Down data obtained through Bspline basis and L_2 distance

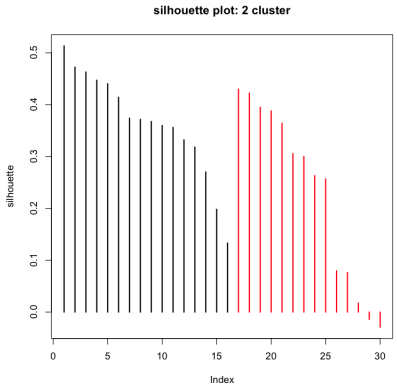

Figure 4.21: Silhouette plot for 2 clusters of O_2Hb Down data obtained through Bspline basis and L_2 distance

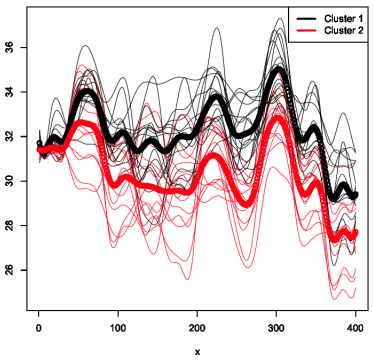

Figure 4.22: Cluster plot for 2 clusters of O_2Hb Down data obtained through Bspline basis and L_2 distance

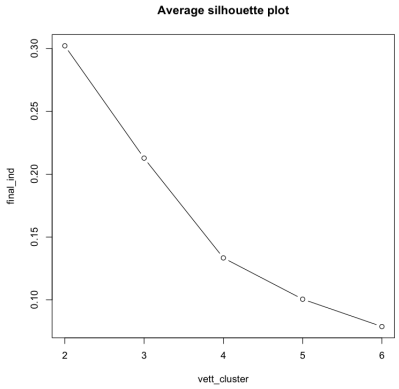

Figure 4.23: Silhouette Average plot for O_2Hb Down data obtained through Fourier basis and L_2 distance

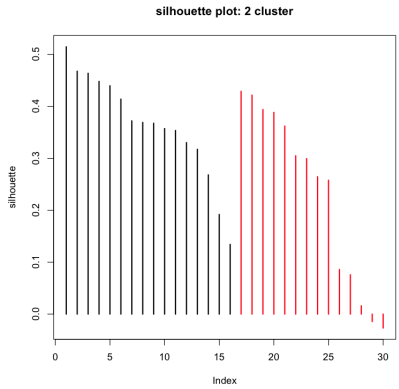

Figure 4.24: Silhouette plot for 2 clusters of O_2Hb Down data obtained through Fourier basis and L_2 distance

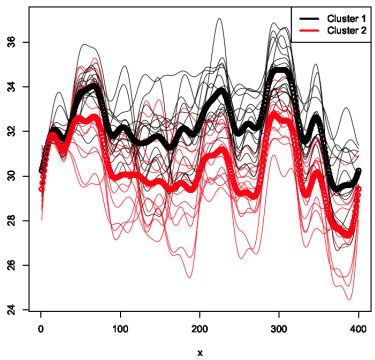

Figure 4.25: Cluster plot for 2 clusters of O_2Hb Down data obtained through Fourier basis and L_2 distance

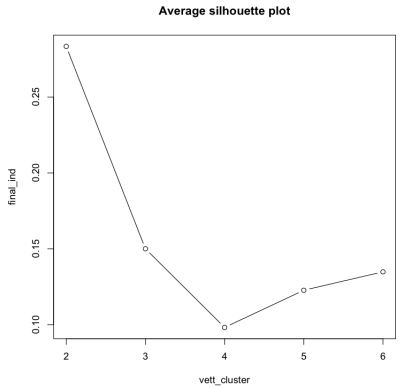

Figure 4.26: Silhouette Average plot for O_2Hb Up data obtained through Bspline basis and L_2 distance

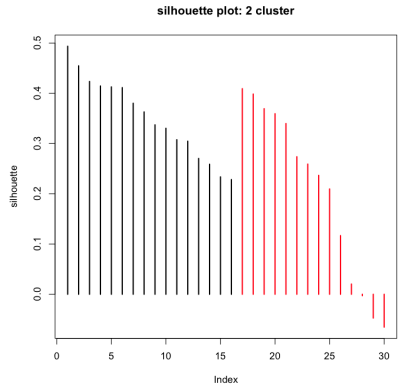

Figure 4.27: Silhouette plot for 2 clusters of O_2Hb Up data obtained through Bspline basis and L_2 distance

4.4 Clustering analysis on tHb and StO_2 data

With the purpose of discussing the analysis on total hemoglobin concentration (tHb) and on the tissue oxygen saturation (StO_2), we develop these clustering analysis to confirm results of the previous K-means analysis on the concentration of O_2Hb and HHb.

As written before, these data are obtained from the functional data of O_2Hb and HHb through the Bspline basis. Since we found that the best cluster classification is obtained for $k = 2$, we ran the K-means algorithm imposing the number of clusters equals to two. Moreover, the silhouette value $s(2) \simeq 0.3$ confirms the goodness of our analysis for two clusters.

The final classification confirms the clusters composition obtained in the previous analysis but what is evident is that for Down data there is a clear separation between the two clusters. In Up data, as explained, this difference is not so accentuated because the O_2Hb and HHb measurements are flat.

In any case, these lasts clustering analysis help us to confirm the results of k means on O_2Hb

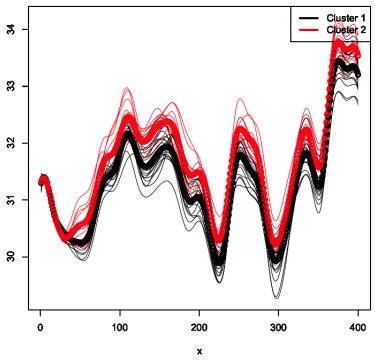

Figure 4.28: Cluster plot for 2 clusters of O_2Hb Up data obtained through Bspline basis and L_2 distance

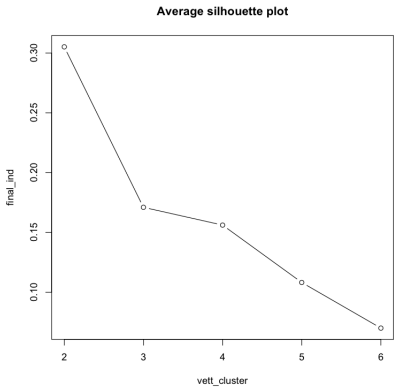

Figure 4.29: Silhouette Average for O_2Hb Up data obtained through Fourier basis and L_2 distance

and HHb concentrations.

At last we can conclude that these results are coherent with the previous one. The clustering results are shown in the table 4.3.

The plots for the final clustering assignment for tHb and StO_2 for DOWN and UP data are in the Figure 4.32, 4.33, 4.34 and 4.35

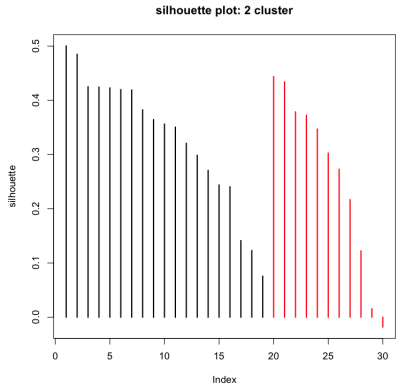

Figure 4.30: Silhouette plot for 2 clusters of O_2Hb Up data obtained through Fourier basis and L_2 distance

	$DOWN$	UP
tHb **data**	3-15-16-17-18-20-21-28-29	3-10-15-16-17-18-21-22-28-29
StO_2 **data**	3-15-16-17-18-20-21-28-29	3-10-15-16-17-18-21-22-28-29

Table 4.3: *Activated* channels resulting in our study for tHb and StO_2 data for datasets DOWN and UP.

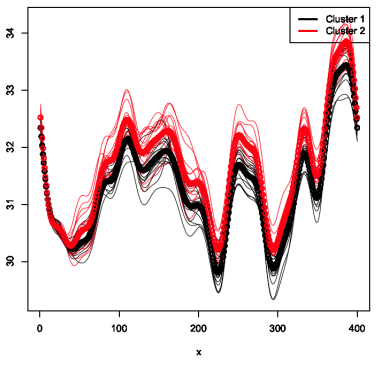

Figure 4.31: Cluster plot for 2 clusters of O_2Hb Up data obtained through Fourier basis and L_2 distance

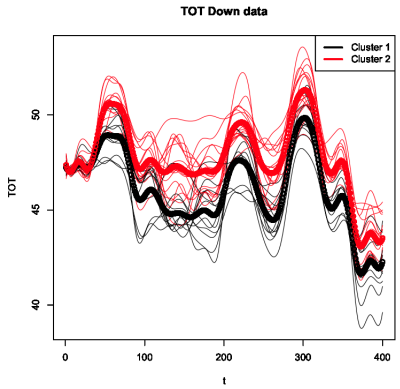

Figure 4.32: Cluster plot for 2 clusters of TOT Down data obtained through Bspline basis and L_2 distance

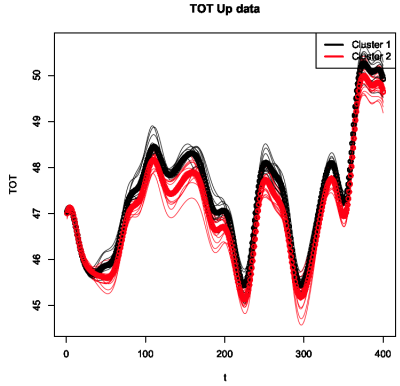

Figure 4.33: Cluster plot for 2 clusters of TOT Up data obtained through Bspline basis and L_2 distance

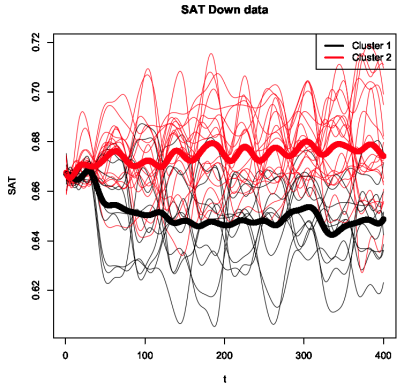

Figure 4.34: Cluster plot for 2 clusters of SAT Down data obtained through Bspline basis and L_2 distance

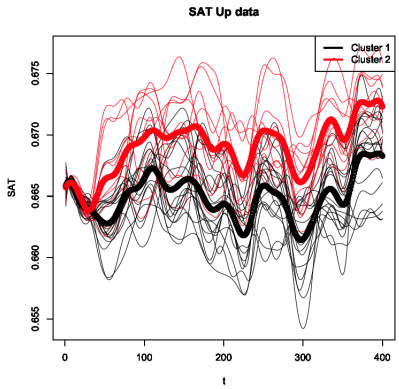

Figure 4.35: Cluster plot for 2 clusters of SAT Up data obtained through Bspline basis and L_2 distance

Chapter 5

Functional Median

In this chapter we want to analyze and discuss the previous results using a new method: the *functional median*.

In order to explain what functional median is, we need to introduce the concept of *depth* for the functional data.

5.1 Introduction to Depth Measures for functional data

The statistical analysis of functional data is a growing need in many research areas. In particular, a robust methodology is important to study curves, which are the output of many experiments in applied statistics.

As a starting point, we need a new definition of depth for functional observations based on the graphic representation of curves. Given a collection of functions, a depth measure establishes the "centrality" of an observation and provides a natural center-outward ordering of the sample curves. Robust statistics, such as the median function or a trimmed mean function, can be defined from this definition.

This new depth is also suitable for complex observations such as microarray data, images, and those arising in market and financial studies. Simulation results show that the corresponding depth based trimmed mean presents better performance than other possible location estimators proposed in the literature. Data depth can be also used to detect outliers.

A fundamental task in functional data analysis is to provide an ordering within a sample curves that allows the definition of order statistics, such as ranks and L-statistics. A natural tool to analyze these functional data aspects is the idea of *statistical depth*.

Given a distribution of probability F in $\mathbb{R}^d$, a statistical depth assigns to each point $\mathbf{x}$ a real nonnegative bounded value $D(\mathbf{x}, F)$.

Several depth definitions for multivariate data have been proposed and analyzed by Mahalanobis (1936), Tukey (1975), Oja (1983), Liu (1990), Singh (1991), Fraiman and Meloche (1999), Vardi and Zhang (2000), Koshevoy and Mosler (1997) and Zuo (2003), among others. Liu (1990) and Zuo and Serfling (2000a) introduced and studied key proprieties a depth should satisfy.

Data depth can widely applied. For example, Liu and Singh (1993) presented a nonparametric multivariate rank test using quality depth index. Yeh and Singh (1997) studied confidence sets based on Tukey's depth. Also, Liu, Parelius and Singh (1999) offered depth tools for multivariate analysis; for instance, they defined trimmed regions, central regions and contours, and constructed

a scale curve to visualize dispersion. In addition, Rousseeuw and Hubert (1999) introduced the idea of regression depth and Li and Liu (2004) designed a graphic tool and a test to check if two multivariate samples come from the same population.

Direct generalization of the multivariate depths to functional data often leads to either depths that are computationally intractable or depths that do not take into account some natural proprieties of the functions, such as shape. Vardi and Zhang (2000) proposed the L_1-depth for multivariate data, which is computationally feasible in high dimensions and therefore can be extended to functional data. This depth was also analyzed by Serfling (2004) in the context of spatial rank functions, and it is also closely related and motivated by the geometric quantiles for multivariate data introduced by Chaudhuri (1996), who indicated that the geometric quantiles, in particular the median, could be extended to Hilbert and Banach spaces. Fraiman and Muniz (2001) defined and studied a concept of depth for functional observations based on the integrals of univariate depth.

Lopez-Pintado and Romo (2006) proposed a new notion of depth for functional data. That is based on the graphic representation of the functions and makes use of the bands defined by their graphs on the plane. This definition has also the advantage of being computationally feasible, which is essential for analyzing high-dimensional data. Some asymptotic results, such as the uniform convergence of the sample depth and the deepest point are established. Most of these proprieties are extended to functions. With this definition is possible to generalize the concept of multivariate L-estimates, in particular trimmed means, to a functional context. Robust methods are even more relevant in a functional setting then in multivariate problems because outliers can affect functional statistics in more ways, and they can be more difficult to detect. For instance, a curve could be an "outlier" without having any unusual large value but it could be a decreasing curve in a set of of increasing functions or a very irregular curve within a set of smooth curves.

The Lopez-Pintado and Romo definition of functional depth is that we use to define the functional median in our work.

5.2 Band depth for functional data

In our work we discuss, without loss of generality, only the empirical aspects of band depth, for the theoretical aspects see Lopez-Pintado and Romo (2006).

The definition of band depth that we use follows a graph-based approach.

Let $C(I)$ be the set of continuous functions defined on the compact interval $I \in \mathbb{R}$. Let $x_1(t), ..., x_n(t)$ be a collection of observations belonging to $C(I)$. The graph of a function x is the subset of $\mathbb{R}^2$ given by

$$G(x) = \{(t, x(t)) : t \in I\}.$$

The band in $\mathbb{R}^2$ defined by the curves $x_{i_1}, ..., x_{i_k}$ is

$$\begin{aligned} B(x_{i_1}, x_{i_2}, ..., x_{i_k}) &= \left\{(t, y) : t \in \min_{r=1,...,k} x_{i_r}(t) \leq y \leq \max_{r=1,...,k} x_{i_r}(t)\right\} \\ &= \left\{(t, y) : t \in y = \alpha_t \min_{r=1,...,k} x_{i_r}(t) + (1 - \alpha_t) \max_{r=1,...,k} x_{i_r}(t), \text{for some } \alpha_t \in [0,1]\right\} \end{aligned}$$

Figure 5.1 shows the band defined by three curves that is the region in the plane enclosed by all of them.

For any function x in $C(I)$,

$$BD_n^j(x) = \binom{n}{j}^{-1} \sum_{1 \le i_1 < i_2 < \ldots < i_j \le n} \mathbb{1}\{G(x) \subset B(x_{i_1}, x_{i_2}, ..., x_{i_k})\}, \quad j \ge 2 \tag{5.1}$$

express the proportion of bands $B(x_{i_1}, x_{i_2}, ..., x_{i_j})$ given by j different curves $x_{i_1}(t), x_{i_2}, ..., x_{i_n}(t)$ containing the graph of x and $\mathbb{1}\{A\}$ is one if A is true and zero, otherwise.

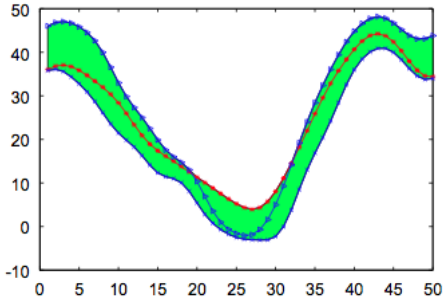

Figure 5.1: Band defined by three curves

Given the sample $x_1, ..., x_n$, the *band depth* of x is

$$BD_{n,J}(x) = \sum_{j=2}^{J} BD_n^j(x) \quad J \ge 2. \tag{5.2}$$

Hence, the band depth is the sum up to J (with $J \ge 2$) of indexes defined as in (5.1). The parameter J is a parameter that indicates the maximum number of curves used to construct a band. Lopez-Pintado and Romo (2006) show that the order induced by $BD_{n,J}(x)$ in the sample is very stable with respect to growing J. Therefore, it is convenient, for computational reason, to use small values of J (e.g., $J = 2$ or 3). On the other hand, when the sample curves are very irregular there might be too many crossovers and it is very unlikely for a curve to be in a band delimited only by two curves. Usually, the band depth are considered with $J = 2$ ($BD_{n,2}(x)$).

5.3 Modified band depth for functional data

Instead of considering the indicator function as in (5.1), Lopez-Pintado and Romo (2006) proposed a more flexible definition of depth, called *modified band depth (MBD)*.

The indicator function in definition is replaced by the Lebesgue measure of the set where the function is inside the corresponding band. For any function x in $x_1, ..., x_n$ and for $2 \le j \le n$, let

$$\begin{aligned} A_j(x) &= A_{i_1,...,i_j}(x) = A(x; x_{i_1}, x_{i_2}, ..., x_{i_j}) \\ &= \left\{ t \in I : \min_{r=i_1,...,i_j} x_r(t) \le x(t) \le \max_{r=i_1,...,i_j} x_r(t) \right\} \end{aligned}$$

be the set in the interval I where the function x is in the band determined by the observation $x_{i_1}, x_{i_2}, ..., x_{i_j}$. If λ is the Lebesgue measure on I, $\lambda_r(A_j(x)) = \lambda(A_j(x))/\lambda(I)$ gives the "proportion of time" that x is in the band. Now,

$$MDB_n^j(x) = \binom{n}{j}^{-1} \sum_{1 \le i_1 < i_2 < \ldots < i_j \le n} \lambda_r(A(x; x_{i_1}, x_{i_2}, ..., x_{i_j})), \quad 2 \le j \le n \tag{5.3}$$

is more flexible version of $BD_n^j(x)$: if x is always inside the band, the value $\lambda_r(A_j(x))$ is one as in the previous notion of depth.

For functions $x_1, ..., x_n$, the modified band depth of one of these curves x is

$$MDB_{n,J}(x) = \sum_{j=2}^{J} MDB_n^j(x). \tag{5.4}$$

The band depth (BD) is less adaptive (or more restrictive) than its modified version (MBD) because, if the curves are very irregular, it is very unlikely for a curve to belong to a band. This provides many ties in depth. On the other hand, it depends strongly on the curves shape. Another important difference between both definitions is their behavior for curves leaving the sample center only for a short interval, i.e., remaining in the interior of the sample most of the time but taking extreme values in short intervals: for these curves the MDB can still be large whereas the band depth (BD) takes small values.

5.4 Functional median

Using the previous definition of band depth (BD) and modified band depth (MBD), it is possible to define the *functional median*.

The sample median of a functional dataset $\widehat{m}_n$ is the element of the functional dataset fulfilling the maximum depths, given a certain definition of depth.

For instance, for DB:

$$\widehat{m}_n = \underset{x \in \{x_1, ..., x_n\}}{\operatorname{argmax}} DB(x)$$

while for MBD:

$$\widehat{m}_n = \underset{x \in \{x_1, ..., x_n\}}{\operatorname{argmax}} MDB(x)$$

In our work we use the R[1] package **roahd**[2] to compute the functional median of the concentration of O_2Hb and HHb data. The function `median_fdata( fData, type = 'MBD', ... )` of the package roahd allows to compute the functional median using both definition of depth. In this thesis we decided to use the modified band depth because it is the default method.

Roahd (RObust Analysis of High dimensional Data) is a package meant to gather recently proposed statistical methods to deal with the robust analysis of functional data. The package contains an implementation of quantitative methods, based on functional depths and indexes, on top of which are built some graphical methods useful to carry out an explorative analysis of a functional dataset. In order to allow the processing of high-dimensional data, the package support

[1] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

[2] Nicholas Tarabelloni, Ana Arribas-Gil, Francesca Ieva, Anna Maria Paganoni and Juan Romo (2016). roahd: Robust Analysis of High Dimensional Data. R package version 1.1. https://CRAN.R-project.org/package=roahd

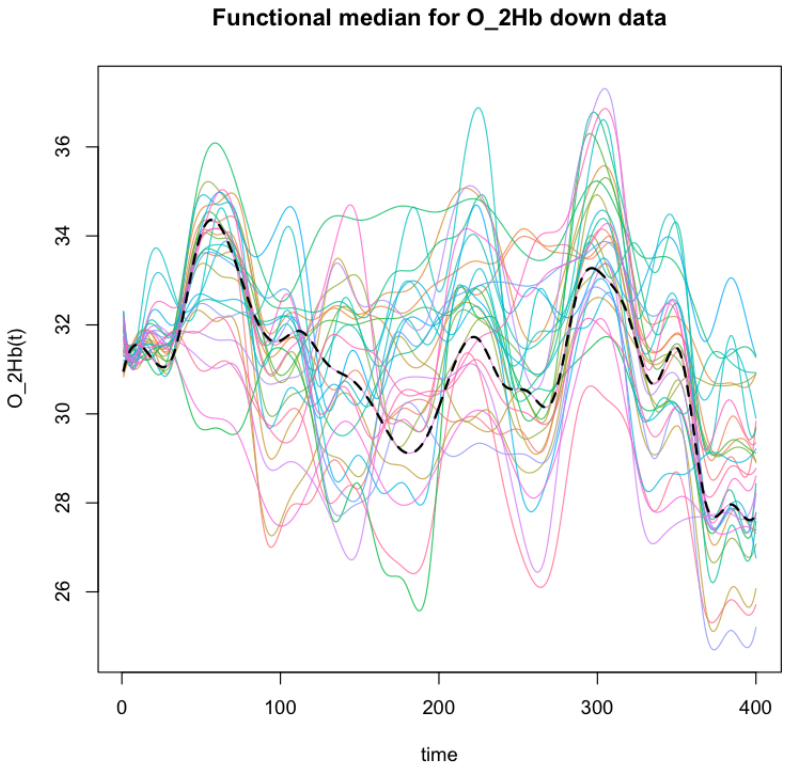

Figure 5.2: Example of functional median for O_2Hb data. The functional median is the dotted line.

univariate and multivariate data and functions have special consideration for the computational efficiency.

The Figure 5.2 is an example of how `median_fdata( ... )` works. The examples refers to O_2Hb concentration for Down data and the dotted line represents the functional median.

5.5 Clustering through functional median

As explained in Chapter 2, the signal analyzed represents the concentration of each channel for O_2Hb or HHb during a trial of length of $400s$. This trial can be split in 10 sub-trials each of length of $40s$ for which we want to estimate the functional median of each channel. Thus, we obtain a new functional data composed by the 30 median functions.

At this stage of the analysis, using the K-means algorithm, we are interested in clustering the 30 functional medians in order to observe if we discover the same cluster assignment found for the entirely signal concentration of O_2Hb and HHb. This interest came form the fact that the functional median is a robust method to summarize and estimate the informations of a functional data.

Figures 5.3- 5.6 represent the results of functional median clustering for O_2Hb and HHb for Up and Down data.

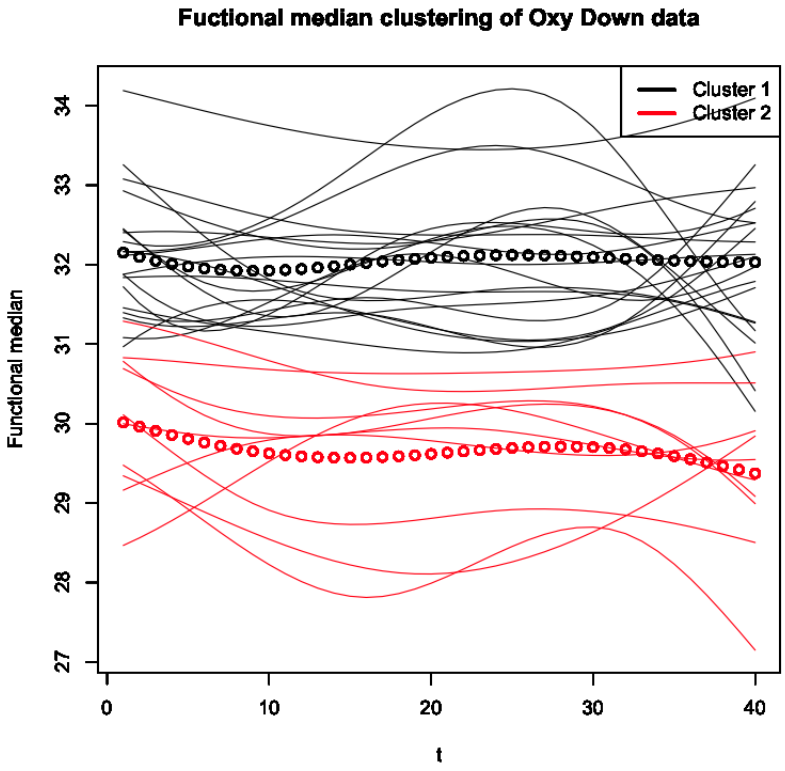

Figure 5.3: Clustering obtained for functional median of O_2Hb Down Data.

In conclusion, we find that channels 16, 17, 18, 21, 28, 29 are gathered together as resulting from the k-mean applied to the entirely signal and from the previous study conducted by Bonomini et al. (2016)

The K-means algorithm applied to functional medians is able to classify channels taking advantage of the differences in signal's amplitude but not in covariance structure of data.

For this reason, in the following chapter we propose a covariance-based clustering algorithm.

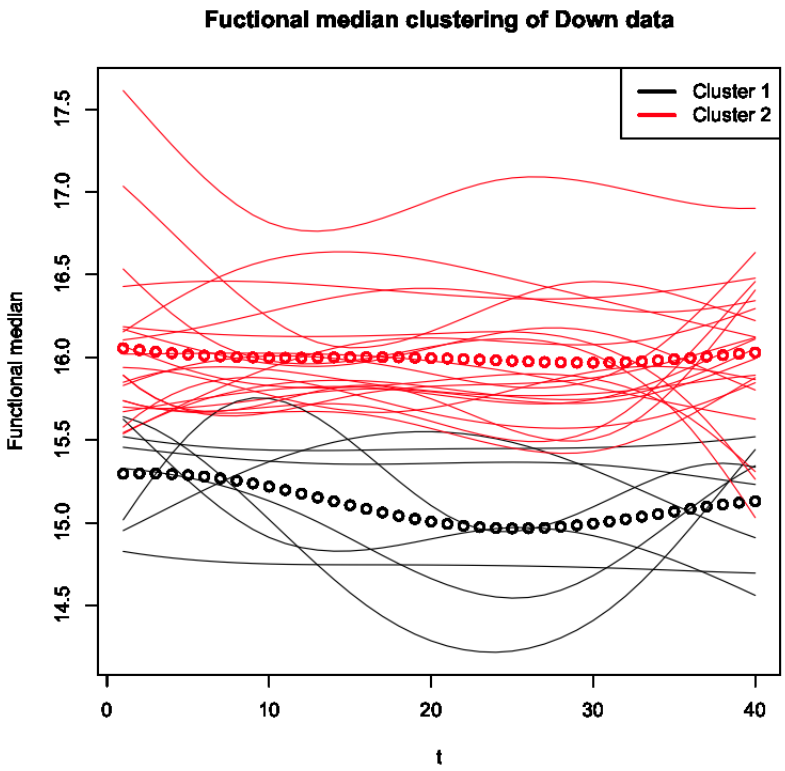

Figure 5.4: Clustering obtained for functional median of HHb Down Data.

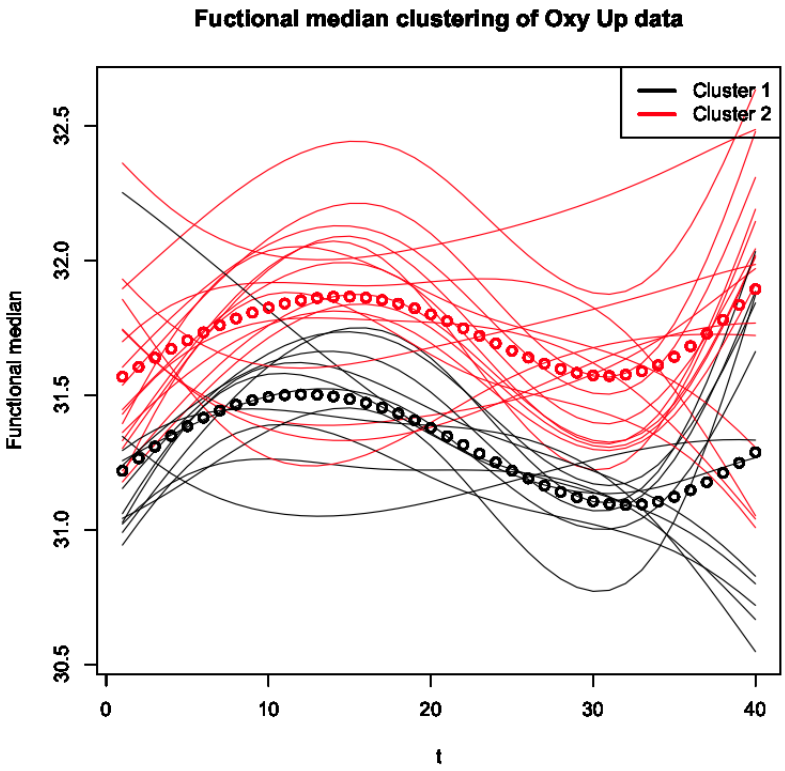

Figure 5.5: Clustering obtained for functional median of O_2Hb Up Data.

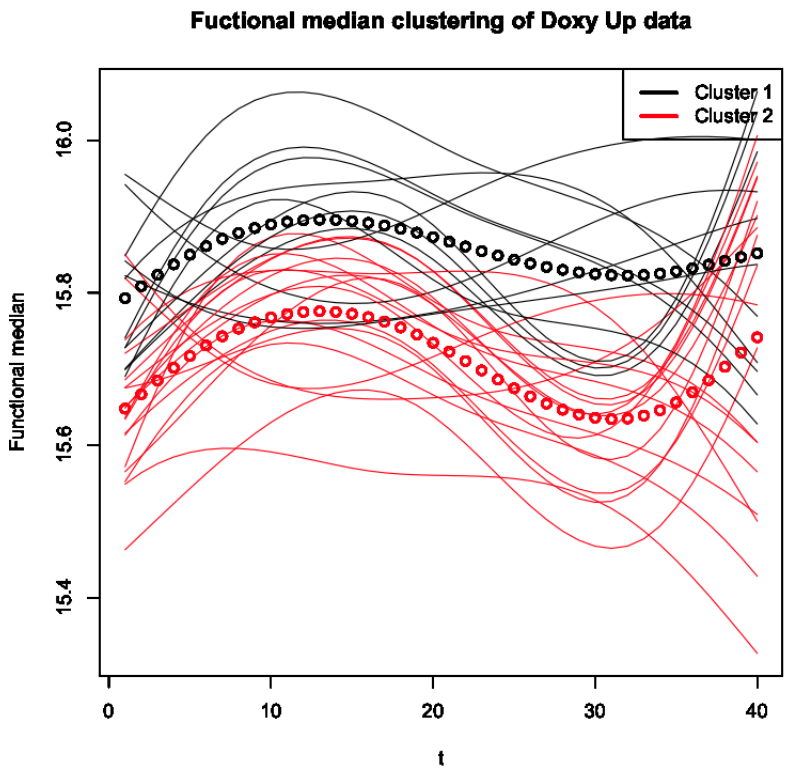

Figure 5.6: Clustering obtained for functional median of HHb Up Data.

Chapter 6

Covariance-based Clustering

In this chapter we introduce and apply a new algorithm to perform clustering of multivariate and functional data. In Chapter 4 we have already seen *K-means* algorithm that aggregates data depending on similarities in location. Our focus now goes to reach the clustering entirely on the basis of differences in covariance structures. The algorithm proposed relies on a proper quantification of distance between the estimated covariance operators of the groups and it subdivides the data in two groups maximizing the distance between the induced covariances.

6.1 Introduction

In literature there are not so much work about covariance based clustering, however this way to gather data is not less trivial since it gives a proper quantification of differential correlation or distance between covariances of data. In fact, it might happen to analyze groups of data that are scarcely distinguishable in terms of location, while showing differences in their variability.

Examples can be found in biostatistics where, for instance, the dichotomy between physiological and pathological features often shows an interesting change in pattern of variability. This is also of great interest in genomics, where instead of focusing on gene expression levels, one could be interested in finding different correlation structures among subsets of data. This is the core task in the analysis of the differential co-expression of genes, namely the differential correlation structure among expression levels in different subsets of experimental conditions. Watson (2006) proposed a method to identify groups of genes that are correlated in a first group of microarrays and are not in a second one. Mitra et al. (2016), provided a more complex modeling strategy that is able to specify the differential dependence structure by using Bayesian graphical models. Moreover, Cai and Zhang (2016) presented, within a supervised framework, a way to estimate the differential correlation matrices of data belonging to two differentially expressed groups. Ieva et al. (2016), instead analyzed the problem from a different point of view, and focused on differences between global covariance structures of data belonging to two unknown groups, which are also identified. In particular, Ieva et al. (2016) centered their work on the specific case of a set of observations from two populations whose probability distributions have the same mean but differ in terms of covariances. This method can be applied both to the traditional case of random vectors and also to functional data as will be done in our work.

6.2 Covariance Operators for Functional Data

In this section we want to define the concept of covariance operator for functional data and its most important proprieties (for more details see Bosq (2000)).

Let $\mathcal{X}$ be a stochastic process tacking values in $L^2(I)$, with $I \subset \mathbb{R}$ a compact interval, having mean function $\mathbb{E}[\mathcal{X}] = \mu$ and such that $\mathbb{E}\|\mathcal{X}\|^2 < \infty$, where we denote by $\|\cdot\|$ the $L^2(I)$ norm induced by the scalar product $\langle\cdot,\cdot\rangle$. Without loss of generality we can assume $\mu = 0$ and define the following covariance operator $C \in \mathcal{L}(L^2(I); L^2(I))$:

$$\langle y, Cx\rangle = \mathbb{E}\left[\langle x, \mathcal{X}\rangle \langle y, \chi\rangle\right], \qquad \forall x, y \in L^2(I).$$

C is a compact, self-adjoint, positive semidefinite linear operator between $L^2(I)$ and $L^2(I)$. Therefor it can be decomposed into:

$$C = \sum_{k=1}^{\infty} \lambda_k e_k \otimes e_k,$$

where $\otimes$ indicates an outer product in $L^2(I)$, $\{e_k\}_{k=1}^{\infty}$ is the sequence of orthonormal eigenfunctions, forming a basis of $L^2(I)$ and $\{\lambda_k\}_{k=1}^{\infty}$ is the sequence of eigenvalues. We assume that eigenvalues are sorted in decreasing order, so that:

$$\lambda_1 \geq \lambda_2 \geq ... \geq 0.$$

By expressing $\mathcal{X}$ with respect to the eigenfunctions basis, $\mathcal{X} = \sum_{k=1}^{\infty} \xi_k e_k$, it holds

$$\lambda_k = \langle e_k, Ce_k\rangle = \mathbb{E}\left[\xi_k^2\right],$$

thus, the covariance operator is nuclear, meaning that

$$\mathbb{E}\|\mathcal{X}\|^2 = \sum_{k=1}^{\infty} \lambda_k = \sum_{k=1}^{\infty} \|\lambda_k\| < \infty.$$

C is also a Hilbert-Schmidt operator (see Bosq (2000)), since it holds:

$$\sum_{k=1}^{\infty} \lambda_k < \infty.$$

We equip the space of Hilbert-Schmidt operators with the Hilbert-Schmidt norm, defined as $\|\mathcal{U}\|_S^2 = \sum_{k=1}^{\infty} \lambda_k^2$, where $\{\lambda_k\}_{k=1}^{\infty}$ are the eigenvalues of $\mathcal{U}$. This is induced by the following scalar product:

$$\langle \mathcal{U}, \mathcal{V}\rangle_S = \sqrt{Tr(\mathcal{U} - \mathcal{V})(\mathcal{U} - \mathcal{V})^*}, \tag{6.1}$$

where $Tr(\cdot)$ denotes the trace operator, and $\mathcal{U}^*$ s the Hilbertian adjoint of $\mathcal{U}$, i.e.,

$$\langle \mathcal{U}(x), y\rangle = \langle x, \mathcal{U}^*(y)\rangle \qquad \forall x, y \in L^2(I).$$

The space of Hilbert-Schmidt operators on $L^2(I)$, endowed with the scalar product (6.1) and the associated norm, becomes a separable Hilbert space itself. Within this theoretic framework, a

natural definition of dissimilarity between Hilbert-Schmidt operators (among which are covariance operators) may be the Hilbert-Schmidt distance:

$$d(\mathcal{U},\mathcal{V}) = \|\mathcal{U}-\mathcal{V}\|_S^2 = \sum_{k=1}^{\infty}\eta_k^2,$$

where $\{\eta_k\}_{k=1}^{\infty}$ is the sequence of eigenvalues of $\mathcal{U}-\mathcal{V}$.

6.3 Covariance-based Clustering

In this subsection we face the problem of classifying observations belonging to two different populations. Let $\mathcal{X}$ and $\mathcal{Y}$ be stochastic process on $L^2(I)$ generated by the laws $\mathcal{P}_{\mathcal{X}}$ and $\mathcal{P}_{\mathcal{Y}}$. We suppose to have a set of N data composed by an equal number of observations deriving from two families. Let call D the data set, $D = \{X_1,\dots,X_K,Y_1,\dots,Y_K\}$ with $K = N/2$ and where $\{\mathcal{X}\}_{i=1}^K$, $\{\mathcal{Y}\}_{i=1}^K$ are i.i.d and follow respectively $\mathcal{P}_{\mathcal{X}}$ and $\mathcal{P}_{\mathcal{Y}}$.

We introduce the following quantities:

$$\mu_1 = \mathbb{E}[X_i], \qquad C_1 = \mathbb{E}[X_i\otimes X_i], \qquad \forall i = 1,\dots,K,$$

$$\mu_2 = \mathbb{E}[Y_j], \qquad C_1 = \mathbb{E}[Y_j\otimes Y_j], \qquad \forall j = 1,\dots,K.$$

Let us consider the vector of indexes of units constituting the two populations in D :

$$I^{(0)} = \big(\overbrace{1,2,\cdots,K}^{I_1^{(0)}},\overbrace{K+1,\cdots,N}^{I_2^{(0)}}\big),$$

which is unique, provided we do not distinguish among permutations of sub-intervals $I_1^{(0)}$ and $I_2^{(0)}$. In the following we shall consider recombinations of these indexes into two subsets:

$$I^{(i)} = \big(I_1^{(i)};I_1^{(i)}\big), \qquad i\in\{1,\cdots,N_C\},$$

where $I^{(i)}$ denotes the i-th combination out of $N_C = \binom{N}{K}$, however enumerated.

We denote with $\widehat{\mu}_1^{(i)},\widehat{\mu}_2^{(i)}$ and $\widehat{C}_1^{(i)},\widehat{C}_2^{(i)}$ the sample estimators of means and covariance operators induced by the dataset subdivision. We point out that, when $i=0$, we recover the estimators of μ_1,μ_2 and C_1, C_2 and for this reason we rename the latter quantities as μ_1^0, μ_2^0 and C_1^0, C_2^0.

An important hypothesis is that observations drawn from families $\mathcal{P}_{\mathcal{X}}$ and $\mathcal{P}_{\mathcal{Y}}$ constituting the dataset D may be distinguish on the basis of their covariances, but not of their means. It means that $\mu_1^0 = \mu_2^0$ and $C_1^0 \neq C_2^0$, and therefore $\big\|\mu_1^{(0)}-\mu_2^{(0)}\big\| = 0$ and $d(C_1^0,C_2^0)\gg 0$.

As consequence of this assumption we conveniently center data and assume they have zero means.

In order to illustrate this clustering method, let us consider a situation where the original data set has been split according to a vector of indexes $I^{(i)} = \Big[I_1^{(i)};I_2^{(i)}\Big]$. For the sake of simplicity, let us encode this through the binary variables $w_{j,g} = \mathbb{1}(X_j\in I_g^{(i)}), j=1,\cdots,K$, $g=1,2$ and

$v_{j,g} = \mathbb{1}(Y_j \in I_g^{(i)}), j = 1, \cdots, K,\ g = 1, 2$. In other words, such variables express the fact that observation j from the original population $\mathcal{X}$ or $\mathcal{Y}$ belongs to the first ($I_1^{(i)}$) or second ($I_2^{(i)}$) group into which data are split. According to the setting previously introduced, it is :

$$K = \sum_{g=1}^{2}\sum_{j=1}^{K} w_{j,g}, \qquad K = \sum_{g=1}^{2}\sum_{j=1}^{K} v_{j,g},$$

$$K = \sum_{j=1}^{K} w_{j,1} + \sum_{j=1}^{K} v_{j,1}, \qquad K = \sum_{j=1}^{K} w_{j,2} + \sum_{j=1}^{K} v_{j,2}.$$

Then, we can re-write the sample covariances $\widehat{C}_1^{(i)}$ and $\widehat{C}_2^{(i)}$ as:

$$\begin{aligned} \widehat{C}_1^{(i)} &= \frac{\sum_{j=1}^{K} w_{j,1} X_j \otimes X_j + \sum_{k=1}^{K} v_{k,1} Y_k \otimes Y_k}{K}, \\ \widehat{C}_2^{(i)} &= \frac{\sum_{j=1}^{K} w_{j,2} X_j \otimes X_j + \sum_{k=1}^{K} v_{k,2} Y_k \otimes Y_k}{K}. \end{aligned} \tag{6.2}$$

The difference $\widehat{C}_1^{(i)} - \widehat{C}_2^{(i)}$ is:

$$\widehat{C}_1^{(i)} - \widehat{C}_2^{(i)} = \frac{1}{K}\sum_{j=1}^{K}(w_{j,1} - w_{j,2}) X_j \otimes X_j + \frac{1}{K}\sum_{k=1}^{K}(v_{k,1} - v_{k,2}) Y_k \otimes Y_k,$$

using the definition of distance between covariance operators, we can write:

$$\begin{aligned} K^2 \big\| \widehat{\mathcal{C}}_1^{(i)} - \widehat{\mathcal{C}}_2^{(i)} \big\|_S^2 &= \left\| \sum_{j=1}^{K}(w_{j,1} - w_{j,2}) X_j \otimes X_j \right\|_S^2 + \left\| \sum_{k=1}^{K}(v_{k,1} - v_{k,2}) Y_k \otimes Y_k \right\|_S^2 \\ &\quad + 2\sum_{j=1}^{K}\sum_{k=1}^{K}(w_{j,1} - w_{j,2})(v_{k,1} - v_{k,2}) \langle X_j, Y_k \rangle^2 \\ &= \sum_{k=1}^{K} \big\| X_j \otimes X_j \big\|_S^2 + 2\sum_{j<k}(w_{j,1} - w_{j,2})(w_{k,1} - w_{k,2}) \langle X_j, X_k \rangle^2 \\ &\quad + \sum_{k=1}^{K} \big\| Y_j \otimes Y_j \big\|_S^2 + 2\sum_{j<k}(v_{j,1} - v_{j,2})(v_{k,1} - v_{k,2}) \langle Y_j, Y_k \rangle^2 \\ &\quad + 2\sum_{j=1}^{K}\sum_{k=1}^{K}(w_{j,1} - w_{j,2})(v_{k,1} - v_{k,2}) \langle X_j, Y_k \rangle^2 \end{aligned}$$

Let us now call $\delta_{j,k}^X = (w_{j,1} - w_{j,2})(w_{k,1} - w_{k,2})$, $\delta_{j,k}^Y = (v_{j,1} - v_{j,2})(v_{k,1} - v_{k,2})$ and $\delta_{j,k}^{XY} = (w_{j,1} - w_{j,2})(v_{k,1} - v_{k,2})$. Now, it is $\delta_{j,k}^X = +1$ if observations X_j and X_k are assigned to the same group, while on the contrary it is $\delta_{j,k}^X = -1$. The same applies for $\delta_{j,k}^Y$ with Y_j and Y_k. Finally, $\delta_{j,k}^{XY} = +1$ if X_j and Y_k are assigned to different groups, otherwise $\delta_{j,k}^{XY} = -1$. From this brief

analysis, it is clear that distance between covariances operator increases when two observations of populations $\mathcal{X}$ or $\mathcal{Y}$ are both assigned to the same group, or when two observations of the opposite populations $\mathcal{X}$ and $\mathcal{Y}$ are assigned to different groups.

Using the expressions (6.2) for the sample covariances $\widehat{\mathcal{C}}_1^{(i)}$ and $\widehat{\mathcal{C}}_2^{(i)}$ and computing their expected values, we obtain:

$$\mathbb{E}\left[\widehat{\mathcal{C}}_1^{(i)}\right] = \mathcal{C}_1^{(i)} = \frac{\sum_{j=1}^{K} w_{j,1}}{K}\mathcal{C}_1\frac{\sum_{j=1}^{K} v_{j,1}}{K}\mathcal{C}_2,$$

$$\mathbb{E}\left[\widehat{\mathcal{C}}_2^{(i)}\right] = \mathcal{C}_2^{(i)} = \frac{\sum_{j=1}^{K} w_{j,2}}{K}\mathcal{C}_1\frac{\sum_{j=1}^{K} v_{j,3}}{K}\mathcal{C}_2,$$

now, denoting by $N_{1,2} = \sum j = 1^K w_{j,2}$ and considering the relations among the variables $w_{j,g}$ and $v_{j,g}$ we get:

$$d(\mathcal{C}_1^{(i)}, \mathcal{C}_2^{(i)}) = \left\|\mathbb{E}\left[\widehat{\mathcal{C}}_1^{(i)} - \left[\widehat{\mathcal{C}}_2^{(i)}\right]\right]\right\|_S^2 = \left(1 - 2\frac{N_{1,2}}{K}\right)^2 \left\|\mathcal{C}_1 - \mathcal{C}_2\right\|_S^2, \tag{6.3}$$

noting that the maximum distance between covariances is obtained when the groupings coincide with the original but unknown indexing of the dataset.

Then, a natural way to find the true indexing can be the recombination of data in two groups maximising the distance function. It is equivalent to solve the optimization problem:

$$[I_1^*; I_2^*] = \arg\max_{i \in R_C}\left\{d\left(\mathcal{C}_1^{(i)}, \mathcal{C}_3^{(i)}\right)\right\}, \quad R_C = \{1, \ldots, N_C\}.$$

Due to the symmetry of the optimization problem we have two optimal solutions that are $(I_1^*, I_2^*) = (I_1^0, I_2^0)$ or $(I_1^*, I_2^*) = (I_2^0, I_1^0)$.

Practically, only approximate estimates of $\mathcal{C}_1^{(i)}$ and $\mathcal{C}_2^{(i)}$ are available, thus we must recast the previous problem into :

$$\left[\widehat{I_1}^*; \widehat{I_2}^*\right] = \arg\max_{i \in R_C}\left\{d\left(\widehat{\mathcal{C}}_1^{(i)}, \widehat{\mathcal{C}}_2^{(i)}\right)\right\}, \quad R_C = \{1, \ldots, N_C\}. \tag{6.4}$$

In general $\widehat{I_1}^*$ and $\widehat{I_2}^*$ may differ from I_1^0 and I_2^0, since they are determined estimating the covariance operators.

6.4 Max-Swap Algorithm

In order to solve the problem (6.4), it would be necessary to test each N_C ($N_C = \binom{N}{K}$) recombinations of indexes to find $\widehat{I_1}^*$ and $\widehat{I_2}^*$. Of course, the number of tests to be performed undergoes a combinatorially-fast growth, as N increase. In consequence of that, the naive approach of performing an exhaustive search in the set of recombinations is not feasible.

To solve this problem Ieva et. al (2016) proposed a greedy algorithm, called *Max-Swap algorithm*, with the aim to reduce the complexity of solving the problem (6.4).

The main idea is to interpret $d(\widehat{\mathcal{C}}_1^{(i)}, \widehat{\mathcal{C}}_2^{(i)})$ as an objective function of i, and, starting from an initial guess $(I_0(1), I_0(2))$, to iteratively increase it by allowing exchanges of units between two

groups, such that each group discards and receives an equal number of units. Ieva et al. (2016) suggested to choose the swapping units in such a way that the distance of estimated covariance operators at the next step is strictly higher than the previous one and, heuristically, the highest possible. Convergence is reached when no further swap can increase that distance.

An important aspect of this algorithm is that ensures that convergence always happens, at least to a local maximum of $d(\widehat{\mathcal{C}}_1^{(i)}, \widehat{\mathcal{C}}_2^{(i)})$ (for the proof see Ieva et al. (2016)).

6.5 Shrinkage estimation of covariance

Ieva et al. (2016) proposed in their work an alternative estimator to use in Max-Swap algorithm for the covariance operators: *the shrinkage estimator*. It is better conditioned and in some circumstances achieves lower MSE then the sample covariance. Below we see how to obtain shrinkage estimator.

Let us consider a generic family of functional data $\mathcal{X} \sim \mathcal{P}_{\mathcal{X}}$, such that $\mathbb{E}[\mathcal{X}] = 0$, $\mathbb{E}\|\mathcal{X}\|^2 < \infty$. Called $\mathcal{C}$ its covariance, we want to estimate it with $\widehat{\mathcal{C}}$ such that:

$$MSE_S(\widehat{\mathcal{C}}) := \mathbb{E}\|\widehat{\mathcal{C}} - \mathcal{C}\|_S^2.$$

This method is interested in solving the follow estimation problem:

$$\widehat{\mathcal{C}}^* = \underset{\widehat{\mathcal{C}}}{\operatorname{argmin}}\, MSE_S(\widehat{\mathcal{C}}) = \underset{\widehat{\mathcal{C}}}{\operatorname{argmin}}\, \mathbb{E}\|\widehat{\mathcal{C}} - \mathcal{C}\|_S^2 \tag{6.5}$$

where the minimum is sought among all possible estimators $\widehat{\mathcal{C}}$ of $\mathcal{C}$. We point out that in MSE_S is used the distance defined in (6.3).

Remember that we are working on a functional dataset, let us indicate by X_i the i-th (out of N) sample realization of process $\mathcal{X}$, i.e:

$$X_i = (X_i(t_j))_{j=1}^P, \qquad I^h = [t_1, \dots, t_P],$$

where we have imagined that the grid I^h to be regularly spaced, i.e, $t_{j+1} - t_j = h$ for $\forall j = 1, \dots, P-1$ for $t_{j+1} - t_j = h$ for $\forall j = 1, \dots, P-1$.

It is clear that, within this habit, covariance estimators of $\mathcal{C}$ are discrete, matrix-type approximations obtained starting from pointwise observations X_i. For instance, standard sample covariance estimator for zero-mean data is:

$$\mathbf{S} = \frac{1}{N} \sum_{i=1}^{N} X_i X_i^T.$$

If we denote the true discrete covariance structure related to each X_i by $\mathbf{C}$ the discrete version of the problem (6.5) is:

$$\mathbf{C}^* = \underset{\widehat{\mathcal{C}}}{\operatorname{argmin}}\, \mathbb{E}\|\widehat{\mathbf{C}} - \mathbf{C}\|_F^2 \tag{6.6}$$

where the minimum is sought inside the set of symmetric and positively defined matrix-type estimators of dimension P. We point out that the subscript F in (6.6) indicates the Frobenius norm, that is the finite-dimensional counterpart of the Hilbert-Schmidt norm for operators.

When the sample size N is low compared to the number of features P, sample covariance may loose in accuracy, meaning that the actual estimate might be quite distant from the true covariance $\mathbf{C}$. A typical remedy to the poor performances of sample covariance, often used in the setting of Large P - Small N problems, is to replace it with a biased, shrinkage estimator. Shrinkage estimation has been explicitly applied to the context of large covariance matrices in Ledoit and Wolf (2003), (2004) and Schafer and Strimmer (2005), turning out in a sufficiently lightweight procedure. In those works, authors started from problem (6.6) and built an estimator that is asymptotically more accurate and better conditioned than sample covariance. Ledoit and Wolf (2004), considered the class of linear shrinkage estimators of the form:

$$\widehat{\mathbf{C}} = \mu\gamma\mathbf{I} + (1-\gamma)\mathbf{S},$$

where $\mathbf{I}$ is $P \times P$ identity matrix and $\gamma \in [0,1]\,, \mu \in \mathbb{R}^+$ and $\mathbf{S}$ is the sample covariance estimator. Obviously, the class contains the sample covariance estimator itself. Then (6.6) is solved with respect to the optimal values of μ and γ:

$$(\mu^*, \gamma^*) = \arg\min_{\mu,\gamma} \mathbb{E} \frac{\left\|\mathbf{C} - \mu\gamma\mathbf{I} - (1-\gamma)\mathbf{S}\right\|_F^2}{P}. \tag{6.7}$$

If we introduce the quantities:

$$\alpha^2 = \frac{\left\|\mathbf{C} - \mu\mathbf{I}\right\|_F^2}{P}, \quad \beta^2 = \frac{\mathbb{E}\left\|\mathbf{S} - \mathbf{C}\right\|_F^2}{P} \quad \delta^2 = \frac{\mathbb{E}\left\|\mathbf{S} - \mu\mathbf{I}\right\|_F^2}{P},$$

and note that these are subjected to $\alpha^2 + \beta^2 = \delta^2$, it is possible to perform the explicit minimization (6.7) problem. The expressions of μ^* and γ^* are:

$$\mu^* = \frac{\langle \mathbf{C}, \mathbf{I} \rangle_F}{P} = \frac{Tr(\mathbf{C})}{P}, \quad \gamma^* = \frac{\beta^2}{\delta^2}$$

where we have used $\mu = \mu^*$ in the computation of δ. The desired shrinkage estimators becomes:

$$\mathbf{S}^* = \mu^* \frac{\beta^2}{\delta^2}\mathbf{I} + \frac{\alpha^2}{\delta^2}\mathbf{S}$$

$\mathbf{S}^*$ depends on the unknown exact covariance matrix $\mathbf{C}$, even though only through four scalar functions. In Ledoit and Wolf, 2004 authors solved this problem by proposing the following estimators for α, β, δ and μ^*:

$$\widehat{\mu^*} = \frac{Tr(\mathbf{S})}{P}, \quad \widehat{\delta}^2 = \frac{\left\|\mathbf{S} - \mu^*\mathbf{I}\right\|_F^2}{P},$$
$$\widehat{\beta}^2 = \min\left(\widehat{\delta}^2; \frac{1}{N^2}\sum_{k=1}^{N} \frac{\left\|X_k X_k^T - \mathbf{S}\right\|_F^2}{P}\right),$$

and $\widehat{\alpha}^2 = \widehat{\delta}^2 - \widehat{\beta}^2$.

Then, the actual shrinkage estimator is:

$$\widehat{\mathbf{S}}^* = \widehat{\mu}^* \frac{\widehat{\beta}^2}{\widehat{\delta}^2}\mathbf{I} + \frac{\widehat{\alpha}^2}{\widehat{\delta}^2}\mathbf{S}$$

In Ledoit and Wolf, 2004 authors showed how estimates $\widehat{\mathbf{S}}^*$ are consistent. It converges to the exact values in quadratic mean, under the general asymptotic limits of P and N, i.e.,when both P and N are allowed to go to infinity but there exists a $c \in \mathbb{R}$ independent on N such that $\frac{P}{N} < c$ (see Ledoit and Wolf, 2004). Moreover, estimator $\widehat{\mathbf{S}}^*$ is an asymptotically optimal linear shrinkage estimator for covariance matrix $\mathbf{C}$ with respect to quadratic loss.

6.6 Results of Covariance-based Clustering

In this section we apply the covariance-based clustering to the concentrations of O_2Hb and HHb for both Up and Down data.

We decide to employ the clustering algorithm not to all entirely measuring concentration of O_2Hb and HHb for each channel but to functional median of each channel for the reasons explained in the Subsection 5.5. Thus, the functional data consists of 30 sampling, each one for a channel, along $40s$ of functional median of O_2Hb and HHb concentration of hemoglobin. Our clustering purpose is to recognize which channels have a widely fluctuations about their mean. Remembering that the aim of our work is to detect activated and non activated areas of the brain, we suppose that this activation is reflected in difference in covariance operators more than in difference in the mean level of O_2Hb and HHb concentration.

To apply the algorithm the set of $N = 30$ signals is subdivided into two groups of $K = 15$ and the sampling rate is $1s$ so that $P = 40$.

We run the Max-Swap clustering algorithm on these data to perform clustering, both using $\mathbf{S}$ and $\widehat{\mathbf{S}}^*$ estimators, finding equal partitions of initial data. The results are shown in Figures 6.1 - 6.8 and highlight how the algorithm is able to answer to our request, in fact, we can clearly detect two clusters of functions that are well distinguishable in terms of their different variability. We can interpreter these two clusters as activated and non activated channels. We point out that these results are in agreement with those obtained by Bonomini et al. (2015).

In the plots below we can also notice how the results of the clustering algorithm are similar using the Shrinkage estimator and the sample covariance.

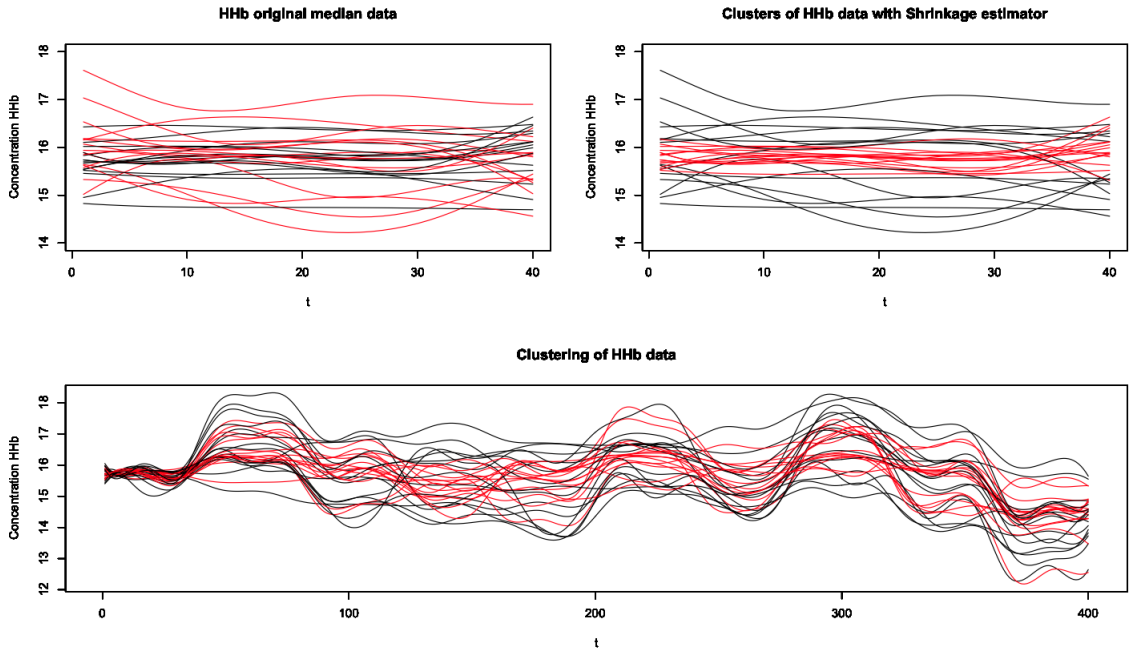

Figure 6.1: Deoxygenated hemoglobin's concentration data Down. In the top left panel are represented the preprocessed data on which the clustering using Shrinkage estimator is carried out. In the top right panel is shown the output. In the bottom panel are represented the clustering obtained on the entirely data

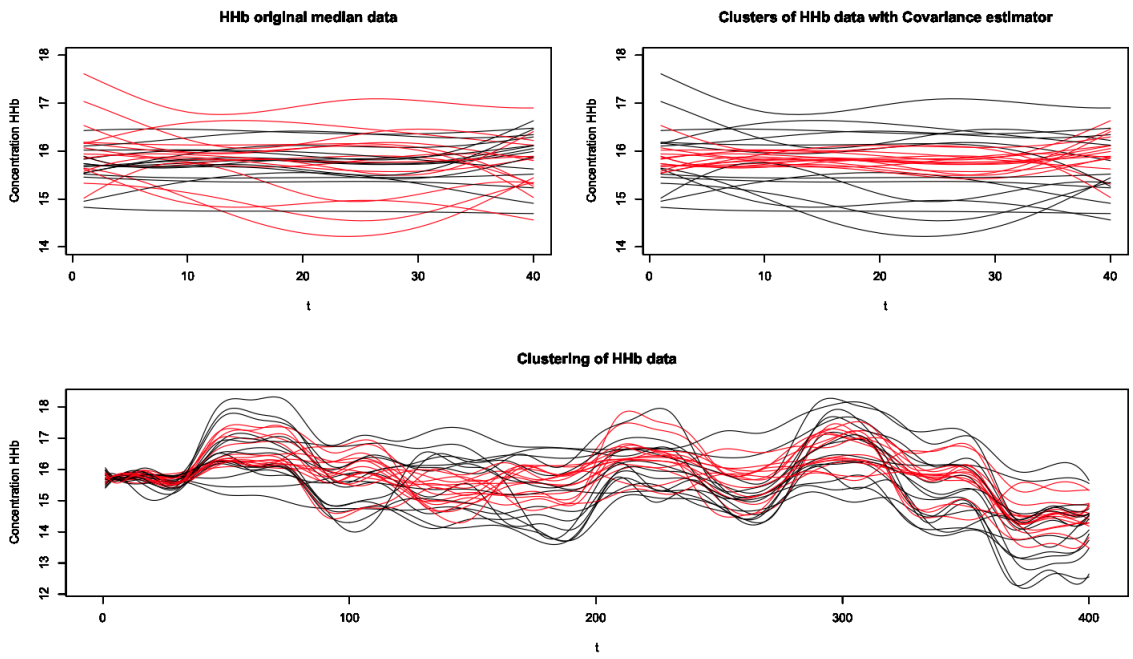

Figure 6.2: Deoxygenated hemoglobin's concentration data Down. In the top left panel are represented the preprocessed data on which the clustering using Covariance estimator is carried out. In the top right panel is shown the output. In the bottom panel are represented the clustering obtained on the entirely data

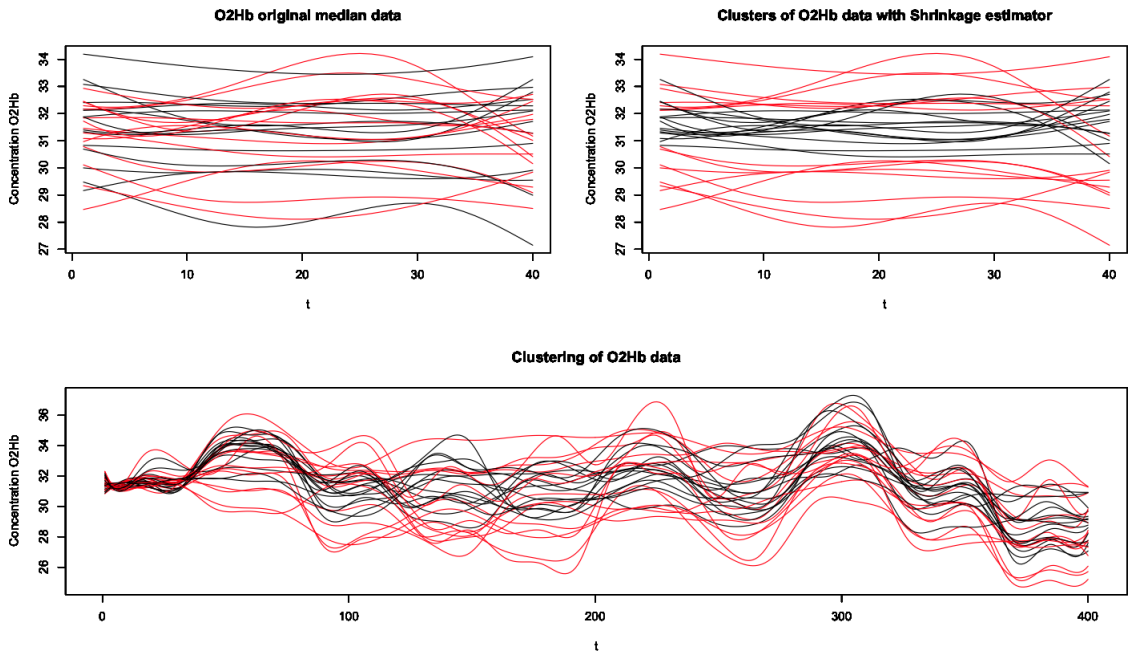

Figure 6.3: Oxygenated hemoglobin's concentration data Down. In the top left panel are represented the preprocessed data on which the clustering using Shrinkage estimator is carried out. In the top right panel is shown the output. In the bottom panel are represented the clustering obtained on the entirely data

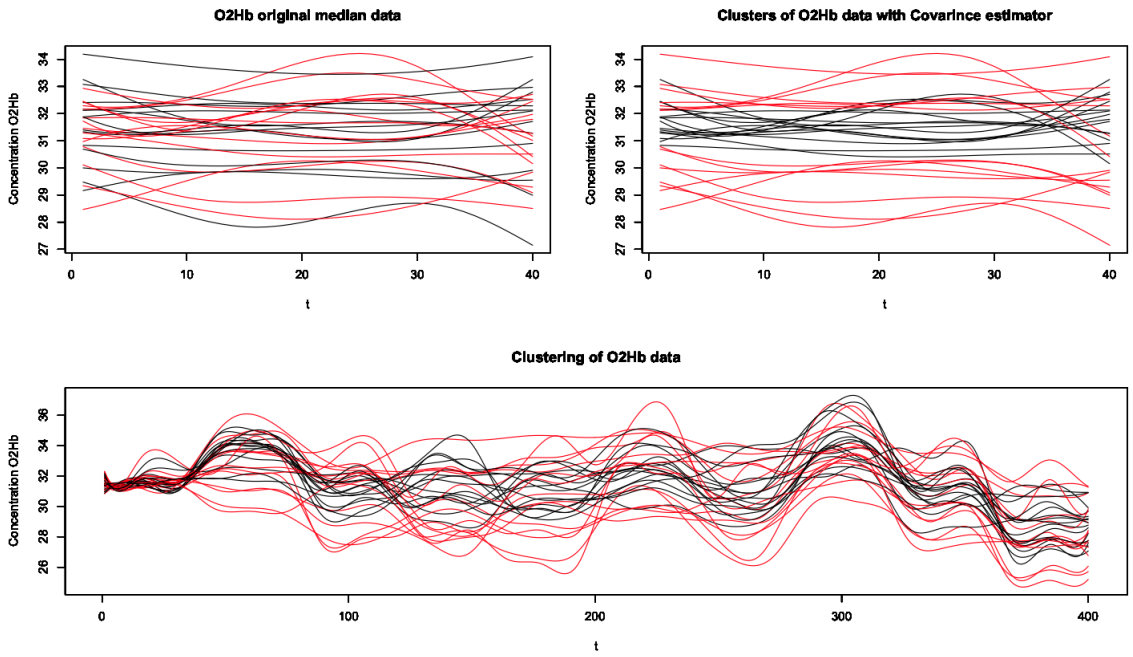

Figure 6.4: Oxygenated hemoglobin's concentration data Down. In the top left panel are represented the preprocessed data on which the clustering using Covariance estimator is carried out. In the top right panel is shown the output. In the bottom panel are represented the clustering obtained on the entirely data

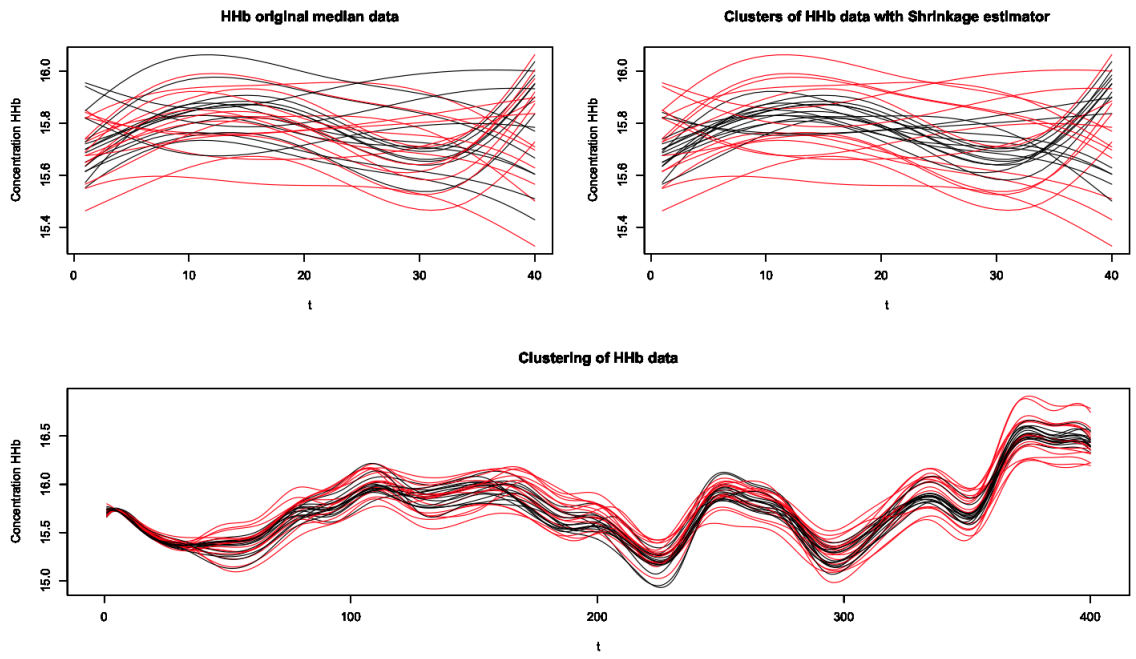

Figure 6.5: Deoxygenated hemoglobin's concentration data Up. In the top left panel are represented the preprocessed data on which the clustering using Shrinkage estimator is carried out. In the top right panel is shown the output. In the bottom panel are represented the clustering obtained on the entirely data

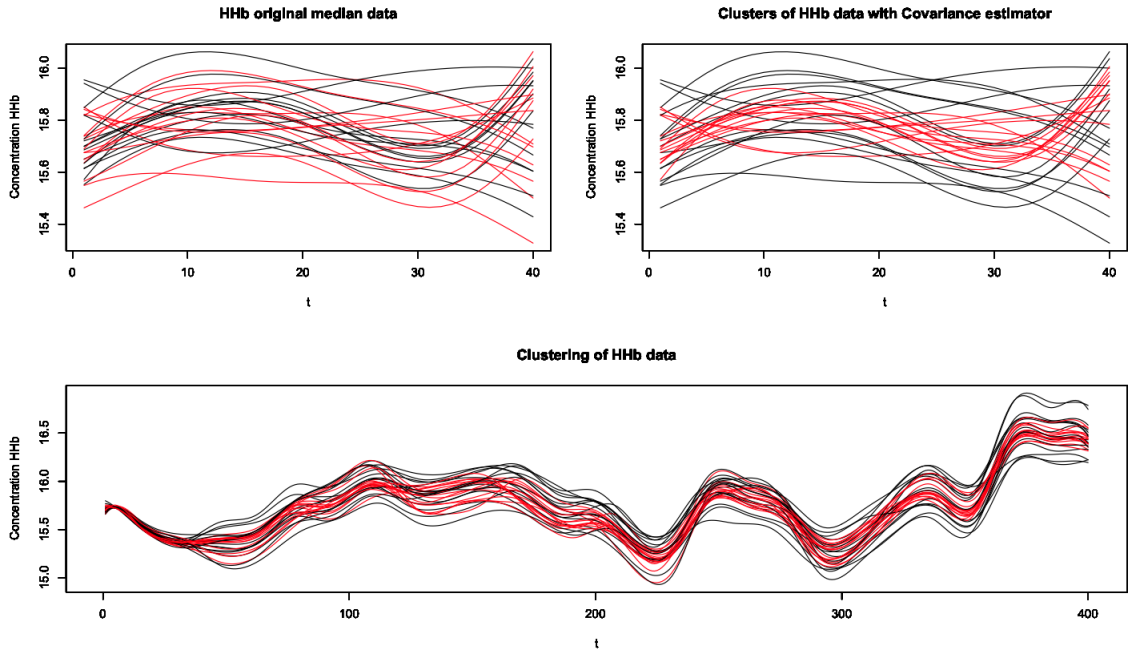

Figure 6.6: Deoxygenated hemoglobin's concentration data Up. In the top left panel are represented the preprocessed data on which the clustering using Covariance estimator is carried out. In the top right panel is shown the output. In the bottom panel are represented the clustering obtained on the entirely data

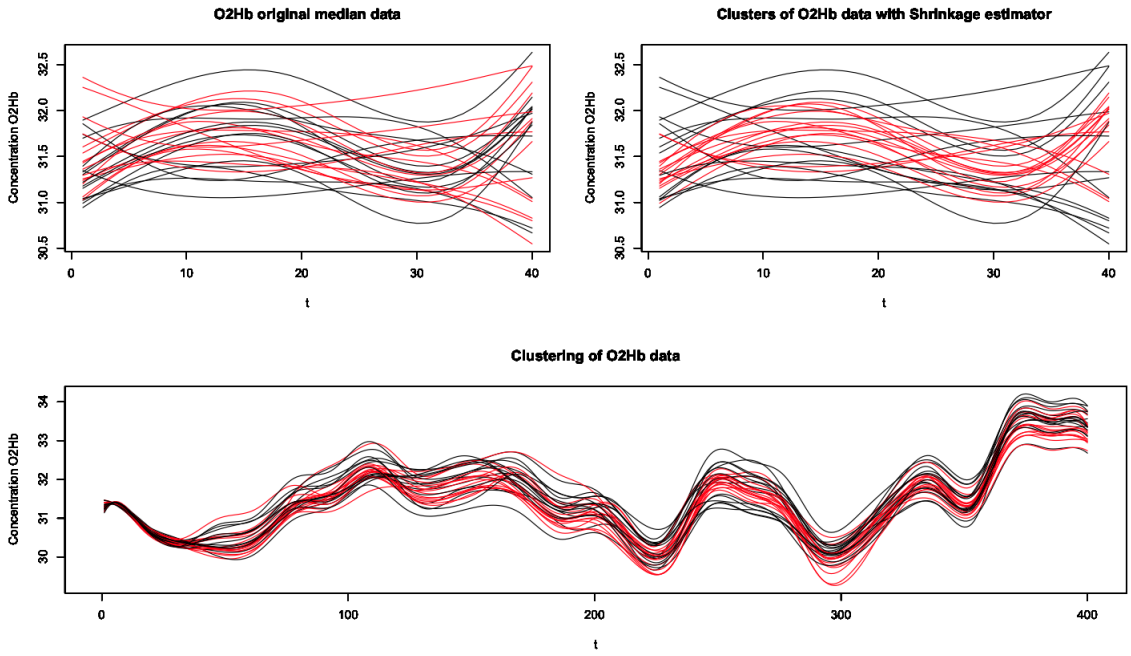

Figure 6.7: Oxygenated hemoglobin's concentration data Up. In the top left panel are represented the preprocessed data on which the clustering using Shrinkage estimator is carried out. In the top right panel is shown the output. In the bottom panel are represented the clustering obtained on the entirely data

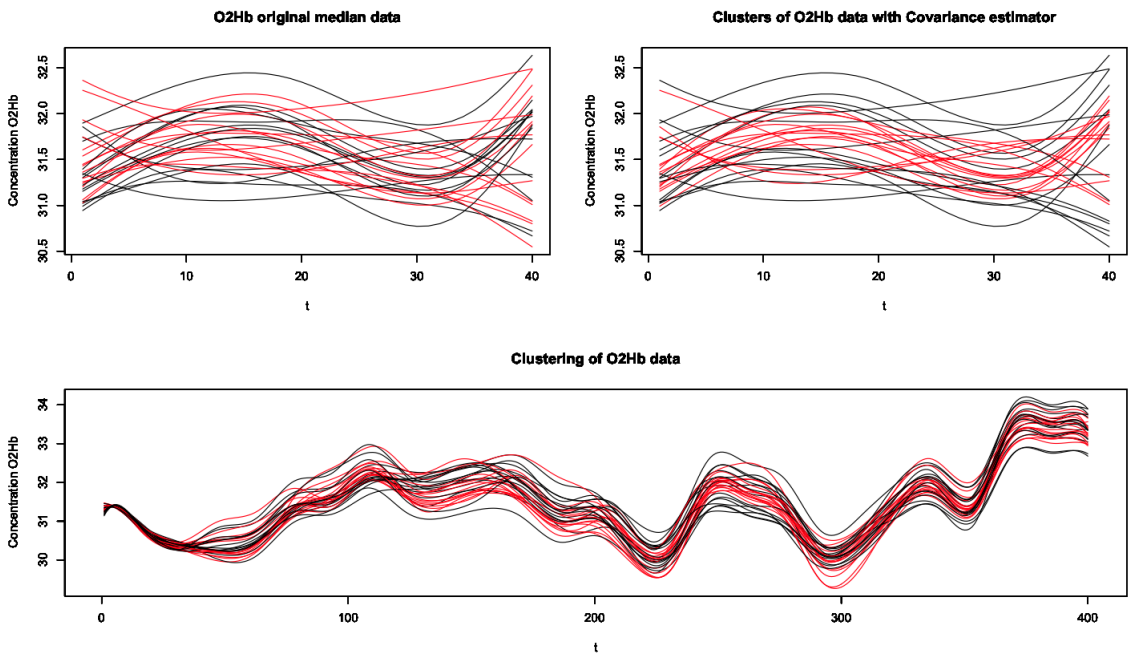

Figure 6.8: Oxygenated hemoglobin's concentration data Up. In the top left panel are represented the preprocessed data on which the clustering using Covariance estimator is carried out. In the top right panel is shown the output. In the bottom panel are represented the clustering obtained on the entirely data

Chapter 7

Conclusions

The present work describes and faces the problem of fNIRS data analysis. Our study is based on the one conducted by Bonomini et al. (2015), where the authors want to classify the cerebral areas of the brain, called channels, in activated and not-activated ones.

The main purpose of this thesis is to repeat this classification using a functional statistical approach.

We decide to use a functional statistical approach to the problem because data derived by time series are conformed to be performed as a functional data. In our case, we consider the concentrations of hemoglobin for a lapse of time of 400s. The basic philosophy of functional data analysis is to think about observed data functions as single entities, rather than as a sequence of individual observations. In order to obtain functional data we are interested in smoothing data. To represent functional data we introduce two basis function: Fourier basis and Bspline basis. Fourier basis is especially useful to describe periodic data. In our study, in fact, we can see data as ten realization of the same event. Bspline basis is a polynomial basis that is used for its fast computation and for its great flexibility.

Once we obtain functional data we concentrate the analysis on clustering methods.

The first method that we adopt is a functional K-means algorithm. Assuming that $F_i(t) : T \to \mathbb{R}$ is the functional data for the channel i, where $T = (0, 400)$ represents the time domain, in order to evaluate the distance between functions, we introduce two distances: the $L^2(T, \mathbb{R})$ and $H^1(T, \mathbb{R})$ distance. Our study is based on the analysis of K-means using both distances and the two different basis expansion seen before.

K-means is a partitioning method, for this reason we need at first a preliminary analysis to discover which is the best number of clusters to fit the data. One of the most common method to detect the best number of partitioning, is silhouette method. Using this last method we set the number of clusters equal to two.

As the fact that we get similar results for the two different distances and for the two different basis, for the remaining analysis we fix the distance $L^2(T, \mathbb{R})$ and the Bspline basis.

The K-means algorithm is also used to classify two interesting quantities from a medical viewpoint: the tHb (total hemoglobin concentration) and the StO_2 (tissue oxygen saturation). In all cases we recover that the activated channel discovered in Bonomini et al. (2015) are gathered together. We find the same cluster assignment for both Down and Up data but the clustering assignment for Up data seems to be artificial because the distance between medoids is small.

We investigate this issue computing the functional median for each cluster in the ten subintervals

of the experiment.

It is interesting to study clusters assignment applied to functional median because it represents a robust method to summarize and estimate the informations of a functional data. Using the K-means algorithm applied to the functional median of each channel we discover that activated channels for Bonomini et al. (2015) are still in the same cluster.

In the last part of our work we propose a covariance-based clustering for the functional median of data. This way of clustering is needed because the K-means algorithm classify channels taking advantage of the difference in signal's amplitude while now we want to investigate the differences in covariance structures of the data.

To solve this problem Ieva et. al (2016) proposed a greedy algorithm, called Max-Swap algorithm. This algorithm allows to divide data in two groups maximising the distance between the estimated covariance operators of the groups. We discuss the Max-swap algorithm using the sample covariance estimator and the shrinkage estimator. The last one it is better conditioned and in some circumstances achieves lower MSE then the sample covariance. Anyway, using the max-swap algorithm for our data the clusters assignment of channels is the same for both estimators.

We obtain that activated channels resulting from the K-mens algorithm are clustered in the same group.

Finally, we can conclude that, also using clustering method for functional data, we recover the same activated channels resulting from Bonomini et al. (2015).

Ringraziamenti

Desidero esprimere la mia gratitudine a tutti quelli che mi hanno aiutato nella stesura della tesi.

Ringrazio anzitutto la Prof.ssa Anna Maria Paganoni, la relatrice, ed il Prof. Alessandro Torriceli, correlatore, per avermi seguito, per l'aiuto sempre attento e preciso che hanno saputo darmi, per la competenza con cui mi hanno guidato ed anche per la pazienza dimostrata nei miei confronti durante questo lungo lavoro di tesi. Sono particolarmente grato alla Prof.ssa Anna Maria Paganoni per avermi fatto conoscere la statistica funzionale.

Ringrazio anche la Prof.ssa Francesca Ieva ed i Proff. Rebecca Re, Davide Contini e Lorenzo Spinelli del Dipartimento di Fisica del Politecnico di Milano che, con la loro disponibilità e professionalità, hanno reso possibile la realizzazione di questo lavoro.

G. D. F.

Bibliography

[1] Benaron D. A., Hintz S. R., Villringer A., Boas D., Kleinschmidt A., Frahm J., Hirth C., Obrig H., Van Houten J. C., Kermit E. L., Cheong W.-F., Stevenson D. K. (2000) Noninvasive functional imaging of human brain using light. *Journal of Cerebral Blood Flow and Metabolism*, vol. 20, 3: 469-477.

[2] Boas D. A., Dale A. M., Franceschini M. A. (2004) Diffuse optical imaging of brain activation: approaches to optimizing image sensitivity, resolution and accuracy. *Neuroimage*, 23, S275-S288.

[3] Bosq D. (2000) Linear Processes in Function Spaces: Theory and Applications, volume 149 of *Lecture Notes in Statistics*. Springer.

[4] Bonomini V., Zucchelli L., Re R., Ieva F., Spinelli L., Contini D., Paganoni A. M., Torricelli A. (2015) Linerar regression Models and k-means clustering for statistical analysis of fNIRS data. *Biomedical Optics Express*, Vol. 6, **2**, 615-630.

[5] Cai T. T., Zhang A. (2016) Inference for high-dimensional differential correlation matrices. *Journal of Multivariate Analysis*, 143:107-126.

[6] Chaudhuri, P. (1996) On a Geometric Notion of Quantiles for Multivariate Data. *Journal of the American Statistical Association*, 91, 862–872.

[7] Contini D., Spinelli L., Caffini M., Cubeddu R., Torricelli A. (2009) A multichannel time-domain brain oximeter for clinical studies. Proceedings of SPIE-OSA European Conference on Biomedical Optics.

[8] Contini D., Torricelli A., Pifferi A., Spinelli L., Paglia F., Cubeddu R. (2006) Multi-channel time-resolved system for functional near infrared spectroscopy. *Optics express*, col.14, 12: 5418-5432.

[9] Cooper R. J., Selb J., Gagnon L., Phillip D., Schytz H. W., Iversen H. K., Ashina M., Boas D. A. (2012) A systematic comparison of motion artifact correction techniques for functional near-infrared spectroscopy. *Frontiers in neuroscience*, vol. 6, 147: 1–10.

[10] Cui X., Bray S., Reiss A. L. (2010) Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlatio between oxygenated and deoxygenated hemoglobin dynamics. *Neuroimage*, 49(4): 3039.

[11] Ferrari M., Quaresima V. (2012) A brief review in the history of human functional near infrared spectroscopy (fNIRS) development and field of application. *Neuroimage* 63, 921-933.

[12] Fraiman, R., and Meloche, J. (1999) Multivariate L-Estimation. Test, 8, 255–317.

[13] Fraiman, R., and Muniz, G. (2001) Trimmed Means for Functional Data. Test, 10, 419–440.

[14] Franceschini M. A., Boas D. A. (2004) Noninvasive measurement of neuronal activity with near-infrared optical imaging. *NeuroImage*, vol. 21, 1: 372-386.

[15] Franceschini M. A., Joseph D. K., Huppert T. J., Diamond S. G. (2006) Diffuse optical imaging of the whole head. *Journal of Biomedical Optics*, vol. 11, 5: 1-22.

[16] Gagnon L., Cooper J. R., Yucel M., A., Perdue K. L., Greve D. N., Boas D. A. (2012). Short separation channel location impact the performance of short channel regression in NIRS. *Neuroimage*, 59, 2518-2528.

[17] Gagnon L., Perdue K., Greve D. N., Goldenholz D., Boas D. A. (2011) Improved recovery of the hemodynamic response in diffuse optical imaging using short optode separations and state-space modeling. *Neuroimage*, vol. 56, 3: 1362-1371.

[18] Gregg N. M., White B.R., Zeff B.W., Berger A. J., Culver J. P. (2010) Brain specificity of diffuse optical imaging: improvements from superficial signal regression and tomography. *Neurogenetics*, 2 (14), 1-8.

[19] Hoshi, Y., Tsou, B., Billock, V., Tanosaki, M., Iguchi, Y., Shimada, M., Shinba, T., Yamada, Y., Oda I. (2003) Spatiotemporal characteristics of hemodynamic changes in the human lateral prefrontal cortex during working memory tasks. *Neuroimage*, 20(3), 1493–1504.

[20] Huppert T. J. (2016) Commentary on the statistical properties of noise and its implication on general linear models in functional near-infrared spectroscopy. *Neurophotonics*, **3**(1).

[21] Huppert T. J., Solomon G. D., Franceschini M. A., Boas D. A. (2009) HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *NIH Public Access*, 48(10): D280-298.

[22] Ieva F., Paganoni A. M. (2013) Depth measures for multivariate functional data. *Communications in Statitics-Theory and Method*, 42:7, 1265-1276.

[23] Ieva F., Paganoni A. M., Pigoli D., Vitelli V. (2013) Multivariate functional clustering for the morphological analysis of electrocardiograph curves. *Journal of Royal Statistical Society*, **62**, 401-418.

[24] Ieva F., Paganoni A. M., Tarabelloni N. (2016) Covariance-based Clustering in Multivariate and Functional Data Analysis. *Journal of Machine Learning Research*, **17** 1-21.

[25] Izzetoglu M., Devaraj A., Bunce S., Onaral B. (2005) Motion artifact cancellation in NIR spectroscopy using wiener filtering. *IEEE Transaction on Biomedical Engineering*, vol. 52, no.5.

[26] Izzetoglu M., Chitrapu P., Bunce S., Onaral B. (2010) Motion artifact cancellation in NIR spectroscopy using discrete Kalman filtering. *Biomedical Engineering Online*, vol. 9, 1: 16.

[27] Jang K. E., Tak S., Jung J., Jang J., Jeong Y., Ye J. C. (2009) Wavelet minimum description length detrending for near-infrared spectroscopy. *Journal of Biomedical Optics*, 14, 034004.

[28] Koshevoy, G., and Mosler, K. (1997) Zonoid Trimming for Multivariate Distributions. *The Annals of Statistics*, 25, 1998–2017.

[29] Leidot O., Wolf M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10(5):603-621.

[30] Leidot O., Wolf M. (2003). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multtivariate Analysis*, 88(2):365-411.

[31] Lina J.-M., Matteau-Pelletier C., Dehaes M., Desjardins M., Lesage F. (2010) Wavlet-based estimation of the hemodynamic response in diffuse optical imaging. *Medical Image Analysis*, vol. 14, 4: 606-616.

[32] Li, J., and Liu, R. (2004) New Nonparametric Tests of Multivariate Locations and Scales using Data Depth. *Statistical Science*, 19, 686–696.

[33] Liu, R. (1990) On a Notion of Data Depth Based on Random simplices. *The Annals of Statistics*, 18, 405–414.

[34] Liu, R., Parelius, J. M., and Singh, K. (1999), Multivariate Analysis by Data Depth: Descriptive Statistics, Graphics and Inference. *Annals of Statistics*, 27, 783–858.

[35] Liu, R., Singh, K. (1993) A Quality Index Based on Data Depth and Multivariate Rank Test. *Journal of the American Statistical Association*, 88, 257–260.

[36] Lopez-Pintado S., Romo J. (2007) Depth-based inference for functional data. *Computational Statistics & Data Analysis*, 51, 4957-4968.

[37] Lopez-Pintado S., Romo J. (2009) On the concept of depth for functional data. *Journal of the American Statistical Association*, 104:486, 718-734.

[38] Mahalanobis P. C. (1936) On the Generalized Distance in Statistics. *Proceedings of National Academy of Science of India*, 12, 49–55.

[39] MATLAB and Statistics Toolbox Release 2015, The MathWorks, Inc., Natick, Massachusetts, USA.

[40] Mitra R., Muller P., Ji Y. (2016) Bayesian graphical models for differential pathways. *Bayesian Analysis*, 11(1):99-124.

[41] Murata Y., Sakatani K., Katayama Y., Fukaya C. (2002) Increase in focal concentration of deoxyhaemoglobin during neuronal activity in cerebral ischaemic patients. *Journal of Neurology, Neurosurgery and Psychiatry*, vol. 73, 2: 182-184.

[42] Pourshoghi A., Zakeri I., Pourrezaei K., (2016) Application of functional data analysis in classification and clustering of functional near-infrared spectroscopy signal in response to noxious stimuli. *Journal of Biomedical Optics*, **21**(10).

[43] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

[44] Ramsay J., Silverman B. W. (2005) Functional Data Analysis. 2nd edition. New York: Springer.

[45] Ramsay J., Silverman B. W. (2002) Applied Functional Data Analysis: Methods and Case Studies. 1st edition. New York: Springer.

[46] Saager R. B., Berger A. J. (2005) Direct characterization and removal of interfering absorption trends in two-layer turbid media. *Journal of the Optical Society of America. A Optics, Image Science, and Vision*, vol. 22, 9: 1874-1882.

[47] Saager R. B., Berger A. (2008) Measurements of layer-like hemodynamic trends in scalp and cortex: implications for physiological baseline suppression in functional near-infrared spectroscopy. *Journal of Biomedical Optics*, 13, 034017.

[48] Saager R. B., Telleri N. L., Berger A. J. (2011) Two-detector corrected near infrared spectroscopy (C-NIRS) detects hemodynamic activation responses more robustly than single-detector. *Neuroimage* 55, 1679-1685.

[49] Sato H., Tanaka N., Uchida M., Hirabayashi Y., Kanai M., Ashida T., Konishi I., Maki A. (2006) Wavelet analysis for detecting body-movement artifacts in optical topography signals. *NeuroImage*, vol. 33, 2: 580–587.

[50] Scarpa F., Brigadoi S., Cutini S., SCatturin P., Zorzi M., Dell' Acqua R., Sparacino G. (2013) A reference-channel based methodology to improve estimation of event-related hemodynamic response from fNIRS measurements. *NeuroImage*, vol. 72: 106-119.

[51] Scholkmann F., Spichtig S., Muehlemann T., Wolf M. (2010) How to detect and reduce movement artifacts in near-infrared imagin using moving standard deviation and spline interpolation. *Physiological Measurements*, 31, 649-662.

[52] Schroeter, M.L., Bucheler, M.M., Muller, K., Uludag, K., Obrig, H., Lohmann, G., Tittgemeyer, M., Villringer, A., von Cramon, D.Y. (2004) Towards a standard analysis for functional near-infrared imaging. *Neuroimage*, 21, 283–290.

[53] Scholkmann F., Wolf M. (2012) Measuring brain activity using functional near infrared spectroscopy: a short review. *Spectroscopy Europe* vol. 24 no. 4.

[54] Struyf A., Hubert M., Rousseeuw P. J. (1997) Clustering in an Object-Oriented Environment. *Journal of Statistical Software*. **4**.

[55] Tak S., Ye J. C. (2014) Statistical analysis of fNIRS dtata: A comprehensive review. *Neuroimage* , 85 72-91.

[56] Tian T. S., (2010) Functional data analysis in brain imaging studies. *Frontiers in Psychology*. **1**. 1-35.

[57] Torricelli A, Contini D., Pifferi A., Caffini M., Re R., Zucchelli L., Spinelli L. (2014) TIme domain functional NIRS imagining for human brain mapping. *Neuroimage*, 85, 28-50.

[58] Ullah S., Finch C. F., (2013) Applications of functional data analysis: A systematic review. *BMC Medical Research Methodology*, **13**, 43.

[59] Umeyama S., Yamada T. (2009) Monte Carlo study of global interference cancellation by multidistance measurement of near-infrared spsctroscopy. *Journal of Biomedical Optics*, 14(6).

[60] Watson M. (2006) Coxpress: differential co-expression in gene expression data. *BMC bioinformatics*, 7(1):1.

[61] Zhang Q., Brown E. N., Strangman G. E. (2007) Adaptive filtering for global interference cancellation and real-time recovery of evoked brain activity: a Monte Carlo simulation study. *Journal of Biomedical Optics*, 12(4).

[62] Zucchelli L., Contini D., Re R., Torricelli A., Spinelli L. (2013) Method for the discrimination of the superficial and deep absorption variations by time domain fNIRS. *Biomedical Optics Express*, vol. 4, 12: 2893-2910.

[63] Zucchelli L. (2013) Development and applications of a medical device based on time domain functional Near Infrared Spectroscopy for brain imaging. Ph.D. Thesis, Politecinco di MIlano.