Analyzing Brain Signals using Functional Geostatistics

Thesis submitted in partial fulfilment of the requirements for the Degree of Master of Science in Geospatial Technologies

Norgith Bibiana Quintero Alonso

Supervised by:

Edzer Pebesma
Institute for Geoinformatics
Universität Münster

Co-supervised by:

Jorge Mateu
Department of Matematics
Universitat Jaume I

Ana Cristina Costa
NOVA Information Management School
Universidade Nova de Lisboa

February 26, 2024

Universität Münster

UNIVERSITAT JAUME I

NOVA UNIVERSIDADE NOVA DE LISBOA

Declaration of Academic Integrity

I hereby declare that this thesis, titled "Analyzing Brain Signals Using Functional Geostatistics", is entirely my own work. I confirm that whenever I have consulted the published work of others, it has consistently been clearly acknowledged and cited. Likewise, I have acknowledged all the main sources of help.

Norgith Bibiana Quintero Alonso
Münster, Germany
February 26, 2024

Acknowledgements

I would like to express my gratitude to my supervisors, Edzer Pebesma, Jorge Mateu and Ana Cristina Costa, for their support, guidance and feedback on this work. Thanks for your kind words in every meeting and email. I would also like to give my deepest thanks to Martha Patricia Bohoquez. Despite not being officially included as a supervisor, this work would not have been possible without you. Thank you for not only sharing with me the dataset and your vast knowledge of functional geostatistics, but also for your kindness, for every video call, and for being an inspiration to encourage more women to join science.

I express my gratitude to the Erasmus Mundus Program and master program coordinators, Joaquín Huerta, Christoph Brox, and Marco Painho, for creating a platform that has allowed me to grow academically and personally. Many thanks to my colleagues and friends that I met on this journey. It was an honor to get to know great people and learn from their cultures and ways of seeing the world.

Finally, I'd like to thank my mom and siblings for always being my strength and supporting me in every phase of my life. Thanks for your unconditional love, for believing every day in myself, and for always being there for me. And to all my friends for always listening; your friendship truly nourishes me. I am so grateful to be surrounded by such kind human beings. Thanks to all for your love.

Abstract

Electroencephalography (EEG) signals represent the brain's electrical activity obtained through placing electrodes on the scalp. These recordings can be analyzed as curves across space and time. In this work, EEG signals are treated as functions or curves using functional data analysis. The construction of the curves is achieved through a set of B-splines basis functions. Due to the inherent spatial and temporal dependencies of EEG signals, various applications in spatial statistics have emerged, especially for modeling brain neuroimaging data. Here, we consider the brain as a random field where the EEG signals are analyzed as spatially correlated curves. Consequently, functional geostatistics methods are employed to predict curves at unsampled sites. The spatial prediction of the curves is performed using functional kriging, which employs the empirical functional principal components' representation of the EEG curves. The predictive performance is assessed through leave-one-out functional cross-validation.

The EEG signals used in this study were obtained from a controlled experiment involving inner speech. The analysis extracted key features from the smoothed curves, revealing that inner speech involves multiple brain regions, including the association area, Broca's and Wernicke's area.

Keywords: Functional data analysis, functional geostatistics, EEG signals, functional principal components.

Contents

List of Figures

List of Tables

Abbreviations

EEG	Electroencephalography
PSD	Power Spectral Density
FDA	Functional Data Analysis
SFDA	Spatial Functional Data Analysis
FPCA	Functional Principal Component Analysis
FPC	Functional Principal Components
EFPC	Empirical Functional Principal Components

Chapter 1

Introduction

1.1 Context

The brain is the most complex organ that executes cognitive functions, synchronized by the never-ending activity of millions of nerve cells. Understanding neural activities is crucial for medical diagnosis because it provides valuable insights regarding body and brain function. Various sensors and techniques have emerged to capture and explore brain activity. Some of them include functional magnetic resonance imaging (fMRI), electroencephalography (EEG), and electrocorticography (ECoG). The most commonly used techniques are EEG and fMRI due to their non-invasive characteristics for real-time monitoring and observation of neural mechanisms (Gui et al., 2010). While fMRI records blood flow as a proxy for brain neural activity, EEG measures the electrical potentials produced by nerve cells within the cerebral cortex. The preference for using EEG instead of fMRI for studying neural activity relies on its affordability and accessibility for obtaining quick brain activity information. EEG signal analysis is primarily focused on detecting neurological and mental disorders (e.g., epilepsy and depression) and developing control and communication technology for Brain-Computer Interfaces (BCI) (Chaddad et al., 2023).

Traditionally, EGG data are analyzed in time, frequency, and time-frequency domains. The time domain relies on the characteristics and properties of signals varying over time, while the frequency domain decomposes signals into their frequency components. The most popular frequency property is the power spectrum density (PSD). Meanwhile, the time-frequency domain integrates both time and frequency domains, providing extra information to recognize dynamical features of non-stationary signals (Hernández et al., 2018). The advent of machine learning algorithms has further enhanced EEG signals research, with classification methods such as support vector machines (Saccá et al., 2018), decision trees (Bastos

et al., 2020), linear discriminant analysis (Nkengfack et al., 2020), and k-nearest neighbors (Sha'Abani et al., 2020), among others.

EEG signals are characterized by their non-stationary and non-linear features, given that these records correspond to densely repeated measurements over time and space. Therefore, each EEG data observation can be represented as a space-time curve. The increased need for handling this data type has led to the development of advanced techniques such as functional data analysis (FDA). FDA extends traditional statistical methods to model data in the form of functions such as curves, surfaces, or images. In the FDA framework, the entire observed curve is modeled as a realization of a random function rather than a series of discrete values, and the analysis is developed on the whole functional dataset. Since its inception in the early 1990s, there has been a growing interest within the statistical community to create models for functional data (Giraldo et al., 2007), as evidence by the pioneering work of Deville (1974) and Ramsay and Dalzell (1991).

The inherent spatial and temporal dependencies present in EEG signals, intrinsic to the brain's nature, have led to various applications in spatial statistics for modeling brain neuroimaging data. The spatial component is given by the location of electrodes on the scalp that capture brain activity, while the changes define the temporal aspect in brain activity over time. For example, McCall et al. (2012) analyzed EEG data for epilepsy detection, introducing spatial and temporal domains for identifying abnormal electrical activity of the cortex in epileptic patients. As expressed before, EEG signals represent cerebral electrical activity that can be analyzed in the form of spatially dependent curves. A recent approach, functional geostatistics, offers a novel perspective for modeling the spatial dependence structures among curves. Geostatistical methods can be applied within this framework to model and predict spatial functional data, including EEG signals.

1.2 Motivation

Conventional EEG analysis is focused on temporal features and involves signals preprocessing, feature extraction and pattern recognition, and classification (Ruchika et al., 2023). Common signal processing methods, such as wavelet transformation and Fourier transformation, determine the frequency value of each frequency signal band. Feature extraction and pattern recognition usually employ a combination of time and frequency domains and wavelet techniques (Guo et al., 2015). Currently, these methods are automated with machine learning techniques to provide better classification results (Tran, 2022). Although various data analysis techniques are available, most are tedious and time-consuming, unable to manage the enormous size of EEG data. Additionally, many techniques presume EEG signals analysis can be conducted through linear statistics (Thivierge, 2007). Because of these limitations, there is an increased need to apply novel approaches to EEG data that consider more than time- and frequency-domain methods (Chandel et al., 2023).

FDA emerges as a valuable framework for understanding the evolution of EEG signals and their connection to human conduct (Yi et al., 2022). A popular application of FDA is brain neuroimaging data, such as EEG signals, which contain information regarding the complex interaction between brain cortical activity and cognitive processes (Thivierge, 2007). As technological advancements result in a significant increase in data collection, functional data analysis has gained prominence across various domains. For instance, FDA finds applications in the climate and meteorology fields when analyzing variations in temperature and precipitation among weather stations (Ramsay & Dalzell, 1991) or PM10 pollutant behavior (ShAADAN et al., 2012). FDA is a statistical technique introduced by Grenander (1950) and Rao (1958), and later formalized by Ramsay (1982) and Ramsay Dalzell (1991) (Wang et al., 2016).

In FDA, many applications correspond to systematic spatio-temporal datasets (Cressie & Wikle, 2015). Functional data presenting inherent spatial dependencies is commonly referred to as spatial functional data (Delicado et al., 2010). The observed curves of spatial functional data are related according to the data proximity. In this paradigm, spatial functional data analysis (SFDA) extends spatial statistical methods to handle functional data objects with spatial dependencies (Guo et al., 2022). In contrast to other spatio-temporal approaches, SFDA techniques broaden spatial methodologies without relying on parametric assumptions about temporal effects. In SFDA, the fundamental unit of information becomes the entire observed function, incorporating the spatial component. The methodologies developed in SFDA can be applied to various spatial data structures, comprising geostatistical data, point patterns, and aerial data. Consequently, much like univariate or multivariate spatial data analysis, SFDA methods can be broadly classified into geostatistical functional methods and purely spatial functional methods (Mateu & Romano, 2017). The SFDA methods are presented by Giraldo et al. (2011); Bohorquez et al. (2017) and Menafoglio and Secchi (2017), among others.

The brain is a spatially multi-dimensional and multi-scale system with continuous dynamic processes (Zaleshina & Zaleshin, 2017). Therefore, it can be studied as a spatial map, allowing the application of geostatistical techniques such as Kriging. Kriging is a set of spatial prediction methods extensively used in geostatistical applications for spatial predictions at unknown locations that minimize mean square prediction errors (Chiles & Delfiner, 2009). The potential of studying EEG signals as functional data, including brain spatial structures, is the main motivation for this work. This research is focused on spatial prediction of EEG signals produced during covert speech using functional geostatistics approach, giving insights about speech perception and production on the left hemisphere of the brain. In the same way, this thesis attempts to indirectly highlight the potential of considering spatially informed data for understanding brain function and cognitive processes.

1.3 Related work

In the past few years, FDA has been used to analyze EEG data for the detection of neurological disorders and the development of BCI. Zhang et al. (2020) employed FDA to analyze the relationship between working memory ability and frontal activity of the brain using EEG signals. Jaramillo-Jimenez et al. (2023) conducted an analysis of the resting-state EEG spectral features between subjects with Parkinson's disease and healthy subjects to find differences in theta and alpha rhythms. The authors used the FDA and averaged epochs approaches to explore the external validity and reproducibility of the findings. FDA allows the extraction of time or frequency information from EEG signals, facilitating the selection of significant channels and frequencies to identify diverse human behaviors (Yi et al., 2022). For instance, A. W. Scheffler et al. (2022) investigated the neurodevelopmental differences between typically developing children and children with autism spectrum disorder through modeling the patterns of alpha spectral variations across the scalp.

Functional principal component analysis (FPCA) is a method used to understand EGG signals patterns and variation by extracting main features from the signals. Vidal Badia et al. (2021) proposed a functional data framework for removing unwanted noise from the signals to estimate brain electrical activity sources. The authors used functional independent component analysis and a principal component expansion using B-spline basis functions. Functional linear regression models are applied to predict working memory ability based on EEG signals, with the B-spline approximation of functional principal components (Sundaram et al., 2022; A. W. Scheffler et al., 2020). Another alternative method involves employing independent component analysis, band-pass filtering, and discrete wavelet transform to extract features from EEG signals (Das et al., 2023). Machine learning algorithms are employed to classify the extracted features, such as linear discriminant analysis, support vector machine, k-nearest neighbor, and artificial neural networks, for effective classification (A. Scheffler et al., 2020).

Diverse methodologies have been proposed to examine the spatial properties of EEG signals. Adaptive spatial filters, such as independent component analysis and principal component analysis, represent one approach to extracting weights from the spatial attributes of the data (Loza & Principe, 2021). The approach entails utilizing knowledge regarding the spatial arrangement of EEG recordings to enhance the classification precision. Another study involves utilizing a two-dimensional representation of EEG channel locations to enhance classification accuracy (Wen et al., 2020).

There is a growing interest in modeling correlated functional data, such as temporally and spatially correlated functional data. Statistical methods for modeling correlated variables have been adapted to the functional context. Specifically, autoregressive Hilbertian models have been used to deal with temporal correlation in functional data, and regression models have been developed to deal with spatially correlated functional data (Giraldo et al., 2010).

Kriging, a geostatistical method known for its high accuracy in spatial prediction, has demonstrated efficacy in diverse applications (Olokodana, Mohanty, & Kougianos, 2020a). Kriging has been employed for spatial filtering of EEG signals (Zhao et al., 2023). A deep neural network (DNN) model has been proposed for rapid and accurate seizure detection using EEG signals (Ibrahim et al., 2022). This model complements kriging computation across different cores, and feeds the correlated data into the DNN for training, exhibiting enhanced efficiency compared to conventional kriging methods and DNN (Olokodana, Mohanty, & Kougianos, 2020b). The authors explored three variations of kriging methods (simple, ordinary, and universal). Experimental results with EEG signals from both healthy and diseased patients indicated that all three kriging methods performed well in terms of accuracy, sensitivity, and detection latency. Simple kriging was the preferred method for seizure detection, outperforming ordinary kriging and universal kriging within a 68.2% confidence interval.

Kriging and cokriging predictors have been considered for spatial prediction of functional data. A nonparametric approach has been introduced to address the spatial prediction of functional data by selecting an appropriate number of basis functions in the modeling process. A practical application is presented with meteorological data from Canada's Maritime Provinces. The spatial prediction of meteorological data holds significance for various models, including hydrological models and forest ecosystem models (Giraldo et al., 2010). Similarly, Nerini et al. (2010) proposed a kriging method for spatially sampled curves with a spatial functional linear model. The method analyzes the temperature profiles of elephant seals in the Antarctic Ocean.

1.4 Aim and objectives

This study aims to use functional geostatistics to predict spatial EEG signals generated during verbal thinking. This research will attempt to answer the following question: How can spatial modeling of EEG signals help better understand neural activities responding to language and expression brain areas?

To address this research question, the following objectives have been established:

- Analyze EEG signals as space-time curves using FDA.
- Model the spatial autocorrelation structure of EEG signals using functional geostatistics.
- Employ simple kriging to estimate the curved signals within the brain area associated with language.

1.5 Outline

This thesis is organized into five chapters, as follows: Chapter 2 defines background on the speech and EEG concepts employed in this thesis. Furthermore, it introduces the background of geostatistics, functional data analysis, and functional geostatistics. Chapter 3 describes the methods and thesis workflow, considering the concepts highlighted in Chapter 2. Chapter 4 presents the results of the exploratory analysis of EEG curves, and spatial modeling and prediction. Finally, Chapter 5 discusses the outcomes, suggests future research directions, and concludes this work.

Chapter 2

Background theory

This chapter provides a background on the relevant concepts for the thesis and includes a literature review of existing techniques. The first section explains the concepts related to the speech and EEG signals. The remaining sections include the statistical background of functional data analysis, and geostatistics for scalar and functional data. The main references for the theoretical concepts were extracted from Ramsay and Dalzell (1991) and Bohorquez et al. (2016).

2.1 Speech

Speech involves an interchange of motor functions originating in the brain and ending in muscular movements, and it has been a focal point of EEG investigations (Schultz et al., 2017). During covert and overt speech, the electrical activity of the brain is triggered by multiple cortex regions, which are involved in speech perception and production. In both forms of speech, EEG has been extensively studied through experiments. Overt speech can introduce potential effects of muscle activity on electrophysiology experiments, which can affect the signal-to-noise ratio and result in poor measurements (Ganushchak et al., 2011). In contrast, covert speech, often called imagined or inner speech, involves self-communication without overt articulation, making it a compelling domain for exploration (Li et al., 2021). Decoding covert speech signals is promising, particularly in applications such as BCI, which aims to enhance the quality of life for individuals who are unable to communicate through conventional means or present severe motor disabilities (Nieto et al., 2022).

Inner speech is associated with increased activity in the left inferior frontal gyrus, left premotor cortex, supplementary motor area and left temporal cortex (Shergill et al., 2003). These areas are involved in speech generation and perception. However, the activation of the areas varies across individuals regarding age and mental

state (Alderson-Day & Fernyhough, 2015). For example, patients with schizophrenia, who often experience auditory verbal hallucinations, show attenuated activation in the left superior temporal gyrus during the processing of inner speech, indicating deficits in self-monitoring (Simons et al., 2010).

Research on brain mapping and inner speech has revealed the specific neural correlates of inner speech, and its role in language ability (Geva et al., 2011). This includes the left pars opercularis in the inferior frontal gyrus and the white matter adjacent to the left supramarginal gyrus. The interaction between areas that generate and perceive inner speech in the frontal and temporal cortex has shown fronto-temporal connectivity. This connectivity is further supported by the modulation of activity in the temporal cortex during inner speech generation (Shergill et al., 2001). The cognitive roles of inner speech involve enabling abstract thought and supporting metacognition, memory, and executive function (Langland-Hassan, 2021).

2.2 Electroencephalography

Electroencephalography (EEG) is the most commonly used technique for recording and interpreting the brain's electrical activity, measured through electrodes on the scalp or directly on the cortex (Blinowska & Durka, 2006). Neuron cells are constantly communicating with each other by sending electrical signals that generate brain waves (Iqbal et al., 2016). These brain waves, also known as brain rhythms, are categorized into distinct bandwidths depending on their functions. Delta waves are characterized by low frequencies (below to 4 Hz) and are generated during sleep and meditation (Botella-Soler et al., 2012). In contrast, Gamma waves (30–80 Hz) associated with active thought, (Ortiz et al., 2020), are fast and sensitive. Theta rhythms, which range from 3 to 8 Hz, are predominant in light sleep and relaxation (Sabourin et al., 1990). Alpha waves (8 - 13 Hz) are present during awake state of relaxation (Cahn & Polich, 2013), and betha waves (13 - 30 Hz) during cognitive activities (Pfurtscheller et al., 2006). The dominance of specific brain waves depends on the subject's activities and emotions, where low frequencies are dominant during states of tiredness or dreaminess, and higher frequencies for states of alertness or hyperactivity. The speed of brainwaves is quantified in hertz (cycles per second) (Ghane, 2015).

The electrical activity is recorded by placing electrodes along the scalp of a subject, also known as EEG channels. The brain signals are generated by the inhibitory and excitatory postsynaptic potentials of cortical nerve cells, which originate in the cortex and extend to the scalp surface. In a normal adult, the frequency of EEG signals varies from 1 Hz to about 100 Hz and the amplitude from about 10 μV to 100 μV (Subha et al., 2010). As the brain's structure is not uniform and the cerebral cortex is functionally arranged, EEG patterns are highly influenced by the position of the electrodes.

The International Federation of Societies for Electroencephalography has devised the 10–20 electrode placement system to specify electrode locations. This system was created to allow comparisons over time between participants in an experiment using EEG, guaranteeing reproducibility. The method is based on the correlation between the position of an electrode and the underlying cerebral cortex area (Subha et al., 2010). In this system, between 16 and 20 electrodes are evenly distributed, with each pair being separated by a distance of 10% - 20% of the total circumference distance of the head (front-back or right-left) (Khalifa et al., 2015).

EEG recordings are mainly used to diagnose and study functional brain disorders and diseases, including epilepsy, tumors, and cerebrovascular lesions, among others. Brain signals are noisy, nonstationary, and nonlinear. The complexity and randomness of brain signals are influenced by subjects age and mental states. To comprehend the complex behavior and dynamics of interconnected neurons, it is imperative to employ a combination of linear and nonlinear signal processing techniques (Kaur & Kaur, 2015).

2.2.1 Signal preprocessing

EEG observations are contaminated with several types of noise. To improve the quality of the data, EEG signals are preprocessed using a filter to remove the noise components. Filtering is a technique that removes unwanted noise from signals by removing contaminated frequencies. Some commonly used filters for EEG signal processing include: spatial filters (e.g., independent component analysis), temporal filters (e.g., moving averages), frequency selective filters (high and low pass filters), and constant filters (Ghane, 2015).

2.2.2 Feature extraction

The most important features and properties of the signals can be extracted using time-domain features, frequency-domain features, and wavelet features. In the time domain, the EEG signals are analyzed in terms of voltage and time. The frequency domain examines signals in terms of power and frequency (Ghane, 2015).

Power spectral density (PSD) is a feature extraction method that quantifies the power of each frequency level of brain activity in EEG signals (Myers et al., 2017). This approach represents the distribution of a signal's power over a given frequency band, facilitating comprehension of the signal's characteristics (Dempster, 2001).

PSD reflects the intensity of variations across different frequencies. It delineates where frequency variations are robust or weak. PSD is measured in energy per frequency, allowing the determination of energy within a designated frequency range through integration. The PSD essentially constitutes the Fourier transform of the signal's autocorrelation function. Calculating the power of a signal within a specific frequency band involves integrating across positive and negative frequencies

(Unde & Shriram, 2014). The Fourier transform $S(w)$ and power spectral density $S_{x,y}(w)$ are computed as shown in equation 2.1 and 2.2 (Ricker, 2012).

$$S(w) = \int_{-\infty}^{\infty} s(t)e^{-jwt}dt \tag{2.1}$$

$$S_{x,y}(w) = \int_{-\infty}^{\infty} R_{x,y}(z)e^{-jwz}dz \tag{2.2}$$

Where $R_{x,y}(z)$ is the autocorrelation function.

2.3 Functional Data Analysis

In diverse fields, many datasets come from repeated measures with variations over a continuum. Functional data analysis (FDA) emerges to analyze functional data, which is data in the form of continuous functions. This statistical technique is focused on the analysis of sample curves, but it can also include vectors or scalars (Kokoszka & Reimherr, 2017). This section provides an FDA background, including some definitions of functional data and its representation using basis expansion. Furthermore, we include descriptive statistics in the exploratory functional data analysis.

2.3.1 Functional Data

The fundamental unit of information in FDA is the functional data (Ramsay & Dalzell, 1991), which denotes a smooth curve or function observed over a continuous domain, such as time and frequency.

Let $\boldsymbol{\chi}(t)$ be a function defined for a set of T. Then, according to Ferraty (2006):

- A functional variable is a random variable defined as $\boldsymbol{\chi}(t)$, $t \in T$, that takes values in an infinite dimensional space (or functional space).
- A functional data refers to an observation an observation $\chi(t)$ of $\boldsymbol{\chi}(t)$, $t \in T$. Let $T \subset \mathbb{R}$, it is common to assume that the observations are in L^2.
- A functional dataset $\chi_1(t), ..., \chi_n(t)$ corresponds to the realizations of n functional variables, $\chi_1(t), ..., \chi_n(t)$ identically distributed as $\boldsymbol{\chi}(t)$.

The reconstruction of the entire function $\chi_i(t)$ can be performed by fitting a model, since the observed functions correspond to a finite number of discrete values. The model can be parametric (Goulard & Voltz, 1993), or non-parametric (Ramsay & Silverman, 2005) using basis functions. The basis function system enables the representation of multiple time points using fewer coefficients with a reduced number

of errors. It is a collection of known functions $\phi_1, ..., \phi_K$ independent of each other (Ramsay & Silverman, 2005). The function $\chi_i(t)$ can be represented as a linear combination of K basis functions as follows:

$$\chi_i(t) = \sum_{j=1}^{K} a_{ij}\phi_j(t) = \Phi^T(t)a_i \tag{2.3}$$

where $\Phi^T(t)$ represents the vector whose elements are the K basis functions, a_{ij} represents the coefficients of the basis function $\phi_j(t),\ j = 1, ..., K$, for $\chi_i(t)$ and a_i is a vector containing K coefficients.

The most commonly used basis functions include Fourier, B-splines, and wavelets, among others. The decision to use one of them relies on the attributes of the data. For instance, the most popular basis systems, Fourier basis and B-splines, are used for periodic and non-periodic data, respectively (Ramsay & Silverman, 2005).

B-splines basis is a smoothing spline method proposed by De Boor and De Boor (1978). This spline method fits a piecewise defined polynomial function to the data, which aims to decrease noise levels by employing a function that represents the variations of the original data without overfitting it. B-splines functions are determined by the spline order and the placement of knots (Thivierge, 2007).

Given a set of L subintervals of $T = [a, b]$ connected by knots $\tau_l, l = 1, \ldots, L-1$. Therefore, the number of basis functions $K = m + L - 1$, where m represents the order and $L - 1$ the number of knots. A spline basis function $S(t)$ is expressed as (Ramsay & Silverman, 2005):

$$S(t) = \sum_{k=1}^{K} c_k B_k(t) \tag{2.4}$$

Where $B_k(t)$ represents the B-splines function.

2.3.2 Exploratory Functional Data Analysis

The exploratory functional data analysis consists of data visualization and summarization of functional data. The exploratory FDA retrieves general information about the structure of the functional data and explores its main features and unusual characteristics. The classical univariate and bivariate descriptive statistics can be applied to functional data, where the calculated objects correspond to curves. The descriptive functions correspond to the point-wise mean, variance, covariance, and correlation.

Given a set of n functions observed at multiple time points $x_i(t)$, the descriptive statistics functions are (Ramsay & Dalzell, 1991):

- Mean:
$$\bar{\chi}(t) = \frac{\sum_{j=1}^{n} \chi_i(t)}{n}$$

- Variance:
$$Var_{\chi}(t) = \frac{\sum_{j=1}^{n} (\chi_i(t) - \bar{\chi}(t))^2}{n-1}$$

- Covariance:
$$Cov_{\chi}(t_1, t_2) = \frac{\sum_{j=1}^{n} (\chi_i(t_1) - \bar{\chi}(t_1))(\chi_i(t_2) - \bar{\chi}(t_2))}{n}$$

- Correlation:
$$Cor_{\chi}(t_1, t_2) = \frac{Cov_{\chi}(t_1, t_2)}{\sqrt{Var_{\chi}(t_1) Var_{\chi}(t_2)}}$$

2.4 Geostatistics

Spatial statistics is an area of statistics that encompasses a set of methodologies for predicting and estimating the spatial dependence structures of data derived from random observations at point locations, or spatial aggregations, within a region. This approach is integrated with geostatistics, regional data, and point patterns. In geostatistics, the main interest is the prediction of unknown values at unsampled locations. Since we consider the brain as a continuous domain, we use geostatistics to model the spatial dependencies of the EEG signals obtained from 21 electrodes. This section contains an overview of the basic concepts in geostatistics.

2.4.1 Scalar geostatistics

Geostatistics is a set of methods applied to spatial data varying in a continuous space, where $D_s \subset \mathbb{R}^d$. D_s is continuous and $Z(s)$ is a random variable with spatial location s, ($s \in D_s$). In this section, we describe the concepts used in geostatistics when dealing with scalar data.

Spatial process

A spatial process is defined as a stochastic process $\{Z(s), s \in D_s \subset \mathbb{R}^d\}$, where $\mathbb{R}^d$ is the spatial index set. A stochastic process, also known as a random field or random function, is based on the first two moments of $Z(s)$. A random function is characterized by the following elements:

- Cumulative distribution function: For any n points $s_1, \ldots, s_n$, the n-dimensional distribution function of a random vector $Z(s) = (Z(s_1), \ldots, Z(s_n))^t$ is

$$F_{s_1,s_2,\ldots,s_n}(z_1, z_2, \ldots, z_n) = P[Z(s_1) \leq z_1, Z(s_2) \leq z_2, \ldots, Z(s_n) \leq z_n]$$

- Mean function: The first moment is the mathematical expectation, usually denoted as $\mu(s) = E(Z(s))$.

- Variance function: The second moment is the variance $Z(s)$ with respect to $\mu(s)$. It is denoted as $Var(Z(s)) = E(Z(s) - \mu(s))^2$.

- Autocovariance function: The autocovariance of a spatial process $Z(s)$, is a function of spatial locations, s_i and s_j. $s_i, s_j \in \mathbb{R}^d$, for all $i, j \in \mathbb{Z}^+$. $Cov(Z(s_i), Z(s_j))$ is defined as:

$$Cov(Z(s_i), Z(s_j)) = C(s_i, sj) = E((Z(s_i) - \mu(s_i))(Z(s_j) - \mu(s_j)))$$

  where $C(.)$ is a positively defined function to ensure a non-negative prediction error variance.

- Semivariance function: The semivariogram $\gamma(Z(s_i), Z(s_j))$ estimates the spatial variance for specific distances, and is defined as:

$$\gamma(Z(s_i), Z(s_j)) = \gamma(s_i, s_j) = \frac{1}{2}E(Z(s_i) - Z(s_j))^2 \tag{2.5}$$

  Therefore, the variogram is $2\gamma(s_i, s_j)$.

Considering $\mu(\boldsymbol{s}) = E(Z(\boldsymbol{s}))$ is known for $\boldsymbol{s} \in D_{\boldsymbol{s}}$, the optimal prediction of a random variable $Z(s_0)$ at an unsampled site $\boldsymbol{s}_0 \in D_{\boldsymbol{s}}$ can be expressed as the following linear combination of the data (Cressie & Wikle, 2011):

$$Z^*(\boldsymbol{s}_0) = \sum_{i=1}^{n} \lambda_i Z(\boldsymbol{s}_i) = \boldsymbol{\lambda}'\boldsymbol{Z} \tag{2.6}$$

Let $\boldsymbol{\mu} = (\mu(\boldsymbol{s}_i), i = 1, ..., n)$, and $\boldsymbol{\lambda}$ is selected to minimize the mean-squared prediction error denoted as:

$$E(Z(\boldsymbol{s}_0) - \boldsymbol{\lambda}'\boldsymbol{Z})^2 \tag{2.7}$$

The main goal of this method is to estimate the covariance matrix by fitting a positive definite function C which guarantees that the predicted variance $Var(Z^*(\boldsymbol{s}_0))$ is not negative. Simple kriging is an optimal linear predictor that does not require stationarity or normality assumptions. It is a weighted average where the weights are found through minimization of the variance of the prediction error, denoted as:

$$Z^*(\boldsymbol{s}_0) = \mu(\boldsymbol{s}_0) + c'\Sigma^{-1}(\mathbf{Z} - \boldsymbol{\mu}) \tag{2.8}$$

with minimized mean-squared prediction error

$$E\left(Z^*(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)\right)^2 = C(0) - c'\Sigma^{-1}c \tag{2.9}$$

where

- $C(0) = Var(Z(s))$
- $\Sigma \equiv Cov(\mathbf{Z})$. Σ is a positive definite matrix. The i,j-th entry of Σ is $(\Sigma)_{ij} = Cov(Z(s_i), Z(s_j))$ $i,j = 1,...,n$. $Cov(Z(s_i), Z(s_j))$ is given by a known positive-definite covariance function $C(Z(s_i), Z(s_j)|\Theta)$ with parameter vector Θ.
- $c' = cov(Z(\boldsymbol{s}_0); \mathbf{Z})$
- $\boldsymbol{\mu} \equiv E(\mathbf{Z})$

Covariance parameter estimation

Least squares methods are employed for selecting the model C and initial values Θ_0 based on graphical exploratory data analysis. These methods involve estimating the variogram 2γ, which under second-order stationarity, is related to the covariance function by the expression:

$$2\gamma(s_i - s_j|\Theta) = Var\left(Z(s_i) - Z(s_j)\right) = 2\left(C(\mathbf{0}|\Theta) - C(s_i - s_j|\Theta)\right) \tag{2.10}$$

for $s_i - s_j \in \mathbb{R}^d$. In geostatistical analysis, it is widely used the weighted least-squares method, which makes use of the matrix $W(\Theta)$, proposed in (Cressie, 1985). The weighting matrix is given by:

$$W(\Theta) \simeq Diag\left(\frac{2(2\gamma(s_i - s_j|\Theta))^2}{|N(s_i - s_j)|}\right)$$

where $N(s_i - s_j) = \{(i,j) : s_i - s_j\}$, for $i,j = 1,\ldots,n$ which produces the first q spatial lags $q = 1,...,Q$. These methods employ spatial lag categories until half of the maximum distance between every pair of sites. The disadvantage of using least squares methods is the need to define lag classes to perform an empirical estimation of the covariance. When there are not many observations, the amount of data in each class may be small.

2.4.2 Functional geostatistics

In this section, the concepts of geostatistics are broadened to handle spatial functional data. Analyzing spatial functional data offers an alternative approach to space-time modeling when the curves are spatially correlated. Functional

geostatistics allows the spatial prediction of the entire curve at unsampled locations. The spatial functional predictors are presented based on functions or curves in terms of their empirical functional principal components. This approach demonstrated that functional kriging depends on the auto-covariance and cross-covariance of the scalar random field associated with the score vectors. This section contains basic concepts for the prediction of spatial functional data. These concepts are further developed by Bohorquez et al. (2016).

Spatial functional data

The concept of functional data in Section 2.3 is extended to define spatial functional data. Considering $D_{\boldsymbol{s}} \subset \mathbb{R}^d$ as the spatial index set, $\chi_s(t),\ s \in D_s$ represents a spatial functional random variable. A spatial functional dataset $\chi_{s_1}(t), ..., \chi_{s_n}(t)$ refers to the observation of n functional variables $\chi_{s_1}(t), ..., \chi_{s_n}(t)$ at spatial locations $S = \{s_1, ..., s_n\}$, $S \subset D_{\boldsymbol{s}}$.

Spatial functional random field

Let $D_{\boldsymbol{s}} \subset \mathbb{R}^d$ be the spatial index set, and let $\chi_s^1(t), ..., \chi_s^P(t),\ s \in D_{\boldsymbol{s}}$, be P spatial functional square integrable random fields, such that, $\chi_s^p(t) \in H = L^2(B)$ $p = 1, ..., P$. Here the consideration is that $t \in B \subset \mathbb{R}$, represents the functional variable as a curve. $L^2(B)$ represents a real separable Hilbert space. Then, according to Bohorquez et al. (2016), a multivariate spatial functional random field is denoted as $\{\Xi_s : s \in D_{\boldsymbol{s}} \subset \mathbb{R}^d\}$ where $\Xi_s = \left(\chi_s^1(t), ..., \chi_s^P(t)\right)$. A multivariate spatial functional dataset is an observation of Ξ_s at specific spatial locations, denoted by $S \subset D_{\boldsymbol{s}}$.

Functional principal components

Assuming the spatial functional random fields are random elements of $L^2(B)$ and that $E(\chi_s(t)) = 0$, then the covariance operator C of $\chi_s(t)$ is defined as follows:

$$C(y) = E\left[\langle \chi_s, y \rangle \chi_s\right] \qquad y \in L^2(B) \tag{2.11}$$

Thus,

$$C(y)(t) = \int c(t,r)y(r)dr, \quad \text{where} \quad c(t,r) = E\left[\chi_s(t)\chi_s(r)\right]$$

with the following estimators

$$\hat{C}(y) = \frac{1}{n}\sum_{i=1}^{n}\left(\langle \chi_{s_i}, y \rangle \chi_{s_i}\right), \qquad y \in L^2(B)$$

and

$$\hat{C}(y)(t) = \int \hat{c}(t,r)y(r)dr, \quad \text{where} \quad \hat{c}(t,r) = \frac{1}{n}\sum_{i=1}^{n} \chi_{s_i}(t)\chi_{s_i}(r)$$

According to Horvath and Kokoszka (2012), the functional principal components (FPC) are the eigenfunctions of the covariance operator (2.11). The empirical functional principal components (EFPC) are the estimators of FPC.

The reconstruction of the curve χ_{s_i} can be done using a basis function system using the EFPC. The EFPC are formed by the eigenfunctions $\xi_j(t),\ j = 1, ..., K$ of the covariance operator C of $\chi_s(t)$ with the basis coefficients associated to principal component scores $f_j(s_i)$ as follows:

$$f_j(s_i) = \langle \chi_{s_i}, \xi_j \rangle, \quad j = 1, ..., K, \ \ i = 1, ..., n \tag{2.12}$$

In Horvath and Kokoszka (2012), the approximation of this basis is uniformly optimal by minimizing $\hat{\mathcal{S}}^2$, denoted as:

$$\hat{\mathcal{S}}^2 = \sum_{i=1}^{n} \left\| \chi_{s_i}(t) - \sum_{k=1}^{K} f_{s_i}^k \xi^k(t) \right\|^2. \tag{2.13}$$

The K is chosen by considering a minimum percentage of accumulated variance, commonly set at 85%. The functional principal components can be either univariate or multivariate, as the dimension is reduced in T and not in the number of functional random variables. It is possible to have a great approximation using only a few EFPCs with a strong temporal correlation.

By employing the Karhunen-Loève expansion (Bosq, 2000), it is assumed that the model takes the following form:

$$Y_s(t) = \mu(t) + \chi_s(t) = \mu(t) + \sum_{j=1}^{\infty} f_j(s_i)\xi_j(t), \quad Y_s(t) \in L^2 \tag{2.14}$$

where $E(Y_s(t)) = \mu(t)$.

The estimation of the mean function $\mu(t)$ is performed by utilizing the sample mean function $\hat{\mu}(t) = \bar{Y}_s(t)$ with $\bar{Y}_s(t) = n^{-1}\sum_{i=1}^{n} Y_{s_i}(t)$. Therefore, $E(\chi_s(t)) = 0$ and hence, we can use the random variable $\chi_s(t)$. Additionally, for $k = 1, ..., K$

$$E\left(f_{s_i}^k\right) = E\left\langle \chi_{s_i}, \xi^k \right\rangle = \left\langle 0, \xi^k \right\rangle = 0 \tag{2.15}$$

It is noted that for each k and $s \in D_s$, f_s^k is a scalar spatial random field. So, the vector $\left(f_s^1, ..., f_s^K\right)$ is a K-dimensional spatial random field.

Simple kriging using functional principal components

The spatial functional prediction is based on the assumption of a known mean $\mu(t)$, and employs a linear combination of the observed functions or curves as follows:

$$\breve{\chi}_{s_0}(t) = \sum_{i=1}^{n} \lambda_i \chi_{s_i}(t) \tag{2.16}$$

This predictor in (2.16) has the same expression as the simple kriging (see 2.6) considering curves. The predicted curve $\breve{\chi}_{s_0}(t)$ is obtained as a linear combination of the observed functions $\chi(t)$.

The minimum least squares method is used to solve $\lambda_1, \lambda_2, ..., \lambda_n$ (Horvath & Kokoszka, 2012),

$$E\left\|\chi_{s_0}(t) - \breve{\chi}_{s_0}(t)\right\|^2 = E\left(\langle \chi_{s_0}, \chi_{s_0} \rangle\right) - 2\sum_{i=1}^{n} \lambda_i E\left(\langle \chi_{s_i}, \chi_{s_0} \rangle\right) + \sum_{i,i'=1}^{n} \lambda_i \lambda_{i'} E\left(\langle \chi_{s_i}, \chi_{s_{i'}} \rangle\right) \tag{2.17}$$

The expression (2.17) depends entirely on the covariance between pairs of curves at different locations. Substituting each curve into its EFPC representation, then:

$$E\left(\langle \chi_{s_i}, \chi_{s_{i'}} \rangle\right) = \sum_{j=1}^{\infty}\sum_{j'=1}^{\infty} E\left(f_j(s_i) f_{j'}(s_{i'})\right)\left(\langle \xi_j(t)\xi_{j'}(t)\rangle\right) = \sum_{j=1}^{\infty} E\left(f_j(s_i) f_j(s_{i'})\right) \tag{2.18}$$

Given the linear combination representation using empirical functional principal components, see (2.13) and (2.14), the functional covariances between two locations $E\left(\langle \chi_{s_i}, \chi_{s_{i'}} \rangle\right)$ are determined by the sum of the spatial auto-covariance of all score components $f_j(s)$ for $(s_i, s_{i'})$.

Chapter 3

Methodology

This chapter describes the methodology for the spatial prediction of the EEG curves. The foundation of this methodology is based on the theoretical framework presented in Chapter 2. This chapter is organized in three sections. The first section presents a description of the EEG signals dataset. The second section summarizes the software and tools. The third section presents the general workflow using functional data analysis and functional geostatistics approaches.

3.1 Data description

The EEG signal dataset corresponds to the realization of a spatial functional random field comprising 21 electrodes for each vowel. The dataset consists of the power spectral density (PSD) that was calculated for all the signal data using the Fast Fourier transform and averaging each respective signal PSD. Figure 3.1 illustrates the observed EEG signal values generated by every vowel task at the 21 electrodes, and Figure 3.2 displays the location of the 21 electrodes and the brain functional areas.

Experimental protocol

The EEG signals were obtained at the Clinical Electrophysiology Laboratory of the *Universidad Nacional de Colombia*. The experiment involved 22 healthy subjects wearing an EEG neuroheadset equipped with 21 georeferenced electrodes (input channels) in the left hemisphere. Throughout the experiment, the light was on and off for a certain time. While the light source was on, participants were asked to think continuously about a specific vowel using imagined speech, ceasing their focus, and relaxing once the light source was off. Between every silent speech vowel task, a 5-minute resting period was given to each participant to facilitate the transition to the subsequent vowel. The vowels considered were the five Spanish vowels recorded in the order: /a/, /e/, /i/, /o/, and /u/.

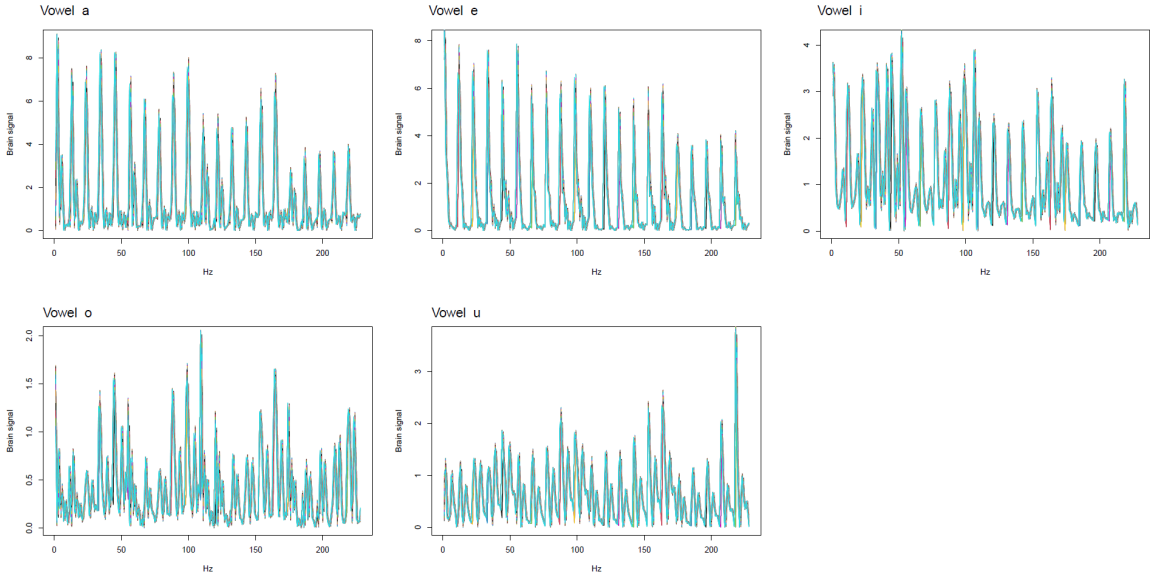

FIGURE 3.1: EEG measurements of the subject thinking about an specific vowel recorded from 21 locations.

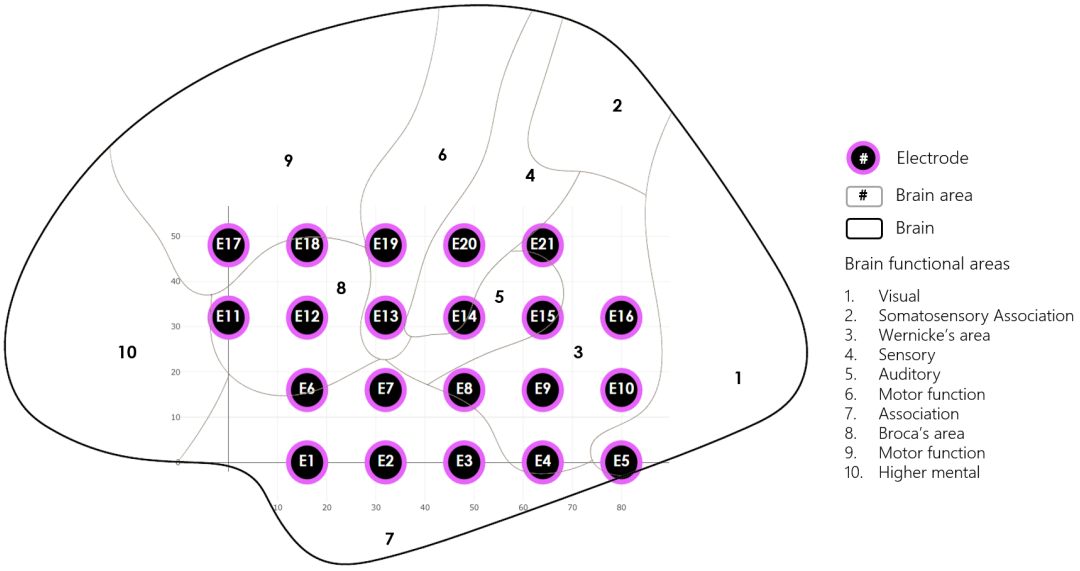

FIGURE 3.2: Left brain hemisphere approximation using Euclidean distance with the location of the 21 electrodes.

When a person thinks about a particular vowel, it generates a noisy signal captured by each electrode with a sampling frequency of 250 Hz. The EEG signals were recorded at the same brightness and noise levels. Nonetheless, these neural signals may be affected by various factors, including the subject's heartbeat, blinking and swallowing.

3.2 Software

In this work, we use the packages fda.usc (Bande et al., 2022), gstat (Pebesma, 2004), and SpatFD (Bohorquez et al., 2016) available in the software R version 4.3.1. The first package is used for transforming the discrete EEG values into curves, exploring functional descriptive statistics, and conducting functional principal component analysis. The second package, gstat, enables the variogram estimation, including the selection of a model C and initial parameters Θ_0. Finally, we employ SpatFD for generating the spatial functional data, and performing the spatial prediction of the EEG curves at unsampled locations using functional kriging. The primary advantage of utilizing functional data is that it represents multiple data points in a single dimension, thereby facilitating efficient computation.

3.3 Methods

This section outlines the workflow for generating the spatial prediction of the EEG curves representing brain activity in the left hemisphere during a specific vowel task using inner speech. The initial two sections focus on the methods developed to select one subject from the 22 participants of the experiment. This choice is driven by our interest in analyzing the spatial functional data for a single random field. The criteria for participant selection depend on the exploration and comparison of the minimum and maximum values of the pointwise descriptive statistics of the signals, as well as the detection of outliers. Consequently, we choose the subject whose EEG curves do not exhibit extreme high, or low, signal values and multiple electrode outliers for each vowel task. The remaining sections correspond to steps for spatial prediction, which include the construction of the spatial functional data with the functional principal component analysis results, the variogram estimation, and functional kriging. Figure 3.3 displays the workflow to predict the spatially correlated EEG curves.

3.3.1 Functional data construction

The construction of functional data involves specifying a set of basis functions and coefficients, and defining a linear combination of these basis functions as follows:

$$\chi_i(t) = \sum_{j=1}^{K} a_{ij}\phi_j(t)$$

The discrete EEG data is transformed into functions by using a set of B-splines of order four with evenly separated knots. The advantages of using B-spline functions include fast computation and great adaptability. Moreover, the basis system is capable of representing data with a significant number of points by employing a

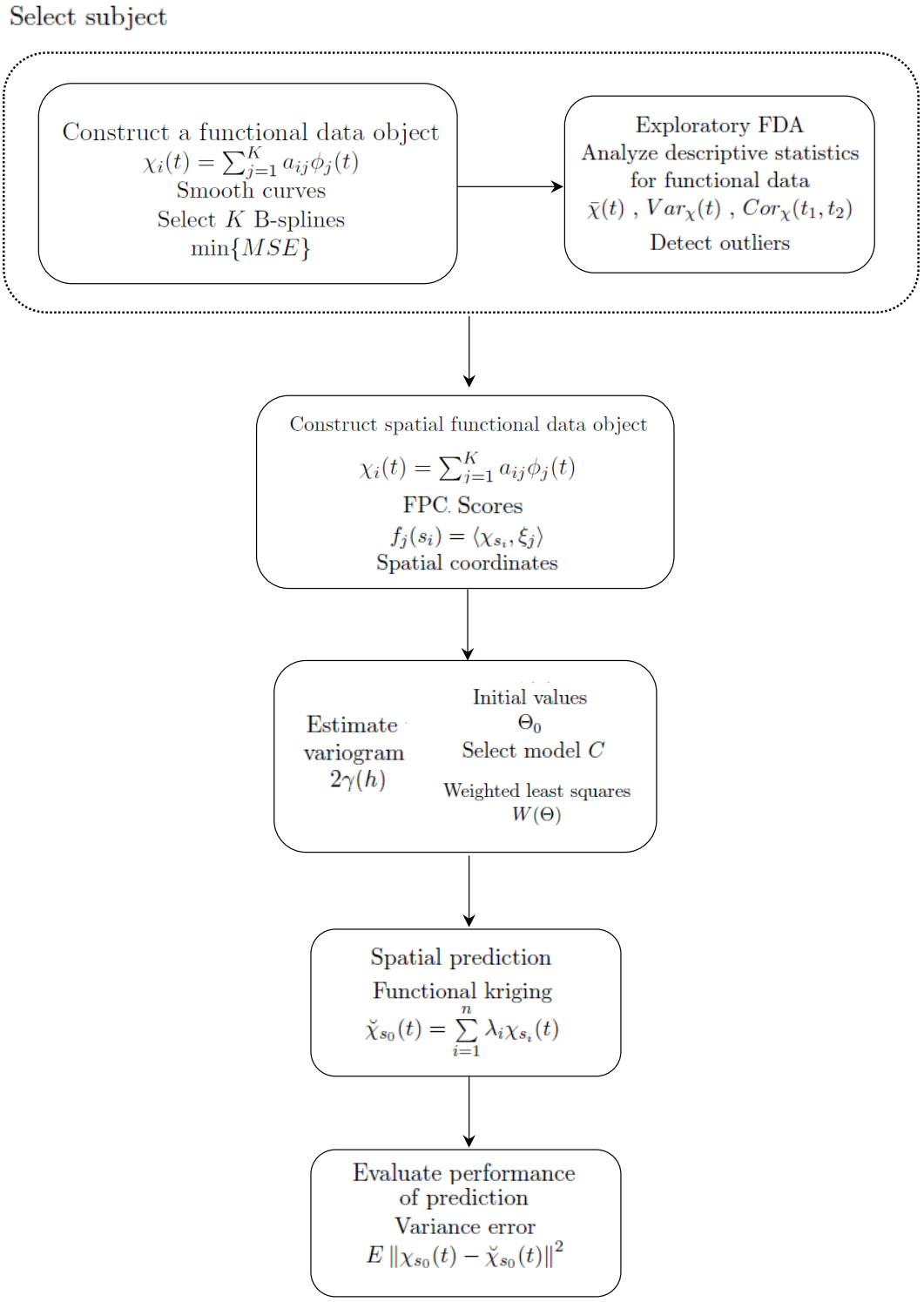

FIGURE 3.3: Workflow to perfom the spatial prediction of correlated EEG curves generated by one participant

smaller number of coefficients. The number of K B-splines was chosen by cross-validation. The selection of basis functions is based on the criteria of minimizing the mean squared error (MSE).

3.3.2 Exploratory FDA

Exploratory functional data analysis is conducted to summarize and explore the general structure and main characteristics of the multivariate functional data for each vowel. We analyze and compare the behavior of every vowel by graphical exploratory analysis using pointwise descriptive statistics such as mean ($\bar{\chi}(t)$), variance ($Var_{\chi}(t)$), and correlation ($Cor_{\chi}(t_1, t_2)$) defined in section 2.3.2. These analog functions from traditional measures have similar interpretations and can be calculated by using the pointwise statistic across the replications.

Furthermore, the trimmed mean and median are computed by employing depth notions to contrast them with the pointwise mean of every vowel. The concept of depth aims to measure the centrality of a given curve within a group of curves (Fraiman & Muniz, 2001). In this work, we use two types of depth notions: mode and random projections. The depth mode function implements the modal depth that selects the deepest and most densely surrounded curves. The random projections' depth function is computed through random projections.

The functional boxplots allow the detection of outliers. Hyndman and Shang (2010) proposed the functional bagplot and the functional highest density region (HDR) boxplot for outlier detection. These functional boxplots are based on the first two robust principal component scores. The functional boxplot shows that the curves outside the outer region represent outliers (Sun & Genton, 2011).

3.3.3 Functional principal component analysis

Functional principal component analysis (FPCA) is employed to analyze and reduce the dimension of functional data. It is an extension of conventional principal component analysis (PCA) specifically designed for functional data (Ramsay & Dalzell, 1991). FPCA explores the variations inherent in functional data, incorporating both spatial and temporal aspects to identify patterns, trends, and relationships within the data. In the FCPA extension, the vectors from the classical PCA are replaced with functions, matrices with linear operators, and summations with integrations (Shang, 2014).

Let $\chi_s(t)$ be a set of spatial continuous EEG curves with mean $\mu = E(\chi_s(t)) = 0$, given the covariance operator C (see 2.11), the FPC are the eigenfunctions $\xi_j(t)$ of the covariance operator C of $\chi_s(t)$, and the estimator of FPC are the EFPC. In FPCA, the functional principal components are associated with eigenvalues that retains the amount of variability explained by the corresponding component. These principal components facilitate the projection of the EEG curves in the

direction of major variability. Each EEG function has an infinite dimension associated with a spatial location. We aim to reduce the infinite dimension $t \in [0, T]$, across the frequency domain to n finite dimensions that explain more than 85% of the variability.

The reconstruction of the curve χ_{s_i} can be done using a basis function system using the EFPC (see 2.12). Therefore, we can use the EFPC representation $f_j(s_i)$ of the EEG curves $\chi_{si}(t)$ to construct the spatial functional data.

The spatial functional data object is constructed from the spatial coordinates of the 21 electrodes and the EEG curves observed at each site. The spatial object storages the smoothed curves, their parameters, the EFPC representation of the EEG curves obtained in the functional principal component analysis, and the coordinates of the electrodes.

3.3.4 Variogram estimation

The variogram 2γ is estimated by using the weighted least squares estimator proposed by Cressie (1985). The initial values Θ_0 are obtained based on the visual exploration of the EFPC representation of the EEG curves using the eyefit tool in R. The spatial structure dependencies are modeled by selecting a model C (e.g., circular, gaussian, spherical, exponential). The available variogram models are displayed in Figure 3.4.

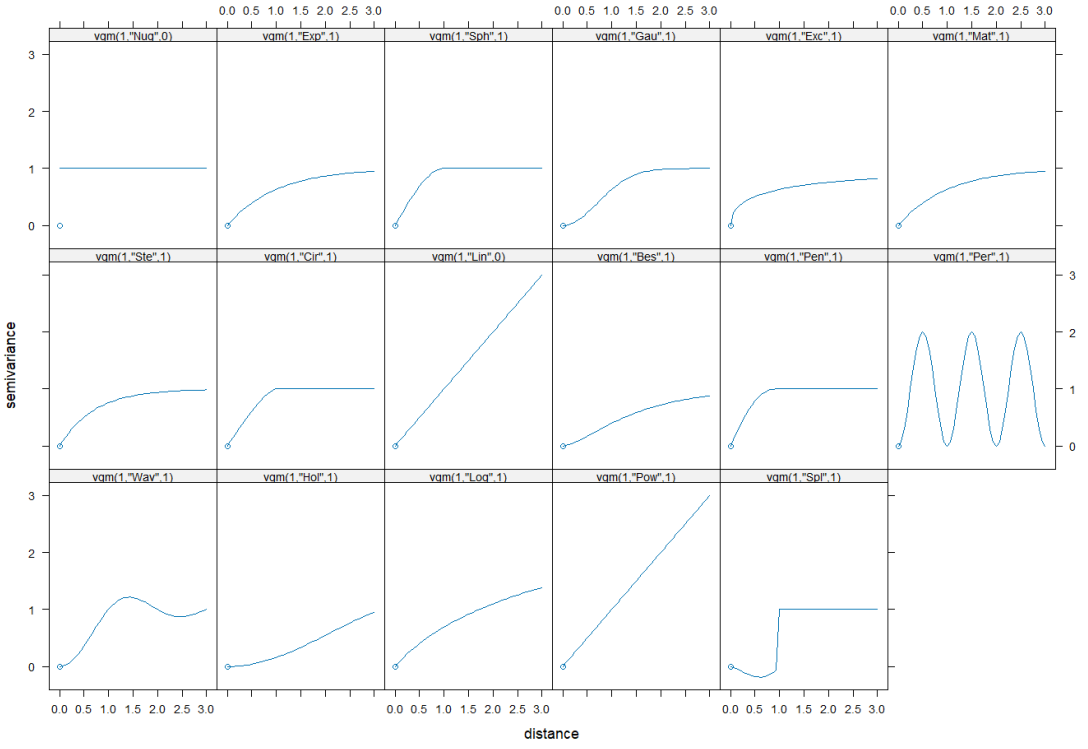

FIGURE 3.4: Variogram models available in gstat

3.3.5 Functional Kriging

Once the spatial EEG curves are represented in terms of their empirical functional principal components and the covariance parameter is estimated, the spatial prediction of the curves at unsampled sites can be done by using functional kriging.

The spatial functional prediction is based on the assumption of a known mean $\mu(t)$, and employing a linear combination of the observed functions or curves as follows (see 2.16):

$$\breve{\chi}_{s_0}(t) = \sum_{i=1}^{n} \lambda_i \chi_{s_i}(t)$$

The performance of the prediction can be assessed by minimizing the prediction variance error $E\,\|\chi_{s_0}(t) - \breve{\chi}_{s_0}(t)\|^2$. We use the leave-one-out functional cross-validation method that removes one EEG observation at a time from the dataset, then fits the model with the remaining data, and then performs the prediction for the left-out observation (Bohorquez et al., 2016).

Chapter 4

Results

This chapter presents and discusses the principal findings when exploring the patterns and characteristics of the signals for the selected participant number 4. The first section presents the results of the exploratory functional data analysis and functional principal component analysis. The second section shows the results of modeling the variogram and brain activity maps.

4.1 From Discrete to Functional Data

Considering $y_{i,j}$ an observation in the location $i = (1, \ldots, 21)$ and frequency $j = (0, ..., 250)$, the model in terms of time and space can be expressed as $y_{i,j} = x_i(t_j) + \epsilon_{i,j}$, where x_i represents the smooth functions and $\epsilon_{i,j}$ the error. The smoothed curves x_i are estimated by using a basis functions based approach. The most common ones are the Fourier basis and the B-splines basis. Due to the EEG signals corresponding to non-periodic data, the B-splines basis was used instead of Fourier, which fits more with periodic data showing repeated cycles.

To transform the discrete dataset into functions, we use a set of B-splines of order four. The number of B-splines was chosen by cross-validation. The basis functions are selected based on the criteria of minimizing the mean squared error (MSE). Figure 4.1 shows the plot of the MSE for a specific number of basis functions. The number of B-splines chosen for each vowel is 11 for vowel /a/, 12 for vowel /e/, 13 for vowel /i/, 12 for vowel /o/, and 13 for vowel /u/. This was also verified by visual inspection to examine the effect of using a different number of B-splines.

Figure 4.2 shows the multivariate functional data with 21 smoothed curves for open vowels (/a/, /e/, and /o/) and closed vowels (/i/, /u/); each curve represents one electrode.

We note that vowels /a/ and /e/ show high signal values in electrode 1 (greater than $7.5\mu V^2/Hz$). This electrode is located in the association region of the brain,

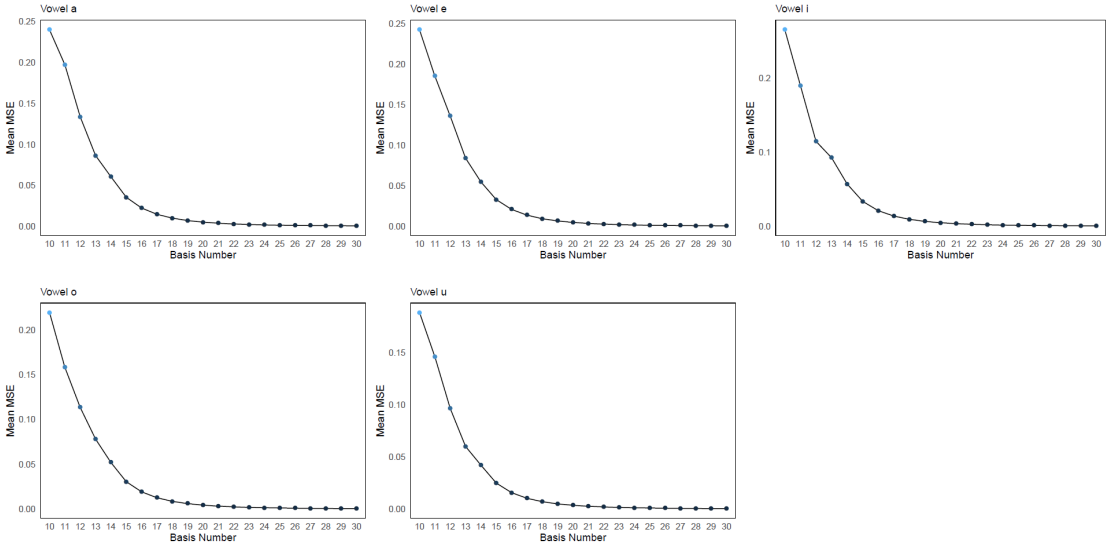

FIGURE 4.1: Plot of the MSE obtained through cross-validation against the corresponding number of basis functions to fit the EEG smoothed curves

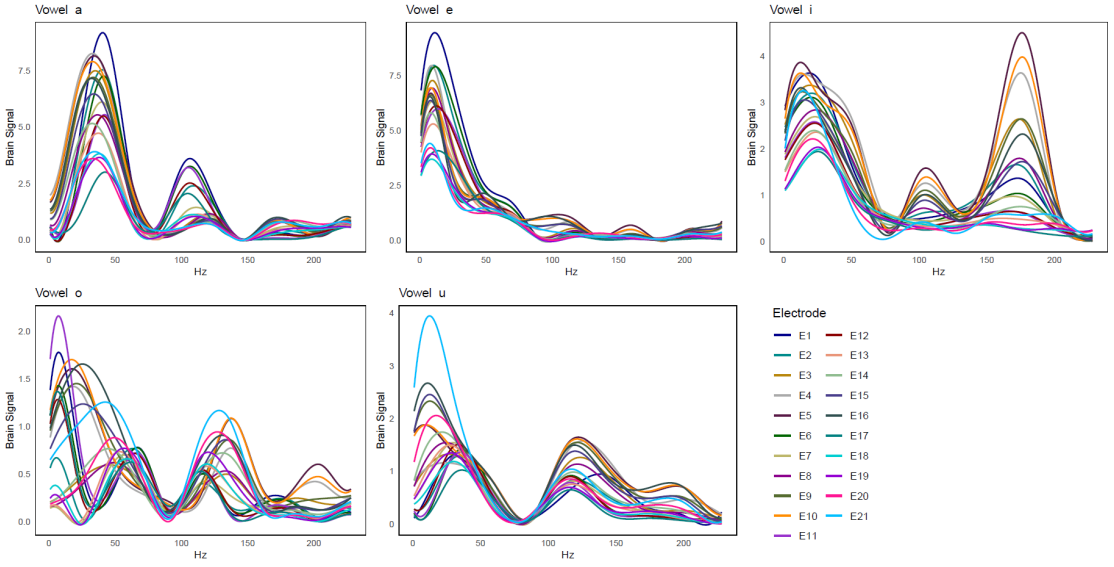

FIGURE 4.2: Smoothed EEG signals data using B-splines basis functions for each vowel

which might indicate that when the task starts, the brain associates the input information with memories of previous experiences. The vowels /a/ and /e/ show similar behavior, with signal values varying from 0 to about 8 $\mu V^2/Hz$. The vowel /i/ shows high brain activity in the intervals 0–50 Hz and 150–250 Hz. However, the signal levels are low, varying from 0 to about 4 $\mu V^2/Hz$. In the case of the vowel /o/, electrode 11 exhibits high signal levels. This electrode is located close to the Broca's area, which activates speech muscles for speech production. For the vowel /u/, the electrode showing high values is 21. This electrode is located near Wernicke's area, which is involved with speech comprehension.

4.2 Exploratory FDA

The exploratory analysis is integrated with descriptive statistics such as functional mean, variance, trimmed mean and median (using different depth notions), and correlation. These analog functions of the classical measures have similar interpretations and can be calculated by computing pointwise measures across the replications. We also identified outliers using functional boxplots.

4.2.1 Descriptive statistics for functional data

The mean and variance curves show the distinctive patterns and variability of the EEG signals in Figure 4.3. The mean and variance curves of vowels /a/ and /e/ exhibit similar behaviors, varying from 0 to 6 $\mu V^2/Hz$. The peaks occur around the frequency range of 0 to 50 Hz. For the vowel /i/, the mean values vary from 0 to 3, while the variance values are between 0 and 2 $\mu V^2/Hz$. Vowel /o/ shows the lowest signal values, reaching a maximum value of 1 in both statistics. Finally, the mean and variance values of the vowel /u/ vary between 0 and 2 $\mu V^2/Hz$.

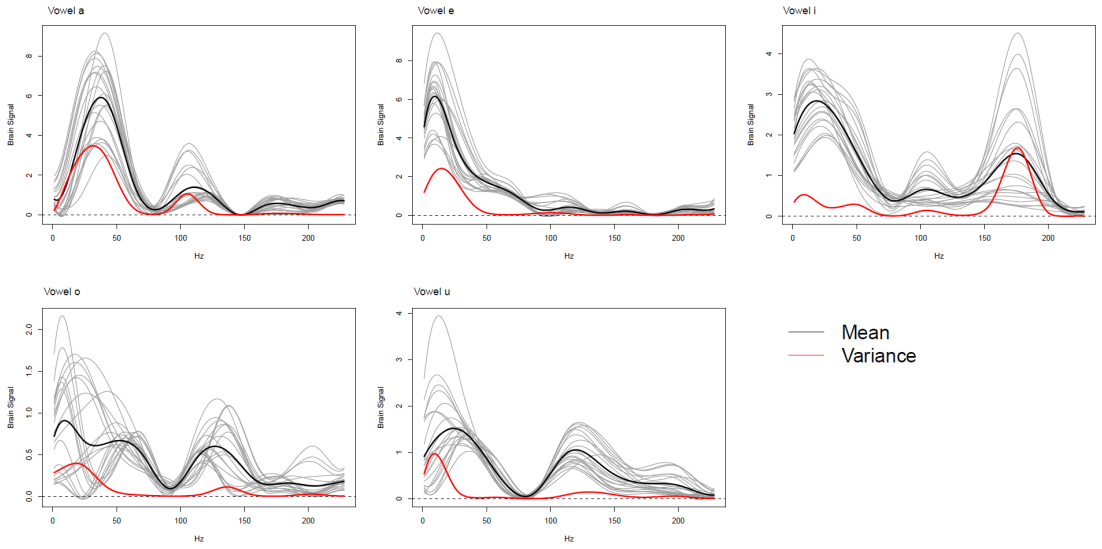

FIGURE 4.3: Smoothed curves representing the functional mean in black and variance in red. The gray lines correspond to the EEG smoothed curves.

Trimmed mean and median curves, using mode and random projections (RP) depth, were also plotted. The concept of depth aims to measure the centrality of a given curve within a group of curves (Fraiman & Muniz, 2001). The depth mode function implements the modal depth that selects the deepest and most densely surrounded by others, and the RP depth function is computed through random projections. Figure 4.4 displays the trimmed mean and median curves. The curves follow similar patterns compared to the mean curve for all five vowels, except for the vowel o, where the trimmed mean varies between about 0.5 and 0.8 $\mu V^2/Hz$ in the interval 0–50 Hz.

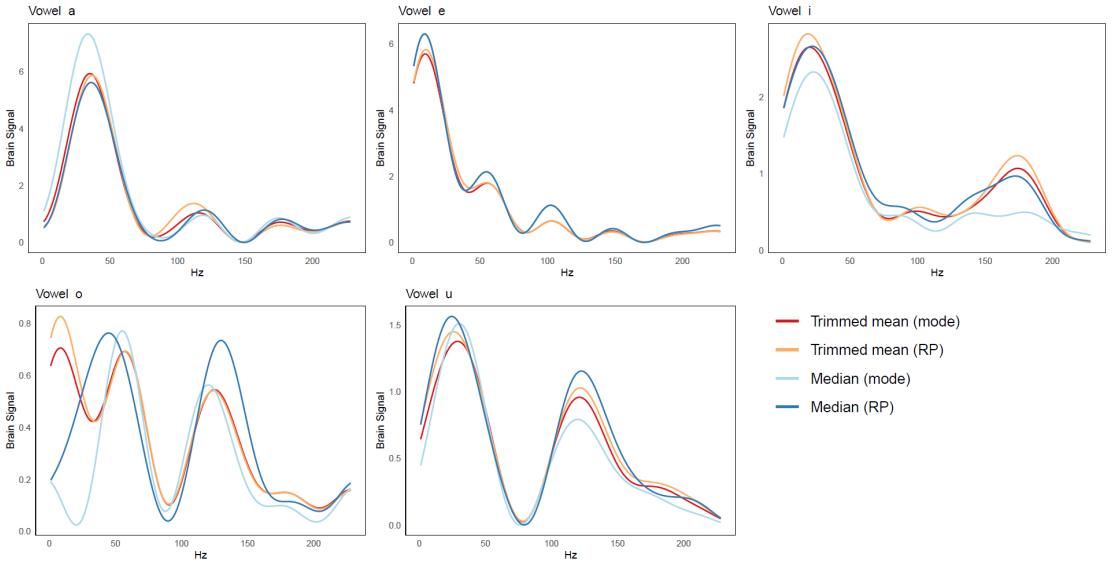

FIGURE 4.4: Measures of central tendency point-wise trimmed mean and median using depth notions for all five vowels

The correlation function displays the temporal dependence within EEG signals. In classical correlation, pairs of data are considered, but in the context of functional data, the correlation is computed between every two time points along the curves. Figure 4.5 presents a symmetrical correlation contour plot, where the color gradient indicates high correlation values in red and low values in blue.

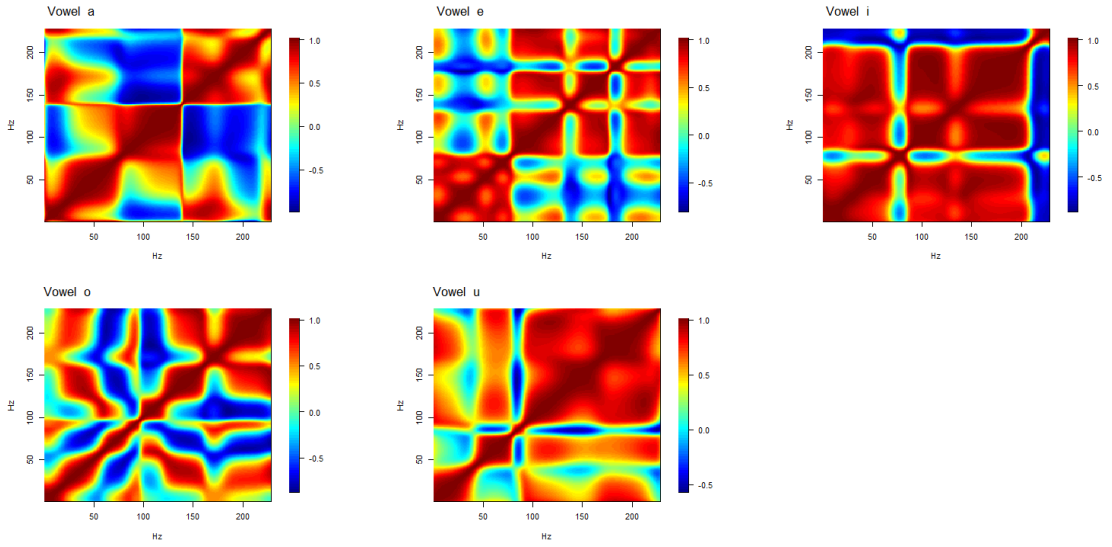

FIGURE 4.5: Contour plot of the correlation between frequency points per vowel. High correlation values are displayed in red and low correlation values in blue

At first glance, the interpretation can be challenging. The diagonal running from the lower left to the upper right corners is equal to one, signifying the correlation of the frequency points with themselves. In the correlation contour plot for the

vowels /i/ and /u/, there is a strong correlation across frequency points in the interval 100–200 Hz.

4.2.2 Outlier detection

To address the presence of outliers in the functional data, we employed two visualization techniques: the functional bagplot and the functional highest density region (HDR) boxplot. Functional boxplots use the first two robust principal component scores to map features into the functional space.

Figure 4.6 shows the functional bagplot which is constructed by inflating the inner region (bag region) with a constant factor and specifying a coverage probability for the outer region (fence region). This plot allows the visualization of outliers for each vowel. The functional bagplot of open vowels /a/, /e/, and /o/ does not exhibit outliers.

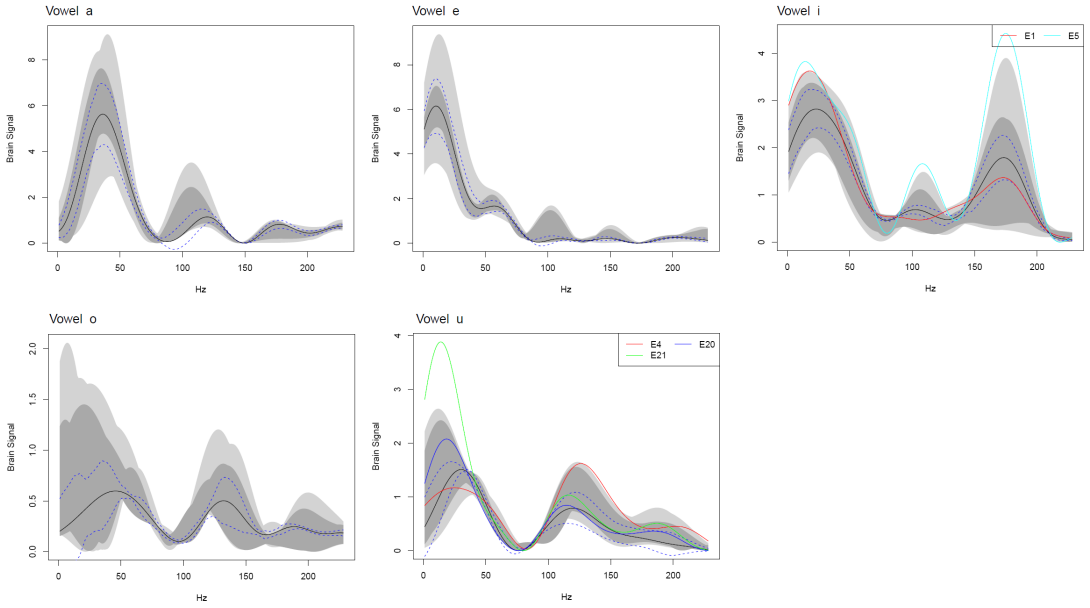

FIGURE 4.6: Functional bagplots for outlier detection, the black line represents the median curve surrounded by 95% pointwise confidence intervals.

The functional boxplot is effective when outliers are considerably distant from the median. However, it may misidentify outliers when they are in proximity to the median. In such situations, the functional HDR boxplot becomes a more suitable outlier detection method. In this boxplot, the dark and light gray regions represent the 50% the highest density region (HDR) and outer HDR, respectively. The black line denotes the modal curve, and curves outside the outer region are identified as outliers. Figure 4.7 reveals that electrode 1 is an outlier for vowels /a/, /e/, and /i/, while electrodes 11 and 21 are outliers for vowels *o*, and *u*, respectively. This suggests that high brain activity is produced in the association, Broca's and Wenicke's areas.

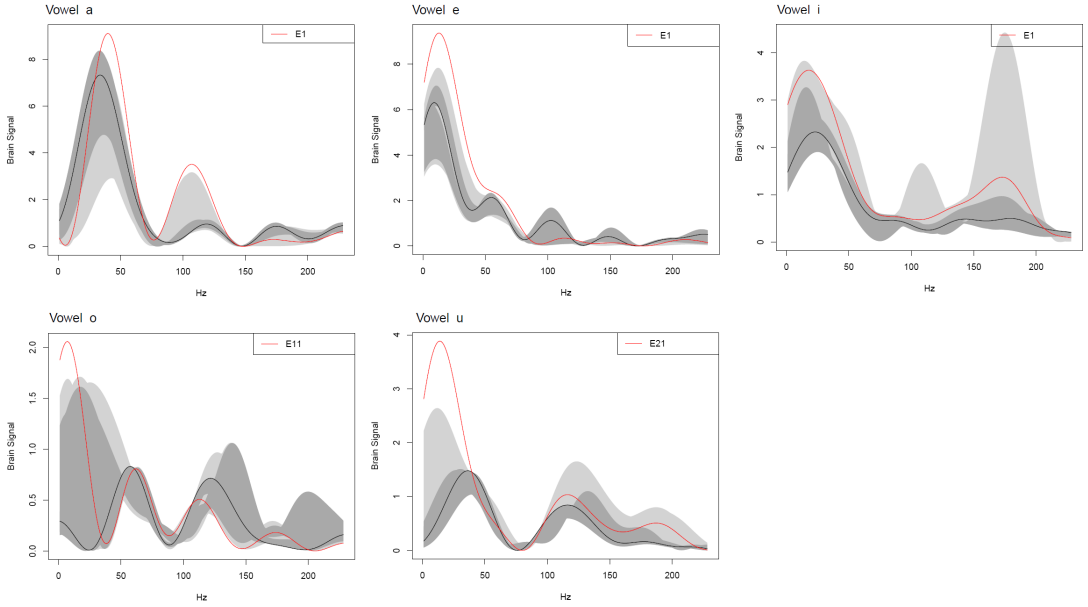

FIGURE 4.7: Functional HDR boxplots for outlier detection. The black line denotes the modal curve, and curves outside the outer region represent outliers

4.3 Functional Principal Component Analysis

Functional principal component analysis is a widely used methodology for exploring and summarizing variability within functional data. FPCA aims to decompose the variability inherent in functions into a few functional principal components. These components represent the dominant patterns of variation present in the functional data. The mean is subtracted from each variable to guarantee that the functional principal components describe the direction of the most significant variation. In this section, we apply the FPCA technique to reduce the dimension of the data and identify outliers and clusters.

Figure 4.8 illustrates the first two functional principal component curves for every vowel, which are the most significant contributors to the variance explanation. The functional principal components show that component 1 is responsible for the highest deviations from the mean in the frequency range 0–50 Hz for all vowels, except for the vowel /i/, where the highest deviation is attributed to component 2. The first two functional principal components explain more than 85% of the total variation for each vowel. The cumulative explained variance in the first two functional components is 97.5%, 97.2%, 96.5%, 91%, and 95.6% for vowels /a/, /e/, /i/, /o/, and /u/, respectively. The rest of the components explain small proportions of the variation in the data. The explained variance of the first three functional principal components is presented in Table 4.1.

Considering the first two principal components, which account for most of the variance in the EEG signal data, we utilize the score plot to assess the underlying data structures. This visual representation helps in identifying clusters and outliers in the EEG dataset. In Figure 4.9, the score plot displays the results for each vowel.

FPC	Vowel a	Vowel e	Vowel i	Vowel o	Vowel u
1	0.685	0.810	0.917	0.607	0.778
2	0.290	0.162	0.048	0.303	0.178
3	0.016	0.020	0.031	0.067	0.023

TABLE 4.1: Explained variance by the eigenfunctions of EEG signals data

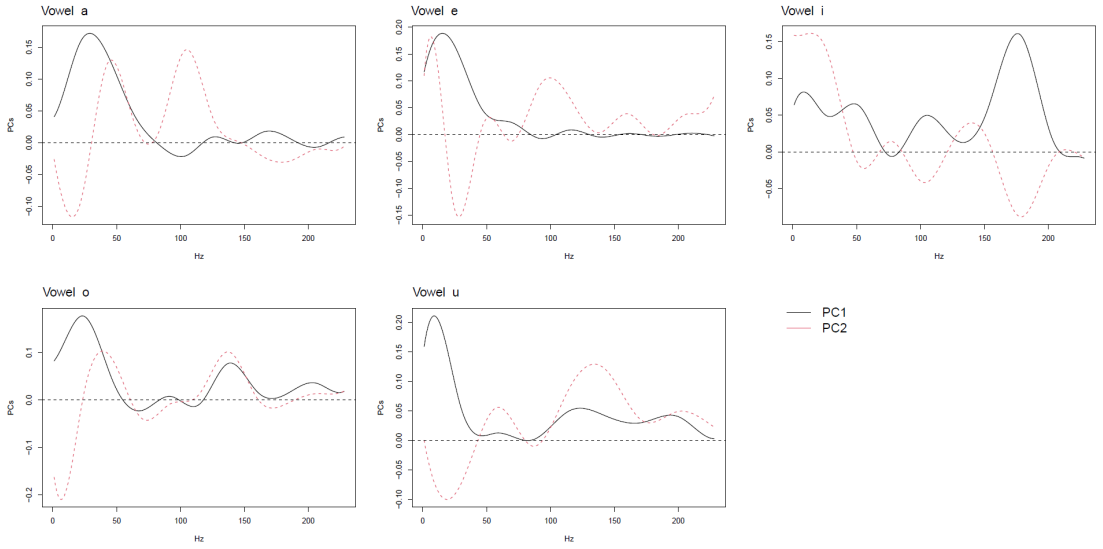

FIGURE 4.8: First two functional principal component curves for every vowel. The black line represents the first FPC, and the red one the second FPC.

The score plot of the vowel /a/ reveals a cluster formed by electrodes 4, 5, and 10, capturing signals in the Wernicke's area. In contrast, electrode 17, located near the brain's motor function area, exhibits low values in the direction of the first functional principal component (FPC1), while electrode 1, associated with the brain's association area, displays the highest values. This discrepancy suggests that electrode 1 is an outlier or an extreme observation.

Moving to the vowel /e/, electrodes 3, 7, and 8, positioned close to the origin in the score plot, indicate values similar to the mean. Electrodes 1, 2, and 6 contribute the most to FPC1, while electrodes 5, 10, and 10 cluster together with high values in the direction of the second functional principal component (FPC2). For the vowel /i/, electrodes 13, 14, 17, 19, and 20 exhibit the lowest contributions to either FPC1 or FPC2.

Analyzing the vowel /o/ electrodes 4, 5, and 10 shows similar patterns with high values in the direction of FPC1. Notably, strong trends are observed in the vowel /o/, with the majority of electrodes either contributing significantly to FPC1 (e.g., electrodes 4, 5, 9, and 10) or FPC2 (e.g., electrodes 3, 8, 14, and 20).

In the case of the vowel u, electrode 21 strongly aligns with the major trend represented by FPC1. Electrodes 5 and 10 consistently appear close to each other in all score plots, indicating similar patterns across the five different vowel values.

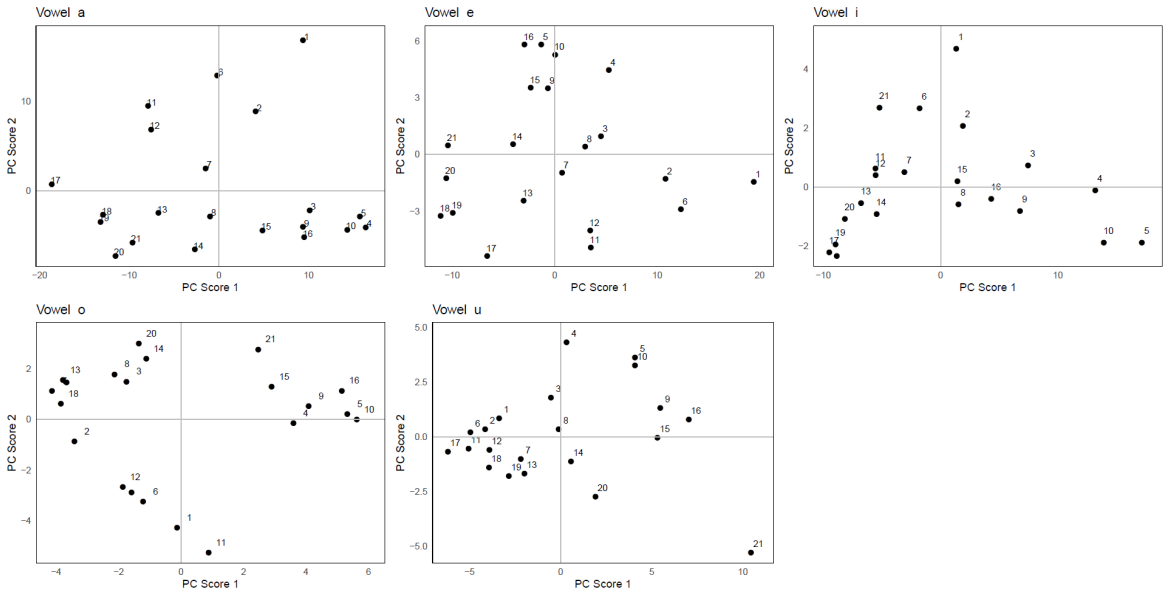

FIGURE 4.9: Score plot of the first two functional principal component displaying the 21 electrodes

4.4 Variogram estimation

We modeled the spatial dependencies of the scores associated with the first and second functional principal components to predict the spatial functional data. As we have seen in the previous section, the first principal component explains 68.5%, 81%, 91.7%, and 60.7% of the variability for vowels /a/, /e/, /i/, /o/, and /u/, respectively, and the second principal component only accounts for 29%, 6.2%, 4.8%, and 1.78%. Therefore, it is enough to include the first two score vectors to model the variogram and then use functional kriging.

	EFPC1			EFPC2		
Vowel	Model	Sill	Range	Model	Sill	Range
a	Hole	105	12	Hole	44.85	16.08
e	Wave	59.89	34.08	Wave	11.99	49.08
i	Wave	55.6	32.24	Wave	3.18	31.17
o	Periodic	10.53	154.67	Hole	4.98	11.60
u	Periodic	19.92	161.17	Periodic	4.55	161.17

TABLE 4.2: Estimated model parameters using the weighted least squares estimator for the EEG curves representation using the first and second EFPC

Figures 4.10 and 4.11 illustrate the weighted least squares estimator for Θ using the first and second EFPC representation of the spatially correlated EEG curves. The nugget parameter is fixed and equal to zero because there is no reason to assume discontinuity at the origin. Table 4.2 displays the fitted model parameters using the weighted least squares estimator. The semivariogram models demonstrate continuous growth, indicating strong spatial autocorrelation structures in the left

hemisphere of the brain, where electrodes with significant distances between them remain correlated.

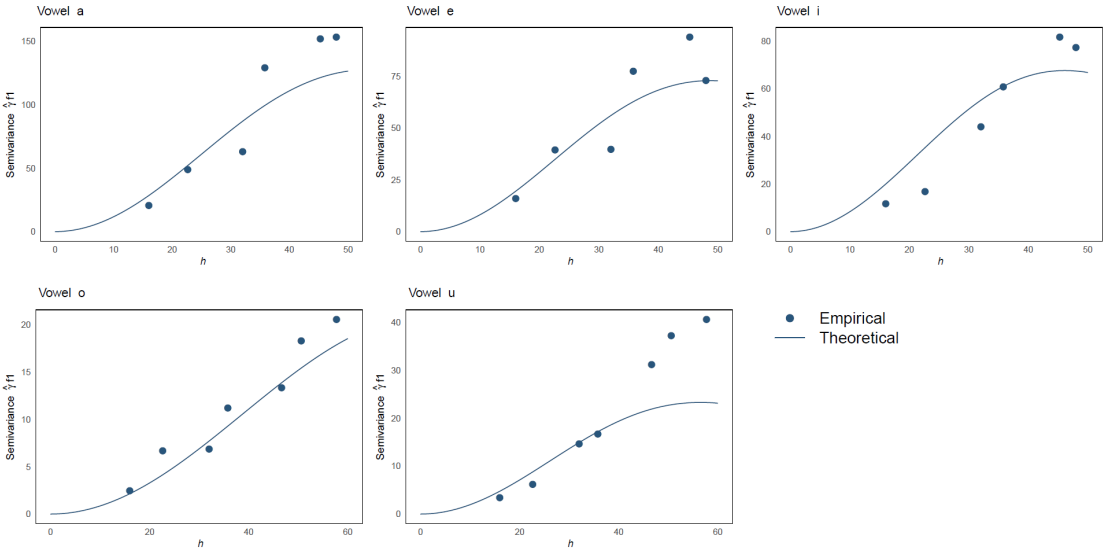

FIGURE 4.10: Weighted least squares estimator for Θ using the first EFPC representation of EEG curves.

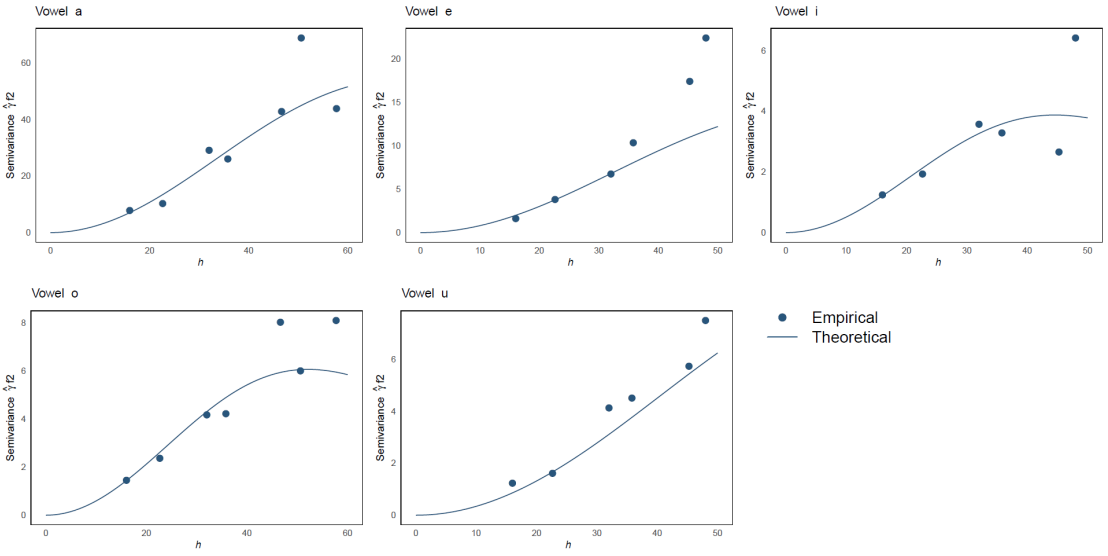

FIGURE 4.11: Weighted least squares estimator for Θ using the second EFPC representation of EEG curves.

4.5 Spatial prediction

Figure 4.12 illustrates the graphical comparison between observation and prediction for electrode 9, located in Wernicke's area, where the predicted curves are similar to the observed functions. The EEG predicted curve of the vowel /u/ shows significant differences due to the low variation range of the signals. The predictions

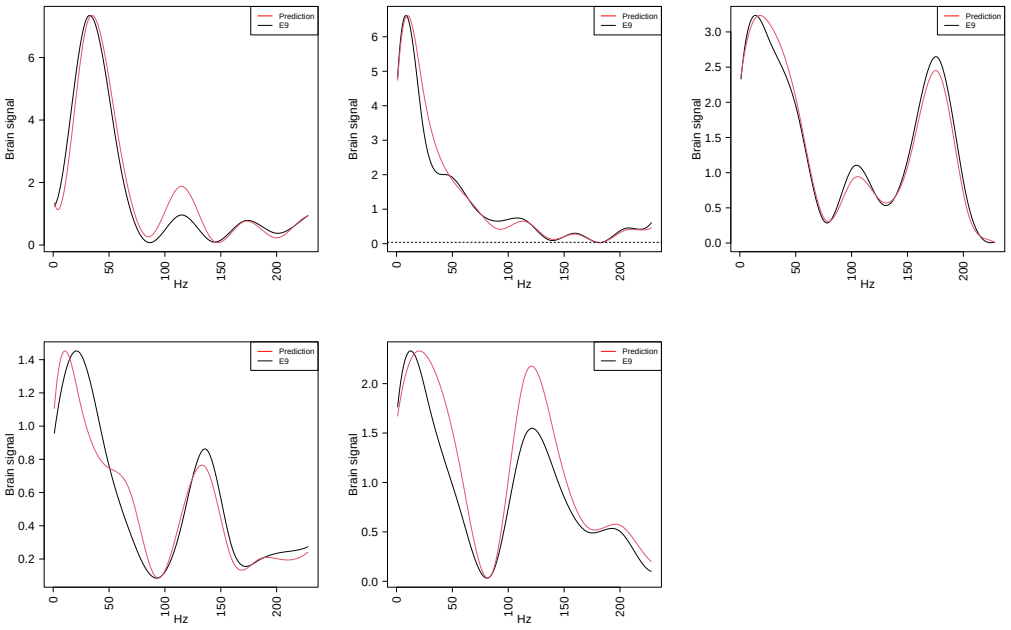

FIGURE 4.12: Predicted vs observed EEG curves for electrode 9 of vowels /a/, /e/, /i/, /o/ and /u/

are obtained using the functional kriging with the scores resulting from the EEG curve EFPC representations. To evaluate the performance of the spatial prediction, we use the leave-one-out functional cross-validation method. This method involves removing each functional observation and using the remaining functional observations to predict a smoothed function at the removed location (Bohorquez et al., 2016). Table 4.3 presents the summary statistics for the prediction error obtained by functional kriging per vowel. The results indicate that the prediction for the vowel /o/ has better performance than other vowel predictions.

Statistic	a	e	i	o	u
Minimum	26.05	12.26	10.86	3.921	7.824
Median	36.37	37.51	23.43	8.254	22.139
Mean	38.84	41.41	22.51	9.497	42.784
Maximum	85.12	94.94	37.63	20.149	256.871

TABLE 4.3: Summary statistics for the prediction error obtained by functional cross-validation for each vowel

The brain maps for each vowel at a certain frequency are shown in Figures 4.13, 4.14, 4.15, 4.16 and 4.17. Figure 3.2 describes the brain areas numbered in these maps. Wernicke's area (#3), which is related to speech comprehension, is activated with great intensity in the vowels /a/ and /o/. However, Broca's (#8) and association's (#7) brain areas are activated with more intensity by the vowels /e/, /i/, and /o/. The distinction of brain activity in the vowel /u/ is challenging due to low signal levels as well as low variation ranges.

Across all the frequency points, the brain activity goes from either the Broca's (#8) or association area (#7) to the Wernicke's area (#3) in every vowel. For instance, for the vowel /a/ task using inner speech, the brain exhibits increased activity in the association region that associates the task with previous experiences. Then, the visual and Broca's areas are activated for image perception, recognition, and speech production. Finally, the region in Wernicke's area is activated. The brain temporarily initiates knowledge regarding the vowel, associates it with a mental representation, and then transmits a command to the speech articulatory muscles, without producing a sound using inner speech.

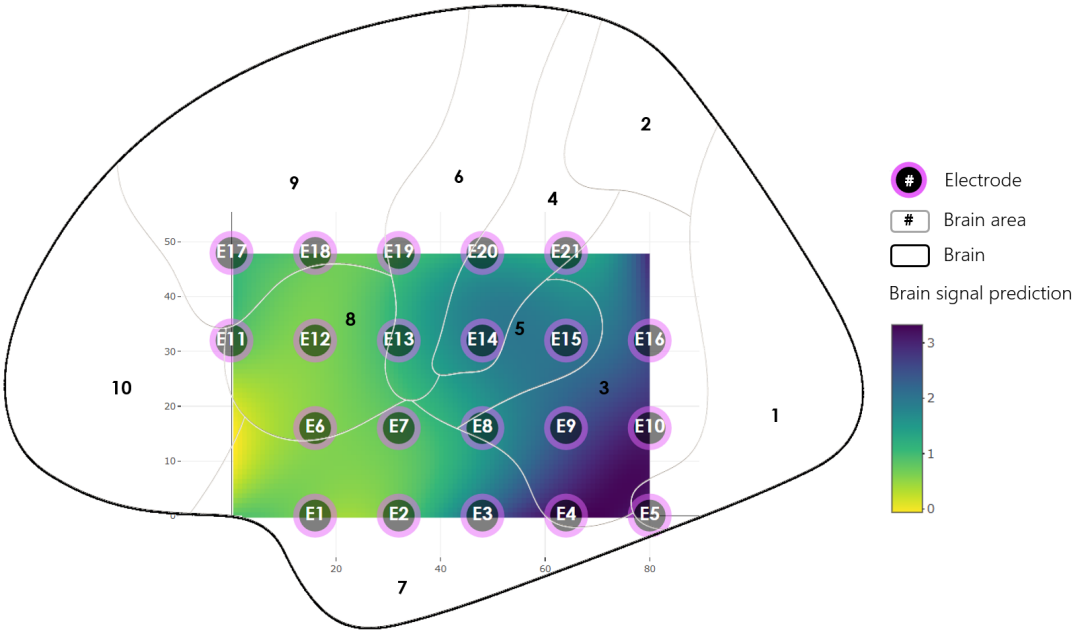

FIGURE 4.13: Brain activity map vowel /a/

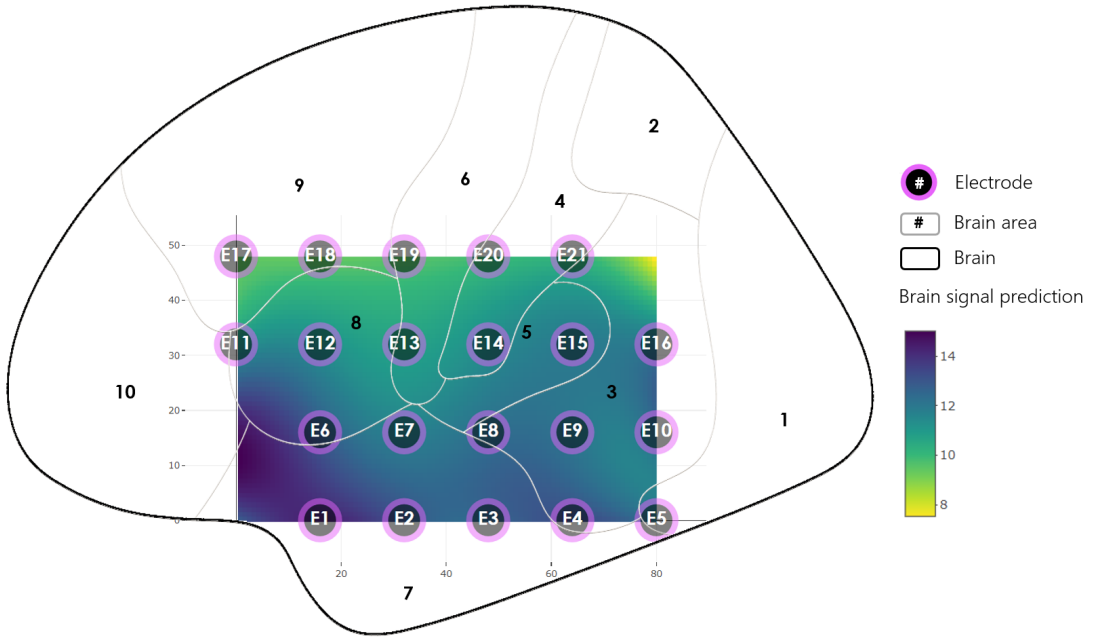

FIGURE 4.14: Brain activity map vowel /e/

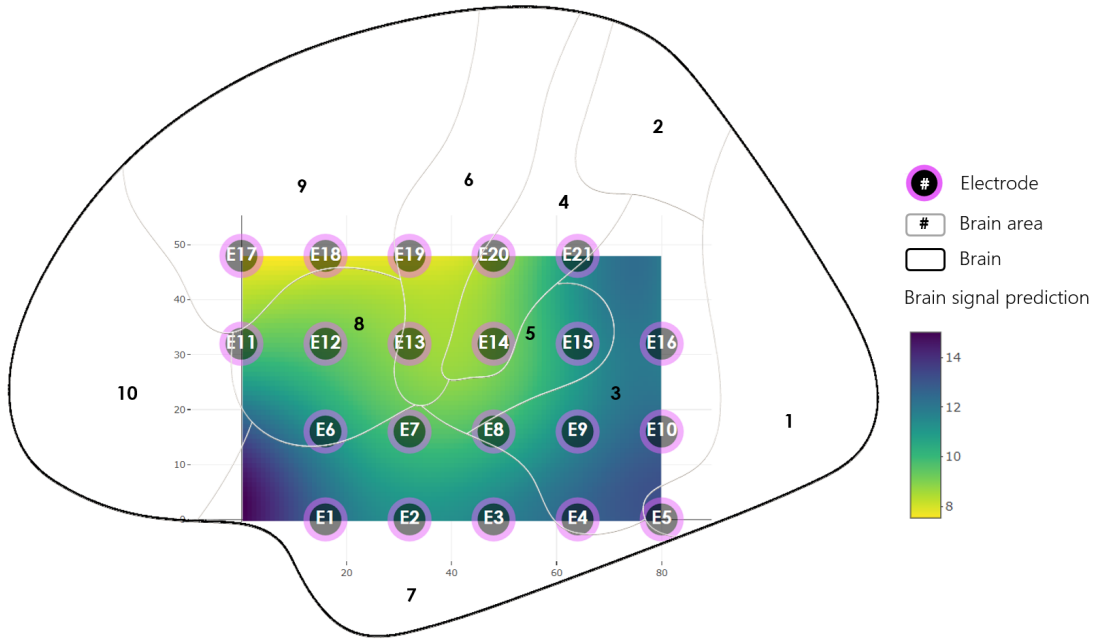

FIGURE 4.15: Brain activity map vowel /i/

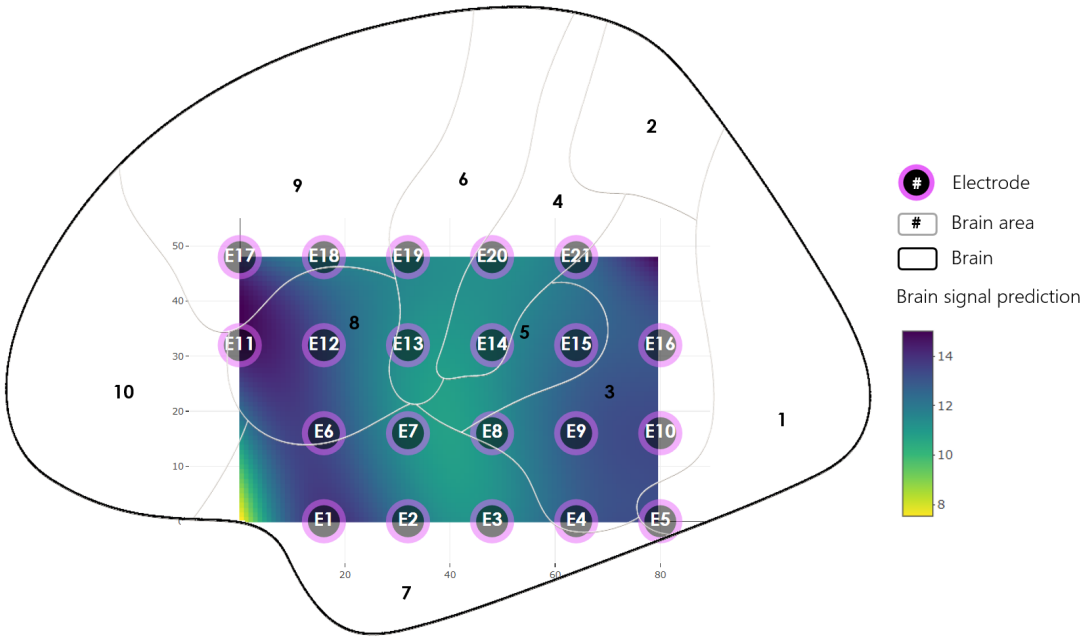

FIGURE 4.16: Brain activity map vowel /o/

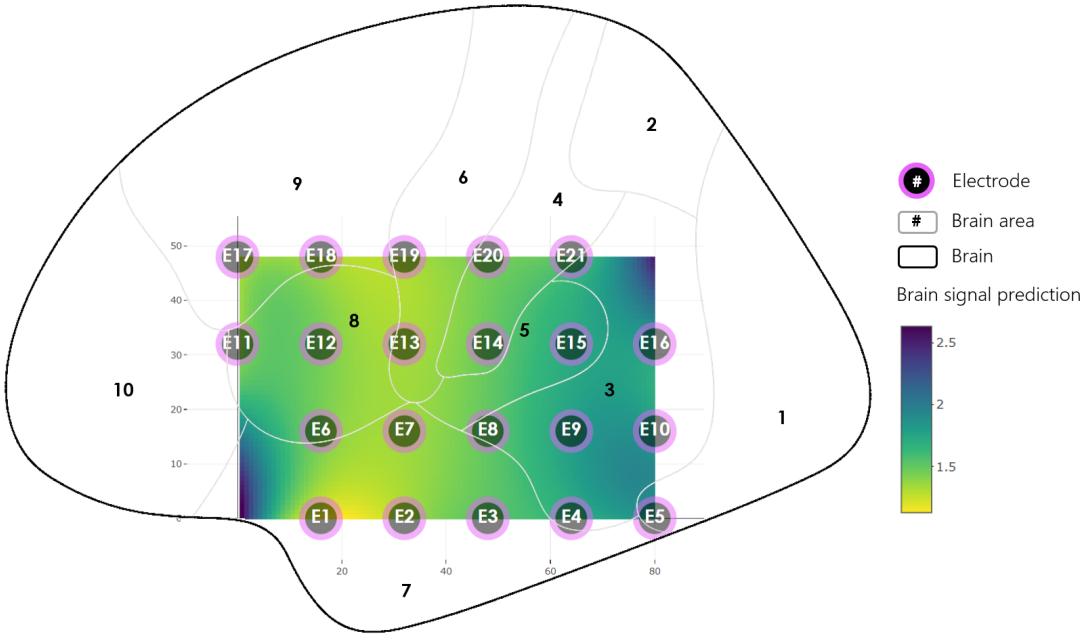

FIGURE 4.17: Brain activity map vowel /u/

Chapter 5

Discussion, future work and conclusions

5.1 Discussion

The electrical activity in the brain is represented by the EEG signals, which can be analyzed as correlated functional data. This work was focused on applying functional data analysis and functional geostatistics methods to explore EEG signals produced during inner speech. The discrete EEG data was converted to functions or curves using B-splines. The analysis was conducted on the resulting smoothed curves to extract the main features and characteristics of the signals. The exploratory functional data analysis revealed distinctive variations in the EEG signals. For instance, the brain activity for the vowels /a/ and /e/ is higher than that for the vowels /i/, /o/, and /u/. Furthermore, in the case of the vowel /a/, the brain areas with high activity are the association and Wernicke areas, and with low brain activity, the motor function area.

EEG recordings are highly correlated across time and space. As demonstrated in the FPCA results, the reduction of the temporal dimension was sufficient by using the first two principal components, which account for more than 85% of the total variation for each vowel. This method also revealed spatial structures within certain electrodes across the five vowels. For example, the spatial clusters formed by electrodes 5 and 10 located in Wernicke's region, and electrodes 18 and 19 in Broca's area.

The empirical functional principal components were used to model the spatial dependencies. The variograms revealed strong spatial autocorrelation structures. The continuous growth pattern in the variograms, even as the distance increases, indicates that electrodes closer together are spatially autocorrelated, and electrodes that are further apart remain correlated. Finally, we employed simple kriging for predicting spatial functional data at unobserved locations. The performance of the spatial prediction was addressed using leave-one-out functional cross-validation. Although the predicted and observed functions were closely aligned,

there were differences between the predicted and observed curves for the vowel /u/. These differences were attributed to the low variation ranges of the signals.

In every vowel task, the brain activity proceeds from either the Broca's or association area to the Wernicke's area across all frequency points. Some studies indicated that imaginary speech primarily involves the Broca and Wenicke areas as well as the supplementary motor area (López-Vargas et al., 2022). The brain activity maps demonstrated that the inner speech is not a single process itself and involves multiple brain regions. The vowel command activates brain regions that are involved in the association, mental representation, and articulation of sounds. The findings revealed comparable contributions to studies that use alternative methods for mapping the brain.

5.2 Future work

It must be noted that this exploratory work can be expanded and enhanced. For instance, the ANOVA test, which aims to determine significant differences between the means of two or more groups, can be extended to the analysis of spatially dependent curves. Aristizabal et al. (2019) proposed a methodology called spatial functional ANOVA (SFANOVA) to analyze the variance and test the hypothesis of equal functional mean for spatially correlated curves. This technique is applied to an evoked potential dataset representing the cerebral activity response in an emotional stimulus task, where the spatial dependencies are estimated by using a trace-variogram estimator. Therefore, for understanding the characteristics of the inner speech of diverse groups classified by gender or age, we can make inferences about the population by using the SFANOVA. However, the gender representation in the EEG signals dataset utilized in this thesis is imbalanced, with only three female subjects present. The dataset does not provide gender or age differences between the participants.

There are several studies proposed for supervised and unsupervised classification methods based on functional geostatistics. Some supervised classification methods include the classification of spatially correlated curves using normal mixed effect models (Ruiz-Medina et al., 2014) and the classification of bivariate functional data using spatial spline regression techniques (Nguyen et al., 2016). In the case of unsupervised classification, some researchers have proposed methods for classifying curves by reducing the spatial variations in clusters (Romano et al., 2010), functional non-parametric clustering for clustering and aligning correlated curves in space (Abramowicz et al., 2017), and clustering spatial functional data by using hierarchical clustering based on univariate statistics such as the functional mode or the functional mean (Vandewalle et al., 2022). All of these methods can be applied to the detection of Parkinson's disease or epilepsy seizures using EEG signals.

Traditional Kriging methods, which are based on best linear unbiased prediction, are extensively employed in spatial prediction. However, they encounter limitations when dealing with non-linear and non-Gaussian data. To address these challenges, Chen et al. (2020) introduces DeepKriging, a deep neural network (DNN) structure for spatial prediction. DeepKriging incorporates an embedding layer of spatial coordinates with basis functions to capture the spatial dependence. This method outperforms traditional Kriging methods as it can handle large datasets integrated with non-linear and categorical data, and provides non-linear predictions with smaller approximation errors.

DeepKriging can improve spatial prediction accuracy and be used in diverse applications, such as environmental monitoring, disease mapping, and resource management. The authors present an application of PM2.5 concentration data to demonstrate the performance of DeepKriging compared to Kriging and other DNN methods. This approach could further enhance the application of EEG signals in the current thesis, since they are non-linear, non-Gaussian, and non-stationary.

5.3 Conclusions

In this work, the EEG signals obtained in an experimental setting were analyzed as space-time curves using functional data analysis and functional geostatistics methods. The exploratory analysis uncovered brain activity patterns associated with the complex interaction between speech and brain function. This study modeled the spatial dependencies inherent to the brain and EEG signals and used functional kriging for prediction of the spatially correlated curves.

The analysis revealed distinct variations in the EEG signals, with high brain activity observed in the regions associated with speech production and comprehension. These findings contribute to the understanding of brain activity involved in inner speech and provide comparable results to other brain mapping methods. Furthermore, these findings reveal distinct brain regions and corresponding patterns of cerebral activity associated with covert speech, providing insights into language processing and cognitive function.

The use of functional data analysis and functional geostatistics to handle space- and time-varying data highlights the significance of incorporating advanced analytical tools into the neuroscientific field to acquire a deeper comprehension of brain function and neural mechanisms. Understanding the neural mechanisms associated with speech is crucial for developing brain-computer interfaces to help individuals with communication difficulties.

This thesis involves analyzing spatial functional data, that is data in form of curves with spatial dependencies. We apply the exploratory functional data analysis over the curves to explore the main characteristics of the functions with descriptive statistics, such as functional mean and variance. The functional principal component analysis reveals the variations and spatial patterns among the curves. Moreover, we used FPCA to reduce the temporal dimension to the components that

explain the most of the variance, with an 85% threshold. The empirical functional principal components were used to model the spatial dependencies and predict the spatial curves using functional kriging.

Currently, technological advancements have led to a significant rise in datasets that correspond to the realization of a functional random field. In this case, we applied geostatistics to functional data obtained under experimental conditions. Nevertheless, the most popular functional datasets are found in geographic observations such as air quality and temperature.

References

Abramowicz, K., Arnqvist, P., Secchi, P., De Luna, S. S., Vantini, S., & Vitelli, V. (2017). Clustering misaligned dependent curves applied to varved lake sediment for climate reconstruction. *Stochastic environmental research and risk assessment*, *31*(1), 71–85.

Alderson-Day, B., & Fernyhough, C. (2015). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin*, *141*, 931 - 965. Retrieved from `https://api.semanticscholar.org/CorpusID:13446717`

Aristizabal, J.-P., Giraldo, R., & Mateu, J. (2019). Analysis of variance for spatially correlated functional data: application to brain data. *Spatial Statistics*, *32*, 100381.

Bande, M. F., de la Fuente, M. O., Galeano, P., Nieto, A., Garcia-Portugues, E., & de la Fuente, M. M. O. (2022). Package 'fda. usc'.

Bastos, N. S., Marques, B. P., Adamatti, D. F., & Billa, C. Z. (2020). Analyzing eeg signals using decision trees: A study of modulation of amplitude. *Computational Intelligence and Neuroscience*, *2020*.

Blinowska, K., & Durka, P. (2006). Electroencephalography (eeg). *Wiley encyclopedia of biomedical engineering*.

Bohorquez, M., Giraldo, R., & Mateu, J. (2016). Optimal sampling for spatial prediction of functional data. *Statistical Methods & Applications*, *25*(1), 39–54.

Bohorquez, M., Giraldo, R., & Mateu, J. (2017). Multivariate functional random fields: prediction and optimal sampling. *Stochastic Environmental Research and Risk Assessment*, *31*(1), 53–70.

Bosq, D. (2000). *Linear processes in function spaces: Theory and applications* (Vol. 149). Springer.

Botella-Soler, V., Valderrama, M., Crepon, B., Navarro, V., & Le Van Quyen, M. (2012). Large-scale cortical dynamics of sleep slow waves. *PloS one*, *7*(2), e30757.

Cahn, B. R., & Polich, J. (2013). Meditation states and traits: Eeg, erp, and neuroimaging studies.

Chaddad, A., Wu, Y., Kateb, R., & Bouridane, A. (2023). Electroencephalography signal processing: A comprehensive review and analysis of methods and techniques. *Sensors*, *23*(14), 6434.

Chandel, G., Sharma, A., Bajaj, S., & Verma, S. (2023). Computer based detection of alcoholism using eeg signals. In *2023 second international conference on*

electrical, electronics, information and communication technologies (iceeict) (pp. 1–7).

Chen, W., Li, Y., Reich, B. J., & Sun, Y. (2020). Deepkriging: Spatially dependent deep neural networks for spatial prediction. *arXiv preprint arXiv:2007.11972*.

Chiles, J.-P., & Delfiner, P. (2009). *Geostatistics: modeling spatial uncertainty* (Vol. 497). John Wiley & Sons.

Cressie, N. (1985). Fitting variogram models by weighted least squares. *Mathematical Geology*, *17*(5).

Cressie, N., & Wikle, C. (2011). *Statistics for spatio-temporal data*. John Wiley & Sons.

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

Das, R. K., Martin, A., Zurales, T., Dowling, D., & Khan, A. (2023). A survey on eeg data analysis software. *Sci*, *5*(2), 23.

De Boor, C., & De Boor, C. (1978). *A practical guide to splines* (Vol. 27). springer-verlag New York.

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

Dempster, J. (2001). *The laboratory computer: a practical guide for physiologists and neuroscientists*. Academic Press.

Deville, J.-C. (1974). Méthodes statistiques et numériques de l'analyse harmonique. In *Annales de l'insee* (pp. 3–101).

Ferraty, F. (2006). *Nonparametric functional data analysis*. Springer.

Fraiman, R., & Muniz, G. (2001). Trimmed means for functional data. *Test*, *10*, 419–440.

Ganushchak, L. Y., Christoffels, I. K., & Schiller, N. O. (2011). The use of electroencephalography in language production research: a review. *Frontiers in psychology*, *2*, 208.

Geva, S., Jones, P. S., Crinion, J. T., Price, C. J., Baron, J.-C., & Warburton, E. A. (2011). The neural correlates of inner speech defined by voxel-based lesion–symptom mapping. *Brain*, *134*(10), 3071–3082.

Ghane, P. (2015). *Silent speech recognition in eeg-based brain computer interface*. Purdue University.

Giraldo, R., Delicado, P., & Mateu, J. (2010). Continuous time-varying kriging for spatial prediction of functional data: An environmental application. *Journal of agricultural, biological, and environmental statistics*, *15*, 66–82.

Giraldo, R., Delicado, P., & Mateu, J. (2011). Ordinary kriging for function-valued spatial data. *Environmental and Ecological Statistics*, *18*(3), 411-426. Retrieved from `http://dx.doi.org/10.1007/s10651-010-0143-y` doi: 10.1007/s10651-010-0143-y

Giraldo, R., Delicado Useros, P. F., & Mateu, J. (2007). Geostatistics for functional data: An ordinary kriging approach.

Goulard, M., & Voltz, M. (1993). Geostatistical interpolation of curves: A case study in soil science. *In A. Soares (Ed.) Geostatistics Tróia'92*, *2*(2), 805–816.

Gui, X., Chuansheng, C., Zhong-Lin, L., & Qi, D. (2010). Brain imaging techniques and their applications in decision-making research. *Xin li xue bao. Acta psychologica Sinica*, *42*(1), 120.

Guo, X., Kurtek, S., & Bharath, K. (2022). Variograms for kriging and clustering of spatial functional data with phase variation. *Spatial statistics*, *51*, 100687.

Guo, X., Zhao, H., Li, X., Li, T., & Dai, M. (2015). Eeg signal analysis based on fixed-value shift compression algorithm. In *2015 11th international conference on natural computation (icnc)* (pp. 959–963).

Hernández, D., Trujillo, L., Z-Flores, E., Villanueva, O., & Romo-Fewell, O. (2018). Detecting epilepsy in eeg signals using time, frequency and time-frequency domain features. *Computer science and engineering—theory and applications*, 167–182.

Horvath, L., & Kokoszka, P. (2012). *Inference for functional data with applications*. Springer.

Hyndman, R. J., & Shang, H. L. (2010). Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational and Graphical Statistics*, *19*(1), 29–45.

Ibrahim, A. A., Farhan, I. M., & Safi, M. E. (2022). A nonlinearities inverse distance weighting spatial interpolation approach applied to the surface electromyography signal. *International Journal of Electrical and Computer Engineering*, *12*(2), 1530.

Iqbal, S., PP, M. S., Khan, Y. U., & Farooq, O. (2016). Eeg analysis of imagined speech. *International Journal of Rough Sets and Data Analysis (IJRSDA)*, *3*(2), 32–44.

Jaramillo-Jimenez, A., Tovar-Rios, D. A., Ospina, J. A., Mantilla-Ramos, Y.-J., Loaiza-López, D., Isaza, V. H., ... others (2023). Spectral features of resting-state eeg in parkinson's disease: A multicenter study using functional data analysis. *Clinical Neurophysiology*, *151*, 28–40.

Kaur, J., & Kaur, A. (2015). A review on analysis of eeg signals. In *2015 international conference on advances in computer engineering and applications* (pp. 957–960).

Khalifa, W. H., Roushdy, M. I., Abdel-badeeh, M. S., & Revett, K. (2015). Ais inspired approach for user identification based on eeg signals. *Recent Advances in Information Science AIS*, 84–89.

Kokoszka, P., & Reimherr, M. (2017). *Introduction to functional data analysis*. CRC press.

Langland-Hassan, P. (2021). Inner speech. *Wiley Interdisciplinary Reviews: Cognitive Science*, *12*(2), e1544.

Li, F., Chao, W., Li, Y., Fu, B., Ji, Y., Wu, H., & Shi, G. (2021). Decoding imagined speech from eeg signals using hybrid-scale spatial-temporal dilated convolution network. *Journal of Neural Engineering*, *18*(4), 0460c4.

López-Vargas, O., Sarmiento-Vela, L., Bacca-Rodríguez, J., Villamizar-Delgado, S., & Sarmiento-Vela, J. (2022). Silent speech of vowels in persons of different cognitive styles. *Suma Psicológica*, *29*(1), 20–29.

Loza, C. A., & Principe, J. C. (2021). Eeg models and analysis. In *Handbook of neuroengineering* (pp. 1–36). Springer.

Mateu, J., & Romano, E. (2017). *Advances in spatial functional statistics* (Vol. 31). Springer.

McCall, P., Cabrerizo, M., & Adjouadi, M. (2012). Spatial and temporal analysis of interictal activity in the epileptic brain. In *2012 ieee signal processing in medicine and biology symposium (spmb)* (pp. 1–6).

Menafoglio, A., & Secchi, P. (2017). Statistical analysis of complex and spatially dependent data: A review of object oriented spatial statistics. *European Journal of Operational Research*, *258*(2), 401–410.

Myers, M. H., Jolly, E., Li, Y., de Jongh Curry, A., & Parfenova, H. (2017). Power spectral density analysis of electrocorticogram recordings during cerebral hypothermia in neonatal seizures. *Annals of Neurosciences*, *24*(1), 12–19.

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

Nguyen, H. D., McLachlan, G. J., & Wood, I. A. (2016). Mixtures of spatial spline regressions for clustering and classification. *Computational Statistics & Data Analysis*, *93*, 76–85.

Nieto, N., Peterson, V., Rufiner, H. L., Kamienkowski, J. E., & Spies, R. (2022). Thinking out loud, an open-access eeg-based bci dataset for inner speech recognition. *Scientific Data*, *9*(1), 52.

Nkengfack, L. C. D., Tchiotsop, D., Atangana, R., Louis-Door, V., & Wolf, D. (2020). Eeg signals analysis for epileptic seizures detection using polynomial transforms, linear discriminant analysis and support vector machines. *Biomedical Signal Processing and Control*, *62*, 102141.

Olokodana, I. L., Mohanty, S. P., & Kougianos, E. (2020a). Distributed kriging-bootstrapped dnn model for fast, accurate seizure detection from eeg signals. In *2020 ieee computer society annual symposium on vlsi (isvlsi)* (pp. 264–269).

Olokodana, I. L., Mohanty, S. P., & Kougianos, E. (2020b). Krig-detect: Exploring alternative kriging methods for real-time seizure detection from eeg signals. In *2020 ieee 6th world forum on internet of things (wf-iot)* (pp. 1–6).

Ortiz, M., Iáñez, E., Contreras-Vidal, J. L., & Azorín, J. M. (2020). Analysis of the eeg rhythms based on the empirical mode decomposition during motor imagery when using a lower-limb exoskeleton. a case study. *Frontiers in Neurorobotics*, *14*, 48.

Pebesma, E. J. (2004). Multivariable geostatistics in s: the gstat package. *Computers & geosciences*, *30*(7), 683–691.

Pfurtscheller, G., Brunner, C., Schlögl, A., & Da Silva, F. L. (2006). Mu rhythm (de) synchronization and eeg single-trial classification of different motor imagery tasks. *NeuroImage*, *31*(1), 153–159.

Ramsay, & Dalzell, C. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *53*(3), 539–561.

Ramsay, & Silverman. (2005). *Functional data analysis*. Springer, New York.

Ricker, D. W. (2012). *Echo signal processing* (Vol. 725). Springer Science & Business Media.

Romano, E., Balzanella, A., & Verde, R. (2010). Clustering spatio-functional data: a model based approach. In *Classification as a tool for research* (pp.

167–175). Springer.

Ruchika, F., Neupane, D., Shah, S., Delawan, M., & Lucke-Wold, B. (2023). Advancing analytics of eeg signals. *Med discoveries*, *2*(4).

Ruiz-Medina, M. D., Espejo, R. M., & Romano, E. (2014). Spatial functional normal mixed effect approach for curve classification. *Advances in Data Analysis and Classification*, *8*(3), 257–285.

Sabourin, M. E., Cutcomb, S. D., Crawford, H. J., & Pribram, K. (1990). Eeg correlates of hypnotic susceptibility and hypnotic trance: Spectral analysis and coherence. *International journal of psychophysiology*, *10*(2), 125–142.

Saccá, V., Campolo, M., Mirarchi, D., Gambardella, A., Veltri, P., & Morabito, F. C. (2018). On the classification of eeg signal by using an svm based algorithm. *Multidisciplinary approaches to neural computing*, 271–278.

Scheffler, A., Telesca, D., Li, Q., Sugar, C. A., Distefano, C., Jeste, S., & Şentürk, D. (2020). Hybrid principal components analysis for region-referenced longitudinal functional eeg data. *Biostatistics*, *21*(1), 139–157.

Scheffler, A. W., Dickinson, A., DiStefano, C., Jeste, S., & Şentürk, D. (2020). Covariate-adjusted hybrid principal components analysis. In *International conference on information processing and management of uncertainty in knowledge-based systems* (pp. 391–404).

Scheffler, A. W., Dickinson, A., DiStefano, C., Jeste, S., & Şentürk, D. (2022). Covariate-adjusted hybrid principal components analysis for region-referenced functional eeg data. *Statistics and its interface*, *15*(2), 209.

Schultz, T., Wand, M., Hueber, T., Krusienski, D. J., Herff, C., & Brumberg, J. S. (2017). Biosignal-based spoken communication: A survey. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *25*(12), 2257–2271.

ShAADAN, N., Deni, S. M., & JeMAiN, A. A. (2012). Assessing and comparing pm10 pollutant behaviour using functional data approach. *Sains Malaysiana*, *41*(11), 1335–1344.

Shang, H. L. (2014). A survey of functional principal component analysis. *AStA Advances in Statistical Analysis*, *98*, 121–142.

Sha'Abani, M., Fuad, N., Jamal, N., & Ismail, M. (2020). knn and svm classification for eeg: a review. In *Inecce2019: Proceedings of the 5th international conference on electrical, control & computer engineering, kuantan, pahang, malaysia, 29th july 2019* (pp. 555–565).

Shergill, S., Brammer, M., Fukuda, R., Bullmore, E., Murray, R., & McGuire, P. K. (2001). Modulation of activity in temporal cortex during generation of inner speech. *NeuroImage*, *6*(13), 600.

Shergill, S., Brammer, M. J., Fukuda, R., Williams, S. C., Murray, R. M., & McGuire, P. K. (2003). Engagement of brain areas implicated in processing inner speech in people with auditory hallucinations. *The British Journal of Psychiatry*, *182*(6), 525–531.

Simons, C. J. P., Tracy, D. K., Sanghera, K. K., O'Daly, O., Gilleen, J., de Gracia Dominguez, M., ... Shergill, S. S. (2010). Functional magnetic resonance imaging of inner speech in schizophrenia. *Biological Psychiatry*, *67*, 232-237. Retrieved from `https://api.semanticscholar.org/CorpusID:206099996`

Subha, D. P., Joseph, P. K., Acharya U, R., & Lim, C. M. (2010). Eeg signal analysis: a survey. *Journal of medical systems*, *34*, 195–212.

Sun, Y., & Genton, M. G. (2011). Functional boxplots. *Journal of computational and graphical statistics*, *20*(2), 316–334.

Sundaram, C. K., Gayathri, S., & Soundarya, P. (2022, May). Discrete wavelet transform based on eeg signal analysis for diagnosing neurological disorder. *International journal of health sciences*, *6*(S1), 9556–9566. Retrieved from `https://sciencescholar.us/journal/index.php/ijhs/article/view/7213` doi: 10.53730/ijhs.v6nS1.7213

Thivierge, J.-P. (2007). Functional data analysis of cognitive events in eeg. In *2007 ieee international conference on systems, man and cybernetics* (pp. 2473–2478).

Tran, Y. (2022). *Eeg signal processing for biomedical applications* (Vol. 22) (No. 24). MDPI.

Unde, S. A., & Shriram, R. (2014). Coherence analysis of eeg signal using power spectral density. In *2014 fourth international conference on communication systems and network technologies* (pp. 871–874).

Vandewalle, V., Preda, C., & Dabo-Niang, S. (2022). Clustering spatial functional data. *Geostatistical functional data analysis*, 155–174.

Vidal Badia, M., Rosso, M., & Aguilera, A. M. (2021). P-spline smoothed functional ica of eeg data. *arXiv: 2101.05769*.

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

Wen, D., Yuan, J., Zhou, Y., Xu, J., Song, H., Liu, Y., ... Jung, T.-P. (2020). The eeg signal analysis for spatial cognitive ability evaluation based on multivariate permutation conditional mutual information-multi-spectral image. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *28*(10), 2113–2122.

Yi, Y., Billor, N., Liang, M., Cao, X., Ekstrom, A., & Zheng, J. (2022). Classification of eeg signals: an interpretable approach using functional data analysis. *Journal of Neuroscience Methods*, *376*, 109609.

Zaleshina, M., & Zaleshin, A. (2017). The brain as a multi-layered map. scales and reference points for pattern recognition in neuroimaging. *European Journal of Geography*, *8*(1).

Zhang, Y., Wang, C., Wu, F., Huang, K., Yang, L., & Ji, L. (2020). Prediction of working memory ability based on eeg by functional data analysis. *Journal of neuroscience methods*, *333*, 108552.

Zhao, L., Liu, R., Li, S., Wang, X., & Bao, D. (2023). Spatio-temporal variable structure graph neural network for eeg data classification. In *2023 6th international symposium on autonomous systems (isas)* (pp. 1–6).

Appendix A Predictions

This appendix provides the curve predictions obtained using functional kriging. The following figures display the observed curves in black and, in red, the predicted curves. Figure 1 illustrates notable differences between the predicted and observed functions of electrode 5 and electrode 17 in the vowel /a/. Figure 2 shows that the predictions are very similar to the observations among all the electrodes in the vowel /e/, except for electrode 17. The predictions of the vowel /i/ in Figure 3 exhibit variances in the frequency range of 100 to 250 Hz in electrodes 7, 17, 18, and 21. Similarly, the predicted curves for the vowel /o/ in Figure 4 exhibit variances in the frequency range from 0 to 10 Hz, particularly for electrodes 3, 7 and 17. In the case of the vowel /u/, the differences between the predicted and observed curves can be seen in electrodes 1, 18, 16, 11, and 5 (Figure 5). These differences can be attributed to the low variation ranges of the signals.

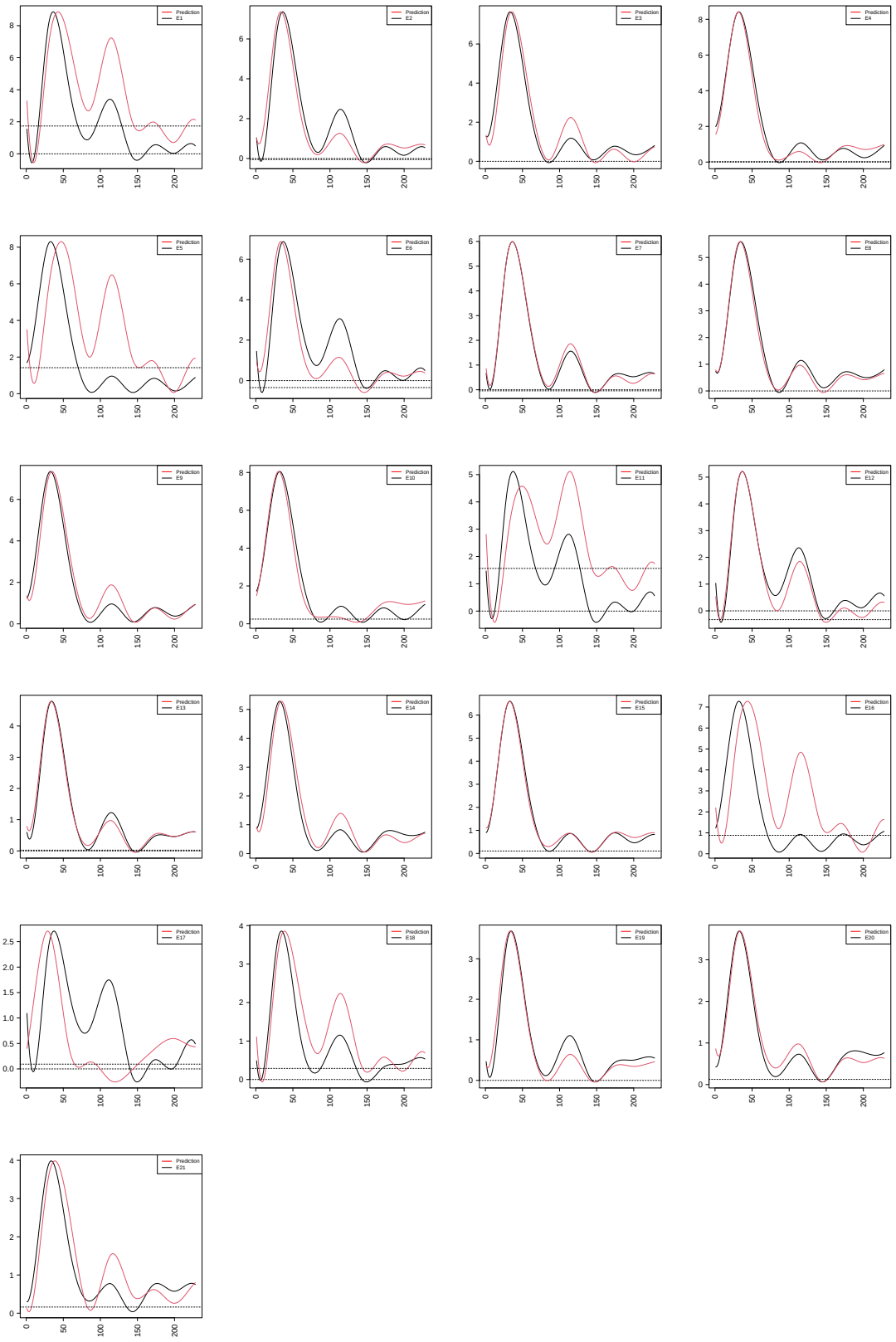

FIGURE 1: Predicted vs observed EEG curves for all 21 electrodes of vowel /a/

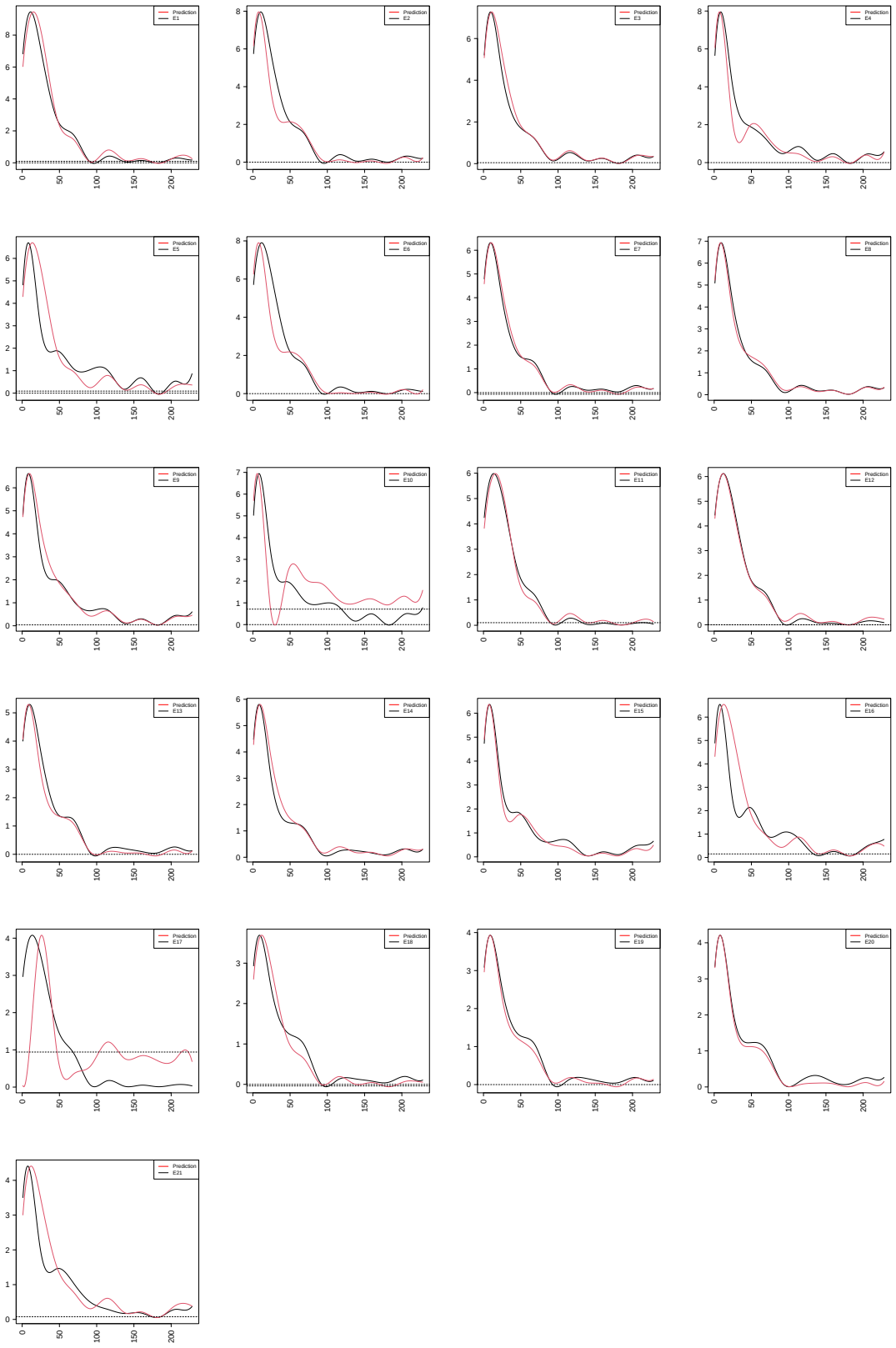

FIGURE 2: Predicted vs observed EEG curves for all 21 electrodes of vowel /e/

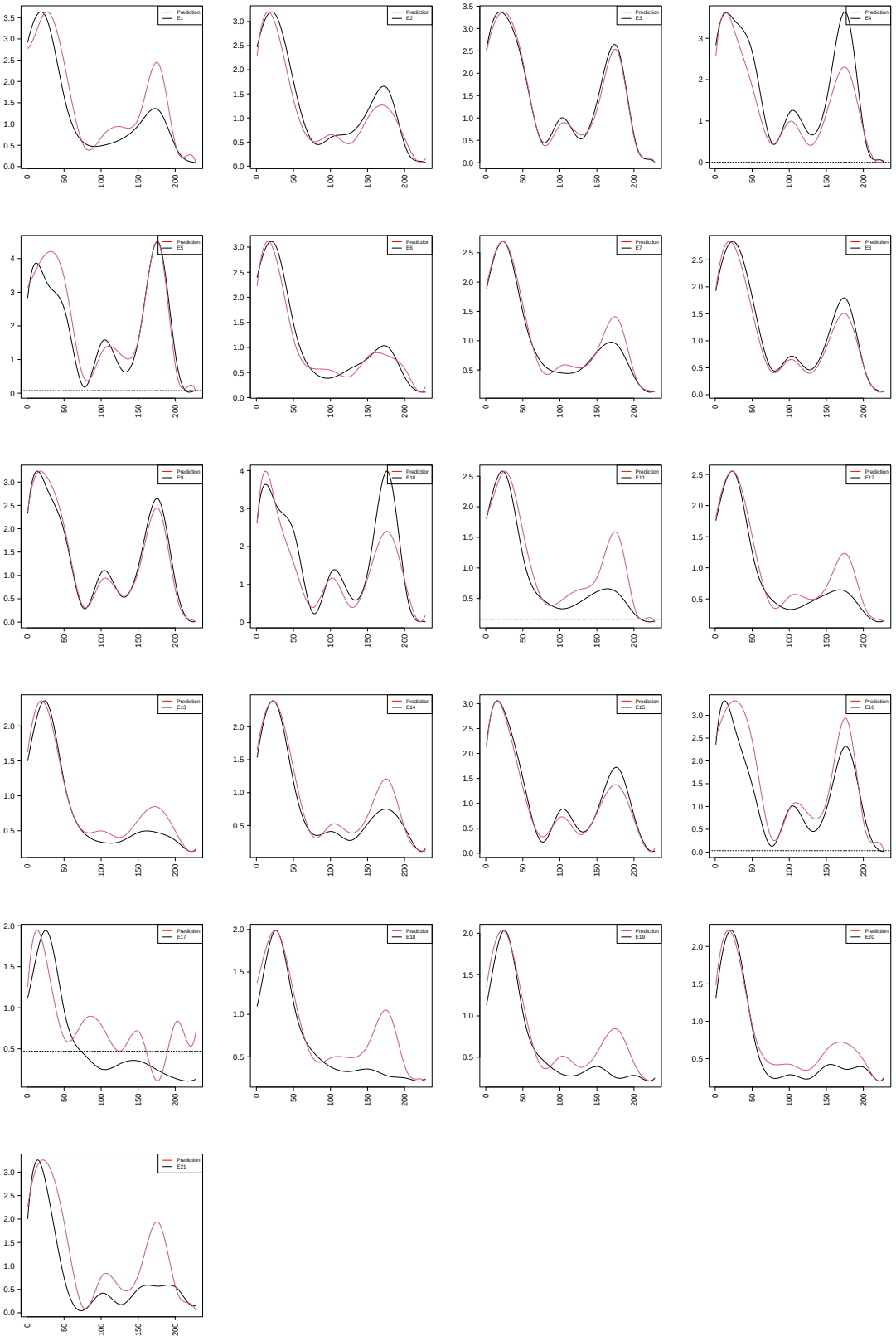

FIGURE 3: Predicted vs observed EEG curves for all 21 electrodes of vowel /i/

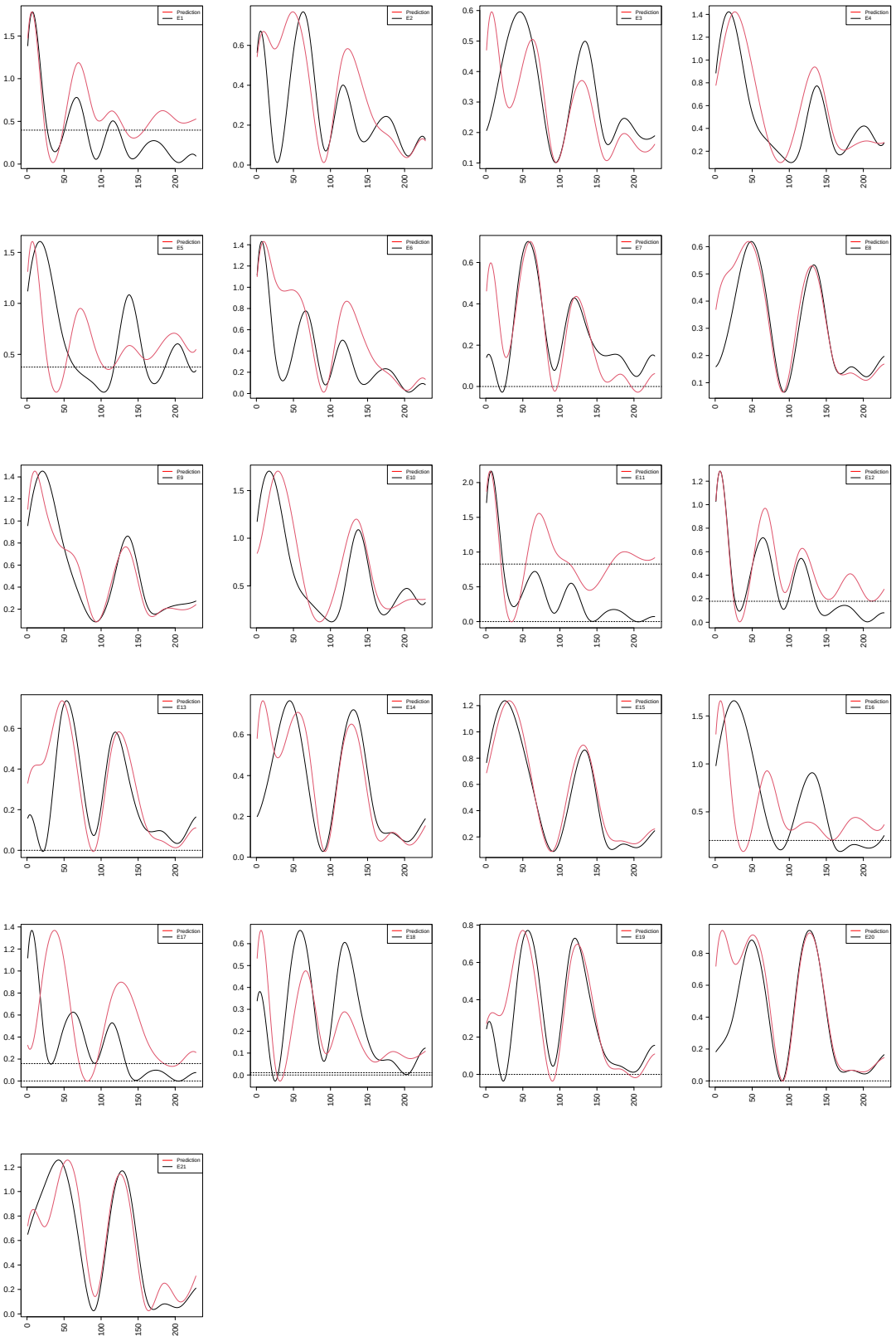

FIGURE 4: Predicted vs observed EEG curves for all 21 electrodes of vowel /o/

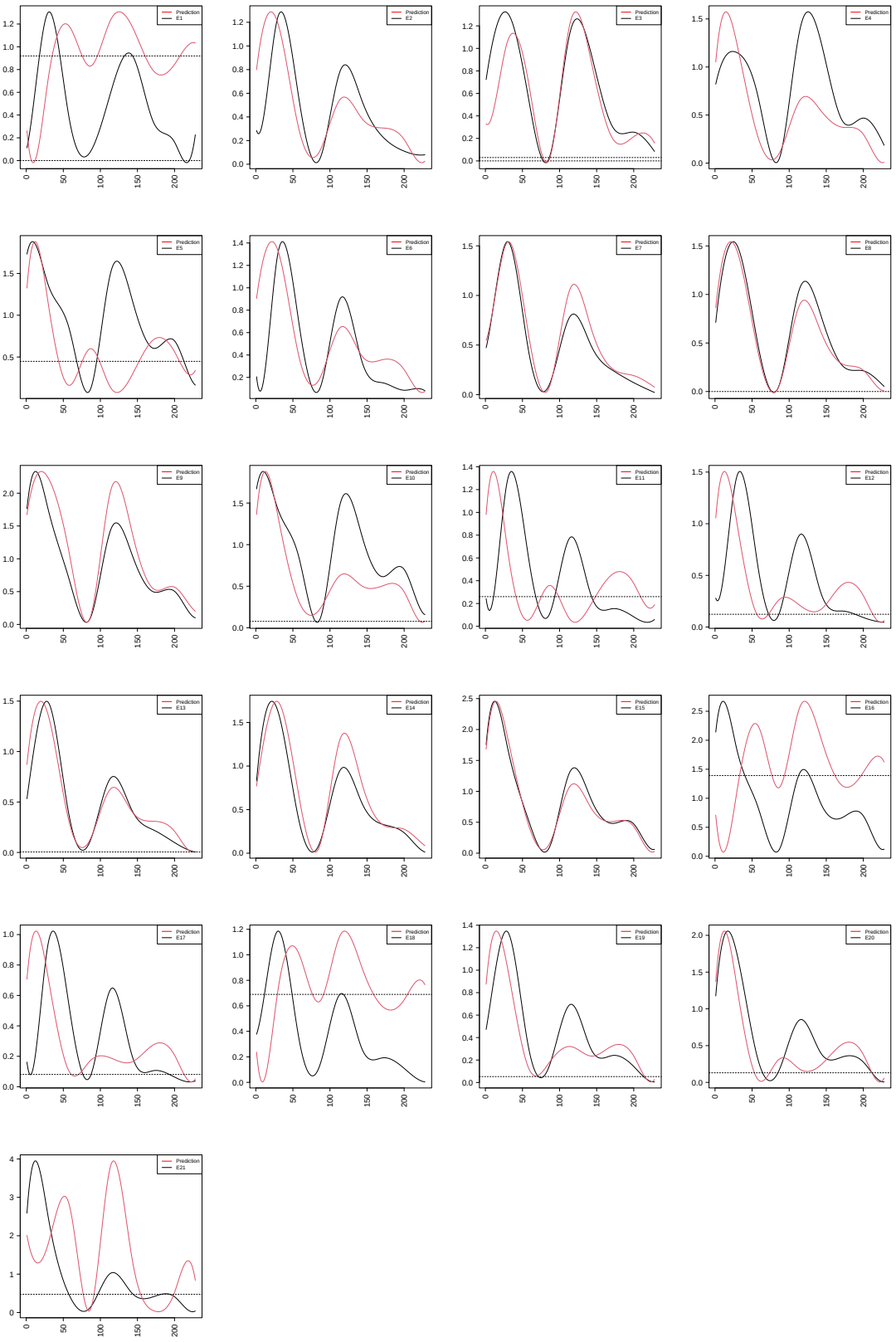

FIGURE 5: Predicted vs observed EEG curves for all 21 electrodes of vowel /u/

Appendix B Code

This chapter presents the source code used in this thesis to perform the exploratory functional data analysis and the spatial prediction of the EEG curves. We use the packages fda.usc, gstat, and SpatFD available in the software R version 4.3.1. The full source code, as well as additional functions, can be found on the GitHub repository at the following link: `https://github.com/bquinteroal/EEGSignals_FunctionalGeostatistics`

```
####### Packages #######
packages = c(
  "Matrix", "fda", "fda.usc", "gstat", "splines", "colorspace", "sp", "geoR", "ggplot2",
  "reshape2", "readr", "readxl", "ggpubr", "ggrepel", "gridExtra", "patchwork", "purrr",
  "tidyr", "dplyr", "ggpubr", "rainbow", "cowplot", "readxl", "tidyverse", "grid", "lattice",
  "roahd", "ftsa", "fields", "SpatFD", "gstat", "sp", "sf", "reshape", "plotly", "processx"
)
sapply(packages, require, character.only = TRUE)

####### Working Directory #######
setwd("D:/Thesis")
source("Package.R")
setwd("D:/Thesis/Data/Dataset")

### Read dataset
for (i in 1:23) {
  if(i!=10){
    assign(paste0("s",i),value = read_excel(paste0("sujetoEpsd",i,".xlsx")))
    assign(paste0("s",i),value=as.matrix(dplyr::select(get(paste0("s",i)),-FRECUENCIA),
    row.names=1))
  }
}

### Create parameters and names for the data.
p = 228 ; nelec = 21 ; nvow = 5
vowels = c("a","e","i","o","u")

### Creation of objects for data functions. Given the functions created for the analysis
for each vowel.
for (i in 1:23) {
  if(i!=10){
    assign(paste0("s",i,".list"),value= data.list(get(paste0("s",i)),p))
    assign(paste0("s",i,".gfdata"),value= gfdata(get(paste0("s",i)),p))
    assign(paste0("names(s",i,".list)"),value= vowels)
    assign(paste0("names(s",i,".gfdata)"),value= get(paste0("names(s",i,".list)")))
  }
}

### RMSE and MSE
nb = seq(10,30,1)
for (i in 1:23) {
  if(i!=10){
    assign(paste0("nb.MSE",i),value= gf.RMSE(get(paste0("s",i,".gfdata")),
    get(paste0("s",i,".list")),seq.nb = nb))
```

```
  }
}

### Plot RMSE n basis selection
x11()
ggplot.RMSE(nb.MSE4,nb)

n.basis<-c(11,13,12,13,11)
for (i in 1:23) {
  if(i!=10){
    assign(paste0("s",i,".gfd"),value=gfd.individual(get(paste0("s",i,".gfdata")),n.basis))}
}

### Change names of the data.
for (i in 1:23) {
  if(i!=10){
    nam<-paste0("names(s",i,".gfd)<-vowels")
    eval(parse(text = nam))
  }
}

plot<-NULL
for (i in 1:length(vowels)) {
    assign(paste0("plots",i),value=ggplot.gfd(s4.gfd,vow = vowels[i],trial = 3,legend = T))
  plot<-paste0(plot,"plots",i,",")
}

### Plots
plots<-paste0("plots<-ggarrange(",plot,"ncol=",3,",nrow=",2,")")
eval(parse(text = plots))
{x11()
  plots}

####### Descriptive statistics #######

### Calculate summary statistics
summary.s4.gfd<-summary.gfd(list.gfd=s4.gfdata)

### Plots mean and variance
plot<-NULL
for (i in 1:length(vowels)) {
  assign(paste0("plots",i),value=ggplot_combinedmean(s4.gfd,summary.s4.gfd,vow = vowels[i],
  trial = 3))
  plot<-paste0(plot,"plots",i,",")
}

plots<-paste0("plots<-ggarrange(",plot,"ncol=",3,",nrow=",2,")")
eval(parse(text = plots))
{x11()
  plots}

### Plots trimmed mean and median
plot<-NULL
for (i in 1:length(vowels)) {
  assign(paste0("plots",i),value=ggplot.gfdata.trim(summary.s4.gfd,vow = vowels[[i]],
  trial = 3, legend= T))
  plot<-paste0(plot,"plots",i,",")
}

plots<-paste0("plots<-ggarrange(",plot,"ncol=",3,",nrow=",2,")")
eval(parse(text = plots))
{x11()
  plots}

### Correlation coefficient
for (i in 1:length(vowels)) {
  assign(paste0("s4.cor.",vowels[i]),value=cor.fd(1:p,get(paste0("s4.gfd"))[[vowels[i]]][['3']]))
}

windows()
par(mfrow=c(2,3))
```

```
for (i in 1:length(vowels)) {
  image.plot(1:p, 1:p, get(paste0("s4.cor.",vowels[i])), xlab='Hz', ylab='Hz')
}

### Outliers
for (i in 1:length(vowels)) {
  assign(paste0("fds.s4.",vowels[i]),value=fds(1:p,get(paste0("s4.list"))[[vowels[i]]][['3']]))
}

### Plot functional boxplot type HDR
windows()
par(mfrow=c(2,3))
for (i in 1:length(vowels)) {
  fboxplot( get(paste0("fds.s4.",vowels[i])), plot.type = "functional",
  type ="hdr", projmethod = "PCAproj",ylab="Brain Signal",xlab="Hz",na.rm = TRUE)
}

### Plot functional boxplot type bag
windows()
par(mfrow=c(2,3))
for (i in 1:length(vowels)) {
  fboxplot( get(paste0("fds.s4.",vowels[i])), plot.type = "functional",
  type ="bag", projmethod = "PCAproj",ylab="Brain Signal",xlab="Hz",na.rm = TRUE)
}


####### FPCA #######

s4.fpca =gfd_pca(s4.gfd)

### Plots PC functions
windows()
par(mfrow=c(2,3))
for (i in 1:length(vowels)) {
  plot(get(paste0("s4.fpca"))[[vowels[i]]][['3']][["harmonics"]], lwd = 1,xlab= "Hz", ylab='PCs')
}

### Plot scores
electrodes=c("1","2","3","4","5","6","7","8","9","10",
"11","12","13","14","15","16","17","18","19","20","21")

plot<-NULL
for (i in 1:length(vowels)) {
  assign(paste0("plots",i),value=ggplot.fpca.scores(s4.fpca,vow = vowels[i],trial=3,electrodes))
  plot<-paste0(plot,"plots",i,",")
}

plots<-paste0("plots<-ggarrange(",plot,"ncol=",3,",nrow=",2,")")
eval(parse(text = plots))
{x11()
  plots}

####### Functional geostatistics #######
coordenadas <- as.data.frame(read_table2("coordenadas.txt"))

for (i in 1:length(vowels)) {
  assign(paste0("vow",vowels[i],"_rep3"),
  value=t(get(paste0("s4.gfdata"))[[vowels[i]]][['3']]$data))
  assign(paste0("vow",vowels[i],"_rep3"),
  value=as.data.frame(get(paste0("vow",vowels[i],"_rep3"))))
}

### Spatial functional data construction
for (i in 1:length(vowels)) {
  assign(paste0("sfd_",vowels[i]),
  value=SpatFD(get(paste0("vow",vowels[i],"_rep3"))%>% select(`E1`:`E21`),
  coords = coordenadas, basis = "Bsplines", nbasis = n.basis[i],
  lambda = 0.00001, nharm = 2,name =  paste0(i)))
  assign(paste0("ptj_",vowels[i]),
  value=data.frame(scores(get(paste0("sfd_",vowels[i])))[[1]]))
}
```

```
colnames(ptj_a)=colnames(ptj_e)=colnames(ptj_i)=
colnames(ptj_o)=colnames(ptj_u)=c("x","y","f1","f2")
coordinates(ptj_a)=coordinates(ptj_e)=coordinates(ptj_i)=
coordinates(ptj_o)=coordinates(ptj_u)=c("x","y")

### Variograms
var_af11=variogram(f1~1,ptj_a,cutoff=50,cressie=T)
var_af21=variogram(f2~1,ptj_a,cutoff=60,cressie=T)

var_ef11=variogram(f1~1,ptj_e,cutoff=50,cressie=T)
var_ef21=variogram(f2~1,ptj_e,cutoff=50,cressie=T)

var_if11=variogram(f1~1,ptj_i,cutoff=50,cressie=T)
var_if21=variogram(f2~1,ptj_i,cutoff=50,cressie=T)

var_of11=variogram(f1~1,ptj_o,cutoff=60,cressie=T)
var_of21=variogram(f2~1,ptj_o,cutoff=60,cressie=T)

var_uf11=variogram(f1~1,ptj_u,cutoff=60,cressie=T)
var_uf21=variogram(f2~1,ptj_u,cutoff=50,cressie=T)

### Models
models_a<- list(vgm(105, "Hol", 12),vgm(44.85, "Hol", 16.08))
models_e<- list(vgm(59.89, "Wav", 34.08),vgm(11.99, "Wav", 49.08))
models_i<- list(vgm(55.6, "Wav", 32.24),vgm(3.18, "Wav", 31.17))
models_o<- list(vgm(10.53, "Per", 154.67),vgm(4.98, "Hol", 11.60))
models_u<- list(vgm(19.92, "Per",161.17),vgm(4.55, "Per", 161.17))

### Spatial functional prediction
for (i in 1:length(vowels)) {
  assign(paste0("KS_SFD_",vowels[i],"_l"),value=KS_scores_lambdas(get(paste0("sfd_",vowels[i])),
  coordenadas, method = "lambda", model = get(paste0("models_",vowels[i]))))
}

class(KS_SFD_a_l)
summary(KS_SFD_a_l)
recons_fd(KS_SFD_a_l)

# Cross Validation
setwd("D:/WWU/Thesis/pred/a")
crossval_loo(KS_SFD_a_l)

setwd("D:/WWU/Thesis/pred/e")
crossval_loo(KS_SFD_e_l)

setwd("D:/WWU/Thesis/pred/i")
crossval_loo(KS_SFD_i_l)

setwd("D:/WWU/Thesis/pred/o")
crossval_loo(KS_SFD_o_l)

setwd("D:/WWU/Thesis/pred/u")
crossval_loo(KS_SFD_u_l)

### Kriging maps
setwd("D:/WWU/Thesis/krig")

dir.create(dirname(paste0('./a/')), recursive=TRUE)
dir.create(dirname(paste0('./e/')), recursive=TRUE)
dir.create(dirname(paste0('./i/')), recursive=TRUE)
dir.create(dirname(paste0('./o/')), recursive=TRUE)
dir.create(dirname(paste0('./u/')), recursive=TRUE)


Sys.setenv(PATH=paste0("C:/Users/bibi/AppData/Local/Programs/orca", Sys.getenv("PATH")))

N_img = 228

t = seq(sfd_a[[1]]$data_fd$basis$rangeval[1],
        sfd_a[[1]]$data_fd$basis$rangeval[2],length.out = N_img)
```

```
for (i in 1:N_img){
  orca(ggmap_KS(KS_SFD_a_l,
                map_path = NULL,
                window_time = t[i],
                method = "lambda")[[1]],  paste0('./a/vowa_',i,'.png'))
}

t = seq(sfd_e[[1]]$data_fd$basis$rangeval[1],
        sfd_e[[1]]$data_fd$basis$rangeval[2],length.out = N_img)

for (i in 1:N_img){
  orca(ggmap_KS(KS_SFD_e_l,
                map_path = NULL,
                window_time = t[i],
                method = "lambda")[[1]],  paste0('./e/vowe_',i,'.png'))
}

t = seq(sfd_i[[1]]$data_fd$basis$rangeval[1],
        sfd_i[[1]]$data_fd$basis$rangeval[2],length.out = N_img)

for (i in 1:N_img){
  orca(ggmap_KS(KS_SFD_i_l,
                map_path = NULL,
                window_time = t[i],
                method = "lambda")[[1]],  paste0('./i/vowi_',i,'.png'))
}

t = seq(sfd_o[[1]]$data_fd$basis$rangeval[1],
        sfd_o[[1]]$data_fd$basis$rangeval[2],length.out = N_img)

for (i in 1:N_img){
  orca(ggmap_KS(KS_SFD_o_l,
                map_path = NULL,
                window_time = t[i],
                method = "lambda")[[1]],  paste0('./o/vowo_',i,'.png'))
}


t = seq(sfd_u[[1]]$data_fd$basis$rangeval[1],
        sfd_u[[1]]$data_fd$basis$rangeval[2],length.out = N_img)

for (i in 1:N_img){
  orca(ggmap_KS(KS_SFD_u_l,
                map_path = NULL,
                window_time = t[i],
                method = "lambda")[[1]],  paste0('./u/vowu_',i,'.png'))
}
```