

Advanced Neuroscience Data Analysis using Functional Data Analysis Methods

by

Yuyan Yi

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama

Dec 9, 2023

Keywords: Functional Data Analysis, Independent Component Analysis,
Signal Processing, Neuroscience Data, Robustness

Copyright 2023 by Yuyan Yi

Approved by

Nedret Billor, Chair, Professor of Mathematics and Statistics
Jingyi Zheng, Co-chair, Assistant Professor of Mathematics and Statistics
Elvan Ceyhan, Associate Professor of Mathematics and Statistics
Roberto Carlo Molinari, Assistant Professor of Mathematics and Statistics
Peng Zeng, Associate Professor of Mathematics and Statistics

Abstract

There has been a growing interest in neuroscience data, encompassing fields like engineering, bioengineering, and neurophysiology. Researchers increasingly utilize neural signals, including electroencephalography (EEG) signals and functional Magnetic Resonance Imaging (fMRI), for various applications like control systems, communication, and medical diagnostics. While numerous models have been developed to explain the relationship between these signals and motor or mental activities, it is still unclear how much information can be decoded from them. Thus, the exploration of neural signal data is still in progress.

Neural signals are inherently characterized as time series data. Most analysis prioritize frequency domain information over time domain. Even when exploring the time domain, these signals are often treated as discrete time points for multivariate analysis, neglecting essential functional dynamics such as continuity and smoothness. Functional data analysis (FDA) facilitates the extraction of information from both the time and frequency domains while considering the temporal dependencies inherent in neural signals. Hence, employing FDA to analyze neural signals is a promising approach.

In this dissertation, we focus on the development of neural signal data analysis with FDA. Specifically, we explore the application of FDA across every phase of signal processing for EEG data, which is a representative data type in neural signal analysis. Firstly, we develop a comprehensive three-stage classification algorithm rooted in functional data analysis, offering the distinct advantage of interpretability. Next, we introduce a robust determination method to automatically identifying the optimal number of ICs. Notably, this method is designed to seamlessly integrate with a variety of ICA techniques, thus ensuring the consistent and reliable generation of results. Furthermore, we propose a robust functional ICA (fICA) method, that significantly enhances the accuracy and reliability of subsequent analysis pertaining to recovered ICs.

In summary, this dissertation encompasses the introduction of a functional classification algorithm, dimension reduction techniques, and robust fICA methods tailored for the neuroscience data analysis.

Acknowledgments

I would like to express my deepest appreciation to my advisors, Dr. Nedret Billor and Dr. Jingyi Zheng. I'm truly thankful for the invaluable mentorship and the inspiring example you've set as statisticians, researchers, and educators. Your unwavering guidance, support, and encouragement have played a pivotal role in my personal and professional growth over the past several years. I am now more self-assured in confronting challenges. I would also like to convey my appreciation to the Department of Mathematics and Statistics, National Science Foundation (NSF), and Center for Clinical and Translational Science (CCTS) for their financial support through research and teaching assistantships, conference travel assistance, and numerous scholarships.

Additionally, I wish to emphasize my profound gratitude to my family. Undertaking this path would have been impossible without their support. As the first member of my family to seize the opportunity to study abroad, I have consistently cherished these hard-earned opportunities. Being a source of pride for my family has been a strong motivator for me to persevere in the face of substantial hardships while living overseas.

In particular, I would like to express my heartfelt gratitude to my parents, Zhaoyan Zhuang and Gengyun Yi. My mother served as my first math teacher, fostering in me a passion of numbers from a young age. I recall mom introducing me to the mathematics game known as the "24 game," which has remained a favorite of mine to this day. When I was about to give up, she would always provide encouragement, inspiring me to endure and persevere. My father, on the other hand, has always been a source of inspiration for me. His steadfast commitment, courage, and precise care throughout my childhood filled me great joy. He steadfastly ensured my safe passage to and from school for a period of twelve years, from elementary to high school, whether it was the scorching days of summer or the chilly winter mornings. Their selfless love and unwavering faith in me have enabled me to pursue my educational goals with courage over the years.

I also want to extend my gratitude to my dearest friends, Jingyi Zhu, Minjia Wang, Qiongyao Zeng, Ruoran Dai and Zhuoyu Liu. Even though we are now dispersed over the world, our united dedication to self-improvement binds us together. I fondly recall how we

supported and cheered each other on during the challenging moments of my Ph.D. journey, and how we celebrated each other's successes. Our decade-long friendship has fostered a deep understanding and a bond that feels like family. Undoubtedly, without their enduring presence in my life, the path would definitely be much lonelier.

Moreover, I must acknowledge the crucial role played by my significant other, Zheren Ou. Our adventure began during our undergraduate years, during which we spent four years traveling through China from south to north. Our academic interests led us to continue our education here as well. We are not only partners who share life's joys, but also a formidable team that faces and overcomes challenges together. I consider myself fortunate to have him by my side, witnessing every pivotal moment in my life.

Finally, I'd like to thank everyone who has crossed my path and provided me with invaluable advice and support. At every stage of my life, someone has pointed me in the direction of mathematics, whether it was my mother, my Mathematical Olympiad teacher, my dedicated high school mathematics instructor, my college classmates, my Ph.D. advisors, and others. I consider myself extremely fortunate to have grown up with the love and guidance of numerous people. As I move forward, I am committed to working hard and passing on the love and care that I have received to others.

Table of Contents

Abstract.....	2
Acknowledgments	3
List of Tables	8
List of Figures.....	9
List of Abbreviations	11
Chapter 1.....	13
General Introduction.....	13
1.1 Neuroscience Data Analysis	15
1.1.1 Electroencephalogram Data.....	15
1.1.2 EEG signal analysis	17
1.2 Functional Data Analysis.....	21
1.2.1 Elements of Functional Data Analysis.....	21
1.2.2 Functional Data Models.....	23
1.3 Independent Component Analysis	25
1.4 Aims and Scope of the Thesis.....	28
Chapter 2.....	31
Interpretable EEG classification algorithm	31
2.1 Introduction	31
2.2 Proposed Algorithm.....	33
2.2.1 Data transformation	35
2.2.2 Initial Feature Selection.....	35
2.2.3 Model Construction	37
2.3 Simulation.....	41
2.4 Implementation of the Three-Stage Algorithm to Scalp EEG Data	45
2.4.1 Various Model Settings.....	47

2.4.2 Results and Interpretations	48
2.5 Discussion.....	52
2.6 Conclusion	54
Chapter 3.....	56
CW_ICA: An Efficient Dimensionality Selection Method for Independent Component Analysis	56
3.1 Introduction	56
3.2 Existing methods	60
3.2.1 Durbin-Watson (DW) criterion.....	60
3.2.2 KMO_ICA_Residuals	62
3.2.3 ICA_corr_y.....	63
3.2.4 ICA-by-Blocks.....	64
3.3 Proposed CW_ICA method	65
3.3.1 Algorithm Development	66
3.3.2 Validation.....	69
3.3.3 Illustrative Example.....	70
3.4 Simulation.....	72
3.4.1 Simulation.....	73
3.4.1.1 Simulated Data Generation.....	73
3.4.1.2 Impact of Correlation Coefficients on Determination Methods	74
3.4.1.3 Accuracy	76
3.4.1.4 Robustness	77
3.4.2 Scalp EEG Data Application	77
3.5 Discussion.....	84
3.6 Conclusion	86
Chapter 4.....	88
Robust Functional Independent Component Analysis for Functional Data	88
4.1 Introduction	88
4.2 Proposed Algorithm.....	90
4.2.1 Algorithm Development	90

4.2.2 Validation.....	93
4.3 Simulation.....	94
4.3.1 Simulated Data Generation.....	94
4.3.2 Simulation Performance Metrics	95
4.3.3 Simulation result analysis	96
4.4 Implementation of the rFICA to Scalp EEG Data	101
4.5 Discussion.....	103
4.6 Conclusion	104
Chapter 5.....	106
General Conclusion	106
5.1 Summary.....	106
5.2 Future Work	107
References	109

List of Tables

Table 2.1 (Summary of three functional feature selection methods used)	37
Table 2.2 (Selected model construction)	48
Table 3.1 (Summary of existing determination methods)	59
Table 4.1 (Contamination Parameter setting).....	94
Table 4.2 (Performance of fICA methods based on CCF metric for varying magnitude of sample outliers)	96
Table 4.3 (Performance of fICA methods based on COS metric for varying magnitude of sample outliers)	96
Table 4.4 (Performance of fICA methods based on CCF metric for varying contamination levels of sample outliers).....	98
Table 4.5 (Performance of fICA methods based on COS metric for varying contamination levels of sample outliers).....	98

List of Figures

Figure 1.1 (The measurement and display of EEG on human brains with electrodes)	1
Figure 1.2 (Procedure of EEG signal analysis Preprocessing)	18
Figure 1.3 (The measurement and display of EEG on human brains with electrodes)..	26
Figure 1.4 (Procedure of ICA algorithm).....	26
Figure 2.1 (Procedure of three-stage algorithm).....	34
Figure 2.2 A (Simulated signals in Group 1).....	42
Figure 2.2 B (Simulated signals in Group 2)	42
Figure 2.3 A (Power spectrum in Group 1 for data in Channel 3).....	43
Figure 2.3 B (Power spectrum in Group 2 for data in Channel 3).....	43
Figure 2.4 (Wavelet power level at different frequency and time points in channel 3) .	43
Figure 2.5 (Data of channel 3 in time-frequency domain at different frequency.)	44
Figure 2.6 (Plot of functional test results of feature that frequency = 8Hz in Channel 3)	45
Figure 2.7 (Plots of collected scalp EEG signal)	46
Figure 2.8 (The structure of the experimental data set at each stage)	46
Figure 2.9 (Plots of AUC for each model).....	50
Figure 2.10 A (Bar plot of the proportion of brain waves).....	50
Figure 2.10 A (Topographical map of selected features).....	50
Figure 2.11 A (Wavelet power spectrum at electrode P5 under long distance task).....	52
Figure 2.11 B (Wavelet power spectrum at electrode P5 under short distance task).....	52
Figure 2.12 (Plots of selected brain waves that generated by the data after wavelet transformation).....	53
Figure 3.1 (Stages of data structures in CW_ICA method)	66
Figure 3.2 A (Correlation plot for the estimated ICs from two blocks (q=5))	71
Figure 3.2 B (Correlation plot for estimated ICs from two blocks (q=8)).....	71
Figure 3.2 C (Signal-correlation plot).....	71
Figure 3.3 (Simulated EEG signals plots).....	72

Figure 3.4 (Signal-Correlation plots for ICA-by-Blocks and CW_ICA).....	74
Figure 3.5 (Accuracy of three determination methods)	75
Figure 3.6 (Estimated number of source signals from the varying mixed signals).....	77
Figure 3.7 (Heatmap of estimated number of ICs for each individual)	78
Figure 3.8 (Precision Comparison of estimated number of ICs for each individual)	79
Figure 3.9 (Extracted EEG source signals from subject 18 with different number of ICs)	80
Figure 3.10 (Precision Comparison of estimated number of ICs for each electrode)....	81
Figure 3.11 (Heatmap of estimated number of ICs for each electrode).....	82
Figure 3.12 (Topographical scalp maps).....	83
Figure 4.1 (Simulated EEG component signals)	93
Figure 4.2 (Simulated EEG signals with sample outliers)	94
Figure 4.3 (Side-by-side boxplot of CCF and COS values for robust and non-robust fICA methods with different with varying magnitude of sample outliers)	97
Figure 4.4 (Side-by-side boxplot of CCF and COS values for robust and non-robust fICA methods with different with varying contamination level of sample outliers.)	99
Figure 4.5 A (Boxplot of classification results based on IC score)	101
Figure 4.5 B (Boxplot of classification results based on Frequency)	101
Figure 4.6 A (Boxplot of classification results by difference fICA methods based on IC score).....	102
Figure 4.6 B (Boxplot of classification results by difference fICA methods based on Frequency)	102

List of Abbreviations

ASR	Artifact subspace reconstruction
AUC	Area under the ROC curve
BCI	Brain-Computer Interfaces
BSS	Blind source separation
CCF	Cross-Correlation Function
COS	Cosine similarity
CV	Cross validation
CW_ICA	Column-wise independent component analysis
DW	Durbin-Watson
ECoG	Electrocorticography
EEG	Electroencephalography
EUC	Euclidean Distance
FDA	Functional data analysis
FFT	Fast Fourier transform
fFOBI	functional Fourth-Order Blind Identification
fICA	Functional independent component analysis
fMRI	functional Magnetic Resonance Imaging
fPCA	Functional principal component analysis
FS	Feature selection
ICA	Independent component analysis
ITP	Interval Testing Procedure
K-L	Karhunen-Loève
KMO	index Kaiser-Meyer-Olkin index
LARS	Least Angle Regression
LASSO	Least Absolute Shrinkage and Selection Operator
MCD	Minimum covariance determinant
MF-ICA	Mean-field independent component analysis

MFLR	Multiple functional logistic regression
ML	Machine learning
MD-FPCA	Multidimensional functional principal components analysis
MSE	Mean Square Error
pMFLR	Penalized multiple functional logistic regression
pFICA	Penalized functional independent component analysis
PSD	Power spectral density
rFICA	robust functional independent component Analysis
SCAD	Smoothly clipped absolute deviation
SE	squared error
SSR	sum-square-root

Chapter 1

General Introduction

The human brain is a complex organ, influenced by various factors ranging from finger movement to heartbeat, and plays a crucial role in behavior control and emotional states. The analysis of neuro signals is of paramount importance in the realm of neuroscience and beyond. These intricate signals, often recorded through methods like EEG or fMRI, provide a window into the workings of the human brain. Understanding and interpreting neuro signals hold the key to unraveling the mysteries of cognition, behavior, and emotions. By studying these signals, researchers can gain valuable insights into neurological disorders, cognitive processes, and emotional states. Moreover, the application of neuro signal analysis extends to fields as diverse as brain-computer interfaces, mental health diagnostics, and the development of assistive technologies. In essence, the analysis of neuro signals not only deepens our comprehension of the brain but also has far-reaching implications for healthcare, technology, and our overall understanding of human nature. Despite the establishment of numerous models aimed at explaining the relationship between neural signals and motor activity, mental states, and cognitive processes, it remains uncertain to what extent information can be decoded from these signals. Hence, the exploration of neural signal data is an ongoing venture.

Neuro signals are inherently treated as conventional time series data. However, in the majority of analytical investigations, researchers tend to pay more attention to information extracted from the frequency domain, compared to the time domain. Furthermore, even when delving into the time domain, neural signals are routinely treated as discrete time series and applied to multivariate analysis. Even though it is widely embraced, the classical multivariate approach overlooks the intricate functional dynamics inherent in the data generation process, including continuity and smoothness. Moreover, it also suffers from highly correlated measurements within each functional signal. Hence, the proposition of employing FDA for the analysis of neural signals emerges as a promising avenue. FDA, as an advanced analytical framework, offers the prospect of extracting comprehensive information from both the time and frequency domains while thoughtfully addressing the inherent temporal dependencies within neural signal dynamics. This approach holds the potential to unveil previously concealed

insights in perspective of time domain and enhance our understanding of the intricate processes underlying neuro signal data.

Specifically, in this dissertation, we introduce three methods centered around addressing the following questions:

1. How do we classify human behaviors based on neural signal data while considering the time dependence of data?
2. How do we improve the preprocessing step for neural signals?

To answer these two questions, we explore the application of FDA across every phase of signal processing for EEG data, which is a representative data type in neural signal analysis.

Firstly, we perform classification analysis on neuroscience data, treating neuro signals as functional data, which allows researchers to leverage the temporal and spatial characteristics of these signals. Specifically, FDA provides a framework to work with continuous functions, enabling researchers and analysts to uncover hidden patterns, trends, and relationships within signals. This approach is particularly useful when dealing with complex and high dimensional data, such as time-series data from various domains including biology, finance, engineering, and more. Therefore, signal analysis methods based on functional data analysis represent a powerful approach to understanding and extracting meaningful information from signals that are inherently dynamic and continuous in nature. Traditional signal analysis techniques often treat signals as discrete data points, disregarding the temporal or spatial continuity present in the data. Functional data analysis, on the other hand, recognizes that signals can be better understood as functions, where each signal is a curve or trajectory that evolves over a continuous domain.

After a comprehensive analysis of the results obtained from our proposed approach when applied to actual scalp EEG signals, we have observed that the neuro signal analysis is sensitive to outliers. Consequently, our focus shifts to the EEG signal pre-processing stage, with the objective of extracting essential information from raw signals. Given the pivotal role that ICA plays in signal analysis, as it excels in uncovering hidden sources of information within mixed signals, our objective is to create an ICA methodology that demonstrates robustness in the face of outliers. This quality is especially critical when working with neuroscience data, where outliers are an inherent and unavoidable challenge. In many real-world scenarios, signals are often convoluted or contaminated by noise, rendering them challenging to interpret accurately. ICA comes to the rescue by enabling the separation of these mixed signals into their underlying, statistically independent components. However, the effectiveness of ICA in signal analysis

hinges significantly on parameter tuning. The choice of parameters, such as the number of components to extract and the algorithm used, can profoundly impact the quality of the results. Careful parameter selection is essential to achieve a meaningful and accurate separation of sources, as incorrect choices may lead to either under- or over-extraction of components, potentially masking important information or introducing artifacts. Thus, the proper calibration of ICA's parameters is a critical step in the signal analysis process, influencing the overall success and reliability of the insights derived from this powerful technique. Hence, we firstly propose a determination method, named CW_ICA, to automatically determine the optimal number of ICs. In addition, we propose a robust functional ICA (FICA) method, named rFICA, to recover the pattern of EEG components while remaining impervious to the influence of outliers.

In the forthcoming sections, I will provide introductory details on neuroscience data, FDA, and ICA algorithms.

1.1 Neuroscience Data Analysis

1.1.1 Electroencephalogram Data

EEG is a powerful neuro-physiological technique that allows us to capture and study the electrical activity of the brain non-invasively[1]. Scalp EEG provides a window into the dynamic patterns of neural activity occurring in the outer layers of the brain. This technique has revolutionized our understanding of brain function, offering insights into cognition, emotion, sleep, and various neurological disorders.

Scalp EEG data is obtained by placing electrodes on the scalp's surface, which then detect and record the electrical signals produced by the firing of neurons in the brain (Figure 1.1[2]). These signals, known as EEG signals or brainwaves, represent the aggregate activity of millions of neurons and provide a time-sensitive view of brain activity. The raw EEG data is a continuous time-series of voltage fluctuations, reflecting the synchronized firing of neurons in response to various cognitive tasks, sensory stimuli, or internal mental processes. Through careful analysis of these data, researchers can extract valuable information about brain states, connectivity patterns, and temporal dynamics. For instance, one of the key features of scalp EEG data is the presence of distinct frequency bands. These frequency bands represent different

brain activities and states. For example, the delta band (0.5-4 Hz) is typically associated with deep sleep and unconsciousness, while the theta band (4-8 Hz) is linked to early stages of sleep, meditation, and memory processes. The alpha band (8-13 Hz) is prevalent during relaxed wakefulness with eyes closed, and the beta band (13-30 Hz) is associated with active thinking, problem-solving, and alertness. The gamma band (30-100 Hz) is thought to be involved in higher cognitive functions. Analyzing EEG data in these frequency bands can provide insights into the brain's functional states.

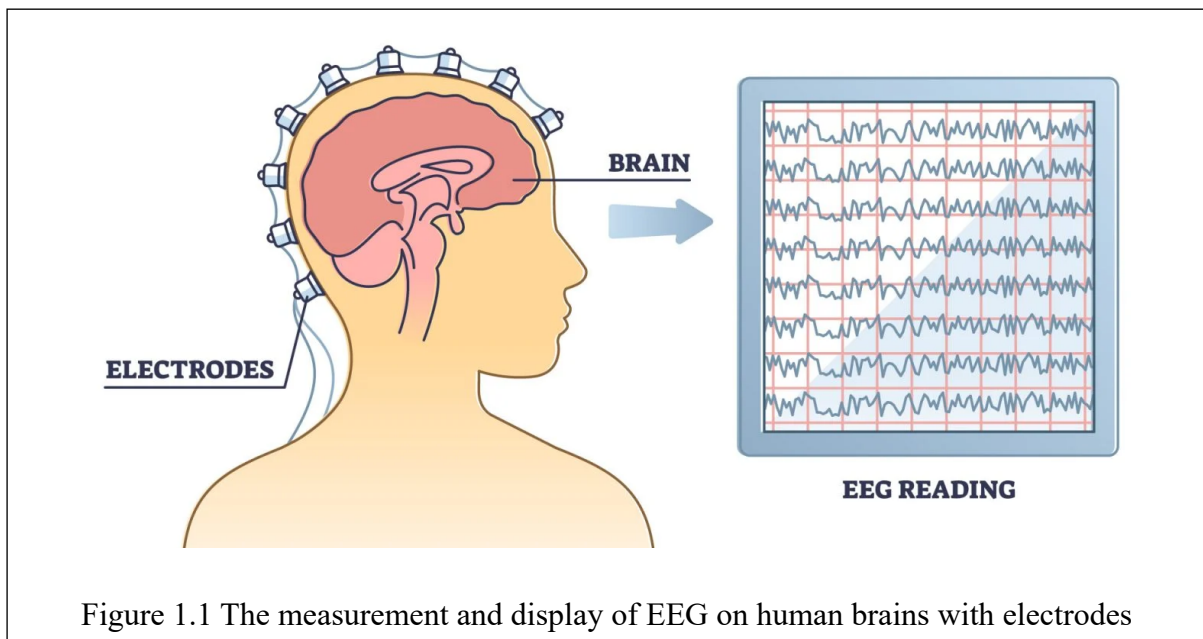


Figure 1.1 The measurement and display of EEG on human brains with electrodes

Advancements in signal processing and computational techniques have facilitated the extraction of meaningful insights from scalp EEG data. Researchers can identify distinct frequency bands (e.g., delta, theta, alpha, beta, and gamma) associated with different cognitive functions and mental states. By studying the changes in these frequency bands, researchers can infer cognitive processes such as attention, relaxation, and alertness. Scalp EEG data has applications in various domains, including cognitive neuroscience, clinical neurology, psychology, and brain-computer interfaces. In cognitive neuroscience, it helps uncover the neural mechanisms underlying perception, learning, memory, and decision-making. In clinical settings, scalp EEG is used for diagnosing and monitoring neurological disorders such as epilepsy, sleep disorders, and brain injuries.

As technology continues to advance, the analysis of scalp EEG data becomes increasingly sophisticated. Advanced techniques such as source localization, connectivity analysis, and

machine learning algorithms enable researchers to uncover deeper insights into the intricacies of brain function and dysfunction.

In this dissertation, we applied our proposed methods to an openly available EEG dataset. The researcher recruited 19 adults (7 females, 12 males) from the University of Arizona and recorded their scalp EEG*. When undergoing EEG recording, participants navigated in virtual reality through teleporters and track the spatial distances travelled inside teleporters. Data from 48 repetitions of distance monitoring were collected. The 48 repetitions were split into two groups with two possible distances travelled: short (100 virtual meters) vs long (200 virtual meters) distances. To extract the time periods when participants were explicitly instructed to monitor the distance, researchers epoched the continuous data surrounding the teleportation periods. Epoch onsets (time 0) were defined as the moment when participants initiated a teleportation, or the moment when they were instructed to start monitoring distances. Epoch offsets (the end of epochs) were defined as the moment when participants exited the teleporter and stopped monitoring spatial distances. Each epoch was cropped for 5.656 seconds long. The scalp EEG signal was recorded with a 64-channel BrainVision ActiCAP system, which included a wireless transmission MOVE module, and two BrainAmp amplifiers (BrainVision LLC, Morrisville, 90 NC). The sampling rate was 500 Hz and the reference electrode was FCz.

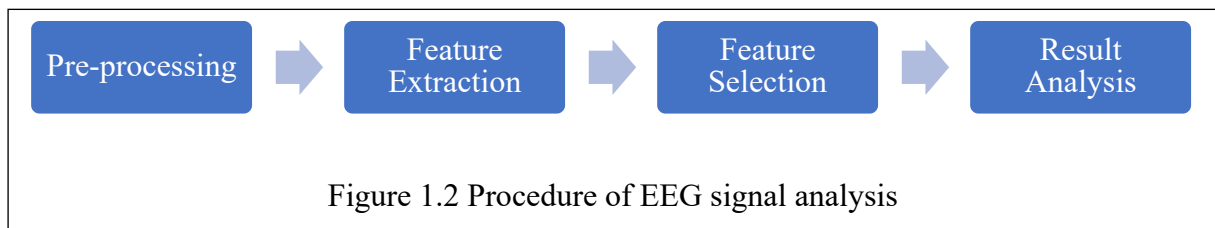
In Chapter 2, our focus lies in applying a functional classification model to neuroscience data. Consequently, we utilized the proposed classification algorithm on the meticulously preprocessed EEG dataset, as handled by neuroscientists. In Chapters 3 and 4, we are interested in the development of signal preprocessing methods. In this pursuit, we employed the proposed ICA dimensionality selection and functional ICA method on the collected raw EEG dataset.

1.1.2 EEG signal analysis

The EEG signal analysis basically performed in four steps: pre-processing, feature extraction, feature selection and finally being at the result analysis phase, the disease diagnosis, or the recognition of the different functional states of the brain is made through models or the statistical tests.

Preprocessing of the EEG signal is essential to obtain only brain activity from the noisy EEG recordings. It aids to eliminate unwanted artifact from the EEG signal and makes it suitable for further processing. Preprocessing of EEG largely includes a number of processes, such as line noise removal, adjustment of referencing, elimination of bad EEG channels, and

artifact removal. Blind source separation (BSS) is one of commonly used signal preprocessing techniques[2]. Petrov et al.[3] proposed basic filtering techniques to remove the unwanted artifacts from the EEG signal. Campbell et al.[4] employed band-pass filtering to eliminate any signal that is not in the range of P300 frequency range of the EEG signal. Maki et al. proposed a new method, multichannel Wiener filter, for enhancement of EEG signal. Jirayucharoensak and Israsena[6] proposed an artifact removal technique that incorporates Lifting Wavelet Transform. Various preprocessing techniques have been discussed by Vidaurre et al.[7] for effective artifact removal from EEG signals.



Signal preprocessing also faces various challenges in modern data-driven applications. Specifically, in the context of biomedical signal processing, noise reduction and artifact removal are critical tasks. These challenges underline the need for robust and adaptive preprocessing techniques to enhance signal quality, which is essential for subsequent analysis, interpretation, and decision-making in various domains. In order to tackle these challenges, we discuss the development of independent component analysis (ICA), which is a widely used blind source separation method for signal pre-processing. Specifically, we propose a method to select the optimal number of ICs in Chapter 3 and explore methods for robustly extracting ICs that accurately recover true signals in Chapter 4. A more comprehensive introduction to ICA can be found in Section 1.3.

Feature extraction of EEG signal is an important step in any BCI-based applications. It helps to extract the most relevant features from the EEG signal, thus giving a more precise description and hence making it suitable for further processing. An EEG is arbitrary and non-stationary signal, thus only fast Fourier transform (FFT) cannot efficiently differentiate EEG signals. Therefore, other nonlinear methods are used to extract features, namely, sample entropy [8], Hurst exponent [9], Lyapunov exponent (LLE) [10], and multi-fractal detrended fluctuation analysis [11] which are popular for feature extraction. Besides, a nondeterministic polynomial (NP) problem often arises out of optimal feature subset selection; hence, for optimal feature subset searching, genetic algorithm (GA) is often used [12]. The discrete Fourier transform

(DFT) [13] is a digital signal that describes signal amplitude versus sampling time constant in time domain frequency. Discrete cosine transform (DCT) [14] is very useful for transforming the encoding video and audio tracks on computers. Continuous WT (CWT) [15] method is used to represent the translation and scale parameter for the wavelet continuously. DWT [16] method is used to transform any wavelet to discretely sample. WT [16],[17] performs a valuable role in the recognition and diagnostic field. As the EEG signal is a time-varying entity, this method is well suited for feature extraction from the raw data in time-frequency domain. In general, feature extraction method can be divided into three types: Time domain signal extraction [18], Spatial domain feature extraction[19] and Feature extraction transformation model. In Chapter 2, features containing time and frequency information are extracted from different frequency band across multiple channels and express as functions of time.

One of the main challenges facing EEG signals is finding the right information for identifying cognitive states. Considering this, feature selection is utilized to reduce the dimensionality impact on the dataset through finding the subset of feature which efficiently define the data [20],[21]. It selects the important and relevant features to the mining task from the input data and removes redundant and irrelevant features [22],[23]. It is useful for detecting a good subset of features that is appropriate for the given problem [24],[25]. The main purpose of feature selection is to construct a subset of features as small as possible but represents the whole input data vital features [25],[26] re-selection provides numerous advantages: reduce the size of data, decrease needed storage, prediction accuracy improvement, overfitting evading, and reduce executing and training time from easily understanding variables. Feature selection algorithm phase is divided into two-phase such as Subset Generation and Subset Evaluation: In subset Generation, we need to generate subset from the input dataset and to use Subset Evaluations we have to check whether the generated subset is optimal or not [27],[28]. Originally evaluation methods in feature selection are divided into four kinds: filter, wrapper, embedded [29],[20], and hybrid [22],[30]. Recently, another type of evaluation method is developed, i.e., ensemble feature selection [31],[32]. In Chapter 2, the functional t-test is utilized to identify "functional" features that exhibit significant overall differences between the means of two trajectories. This test not only extracts signals that differ significantly between the two tasks but also provides interpretable insights into the temporal characteristics of cognitive components. The results of the functional test pinpoint the time intervals where significant differences exist between the two classes.

With the rapid growing application of Internet of Thing, diversity in the process of classification is being observed from EEG signal. It can be divided into two parts: Linear and Nonlinear. k-Nearest neighbor (k-NN) is a classifier [33] which follows a non-parametric approach; it used to classify a given data point according to the neighbor's majority. For real-time BCI applications, linear discriminant analysis is suitable for its lower time and space complexity [34]. In addition, this algorithm is quite simple from the users' perspective. The decision tree (DT) is one of the most powerful and useful tools used for classification of data and prediction is the decision tree [35]. For instance, it can be used for classification of emotions from EEG signals, where each internal node of decision tree defines an attribute of test, each branch tree depicts a result of the test, and each leaf node shows a class label. Adaptive Boosting follows metadata algorithms of machine learning, which is used for weak learner and is adaptive in the sense. It is weak for the instances that are mis-classified by the previous classifiers. A multi-layer perceptron (MLP) follows feed forward method of artificial neural network. In general, propagation method for training in a supervised learning is used for MLP. Naive Bayes classifiers [36] belong to the family of classifier where the concept of probability is used. They require a number of parameters of linear variables that are highly scalable to a learning problem. In this dissertation, our focus centers on the classification of human behaviors based on scalp EEG signals.

Currently, there is a lack of a systematic approach for analyzing EEG data through FDA, despite numerous studies on EEG signal classification. Existing FDA methods for EEG analysis typically focus on either time-related or frequency-related information. While some research has explored FDA with frequency-related data, it often implicitly incorporates spatial information among adjacent voxels in three-dimensional brain image data. To address this gap, we present a three-stage algorithm in Chapter 2. This algorithm leverages the power of FDA, taking into account both time and frequency information, to achieve an interpretable classification of scalp EEG signals. A more detailed introduction to FDA will be presented in Section 1.2.

Furthermore, we employ the random forest classification method on IC scores and frequency information of ICs, extracted through various ICA techniques, to assess the robustness of the proposed rFICA method.

1.2 Functional Data Analysis

Functional Data Analysis (FDA) is a specialized branch of statistics that deals with datasets consisting of continuous functions rather than traditional scalar or vector observations. In FDA, data is represented as curves or functions that vary smoothly over a continuous domain, such as time, space, or wavelength. This approach is particularly valuable in various fields, including economics, biology, neuroscience, and engineering, where data are naturally observed as functional observations. Over last two decades, many FDA techniques have been developed as extensions of multivariate data analysis techniques, such as functional principal component analysis (fPCA), functional classification and functional regression. FDA enables researchers to extract meaningful insights from complex datasets, uncover underlying patterns, and make accurate inferences from the continuous nature of the data.

1.2.1 Elements of Functional Data Analysis

Consider sample curves of the form x_{it} , where $i = 1, \dots, p$, $t = 1, \dots, T$, denotes the i^{th} trial at t^{th} time point. The idea of FDA is to estimate the continuous function $x_i(t)$ using the observed signal x_{it} . By definition, a basis function system is a set of known functions, $\{\phi_j(t)\}, j = 1, \dots, N$, that are independent from each other with the property that one can approximate arbitrarily well any function by taking a linear combination of a sufficiently large number of these functions. Therefore, the assumed smooth functional observation, or linear expansion, $x_i(t)$, can be expressed as

$$x_i(t) = \sum_{j=1}^N a_{ij} \phi_j(t) \quad (1.1)$$

where, each a_{ij} is called a basis coefficient.

Two important issues arise from this basis representation of a function. One of the important issues is the choice of the basis function system. The determination of the appropriate basis function system is based on the characteristics that are inherent in the data. There are many different basis systems that can be used in a basis expansion. In general, Fourier basis functions are used to model periodic data and B-spline basis is used for non-periodic data. Different basis

choices, such as wavelets, trigonometric functions, or even polynomial functions, can be considered if they align better with the inherent properties of the data [37].

The second important aspect in this representation is the dimensionality of the chosen basis, N . Since the degree to which the data X_i is smoothed, is determined by the number of basis functions, therefore, N can also be viewed as a smoothing parameter. When an appropriate basis function system is selected for the observed data, fewer basis functions are required to make a good approximation of the function. There are two opposing discussions when it comes to choosing the number, N , of basis functions. On the one hand, the larger the N , the better the fit. However, this comes with the expense of fitting unnecessary noise that should ideally be ignored. Conversely, a smaller value of N results in a smoother function, as noted by Ramsay et al.[37] However, the drawback here is that if N is set too low, there is a risk of overlooking significant features within the function being estimated. Therefore, a trade-off between fit and smoothness is necessary. A variety of algorithms are available for this purpose, encompassing methods like cross-validation and generalized cross-validation, as discussed by Craven and Wahba .

The functional variable $X(t) \in L^2(T)$ is endowed with the inner product $\langle f, g \rangle = \int f(t)g(t)dt$ and the induced norm $\|f(t)\| = \langle f, f \rangle^{1/2}$. For $s, t \in T$, the covariance operator Cov is an integral operator with kernel $Cov(s, t) = \frac{1}{p} \sum_{i=1}^p x_i(s) x_i(t)$ admitting the Mercer decomposition.

$$Cov(s, t) = \sum_{k=1}^{\infty} \eta_k \gamma_k(s) \gamma_k(t) \quad (1.2)$$

where η_k and $\gamma_k(t)$ are eigenvalues and associated orthonormal eigenfunctions. The functionas $x_i(t)$ can also be approximated represented by a truncated series of the Karhunen-Loève (K-L) expansion.

$$x_i^K(t) = \sum_{k=1}^K z_{ik} \gamma_k(t) \quad (1.3)$$

where $z_{ik} = \langle x_i(t), \gamma_k(t) \rangle$ are vairables with $E[z_{.k}] = 0, Var(z_{.k}) = \eta_k, Cov(z_{.k}, z_{.k'}) = 0$ for $k' \neq k$. These variables are referred to as the principal components scores and are uncorrelated generalized linear combinations of the functional variable with maximum variance.

In FDA problems, the initial and fundamental step involves estimating mean functions [39], [40]. Rice and Wu [40] adopted the mixed effect models where the spline coefficients were estimated by the EM algorithm. The mean and covariance functions are estimated by application of the local linear smoothers in [41]. The functional mixed model framework, with model fitting done by using a Bayesian wavelet-based approach, is generalized from the linear mixed model by Morris et al.[42] A polynomial spline estimator is proposed for the mean function of functional data together with a simultaneous confidence band by Cao et al.[43]

Moreover, the functional principal component analysis (fPCA), an extension of Principal Component Analysis (PCA), is a key dimension reduction tool for multivariate and functional data. A comprehensive framework for statistical inference in fPCA was initially developed by Dauxois et al.[44] Since then, fPCA has become a popular technique in FDA. It converts infinite-dimensional functional data into a finite-dimensional vector of random scores, referred to as functional principal components (FPCs). These FPCs, frequently truncated in practical applications, enable the utilization of established multivariate data analysis methods, effectively achieving dimension reduction [45],[46].

Additionally, outliers in functional data present various challenges due to the high dimensionality of such data. These outliers can appear as unusual measurements at either single or multiple time points or as irregular patterns across an entire function. In order to handle outliers and contamination, there are many robust adaptations of FDA techniques [46]–[50].

In this dissertation, we conduct the functional data analysis with B-spline basis function, since the EEG data is non-periodic. Moreover, we incorporate the algorithm of functional ROBPCA [46] and KFPCA [50] in the development of a robust functional ICA method in Chapter 4.

1.2.2 Functional Data Models

There are various methods that nonparametrically estimate the regression function. In general, most functional regression models can be viewed as extension of the traditional regression models. Numerous functional regression models are accessible, which can be categorized into three classes: function-on-scalar regression, function-on-function regression, and scalar-on-function regression. Most work in functional predictor regression is based on a variant of the Functional Linear Model (FLM), first introduced by Ramsay et al. [51] and first written in its commonly encountered form by Hastie et al.[52]

1. Function-on-scalar regression. Supposed there are multiple functional covariates $x_{il}(t)$ and coefficient function $\beta_l(t)$ with domain T where $l = 1, \dots, L$, we obtain the functional linear model as below:

$$Y_i = \beta_0 + \sum_{l=1}^L \int_T x_{il}(t) \beta_l(t) dt + \epsilon_i \quad (1.4)$$

where $Y_i, i = 1, \dots, p$, is the scalar response, β_0 is the intercept, $x_{il}(t), l = 1, \dots, L$, is the functional predictor, $\beta_l(t)$ is the functional coefficient associated with $x_{il}(t)$, ϵ_i is the residual error that assumed to be gaussian distribution with mean zero.

2. Function-on-function regression. Given a sample of functional responses $Y_i(t)$ and functional predictors $X_{il}(t)$, a general linear function-on-function regression model can be expresses as below:

$$Y_i(t) = \beta_0(t) + \sum_{l=1}^L \int_T x_{il}(t) \beta_l(t) dt + \epsilon_i(t) \quad (1.5)$$

where $Y_i(t)$ is the functional response, $\beta_0(t)$ is the intercept function, $\beta_l(t)$ is the coefficient function and $\epsilon_i(t)$ is the residual function.

3. Scalar-on-function regression. Functional response regression involves the regression of functional responses on a set of scalar predictors. Suppose there is a sample of functional responses $Y_i(t)$ and scalar predictors x_{il} , a general linear scalar-on-function regression model can be expresses as below:

$$Y_i(t) = \beta_0(t) + \sum_{l=1}^L x_{il} \beta_l(t) + \epsilon_i(t) \quad (1.6)$$

where functional coefficient $\beta_l(t)$ represents the partial effect of predictor x_{il} on the response at time t . The goal of scalar-on-functional response regression is often estimation of $\beta_l(t)$ and then either testing whether $\beta_l(t) = 0$ or assessing for which t is it true that $\beta_l(t) \neq 0$. The $\epsilon_i(t)$ denotes the curve-to-curve residual error deviation.

In this dissertation, we will restrict the discussion of functional regression with the first case, that is, with functional predictors and a scalar response.

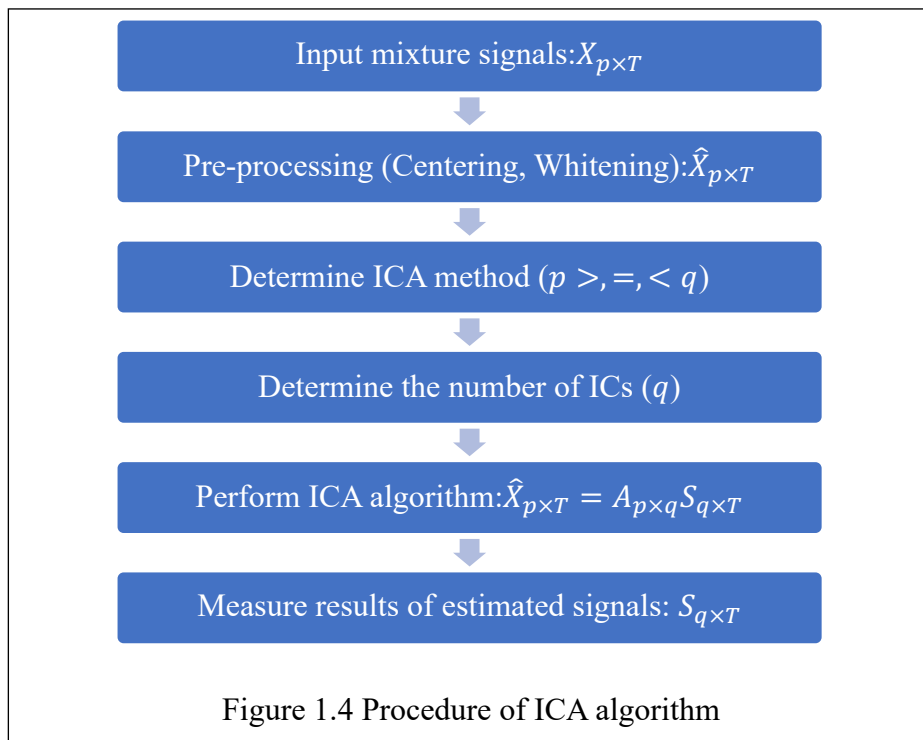
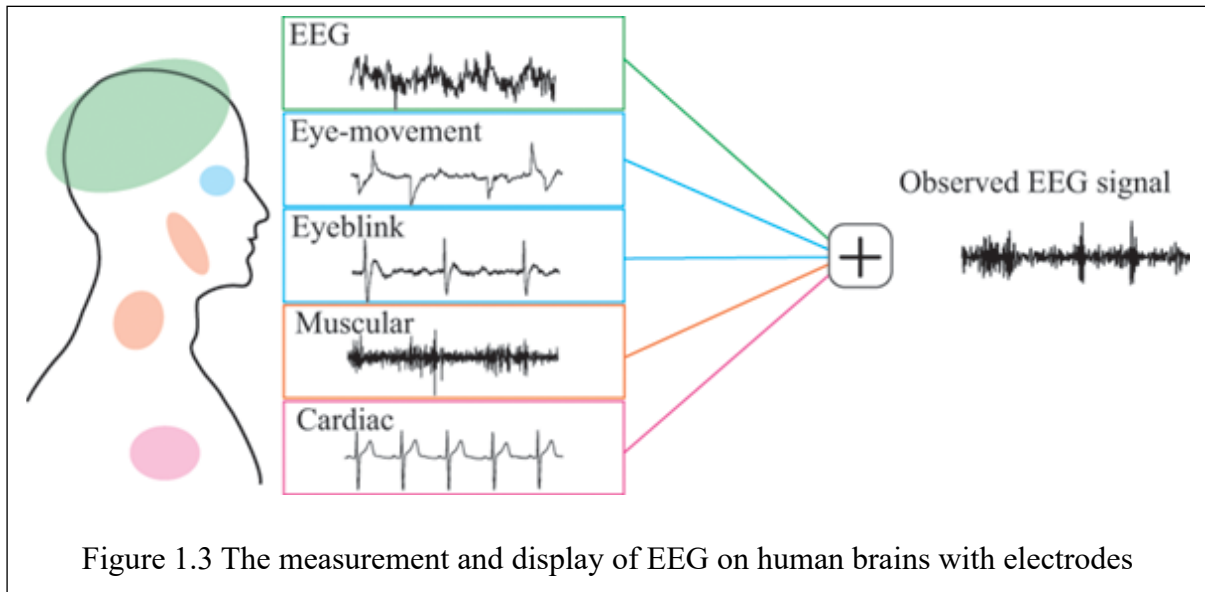
Moreover, another fundamental problem in FDA is classification of functional data. In functional classification, the goal is to categorize or classify these functions into different

groups or classes based on their shape, patterns, or other characteristics. Functional classification models are particularly useful in fields like biomedical signal analysis (e.g., classifying EEG or ECG signals), speech processing, image analysis, and more. They take into account the entire functional nature of the data, considering how it evolves over a range, rather than treating individual data points as independent entities. Common techniques for functional classification include functional regression models, support vector machines, decision trees, and deep learning approaches adapted for functional data. These models can provide valuable insights and predictions in a wide range of applications where functional data is prevalent.

Specifically, we construct the penalized multiple functional logistic regression model in Chapter 2 to interpretably classify scalp EEG signals based on extracted time-frequency information.

1.3 Independent Component Analysis

Independent component analysis is a statistical tool to extract hidden information from observed signals. With the assumption that observed signals are linear combination of mutually independent non-Gaussian source signals, ICA is utilized to find a linear transformation of these mixed signals to recover source signals. Based on assumptions of ICA problem, the performance of ICA is measured by independence or non-Gaussianity of estimated ICs. ICA has been widely applied in various fields. In biomedical fields, ICA has been used to study the brain function via extracting temporal and spatial information from fMRI [53], [54] and EEG [55]. In pharmaceutical fields, ICA is used to examine the distribution of actives and major excipients within the tablet by comparing the calculated signals with the pure spectra of the formulation compounds [56]. In chemistry, ICA is widely applied for separation of unknown sample mixtures [57] via peak detection and matching in high-performance liquid chromatography. Besides, ICA is also widely used in the preprocessing of time series (or signals) to identify and remove noise and contamination to better extract qualified information. In this dissertation, we investigate the application of ICA methods on EEG data.



Suppose that observed signals are organized into a data matrix $X_{p \times T}$ with p measured signals and signal length T , and assuming that these measured signals are linear mixtures of source signals. After preprocessing, the general model of ICA can be described as $\hat{X}_{p \times T} = A_{p \times q} S_{q \times T}$, where $\hat{X}_{p \times T}$ is preprocessed mixed signal, $A_{p \times q}$ is the mixing matrix, which specifies contributions of the source signals to each mixture, and $S_{q \times T}$ is the matrix of source signals. ICA aims to determine both mixing matrix A and source signal matrix S , knowing only

observed signal matrix X . According to the number of observations, p , and source signals, q , ICA methods can be divided into two cases: (over)determined ICA ($p \geq q$) (e.g., FastICA [58], JADE [59], Infomax [60], etc.), and underdetermined ICA ($p < q$) (i.e., single channel ICA [61], MAICA [62], EMD_ICA [63], etc.). In this study, we focus on the case of (over-)determined ICA, where the mixing matrix A is invertible. Therefore, the object of ICA can be achieved by estimating the de-mixing matrix $W = A^{-1}$, so that source signals (i.e., ICs) can be obtained by projecting the whitened data onto the de-mixing matrix.

The procedure of ICA consists of five steps: pre-processing of ICA, determine the ICA method, determine the number of ICs, perform ICA and measure the result of ICA with cross validation. Firstly, we centered observed signals by subtracting the mean value of signals at each time point and transform the centered mixture signals to be uncorrelated and then scaled them to be with a unit variance. Next, the appropriate ICA method is chosen by the property of mixed and source signals, i.e., (over-)determined case or underdetermined case. The main goal of these methods is to extract independent components by maximizing the non-Gaussianity, minimizing the mutual information, or using maximum likelihood (ML) estimation method [64]. Besides the choice of ICA model, the main parameter in the ICA method is the number of ICs (i.e., number of dimensions), which is an important step before any modeling. Then, with determined ICA method and number of ICs, we can perform ICA to separate mixture signals. The results of ICA can be measured by comparison between estimated and true source signals if there exists or between models with different number of ICs.

The determination of the number of independent components (ICs) is crucial for achieving optimal performance, as an incorrect choice can result in either under-decomposition or over-decomposition. Commonly used determination techniques include information criteria, eigenvalue spectrum (ES), bootstrap resampling (BS), and cross-validation (CV), among others. Nevertheless, these methods have their drawbacks. For instance, information criteria may suffer from overfitting when the sample size is small or strict model assumptions are made. Eigenvalue spectrum methods can be subjective in the choice of threshold and may be affected by noisy signals. Bootstrap resampling techniques, although comprehensive, can be computationally expensive. Cross-validation, while generally reliable, may introduce data partition bias and incur computational costs. To address the issues with these determination methods, researchers have proposed several alternatives. However, there still exists some challenges, i.e., restriction of data structure, robustness, capability of diverse ICA methods, etc.

Giving these limitations, we propose a novel method to automatically determine the optimal number of ICs in Chapter 3.

In recent years, researchers have expanded the application of ICA to functional data, referred to as Functional ICA (fICA). This extension is designed to decompose a collection of functional observations into statistically independent components or sources. fICA is particularly valuable when handling data characterized by intricate temporal dependencies, revealing hidden sources of variability and information with consideration of time-continuity. It's important to highlight a significant gap in the existing literature, which lacks emphasis on evaluating the recovery of extracted ICs in comparison to the true ICs. Moreover, there is a notable absence of discussions regarding the robustness of these fICA methods. In response to this gap, in Chapter 4, we propose a robust fICA method designed to investigate the analysis of EEG signals, particularly in the presence of contamination.

1.4 Aims and Scope of the Thesis

The overarching objective of this thesis is to improve the performance of neuroscience signal analysis. To achieve this goal, we involve a comprehensive investigation of the FDA methodology at each stage of processing EEG signals. By incorporating FDA techniques throughout the EEG signal processing pipeline, we aim to tackle traditional analytical constraints. Our focus is not only on improving the quality and robustness of the extracted insights but also on rendering the analytical models more interpretable, ensuring that the findings are not only accurate but also readily comprehensible to facilitate a deeper understanding of the underlying neural processes. Ultimately, this thesis aspires to equip the field of neuroscience signal analysis with cutting-edge tools and methodologies, enabling researchers to extract more reliable, meaningful, and interpretable insights from this complex and dynamic data domain.

In Chapter 2, we introduce a three-stage algorithm designed for the classification of EEG signals, emphasizing interpretability through FDA and dimension reduction. Based on the extracted time-frequency information of EEG signal, we conduct functional testing to select features that are significantly different signals with specific time interval under the two tasks. While functional testing provides valuable temporal insights into EEG signals, it is important to acknowledge that its computational complexity increases with a growing number of features

and extended recording time. Additionally, EEG signals are susceptible to the influence of outlier factors, it is necessary to denoise the original signal to improve the signal-to-noise ratio as much as possible before formally analyzing the data. Thus, we move back to investigate the signal pre-processing. Consequently, we revert to the exploration of signal preprocessing methodologies, specifically, ICA methods.

In Chapter 3, our primary focus lies in addressing a pivotal aspect of ICA – the determination of the optimal number of ICs. We introduce a robust and automated method for this critical task, which contributes significantly to the reliability of subsequent data analysis. Our method leverages the concept of a rank-based correlation matrix, incorporating specific quantitative measurements to identify the optimal number of ICs. The results affirm CW_ICA as a reliable, robust, and computationally efficient technique for determining the optimal number of ICs, that is required for accurate and meaningful analysis of the data. This method not only enhances the quality of the extracted components but also streamlines the data processing pipeline, ensuring its suitability for various ICA applications in the field of neuroscience signal analysis.

Following the accurate determination of the true number of ICs, our exploration ventures into the realm of ICA with a specialized focus on the functional domain. In Chapter 4, we introduce a robust and innovative approach to fICA. This direction is motivated by a profound awareness of the susceptibility of neuroscience data to potential contamination stemming from both internal and external sources. Our novel approach incorporates robust covariance estimation techniques, with a specific emphasis on the Minimum Covariance Determinant (MCD) method, which enhances the overall robustness of the analysis. To further robustify our method against potential outliers, we introduce Kendall's τ function. This additional layer of analysis empowers our approach with a higher degree of resilience against anomalies, such as bumps and sample outliers that might exist in the data. Our approach not only enhances the reliability and robustness of the analysis but also represents a significant advancement in the field of neuroscience signal analysis. This advancement holds the promise of elevating the quality and reliability of the insights derived from neuroscience data, making it a valuable contribution to the field.

In summary, this dissertation presents a comprehensive contribution to the field of neuroscience data analysis. It introduces a novel functional classification algorithm rooted in FDA, enabling accurate and interpretable classification of neural behaviors. Furthermore, we propose a robust dimension selection method that leverages rank-based correlation and

quantitative measurements. This approach not only streamlines subsequent analyses but also enriches the quality of insights obtained from neuroscience data. Additionally, the research extends into the development of robust fICA method, designed to address contamination in neuroscience data. By incorporating the Minimum Covariance Determinant (MCD) and Kendall's τ function, the method bolsters the accuracy and resilience of brain activity analysis. Collectively, these contributions aim to elevate the quality, interpretability, and reliability of neural signal analysis, marking a significant advancement in neuroscience research.

Chapter 2

Interpretable EEG classification algorithm

2.1 Introduction

Scalp EEG signals can be analyzed in time domain, frequency domain or time-frequency domain. Time domain analysis¹ is conducted based on how signals change with time whereas frequency domain analysis often gives an intuitive understanding of frequency components of EEG signals. Besides, time-frequency domain analysis studies EEG signals in both time and frequency domains simultaneously. Regardless of the domain of signals, EEG analysis methods can be roughly divided into traditional and modern models. The choice of traditional analysis method is based on properties of EEG signals. For instance, non-linear methods is used to study self-organization and pattern formation in the complex neuronal networks of the brain since the EEG data is nonlinear and non-stationary [66]. With the rapid development and popularization of machine learning techniques, EEG analysis has adopted modern machine learning (ML) methods such as deep neural networks [67], support vector machines [68], random forest [69], and other ML algorithms that use high-dimensional models [70],[71],[72] to better decode signals.

Since the scalp EEG reflects the cortical electrical activity by continuous flow of voltages, it can be treated as functional data and analyzed using FDA techniques. Also, FDA is well suited for EEG signals which are recorded over long periods of time because it uses fixed or data-driven basis functions to handle high-dimensional data through basis expansion(e.g., B-spline basis, Fourier basis and Wavelet basis [73]). Besides, FDA identifies the relationship between EEG signals recorded from multiple channels and human behaviors which can be characterized as a scalar (e.g., N-back task score [74]), binary (e.g., long/short distance task in our study), categorical (e.g., disease status [75]) variable. Most importantly, FDA helps to effectively describe and interpret the internal change of EEG signals and its relationship with human behaviors, which makes it an interpretable approach.

At present, researches have been using FDA to study EEG data, but most of them are only based on time domain or frequency domain [76],[77],[74]. Several papers have performed functional data analysis in the time-frequency domain [78],[79], but there is no thorough comparison of these methods in the literature. However, as mentioned earlier, the time domain information and frequency domain information contained in the signal are both very important for EEG analysis. Hasenstab et al.[76] proposed a multidimensional functional principal components analysis (MD-FPCA) technique which is shown to be useful for modeling longitudinal trends in the event-related potential functions. Happ and Greven [77] proposed the multivariate fPCA and applied to the neuroimaging study. Both studies focus on multidimensional data only in time domain with scalar response. Zhang et al. [74] proposed a multiple functional linear regression model based on EEG signals to predict working memory ability which is a scalar outcome variable. Shangguan et al. [80] recognized the driver fatigue state by conducting signal feature extraction based on FDA and performing classification via random forest and decision tree. The main advantage of using FDA on EEG signals is that it utilizes the information of functional covariates to interpretably explain the changes in time-frequency domain in EEG signals and the relationships among multiple channels and response [82].

In this chapter, we proposed a three-stage algorithm based on functional data analysis, with the advantage of interpretability. Specifically, the time and frequency information are extracted by wavelet transform in the first stage. Then, functional testing is utilized to select channels and frequencies that show significant differences for different human behaviors. In the third stage, we propose to use penalized multiple functional logistic regression to interpretably classify human behaviors. With simulation and scalp EEG data as validation set, we show that the proposed three-stage algorithm provides an interpretable classification of the scalp EEG signals. Our primary contributions can be summarized as follows:

1. **Extract both time and frequency information from scalp EEG signals and treat them as functions.** In this way, the information of EEG signal from different frequency band on multiple channels can be obtained and expressed as a function of time.
2. **Select features by functional testing methods.** They can not only extract significantly different signals under the two tasks, but also interpretably identify the temporal characteristics of cognitive components. The result of functional testing points out the time interval which is significantly different among two classes.

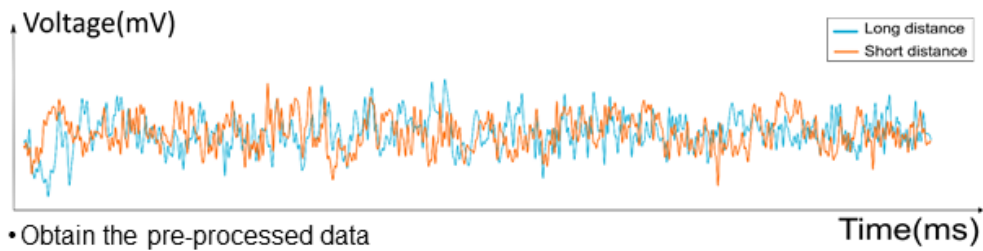
3. **Generate the functional classification model with group counterparts of penalty terms.** It is more suitable for factor selection with multiple predictors, because group penalty terms allow for the joint selection of groups of covariates from the model [83].

The structure of Chapter 2 is organized as follows. We provide the background of application of functional data analysis in EEG signals in Section 2.1. Then, we propose the three-stage algorithm in Section 2.2 and validate our method in simulated EEG data in Section 2.3. The implementation of proposed methods in scalp EEG data can be found in Section 2.4. In Section 2.5, we discuss our findings and the proposed methods, then we conclude this chapter in Section 2.6.

2.2 Proposed Algorithm

In order to conduct interpretable classification in EEG analysis, we propose a method that consists of three stages. In the first stage, both time and frequency-related information are extracted by transforming EEG data into related domain as $\mathbf{X}(t)$ of $n \times L_1 \times T$, where n is the number of subjects, L_1 is the number of features, and T is the time length. The next stage follows the feature selection for initial dimension reduction via functional testing. Penalized multiple functional logistic regression model is used to interpretably classify specific human behaviors at the final stage. The procedure of the proposed algorithm is summarized in Figure 2.1. While the number of trials and recording duration are constant throughout the algorithm, the number of features changes at each step. Therefore, the structure of data frame also changes at each step. Our proposed algorithm is implemented through R language and the R code is freely accessed at <https://github.com/yzy0080/pMFLR.git>.

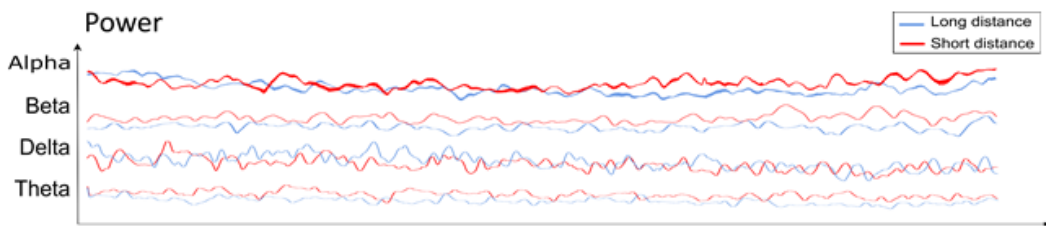
❖ Start with Raw EEG signals



- Obtain the pre-processed data
- $X_{il}(t_j)$, $i=1, \dots, n$; $l=1, \dots, L_0$; $j=1, \dots, T$
- Structure: $n \times L_0 \times T$



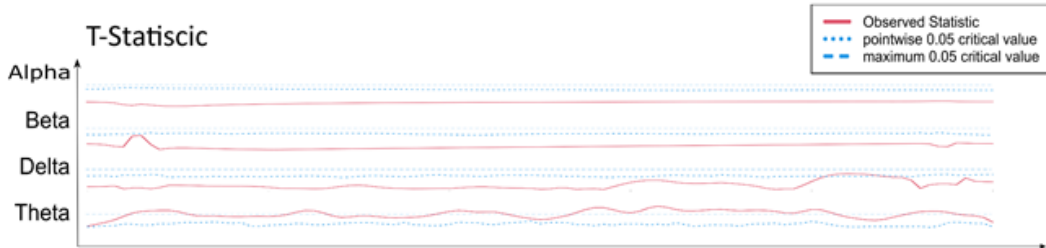
❖ Data Transformation



- Present functional data in Time, Frequency, or Time-frequency Domain
- $X_{il}(t_j)$, $i=1, \dots, n$; $l=1, \dots, L_1$; $j=1, \dots, T$
- Structure: $n \times L_1 \times T$



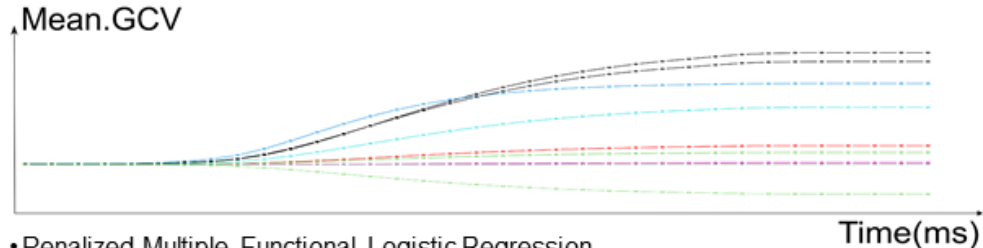
❖ Initial Feature Selection via Functional testing



- $H_0: \mu_{l1}(t) = \mu_{l0}(t), l = 1, \dots, L_1$
- $X_{il}(t_j)$, $i=1, \dots, n$; $l=1, \dots, L_2$; $j=1, \dots, T$
- Structure: $n \times L_2 \times T$



❖ Final Model Construction



- Penalized Multiple Functional Logistic Regression
- $X_{il}(t_j)$, $i=1, \dots, n$; $l=1, \dots, L$; $j=1, \dots, T$
- Structure: $n \times L \times T$

Figure 2.1 Procedure of three-stage algorithm.

2.2.1 Data transformation

Time-domain analysis primarily examines how a signal changes over a duration, while frequency-domain analysis focuses on the mathematical representation of signal characteristics in relation to frequency, rather than time. It reveals the distribution of signal energy across different frequency bands within a defined frequency range.

The most widely used analysis in frequency domain is power spectral density (PSD). PSD quantifies the power associated with each frequency component, contributing to the overall signal, providing valuable insights into signal characteristics. However, it lacks the ability to pinpoint when specific frequencies of interest occur within the signal's specific time point.

Generally, time domain analysis provides better time information, but it is inferior in terms of frequency content information (such as amplitude and phase information) required for EEG classification, while frequency domain analysis provides sufficient frequency information, but ignores the property of EEG signals that the frequency spectrum changes over time, unless it can be performed at each sub-interval time. In order to extract the sufficient information from EEG oscillations, a new approach called "time-frequency analysis" is introduced, which includes decomposing the EEG signal into the amplitude and phase information of each frequency in the EEG and characterizing its change over time [84]. Since the wavelet analysis is especially suitable for non-stationary signal processing, we apply the Morlet wavelet transformation on EEG. Therefore, we can obtain the intensity information of each frequency band at different time positions.

The value in data set is converted from representing voltage to representing power at different frequencies. As we know, each collected data contains information from low-frequency to high-frequency oscillation and oscillations on different frequencies have different trend at different electrodes. Therefore, we also consider feature splitting based on the four brainwaves of the EEG signal, namely δ (0.5-4Hz), θ (4-8Hz), α (8-12Hz), β (12-28Hz).

2.2.2 Initial Feature Selection

As mentioned earlier, the human brain can be divided into several regions, each of which controls different behavioral state, such as movement, sleep, emotion, or memory. In other words, a change in a certain behavioral state has a more or less impact on each area and the magnitude of the change in the recorded EEG signal also varies depending on the position of the electrode. Therefore, one of the main purposes of EEG signal analysis is to find the efficient

information to identify or describe a certain state. Feature selection (FS) algorithm is developed to find a set of features that identifies or describes the relevant event to achieve best predictability.

In order to determine the features which are significantly different across relevant events, several functional testing methods have been performed to choose the most suitable one for the given data, including functional t-test [85], Interval testing procedure (ITP) [86] and one-way ANOVA for functional data via globalizing the point-wise F-test(GPF test) [87]. For instance, we are interested in distance judgement in this paper, then we could claim two classes, which are long distance and short distance, for collected data and determine the features which are significantly different across classes by functional testing. Let $\mu^1_l(t)$ and $\mu^2_l(t)$, $l = 1, \dots, L_1$ denote the trajectory means for feature l under two distance classes, respectively. The statistical problem of interest in this stage focuses on hypothesis testing for each feature:

$$H_0: \mu^1_l(t) = \mu^2_l(t) \quad vs. \quad H_a: \mu^1_l(t) \neq \mu^2_l(t), l = 1, \dots, L_1 \quad (2.1)$$

where H_0 and H_a corresponding to the null and alternative hypothesis, L_1 denotes the number of features after data transformation.

Functional t-test detects the existence of a significant overall difference between two trajectory means by calculating the test statistic, which is expressed as a function of time. The location of these differences can be identified by point-wise testing based on a permutation method. This procedure is distribution-independent and an exact level α test because of the nature of permutation test. However, it is time-consuming for high dimensional data since functional test is performed on each feature.

Interval Testing Procedure (ITP) is a novel technique for functional data which checks the p-values of each basis coefficient and generates a matrix of p-values for each basis function. Its power is comparable to those provided by global inference techniques (e.g., GPF test) and keeps the interpretability of test results, even with large number of components. However, it is time-consuming, especially for the experiments with large number of features.

GPF test is a functional version of one-way ANOVA problem, which globalizes the usual pointwise F-test. It is implemented when the k functional samples are sampled densely and noisily, irrespective of whether the sampling design is regular or irregular across subjects. Comparison of the three functional testing methods is summarized in Table 2.1.

FS Method	Pros	Cons	R package-function
Functional t-test	<ul style="list-style-type: none"> • Distribution independent • Exact p-values for each time point 	<ul style="list-style-type: none"> • Time consuming for high dimensional data 	Package: 'fda' Code: tperm.fd()
ITP test	<ul style="list-style-type: none"> • Distribution independent • Comparable power to the global inference techniques 	<ul style="list-style-type: none"> • Time consuming for high dimensional data 	Package: 'fdatest' Code: ITP2bspline()
GPF test	<ul style="list-style-type: none"> • Much less computation • Suitable for multi-class case 	<ul style="list-style-type: none"> • Less power for highly correlated functional data 	Package: 'fdANOVA' Code: fanova.tests()

Table 2.1 Summary of three functional feature selection methods used

2.2.3 Model Construction

Suppose there are n trials of EEG signal collected from L features with time length N for each subject. We conduct functional data analysis to estimate the function $x_{il}(t)$ based on observed value x_{ijl} , which denotes the observed EEG signal value collected from the l^{th} feature of the i^{th} trial at the j^{th} time point. In this event, we assume that the functional covariates are observed with error, then the x_{ijl} can be expressed as $x_{ijl} = x_{il}(t_j) + \epsilon_i$, $i = 1, \dots, n$; $j = 1, \dots, N$; $l = 1, \dots, L$, where ϵ_i is random error that contributes to the roughness to the raw data.

2.2.3.1 Multiple Functional Logistic Regression

Logistic regression is a commonly used statistical model to perform binary classification. We can simply generate a logistic function to estimate the success probability by inputting predictors. Suppose we have a functional covariate as $x_i(t)$, where $i = 1, \dots, n$; $t \in \Gamma$ and Γ is the support of the covariate; and a random sample of a binary response variable Y to be $Y_i \in \{0,1\}$, $i = 1, \dots, n$. Then the random variable Y is such that $Y_i \sim \text{Bernoulli}(\pi_i)$, where $\pi_i = P(Y = 1 | x_i(t), t \in \Gamma)$.

In this paper, data is recorded over a period at multiple electrodes, a more complex model is considered, which named Multiple Functional Logistic Regression (MFLR) model.

Specifically, the conditional success probabilities can be expressed in terms of functional predictors $x_{il}(t)$ for $t \in \Gamma$ as

$$\eta_i = \text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \sum_{l=1}^L \int_{\Gamma} x_{il}(t) \beta_l(t) dt \quad (2.2)$$

where β_0 is an intercept parameter and $\beta_l(t)$ is the regression slope function for signal function $x_{il}(t)$. Both unknown functional covariates $x_{il}(t)$ and functional parameters $\beta_l(t)$ are assumed to be smooth functions, which can be estimated by taking linear combination of basis functions $\phi_k(t)$, $k = 1, \dots, K_x$. We consider $x_{il}(t) \in L^2(\Gamma)$ of squared integrable function with the inner product, such that

$$x_{il}(t) = \sum_{k=1}^{K_x} a_{ilk} \phi_k(t) \quad (2.3)$$

where the number of basis function K_x is selected based on the features and characteristics of the data. In this case, where the functional covariate is observed with some noise, estimates of coefficients $\{a_{il1}, \dots, a_{ilL_x}\}$ are obtained from the discrete observations through the least-square approximation approach to get the functional form of the covariates.

Further, we also define the regression slope functions, $\beta_l(t)$, in (2.4), as

$$\beta_l(t) = \sum_{k=1}^{K_b} b_{lk} \phi_k(t) \quad (2.4)$$

where $\phi_k(t)$, $k = 1, \dots, K_b$ are basis functions; and $K_x \geq K_b$. For simplicity, we assume that $K_x = K_b = K$.

By inserting (2.3) and (2.4), we further reduce the model defined in (2.2) to a standard multiple one

$$\eta_i = \beta_0 + \sum_{l=1}^L \int_{\Gamma} x_{il}(t) \beta_l(t) dt = \beta_0 + \sum_{l=1}^L [\sum_{k=1}^K \sum_{j=1}^K a_{ilk} (\int_{\Gamma} \phi_k(t) \phi_j(t) dt) b_{jl}] \quad (2.5)$$

where a_{ilk} are the basis coefficients to be estimated for each functional predictor $x_{il}(t)$ and b_{jl} are the basis coefficients to be estimated associated with each $\beta_l(t)$.

Generally, B-spline basis and Fourier basis are used for non-periodic and periodic data, respectively. Besides, there are some other basis, like wavelet, polynomial functions, also suitable for functional data with different characteristics [88].

Besides, the number of basis functions K determines the appropriate estimation. The larger K is, the better fit but results in more complex computations as well as over-fitting. Thus, the trade-off between fit and computation is necessary to approach an ideal approximation, which can be realized by cross-validation.

2.2.3.2 Principal Component Analysis on Multiple Functional Predictors

Since correlation-ship between observations at adjacent time points in each predictor will lead to the potential problem of severe multicollinearity, fPCA has been proposed to alleviate this issue [89]. Besides, the existence of multicollinearity will affect the participation of penalty term in the next step, which is found using a block co-ordinate gradient descent minimization algorithm. Therefore, it is necessary to remove the redundant information to obtain a more reliable estimation of regression slope function.

Model (2.5) can be expressed in the form of vector and matrix multiplication

$$\mathbf{E} = \beta_0 \mathbf{1}_{n \times 1} + \sum_{l=1}^L \mathbf{A}_{n \times K}^l \boldsymbol{\Psi}_{K \times K}^l \mathbf{b}_{K \times 1}^l \quad (2.6)$$

where $\mathbf{E} = (\eta_1, \dots, \eta_n)^T$, $\mathbf{A}_{n \times K}^l$ denotes the matrix of basis coefficients for each l^{th} predictor, $\boldsymbol{\Psi}_{K \times K}^l$ denotes the inner product of basis functions $\phi_k(t)$, $\mathbf{b}_{K \times 1}^l = (b_{1l}, \dots, b_{Kl})^T$ is the vector of basis coefficients for each regression slope function. Next, principal component analysis is performed on each $\mathbf{A}_{n \times K}^l \boldsymbol{\Psi}_{K \times K}^l$ to reduce the dimension of each predictor.

Let $\mathbf{G}^l = \{\xi_{ik}^l\}_{n \times K} = \mathbf{A}^l \boldsymbol{\Psi}^l \mathbf{V}^l$ be the matrix of principal components of the design matrix, where \mathbf{V}^l is a $K \times K$ matrix whose columns are eigenvectors of correlation matrix of $\mathbf{A}^l \boldsymbol{\Psi}^l$. The model (2.6) becomes

$$\mathbf{E} = \beta_0 \mathbf{1}_{n \times 1} + \sum_{l=1}^L \mathbf{G}_{n \times K}^l \mathbf{B}_{K \times 1}^l \quad (2.7)$$

where $\mathbf{B}_{K \times 1}^l = (\mathbf{V}_{K \times K}^l)^T \mathbf{b}_{K \times 1}^l$.

The number of PCs is determined by cumulative variance of each predictor. For simplicity, we select the same number of PCs, s , for each predictor. Then, model (2.7) can be expressed in terms of principal components

$$\mathbf{E} = \beta_0 \mathbf{1}_{n \times 1} + \sum_{l=1}^L \mathbf{G}_{(s)}^l \mathbf{B}_{(s)}^l \quad (2.8)$$

where s denotes the number of selected PCs retained in the model.

Therefore, model is transformed to a linear combination of FPC scores and the dimension is reduced in the within-variable level.

2.2.3.3 Penalized Multiple Functional Logistic Regression

Even though the initial feature selection has been conducted through functional testing to reduce the original large number of predictors by selecting the features which are significantly different between the two groups, there still exists some predictors providing similar information. Especially, in EEG data, signals that are time series of adjacent electrodes are expected to be similar. Therefore, in order to obtain correct interpretability and improve the prediction accuracy of the models, it is common to employ penalized regression techniques. In this study, we use three most commonly used penalized regression techniques with grouped predictors: Group Least Absolute Shrinkage and Selection Operator (LASSO), Group Least Angle Regression (LARS) and Group Smoothly Clipped Absolute Deviation (SCAD).

LASSO shrinks some of the coefficients all the way to zero, thereby delivering a sparse solution with just a few non-zero coefficients. LARS can be viewed as a version of stage-wise procedure that uses mathematical formulas to accelerate the computations. SCAD aims at solving the problem with LASSO that the penalty term is linear in the size of the regression coefficient, which tends to give substantially biased estimates for large regression coefficients.

Firstly, we consider a linear regression model with grouped predictors

$$\mathbf{Y} = \sum_{l=1}^L \mathbf{X}^l \boldsymbol{\beta}^l + \boldsymbol{\epsilon} \quad (2.9)$$

where \mathbf{Y} is an $n \times 1$ vector of response, $\boldsymbol{\epsilon} \sim N_n(0, \sigma^2 \mathbf{I})$, \mathbf{X}^l is the $n \times p_l$ matrix of the l^{th} grouped predictors, $\boldsymbol{\beta}^l$ is the regression coefficient vector associated with the l^{th} grouped

predictors. In order to eliminate the intercept from equation (2.9), response variable and each input variable can be centered, and the observed mean is 0. By Yuan and Lin [83], we conduct group LASSO which aims at selecting important grouped predictors which corresponds to decide which the $\boldsymbol{\beta}^l$ regression coefficient vector is set to zero vectors for each l . This aim can be achieved by optimizing the objective function

$$\arg \min_{\boldsymbol{\beta}^l} \left(\frac{1}{2} \|\mathbf{Y} - \sum_{l=1}^L \mathbf{X}^l \boldsymbol{\beta}^l\|_2^2 + \lambda \sum_{l=1}^L \|\boldsymbol{\beta}^l\|_2 \right) \quad (2.10)$$

where $\lambda \geq 0$ is a tuning parameter. The group LASSO penalty is a mixture of L_1 and L_2 regularization methods, that is, the LASSO and the Ridge regression penalties.

Group LASSO can be applied to multiple functional logistic regression model [90], which performs single variable selection in the functional logistic regression model with multiple functional predictors. The estimation of coefficients can be obtained by minimizing

$$S_\lambda(\boldsymbol{\beta}) = -l(\boldsymbol{\beta}) + \lambda \sum_{l=1}^L df_l^{1/2} \|\boldsymbol{\beta}^l\|_2 \quad (2.11)$$

where the tuning parameter $\lambda \geq 0$ controls the amount of penalization, df_l is the degrees of freedom of the l^{th} group of predictors, $l(\boldsymbol{\beta})$ is the log-likelihood of the multiple functional logistic regression model and we denote by $\boldsymbol{\beta}$ the whole parameter vector, i.e. $\boldsymbol{\beta} = (\boldsymbol{\beta}^1, \dots, \boldsymbol{\beta}^L)'$. Similarly, group LARS and group SCAD can be defined for multiple functional logistic regression model.

2.3 Simulation

To validate the proposed algorithm, we firstly simulate two groups of data for 64 channels with 48 trials and the same time length as the collected data (around 5.656 seconds). Compared with the control group, an extra oscillation is added into channel 3 between 500 ~ 1500ms in the experiment group. Based on settings, we generate data from a series of sine and cosine waves as below:

$$\text{Group1: } y = 4 \sin(4\pi t) + 5 \sin(8\pi t) + \epsilon \text{ for all 64 channels} \quad (2.12)$$

$$\text{Group2: } y = 4 \sin(4\pi t) + 5 \sin(8\pi t) + 3 \cos(16\pi t) + \mathbf{1}_{1 \leq t \leq 2, \text{channel } 3,5,12} + \epsilon \quad (2.13)$$

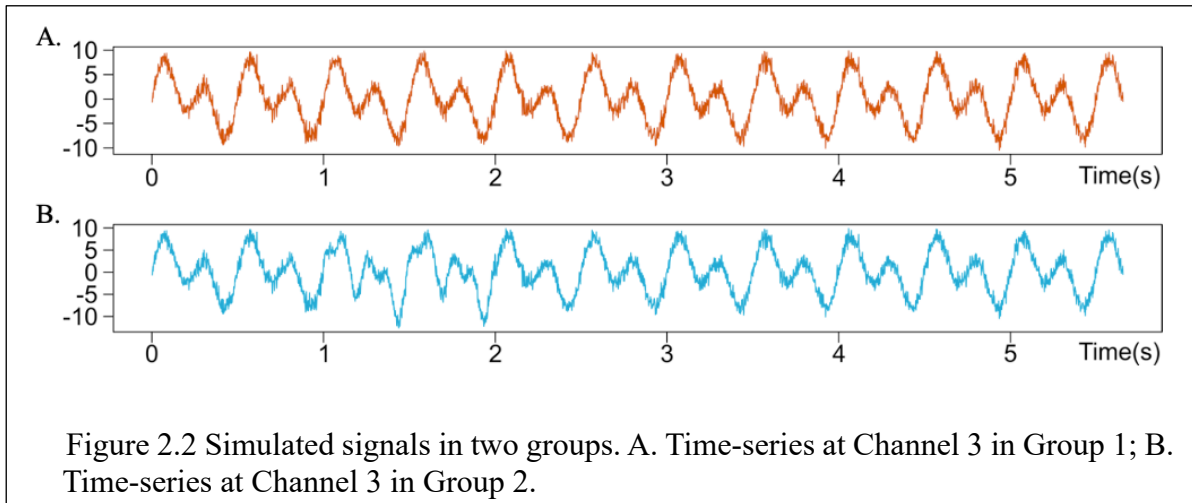
From Figure 2.2, there is no obviously difference between simulated signals from two groups.

In the first step, each time series is transformed into time-frequency domain. The value in new data describes the power at specific frequency and time. From Figure 2.3, we observe that both two plots have peaks at 2 Hz and 4 Hz, the bottom one has one flatter peak at around 8 Hz. From Figure 2.4, we find that in the plot for data in group1, higher power comes out at 2 Hz and 4 Hz during the whole-time interval. Whereas in the plot for data in group2, higher power comes out at the same frequency as in group1 and it is higher at around 8 Hz from 500~1000 time points. Moreover, from Figure 2.5, we notice that there is significant difference between 1~2s at 8 Hz among two groups, whereas no significant difference is observed at other frequency.

Next, functional testing method is applied to conduct the feature selection and features with extra 8Hz oscillations are successfully identified and selected. For instance, there exists significant different during time 1~2s among two groups in Figure 2.6.

Finally, we perform penalized Multiple Functional Logistic Regression on the data.

With consideration of both time and frequency information, the channel and frequency with significant differences between the two groups during a specific time interval are successfully identified. Therefore, the result of simulation validates the interpretability of the proposed three-stage algorithm.



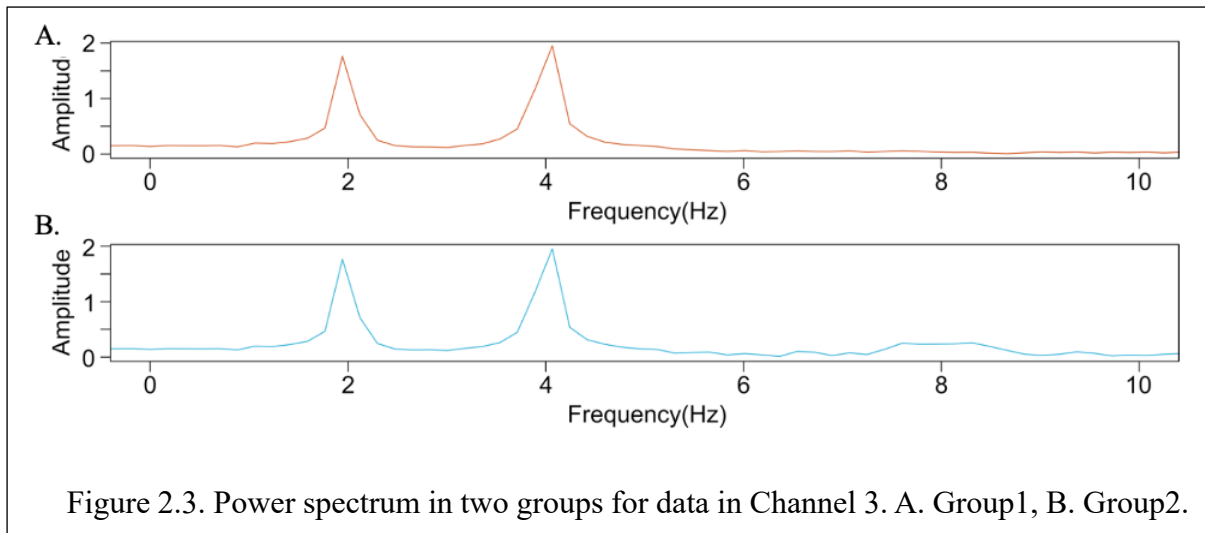


Figure 2.3. Power spectrum in two groups for data in Channel 3. A. Group1, B. Group2.

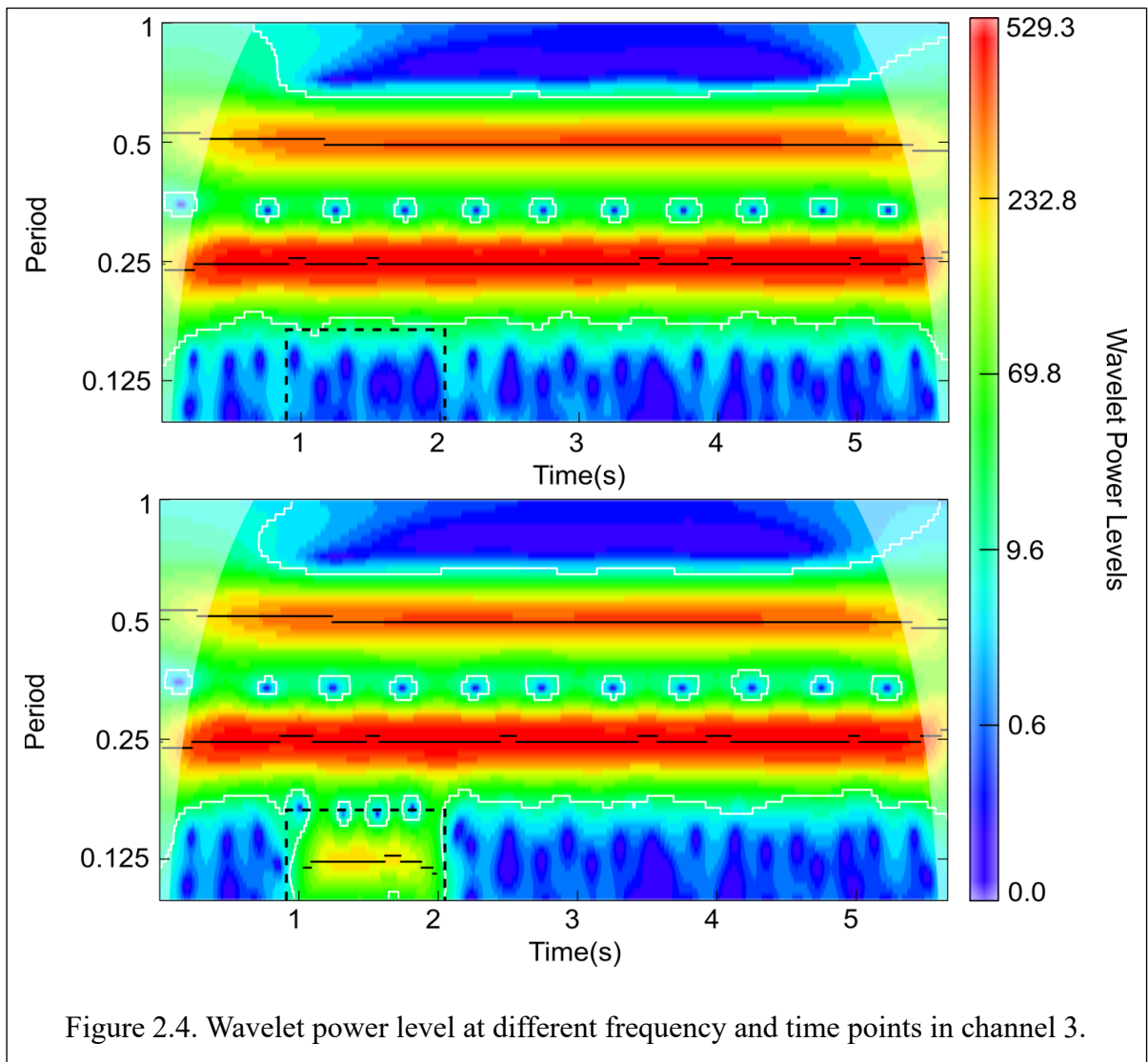


Figure 2.4. Wavelet power level at different frequency and time points in channel 3.

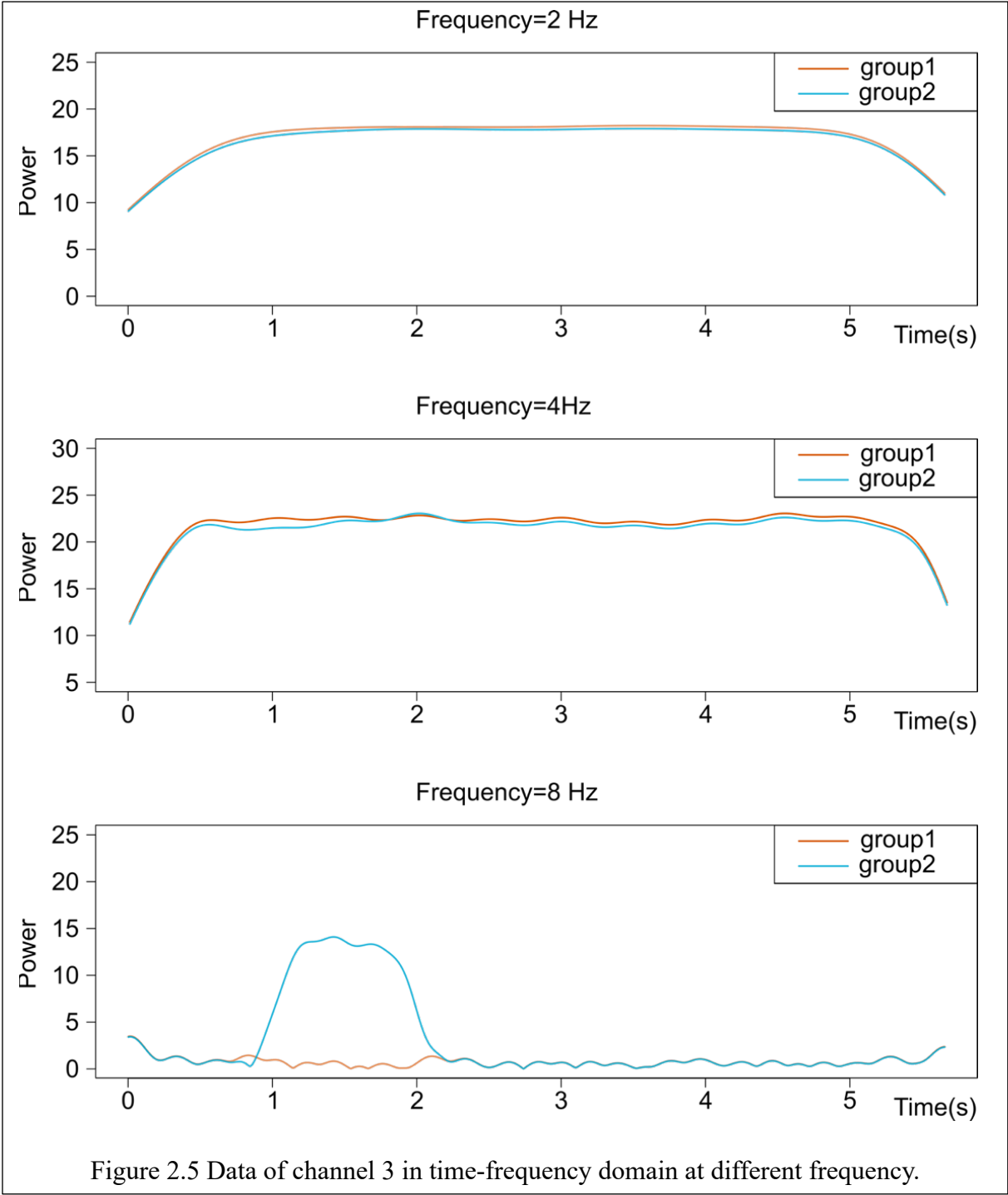
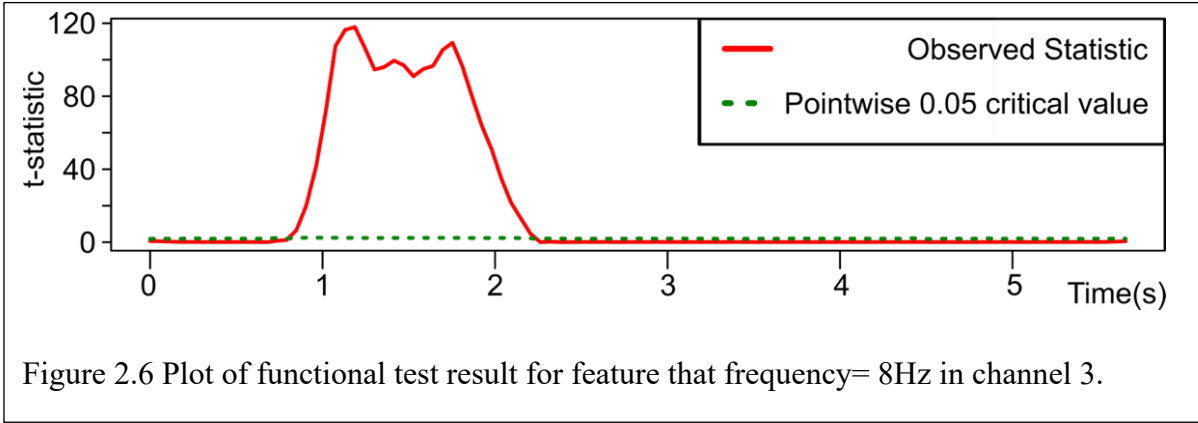


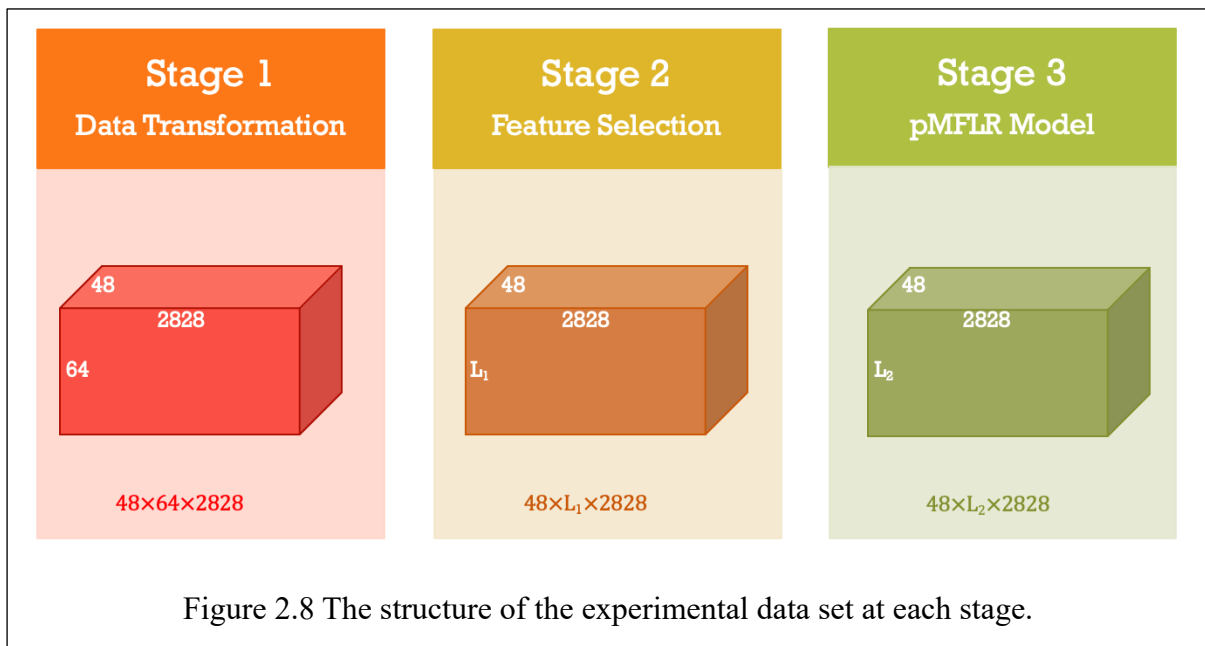
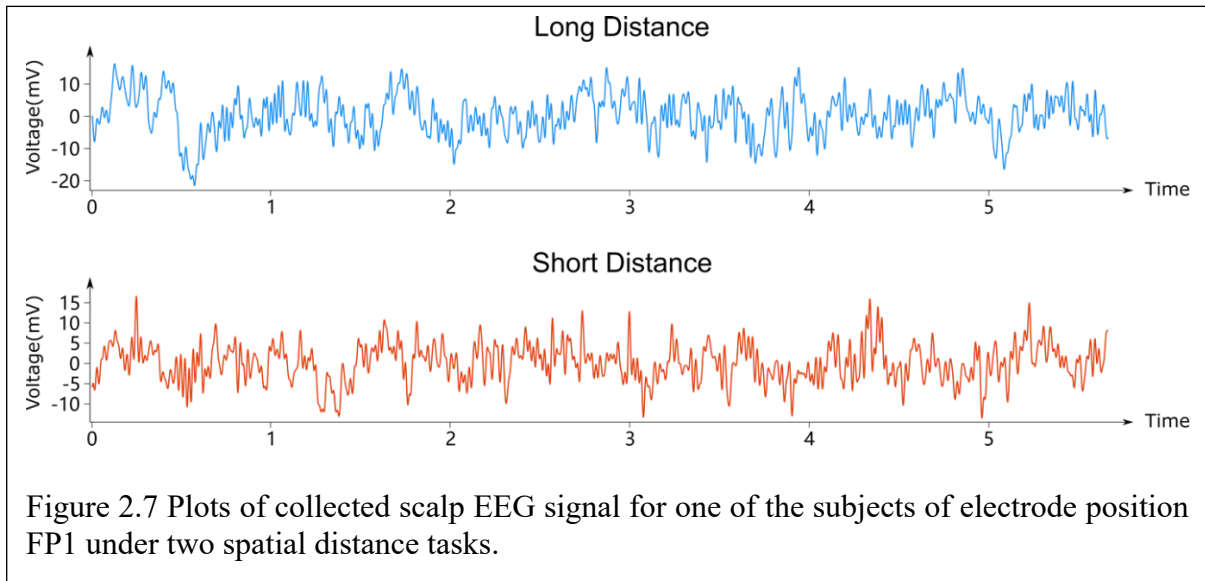
Figure 2.5 Data of channel 3 in time-frequency domain at different frequency.



2.4 Implementation of the Three-Stage Algorithm to Scalp EEG Data

In the proposed three-stage algorithm, various methods are considered to assess the best and interpretable model based on some performance measures. In this study, we apply the proposed algorithm to the scalp EEG dataset, which has been preprocessed by neuroscientists. Specifically, 1-50Hz bandpass filter was applied to the continuous data. The 1Hz high pass edge was used to remove slow drift artifacts. The 50Hz low pass edge was selected to isolate oscillatory activities including delta, theta, alpha and beta bands, in accordance with the researchers' hypothesis. Artifact subspace reconstruction (ASR) was then applied for artifact amelioration. Independent component analysis and an automatic component selection procedure, ICLabel [91] were used to correct eye/muscle artifacts through EEGLAB. The data is available at <https://osf.io/3vxkn/>. For more details in experimental design and data preprocessing, please see [92]. In short, the structure of EEG data set for each person can be expressed as $48 \times 64 \times 2828$, which means that EEG signals for each person are collected from $n = 48$ trials and $L_0 = 64$ electrodes with $T = 2828$ time points.

In the first stage, data is transformed through Morlet wavelet transformation from time domain to time-frequency domain. Next, each feature is split into four according to different frequency bands [93], and the new number of feature increases to be L_1 . In the second stage, features that are significantly different among two levels of response are extracted by functional testing. Therefore, we obtain a new dataset containing 48 observations, each of observation contains L_2 time-series with 2828 time points, which denotes the power of different frequency. In the last stage, the penalized Multiple Functional Logistic Regression (pmFLR) model is fitted to this new dataset for classification.



To prevent the model from overfitting and evaluate the model accurately, cross-validation method (CV) is required, which randomly splits data into two parts: training set and testing set. Training set is used to generate the model through the estimation of coefficients based on pre-set conditions and testing set is used to validate the model and compare the output value with true value from the set. In addition, due to the insufficient sample size of the original data, we consider bootstrap method that uses repeated samples in the original data sample to increase the sample size of the training set and the testing set separately. Specifically, we randomly split

data into training and testing set firstly, then conduct bootstrap on each set, respectively, to increase the sample size. Furthermore, we repeat cross-validation with bootstrapping 300 times to avoid the test results contingency. The whole process of implementation of three-stage algorithm is summarized in algorithm below.

Algorithm Implementation of Three-Stage Algorithm

Preprocess the raw data

- a) Apply artifact subspace reconstruction (ASR) for artifact amelioration.
- b) Apply Independent component analysis (ICA) and an automatic component selection procedure to correct eye/muscle artifacts.
- c) Obtain the new dataset with structure: $\mathbf{n} \times \mathbf{L}_0 \times \mathbf{T}$.

First Stage: Data Transformation

- a) Extract both time and frequency related information by Morlet wavelet transformation.
- b) Split features based on four brainwaves of EEG signals.
- c) the new dataset with structure: $\mathbf{n} \times \mathbf{L}_1 \times \mathbf{T}$.

Second Stage: Initial Feature Selection via Functional Testing

- a) Set up hypothesis test to check the significance of difference between different class for each feature.
- b) Obtain the new dataset with structure: $\mathbf{n} \times \mathbf{L}_2 \times \mathbf{T}$.

Third Stage: Penalized Multiple Functional Logistic Regression Model Construction

- a) Generate MFLR model.
- b) Perform functional PCA on each predictor to alleviate the problem of multicollinearity.
- c) Employ penalized regression techniques with grouped predictors.
- d) Obtain the new dataset with structure: $\mathbf{n} \times \mathbf{L} \times \mathbf{T}$.

Evaluate the model with cross-validation

- a) Randomly split dataset into training and testing set.
 - b) Conduct bootstrapping onto two sets, respectively, to increase the sample size
 - c) Repeat step a) ~ b) 300 times.
-

2.4.1 Various Model Settings

Since EEG data is non-periodic, we conduct the functional data analysis with B-spline basis function. To select the most suitable model for the given EEG signal data, various models are generated based on the combinations of different methods for each stage.

In this experiment, we are interested in the time-related information, thus, we use data either in the time-domain or transformed to a time-frequency domain at the first stage. There are three functional testing methods to be considered to perform initial feature selection at the second stage. Finally, at the last stage, we present three options for penalty terms in penalized MFLR model. Consequently, there are total $2 \times 3 \times 3 = 18$ combinations considered in this setting.

For the sake of simplicity, we only list six representative models for comparison, since the performance of remaining models are not good enough to report. More details about six models are listed in Table 2.2. Models 1 and 2 are set up to analyze data transformation. Models 2, 3 and 4 are set up to compare different functional testing methods. Models 3, 5 and 6 are set up to compare three penalty terms.

Model	Domain	Initial Feature Selection	Penalties
Model1	Time	GPF	grpLASSO
Model2	Time-Frequency	GPF	grpLASSO
Model3	Time-Frequency	ITP	grpLASSO
Model4	Time-Frequency	FT	grpLASSO
Model5	Time-Frequency	GPF	grpLARS
Model6	Time-Frequency	GPF	grpSCAD

Table 2.2 Selected model construction.

2.4.2 Results and Interpretations

To assess the performance of the six models, we use three numerical measures: sensitivity, squared error (SE), and area under the ROC curve (AUC). Results for overall 19 subjects are summarized in the Table 2.3. Predictors in the final model provide the information about task related brain region. Sensitivity, squared error and AUC is helpful to compare models.

In order to determine whether the information from the time-frequency domain further contribute to the analysis, we compare Models 1 and 2. As shown in the Table 2.3, two models have similar sensitivity and AUC values while Model 2 in time-frequency domain results in smaller squared error than Model 1. Therefore, we select Model 2, which is discussed in time-frequency domain.

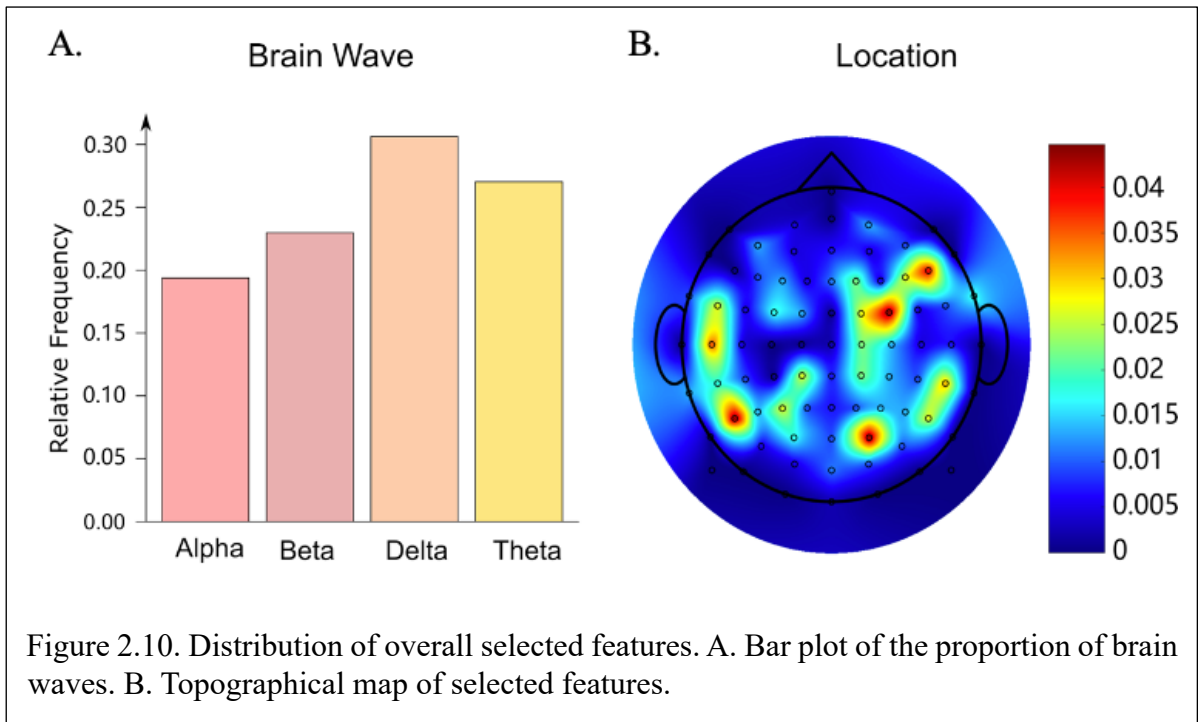
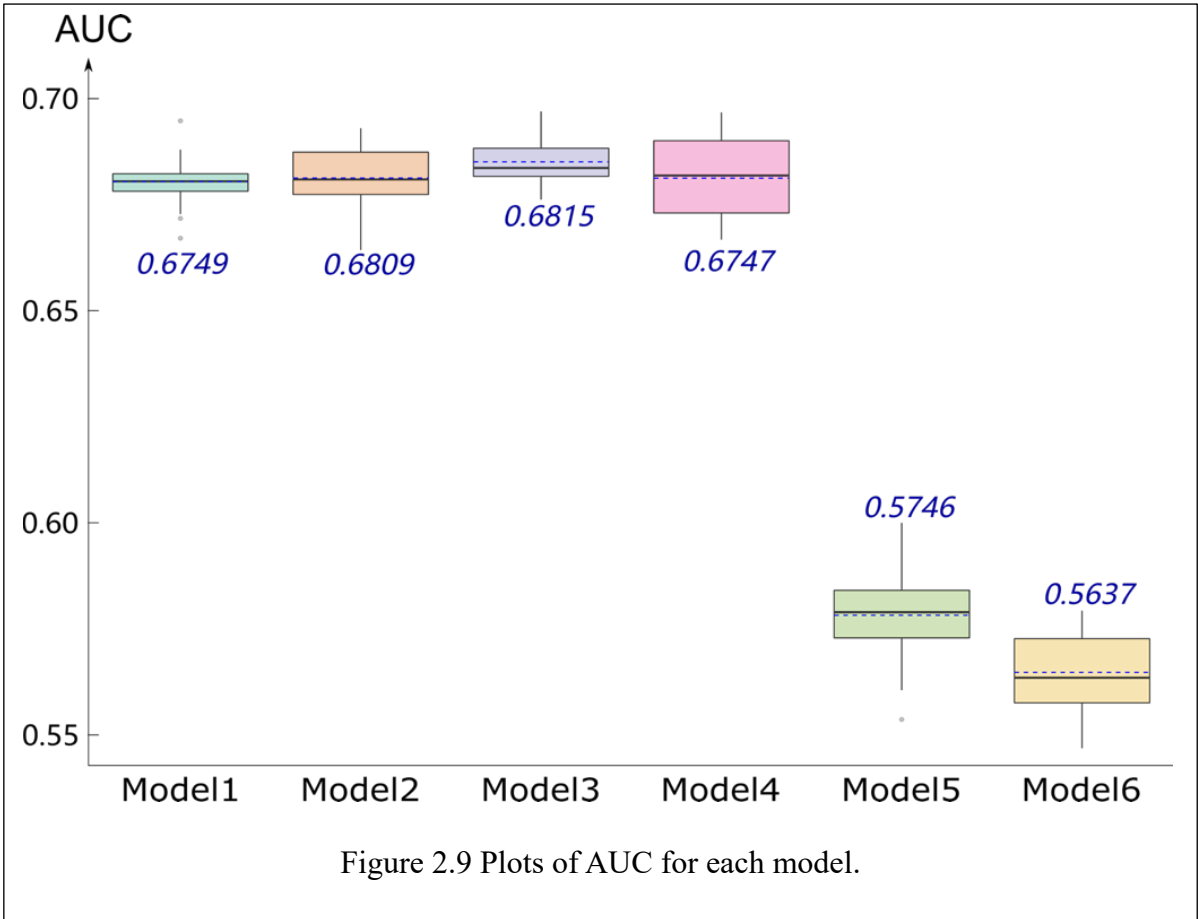
Model	# of Predictors	Sensitivity	Squared Error	AUC
Model1	9	0.63	0.41	0.67
Model2	12	0.65	0.28	0.68
Model3	10	0.62	0.32	0.68
Model4	12	0.63	0.41	0.67
Model5	6	0.50	0.49	0.57
Model6	10	0.51	0.50	0.56

Table 2.3 Performance of model is measured from several aspects.

From the results of models 2, 3, and 4, we also conclude about the comparison of three functional testing methods. As we can see, model 2 with the GPF test provides better performance in model 2 among three models, which indicates that GPF test performs best among three methods. Besides, the functional t and ITP tests are much more computationally expensive than the GPF test because the ITP test creates a p-value matrix for each basis function and the functional t-test generates test statistics for each feature. Hence, these two methods are particularly time-consuming when analyzing the data with large amounts of features. As a result, the GPF test is selected to conduct feature selection at this stage.

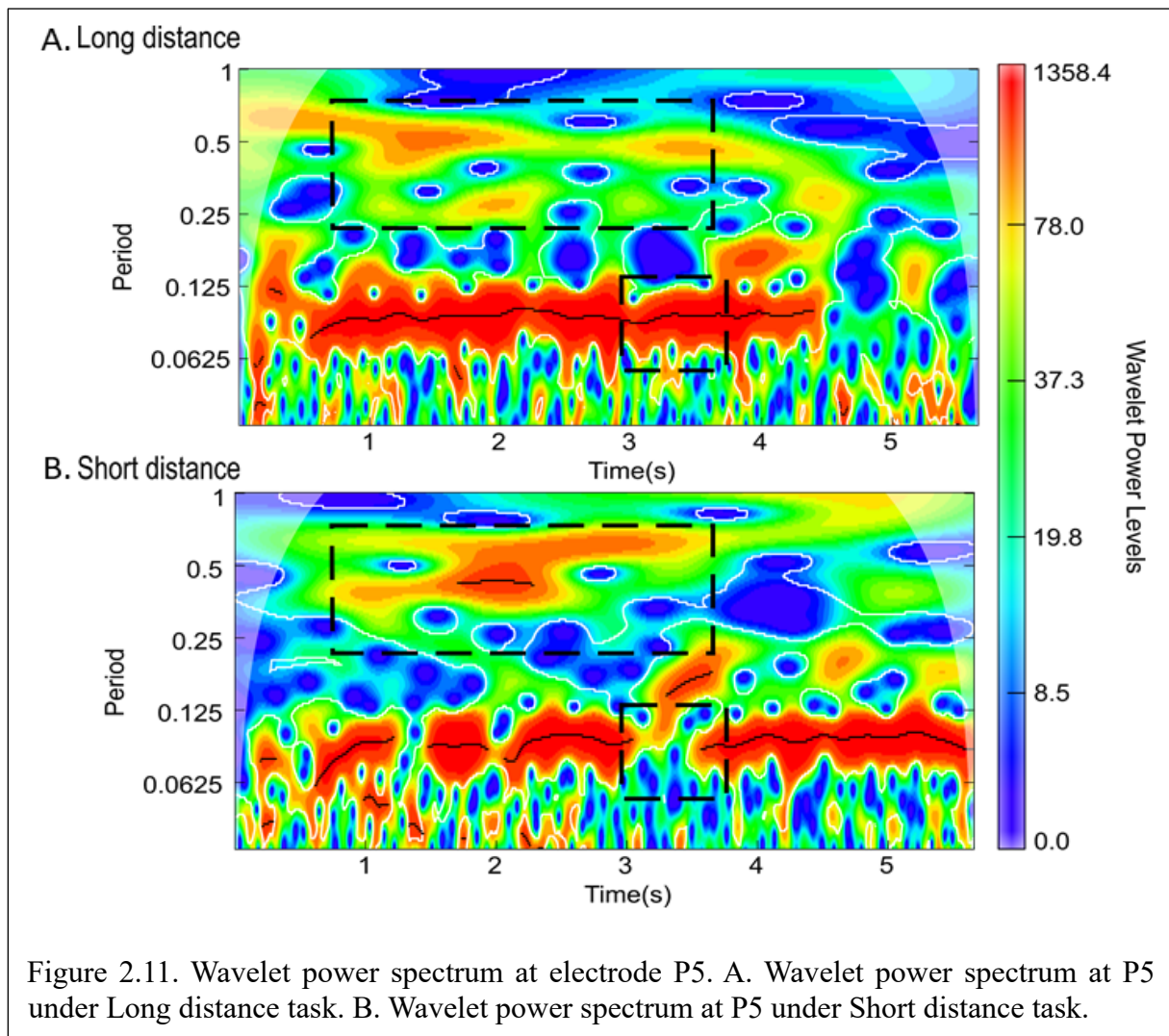
Some claims about penalty terms are made by comparing the results of models 2, 5 and 6. The non-orthogonality of the selected features and the efficiency of the LARS algorithm make it an attractive method for group variable selection. In addition, considering the three ideal attributes of penalty estimation: sparsity, unbiasedness, and continuity, the SCAD is also within the scope of our discussion. However, as we observe in Table 2.3, model 2 has the highest sensitivity and AUC values compared to the rest of models and further, the squared error of model 2 is almost 40% lower than the other two. Therefore, it is evident that model 2 which is obtained from the group LASSO penalization method performs better than the rest based on the numerical measures of sensitivity, squared error and AUC values.

We further investigate the AUC of each model based on the overall 19 individuals. As we can see (Figure 2.9), AUCs of models 5 and 6 are clearly lower than the rest, which might be a result of the use of different penalties. First four models are all obtained based on group LASSO used at the third stage of the algorithm, whereas the other two obtained from group LARS and group SCAD, respectively.



Based on the application of proposed three-stage algorithm, we find that the model, which transforms the original data to the time-frequency domain and selected features by GPF method and applied pMFLR model with group LASSO, is a suitable approach for collected EEG signals to conduct classification. From Figure 2.10 A, the proportion of Delta-Theta oscillation are slightly higher than the other two when we discuss data in time-frequency domain. The proportion of selected electrodes belonging Parietal and Frontal is obviously higher than the other regions from Figure 2.10 B. Thus, we conclude that results from the selected model showed that Frontal delta-theta and Parietal delta-theta oscillations are more related to distance judgment. If we use PSD technique, which is a typical method in EEG signal analysis, to conduct the feature selection, the results show that occipital alpha oscillation is more related instead of the Parietal delta-theta oscillation. This is because that the PSD technique focuses on the power and ignore the change of spectrum of oscillation over time, whereas the functional testing considers the time-related information in the feature extraction process.

Besides, Figure 2.11 presents wavelet power spectrum at electrode P5, which provides both temporal and frequency information. Comparing these two plots, the power of low frequency is higher in Short distance task during the time 1~2s, the power of high frequency is higher in Long distance task during the time 3~4s. Thus, the data after time-frequency transformation can identify the temporal characteristics of cognitive components. Moreover, selected brain waves based on data after wavelet transformation can be found from Figure 2.12, some waves are obviously different under two tasks.



2.5 Discussion

In this chapter, we propose to use FDA method with consideration of both time and frequency information to interpretably classify scalp EEG signal. There is no systematic method for analyzing EEG data via FDA, although there have been many studies on the classification of EEG signals. Currently, existing FDA methods for EEG analysis [74],[82] consider only either time or frequency related information. Even though some research [78],

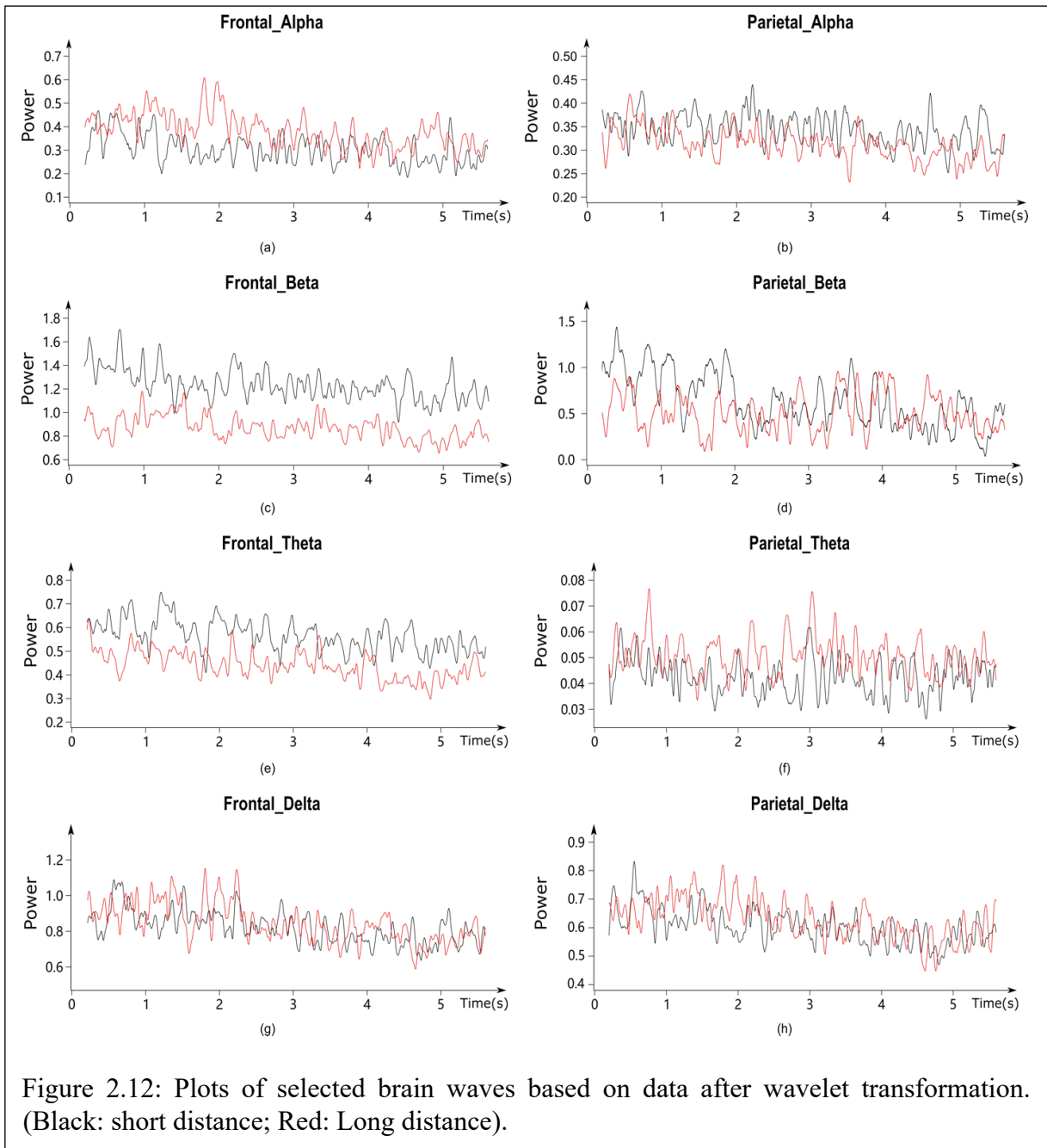


Figure 2.12: Plots of selected brain waves based on data after wavelet transformation. (Black: short distance; Red: Long distance).

[79] have discussed FDA with frequency-related information, they automatically take into account the spatial information among neighboring voxels due to the property of collected three-dimensional brain image data. As mentioned before, high classification rate of the EEG signal can be achieved by some machine learning methods [80]. However, such high-quality classification performance comes with the sacrifice of smoothness and interpretability. Therefore, we propose this novel three-stage algorithm to conduct interpretable classification on EEG signals via FDA considering both time and frequency.

We apply the proposed algorithm on collected EEG signals for distance judgement. Throughout the past decades, research has supported a role for low frequency oscillations in

coding movement speed during navigation [94],[95],[96]. In order to extract both time and frequency-related information from EEG signals, data is transformed into time-frequency domain and expressed as functions of time in the first step. Time series from different frequencies for each channel are treated as different functions. Next, features which are significantly different among two distance tasks are selected by functional testing method. The testing results also point out the specific time interval when such differences happen. This temporal information can only be obtained by interpretable FDA method. Finally, functional logistic regression model is constructed with group counterparts of penalty terms to conduct the classification. It is more suitable for factor selection with multiple predictors (e.g., channels), because group penalty terms allow for the joint selection of groups of covariates from the model. Results indicate that frontal delta-theta and parietal delta-theta have important implications in spatial coding (Figure 2.10, 2.11, 2.12).

To better demonstrate the advantages of our proposed method, we also simulate two groups of data (Figure 2.2). Results show that the proposed three-stage algorithm successfully identifies the channels and the frequency which are significantly different in two groups. It also points out the specific time interval which are significantly different in two groups. Therefore, FDA techniques make the classification become more interpretable since we can directly see what the selected features look like by the estimations of smooth functions.

Although functional testing provides good temporal information of EEG signals, the computing complexity increases as the number of features and recording time length increase, because it is performed on each feature. Besides, EEG signal is easily be affected by outlier factors, it is necessary to denoise the original signal to improve the signal-to-noise ratio as much as possible before formally analyzing the data. Even though the EEG data in this chapter has been pre-processed, there still exists measurement errors in discrete data [97]. In next two chapters, we will propose new methodology for signal preprocessing, to improve the classification performance. Further, three-dimensional functional data analysis and modeling can be carried out from the perspectives of time, space, and frequency, which considers the spatial correlations of electrodes.

2.6 Conclusion

In order to reflect the continuity and internal dynamic changes of the EEG signal, we take advantage of the functional nature of it. Therefore, the development of an interpretable EEG classification method based on functional data analysis has significance and application value.

In this Chapter, we propose a novel three-stage algorithm which provides a systematic way of analyzing EEG data via FDA techniques. To capture desired information from the EEG signals, we first propose using the FDA methods on EEG signals transformed in time-frequency domain. While using data transformed in time-frequency domain helps to extract more information from the EEG signals, it increases the dimensionality of the data. Therefore, applying FDA techniques directly to this transformed data is not feasible. To overcome this challenge, we propose the second stage in this algorithm, which employs the functional testing methods for the initial dimension reduction (i.e., feature selection) so that the number of features that will be used for modeling would be manageable for the functional modelling techniques. Further, we show that the proposed algorithm applied to the Scalp EEG data leads us to an interpretable classification and we conclude that frontal delta-theta and parietal delta-theta oscillations are more related to distance judgment than other oscillations.

Chapter 3

CW_ICA: An Efficient Dimensionality Selection Method for Independent Component Analysis

3.1 Introduction

Independent component analysis is a statistical tool to extract hidden information from the observed (mixed) signals. Assuming that these observed signals are linear combinations of mutually independent and non-Gaussian source signals, ICA seeks to discover the linear combination of these mixed signals to recover the original source signals. The performance of ICA is measured by the independence or non-Gaussianity of the estimated ICs.

According to the number of observed signals (p) and source signals (q), ICA methods can be divided into two cases: (over)determined ICA (i.e., $p \geq q$) (e.g., FastICA[98], JADE[59], Infomax[99], etc.), and underdetermined ICA ($p < q$) (e.g., FastFCA[100], MAICA[62], OICD[101], etc.). In this paper, we focus on (over)determined ICA ($p \geq q$), where the mixing matrix \mathbf{A} is invertible. The object of ICA can be achieved by estimating the de-mixing matrix $\mathbf{W} = \mathbf{A}^{-1}$, and the estimated source signals (ICs) can be further obtained by projecting the whitened data onto the matrix \mathbf{W} .

For comparison purposes, we employed the three most commonly used ICA methods - FastICA, Infomax and JADE when combined with the determination methods. A brief description for each of these methods is given in the following.

FastICA: Hyvärinen and Oja [102] proposed the FastICA algorithm, the key point of which is to find the vector in an unmixing matrix by a fixed-point iteration that maximizes non-Gaussianity measured by either kurtosis or negentropy. FastICA converges quickly as it seeks out each component one by one. However, it requires the selection of an appropriate nonpolynomial function and the random nature of initial values might cause the instability of FastICA result [103].

JADE: This is another representative ICA method for extracting IC by maximizing non-Gaussianity, which is based on the diagonalization of fourth-order cumulative tensors that contain all the data's fourth-order information. The key to JADE is to perform a joint diagonalization of the eigenmatrices computed from the fourth-order cumulants of the whitened signal and optimize them via plane rotation. JADE was first proposed by Cardoso and Souloumiac . This method usually performs well in low dimensional space. However, the computational load of JADE grows rapidly with the number of components [103] and the computation of eigenmatrices limits the setting of the maximum number of ICs when one attempts to determine the optimal number of ICs.

Infomax: Another approach to estimating the independent components is based on mutual information, which is given by the Kullback-Leibler (KL) divergence of the multivariate density from the product of the marginal (univariate) densities [105]. The mutual information is always positive and the value of zero indicates the independence of components. Amari et al. proposed the Infomax to compute the unmixing matrix, which presents a natural gradient form for the IC computation and can separate high-dimensional signals. However, Infomax uses a stochastic gradient algorithm [107], which may require some manual tuning and often fails to converge [108], or only converges slowly.

Both under-decomposition (too few ICs) and over-decomposition (too many ICs) can hamper effective source separation. Commonly used determination techniques include information criteria, eigenvalue spectrum (ES), bootstrap resampling (BS), and cross-validation (CV), among others. Nevertheless, these methods have their drawbacks. For instance, information criteria may suffer from overfitting when the sample size is small or strict model assumptions are made. Eigenvalue spectrum methods can be subjective in the choice of threshold and may be affected by noisy signals. Bootstrap resampling techniques, although comprehensive, can be computationally expensive. Cross-validation, while generally reliable, may introduce data partition bias and incur computational costs.

To address the issues with these determination methods, researchers have proposed several alternatives, that can be categorized into three classes. Firstly, methods leveraging original signal information have been introduced. Wang et al. [109] introduced Mean-field ICA (MF-ICA), a Bayesian-based approach that determines the optimal number of ICs by evaluating the square-root sum of the residual between original and reconstructed data. This method excels in separating complex mixtures, such as those encountered in chemistry. Other approaches in this class, such as those proposed by Monakhova et al. [110] and Kassouf et al. [111] leverage

different metrics for the determination of the optimal number of ICs. Monakhova et al. introduced an index, known as the Amari Index, whereas Kassouf et al. employed a correlation method, referred to as ICA_corr_y. Both methods, however, necessitate prior knowledge of the ground truth or mixing matrix. Secondly, there are methods employing visual analysis. Bouveresse et al. [112] proposed two techniques, one of which is the ICA-by-Blocks method, employing a "signal-correlation" plot to determine the optimal number of ICs. The other method uses a heatmap generated by the Durbin-Watson criterion. Thirdly, there are methods that require specific data structures. Bach et al. [113] suggested a determination method based on a forest-structured graphical model, which is limited to dependencies among sources within a forest structure and may not apply to broader classes of dependencies. Kassouf et al. [111] presented a determination method using the Kaiser-Meyer-Olkin (KMO) index, a measure indicating the presence of a partial correlation among at least two residual signals. Nonetheless, if there is a small number of ICs, this method encounters a potential pitfall in cases of complete or high correlation among variables, which makes the correlation matrix non-invertible and further poses challenges to the analysis. While these methods offer computational efficiency and require fewer assumptions compared to other techniques, they do have certain drawbacks. For instance, they require enough mixed signals and structured signals (such as sparse, periodic, linear, etc.). Additionally, in some methods, the optimal number of ICs must be visually identified from a plot, which can be subjective. Furthermore, the existing methods for determining the number of ICs may not be universally applicable to all ICA methods, which introduce additional challenges, such as uncertainty and instability. It is also worth noting that the robustness of the determination method for signals with different characteristics plays an important role in determining the optimal number of ICs. However, this factor has not been extensively studied in the context of existing determination methods. (See Table 3.1 for a summary of advantages and disadvantages of these techniques categorized into three groups.)

Given the limitations of current determination methods, we propose a method called column-wise independent component analysis (CW_ICA) to automatically determine the optimal number of ICs. Inspired by the ICA-by-Blocks approach, the proposed method addresses challenges related to computational efficiency, consistency, and robustness. Instead of using Pearson correlations, CW_ICA employs Spearman correlations among the ICs obtained from the different blocks. This choice offers advantages in terms of capturing monotonic relationships between ICs, which can be valuable in various scenarios. Moreover, we introduce a novel metric based on the column-wise maximum rank-based correlations

between the extracted ICs in CW_ICA. This metric serves as a criterion for determining the optimal number of ICs. Therefore, compared to the existing determination methods, the major advantages of CW_ICA are:

1. **Efficiency:** the computational cost is significantly less than existing methods.
2. **Consistency:** the optimal numbers of ICs obtained by CW_ICA are consistent when it is coupled with different ICA methods.
3. **Robustness:** CW_ICA is robust for signals with different characteristics.

The rest of this Chapter is organized as follows. In Section 3.2, we provide a brief introduction to the ICA method and review the current determination methods. Section 3.3 presents the CW_ICA method, using a simple example for illustration. We also compare the CW_ICA method with existing determination methods using both simulation and real data in order to evaluate its performance in Section 3.4. A detailed discussion of our findings can be found in Section 3.5, and our conclusion is outlined in Section 3.6.

3.2 Existing methods

The selection of determination methods is contingent upon the specific structure and characteristics of the data under consideration. In this section, we provide a concise summary of the current state-of-the-art in this field, highlighting key differences among these methods. For a detailed comparison of the pros and cons of the existing determination methods, please refer to Table 4.1.

In each of the following algorithms, $\mathbf{X}_{p \times n}$ denotes p mixed signals and signal length n . The residual signal matrix is calculated by subtraction of mixed signals and estimated mixed signals, $\hat{\mathbf{X}}$, that is, $\mathbf{R} = \mathbf{X} - \hat{\mathbf{X}}$. A_{max} represents the maximum number of ICs.

3.2.1 Durbin-Watson (DW) criterion

The DW statistic is a well-known test statistic used for detecting the presence of autocorrelation in the residuals from a regression analysis [114]. It is also used as a measure of

Category	Selected Methods	Pros	Cons
Require original source signals information	MF-ICA	Effective for separating complex mixtures, particularly in Chemistry	Requires original source signals
	Amari index based	A quantitative measure with intuitive interpretation	Requires true mixing matrix
	ICA_corr_y	A data-driven approach with computational efficiency	Requires at least one original source signal
Visual determination	ICA-by-Blocks	Flexible in Block Size, easy interpretation via plot	High computational complexity
	DW criterion	Determine the number of IC based on signal/noise ratio	fails if variance of DW values among mixed signals is large
Require specific data structure	FCA	Model both inter-cluster independence and intra-cluster dependence	Requires forest structured signals
	KMO index based	A quantitative measure with intuitive interpretation	fails for cases where small number of ICs occurs.
	DW criterion	Determine the number of IC based on signal/noise ratio	Requires structured signals

Table 3.1. Summary of existing determination methods.

the signal/noise ratio in signals, which provides a method for determining the number of ICs [115]. The value of DW criterion for i^{th} mixed signal, X_i , is defined as:

$$DW_i = \frac{\sum_{t=2}^n (r_{i,t} - r_{i,t-1})^2}{\sum_{t=1}^n (r_{i,t})^2}, i = 1, \dots, p \quad (3.1)$$

where $r_{i,t}$ denotes the value of i^{th} residual signal at time point t . If the DW_i is close to 0, the signal is noise-free, implying that the extracted IC necessitates further decomposition, However, if it is near 2, the signal is inundated with noise, which indicates that the signal is over-decomposed. The average of DW_i values over all p signals is employed as a measurement for determining the number of ICs. Nevertheless, the variance of DW_i 's tends to be large in real datasets due to the non-linear behavior exhibited in real-world signals, which contradicts the linearity assumption inherent to ICA methods. In practice, heatmaps are used to depict DW values for each mixed signal produced by models with varying numbers of ICs. From the heatmap plot, q_{opt} is determined to be the optimal number of ICs whenever a sudden increase occurs in the DW values of all mixed signals. The procedure that determines the optimal number of ICs based on the heatmap is summarized as below:

Algorithm Method based on the Durbin-Watson criterion

Input: Observed signals X , maximum number of ICs A_{max}

Output: Optimal number of ICs q_{opt}

Initial: Number of ICs, $q = 2$

While $2 \leq q \leq A_{max}$ **do**

Perform ICA with presetting number of ICs, q , on mixing matrix, X .

Calculate residual matrices by subtracting estimated signals from initial signals,

$$R_q = X - \hat{X}.$$

Calculate DW criteria for each signal: $DW_{q,1}, \dots, DW_{q,p}$.

$q = q + 1$.

End While

Generate heatmap for $\{DW_{i,j}\}, i = 1, \dots, A_{max}, j = 1, \dots, p$. The larger the value, the lighter the color.

If for most of the columns, blocks in row $q + 1$ is significantly lighter than the one in row q then $q_{opt} = q$

Return q_{opt}

Due to the large variance of DW values among mixed signals, especially for the large number of mixed signals, it becomes challenging to visually select the optimal number of ICs. Further, another limitation of Durbin–Watson criterion is that it can only be used for structured signals (i.e., sparse, periodic, linear, etc.).

3.2.2 KMO_ICA_Residuals

The Kaiser-Meyer-Olkin (KMO) index was first developed to check whether the factorial analysis of a data set is pertinent [116]. This measure is also used to measure the independence of extracted ICs [111]. The KMO index is calculated from the residual signal matrix \mathbf{R} . The obtained value indicates whether at least two residual signals still have a partial correlation. The KMO index is calculated as follows:

$$KMO = \frac{\sum_i \sum_{i \neq j} b_{ij}^2}{\sum_i \sum_{i \neq j} b_{ij}^2 + \sum_i \sum_{i \neq j} a_{ij}^2} \quad (3.2)$$

where b_{ij} is the (i, j) element in the correlation matrix of residual signals and a_{ij} is defined as

$$a_{ij} = -\frac{v_{ij}}{\sqrt{v_{ii} + v_{jj}}} \quad (3.3)$$

where v_{ij} is the (i, j) element of the inverse of the correlation matrix. If the KMO index is close to one, it indicates that not all pure sources have been extracted and additional ICs must be computed. If the KMO value is close to zero, it indicates that all source signals have already been extracted and the model might be over-decomposed.

The algorithm of KMO_ICA_Residuals is summarized as below:

Algorithm Method based on the KMO Index

Input: Observed signals X , maximum number of ICs A_{max}

Output: Optimal number of ICs q_{opt}

Initial: Number of ICs, $q = 2$

While $2 \leq q \leq A_{max}$ do

Perform ICA with presetting number of ICs, q , on mixing matrix, X .

Calculate residual matrices by subtracting estimated signals from initial signals,
 $R_q = X - \hat{X}$.

Calculate KMO index for ICA model with q ICs, KMO_q .

If $KMO_q \approx 0$ then $q_{opt} = q - 1$.

Else $q_{opt} = q$.

$q = q + 1$.

End While

Return q_{opt}

However, the correlation matrix is not invertible if there exists completely or highly correlated variables, which might occur with small numbers of ICs. Then, the KMO index method is inapplicable in this case.

3.2.3 ICA_corr_y

ICA_corr_y was proposed to select the optimal number of ICs, which requires a known source signal (y) [111]. The key point of this method is to measure the correlation between the estimated source signals (\hat{S}) and the known source signal (y) of the mixed signal. For all models with different number of extracted ICs, if the model with q_{opt} ICs contains the IC which has the highest correlation to the known signal y , q_{opt} is claimed to be the optimal number of ICs. The highest correlation should be observed when the optimal IC number is extracted, even if there are some experimental errors.

The algorithm to determine the optimal number of ICs using ICA_corr_y is displayed as below:

Algorithm ICA_corr_y

Input: Observed signals X , maximum number of ICs A_{max} , a source signal y

Output: Optimal number of ICs q_{opt}

Initial: Number of ICs, $q = 2$

While $2 \leq q \leq A_{max}$ do

Perform ICA with presetting number of ICs, q , on mixing matrix, X .

Calculate correlation coefficient between y and each extracted ICs, respectively, r_1, \dots, r_q .

Record the maximal correlation coefficient $corr_q = \max_{1 \leq i \leq q} (r_i)$ for ICA model with q ICs.

$q = q + 1$.

End While

If $corr_q = \max_{2 \leq i \leq A_{max}} (corr_i)$ then $q_{opt} = q$

Return q_{opt}

However, having at least one known source signal is a strong requirement, and uncommon in many scientific fields e.g., the scalp EEG signal. Although, it is a simple method, it is not widely used for this strong requirement.

3.2.4 ICA-by-Blocks

ICA-by-Blocks was proposed to determine the number of ICs based on the correlation of ICs between blocks [112]. The original data matrix is splitted into B blocks, which is determined in advance. Then, A_{max} ICA models with 1 to A_{max} ICs are computed for each of these predefined blocks. ICs corresponding to “true” source signals are assumed to be found in all blocks. Such “true” ICs derived from different blocks should be highly correlated with each other. If all extracted ICs in each block are “true” ICs, the correlation between these ICs in different blocks will be close to 1. If too many ICs are extracted from the blocks, the extraneous ICs will contain a significant contribution related to noise, and so they will be significantly less correlated with all the ICs from other blocks.

The ICA-by-Blocks algorithm is summarized below:

Algorithm ICA-by-Blocks method

Input: Observed signals X , maximum number of ICs A_{max} , number of blocks B

Output: Optimal number of ICs q_{opt}

Initial: Number of ICs, $q = 2$

Randomly separate observed signals X into B blocks.

While $2 \leq q \leq A_{max}$ do

Perform ICA with presetting number of ICs, q , on each block, respectively.

Generate Pearson correlation coefficient matrix $P_{qB \times qB}$ for all $q \times B$ ICs from all B blocks.

Vectorize correlation matrix P and sort elements from the largest to the smallest, $V_{(qB)^2 \times 1}$

Extract values from $(q \times B + 1)^{th}$ to $(q \times B + q \times (B^2 - B))^{th}$ in the vector V , denote as $L_{q(B^2 - B) \times 1}$

Generate the signal-correlation plot for visualization based on every second value in the vector L .

$q = q + 1$

End While

If most significant drop occurs at q^{th} point in the signal-correlation plot, then $q_{opt} = q$

Return q_{opt}

However, this method is constrained by the number of mixed signals, p , since both the choice of the number of blocks, B , and maximum number of ICs, A_{max} , are depending on the sample size (i.e., $A_{max} \leq \frac{p}{B}$). While multiple blocks are desired to better measure the correlation between ICs extracted from different blocks, too many blocks will restrict the maximum number of ICs and increase the computational complexity. Further, this method does not use a quantitative measure to determine the optimal number of ICs, instead a signal-correlation plot is used visually to determine it, thus, is rather time-consuming and prone to subjective errors.

3.3 Proposed CW_ICA method

3.3.1 Algorithm Development

The CW_ICA method starts by randomly splitting the initial data matrix into two sample blocks (\mathbf{B}_1 and \mathbf{B}_2) with approximately equal numbers of signals. ($A_{max} - 1$) ICA models are then performed with different number of ICs (from 2 to A_{max}) on each block. A_{max} is preset as the maximum number of computed ICs and assumed to be less than the number of signals in each block, ($\frac{p}{2}$). When p is an odd number, the mixed signals are randomly divided into two blocks, with one containing $\frac{p+1}{2}$ signals and the other containing $\frac{p-1}{2}$ signals. ICs corresponding to the true source signals are expected to be in both two blocks. We assume that for each true IC extracted from Block 1, there must be a highly correlated IC extracted from the Block 2. In order to measure the relationship between ICs from different blocks, we use the rank-based correlation matrix, $\mathbf{P}_{2q \times 2q}$, that measures monotonic relationship, not restricted to linear relationship as the Pearson correlation matrix does. Further, due to the symmetry of the correlation matrix, we only need to perform further analysis on any one of off-diagonal blocks, $\mathbf{P}'_{q \times q}$ (see the steps of this process in Figure 3.1).

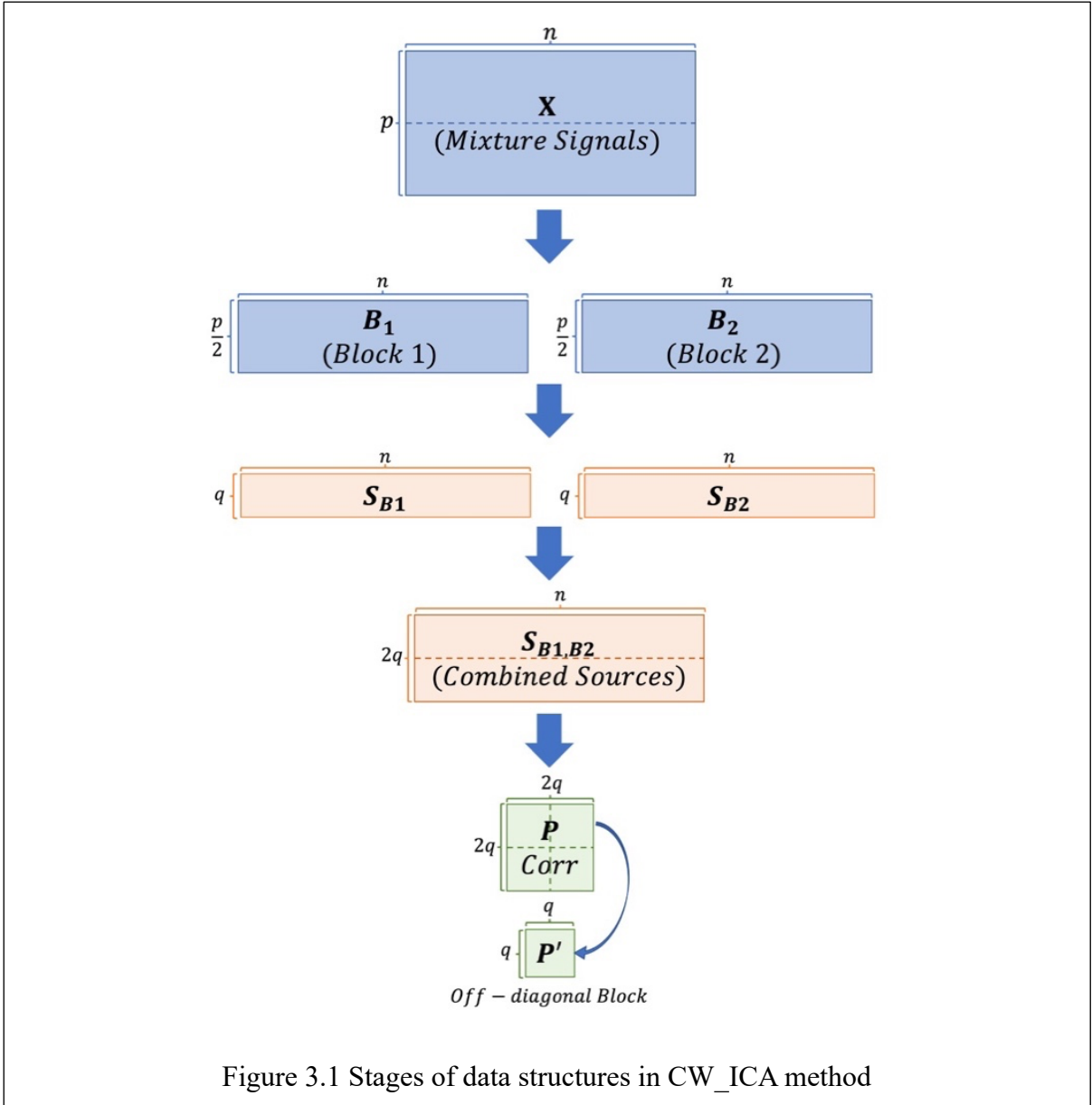


Figure 3.1 Stages of data structures in CW_ICA method

For each model with q ICs, we record the maximum absolute value of each column in \mathbf{P}' , indicating the strongest correlation between each pair of ICs ($\rho_1, \rho_2, \dots, \rho_q$). Then, we record the smallest of these q values to represent the least absolute correlation coefficient between a pair of ICs. This leads to a quantitative measurement for the ICA model with q ICs, defined as

$$R_q = \min_{1 \leq i \leq q} \{\max\{|\rho_i|\}\} \quad (3.4)$$

where r_i is the i^{th} column of the matrix \mathbf{P}' . When q is small, it is likely to underestimate the source signals, resulting in highly correlated extracted ICs, i.e., R_q is closed to 1. As q

increases, it tends to overestimate, introducing noise signals and causing some extracted ICs to be uncorrelated, i.e., R_q is closed to 0. Therefore, to identify the optimal number of IC, we observe the changes of R_q as q grows. The q_{opt} is selected based on the pattern of R_q , where R_q is relatively high while R_{q+1} is significantly lower. Specifically, when the number of ICs exceeds q_{opt} , if R_q decreases significantly and remains consistently low as q increases, we claim that the optimal number of ICs is q_{opt} .

To quantify the “significant drop”, we calculate the first-order difference of $R_2, \dots, R_{A_{max}}$

$$D_{i,q} = R_{i,q} - R_{i,q-1}, q = 3, \dots, A_{max}, i = 1, \dots, Rep \quad (3.5)$$

where $R_{i,q}$ is the smallest column-wise maximum absolute correlation value, Rep is the number of repetitions. A negative value of $D_{i,q}$ signifies a decrease in the measurement, with a smaller value indicating a more substantial decrease. Thus, the optimal number of ICs is automatically selected out according to the index of the smallest first-order difference, that is

$$\min_{3 \leq q \leq A_{max}} \{D_{i,q}\}.$$

We repeat this procedure multiple times, record the detected optimal number of ICs each iteration. The optimal number of ICs is the one that occurs most frequently over all repetitions. The steps are summarized below:

Algorithm CW_ICA method

Input: Observed signals X , maximum number of ICs A_{max} , number of repetitions

Rep

Output: Optimal number of ICs q_{opt}

Initial: Number of ICs, $q = 2$

For $1 \leq i \leq Rep$ **do**

Randomly and evenly split mixed signals into two blocks, B_1 and B_2 .

While $2 \leq q \leq A_{max}$ **do**

Perform ICA with presetting number of ICs, q , on each block, respectively.

Generate Spearman correlation coefficient matrix $P_{2q \times 2q}$ for all $2q$ ICs from two blocks.

Calculate the smallest column-wise maximum value, $R_{i,q}$, in one off-diagonal block matrix $P'_{q \times q}$.

Calculate $D_{i,q} = R_{i,q} - R_{i,q-1}$.

$q = q + 1$

End While

If $D_{i,j} = \min_{2 \leq q \leq A_{max}} \{D_{i,q}\}$ then record $Q_i = j - 1$ as the optimal number of ICs for i^{th} replicate.

Generate the signal-correlation plot for visualization based on recorded $R_{i,2}, \dots, R_{i,A_{max}}$.

$i = i + 1$

If $q_{opt} = mode(Q_1, \dots, Q_{rep})$ then

Return q_{opt}

The signal-correlation plot shown in Figure 3.2C, which depicts the change in R_q as q increases, can also be used to visually identify the optimal number of ICs via detecting a significant drop. However, this method could be time-consuming and may introduce subjective bias.

3.3.2 Validation

CW_ICA starts by randomly partitioning mixed signals into two blocks. This is pivotal as excessively dividing the data into numerous blocks might result in each block containing only a limited number of ICs[111]. In this method, we suggest using a rank-based correlation coefficient, specifically, the Spearman correlation coefficient, as a measure for determining the correlation between ICs. The Spearman correlation coefficient is a statistical measure used to assess the strength and direction of the relationship (i.e., the monotonic relationship) between two variables. As a nonparametric measure, the Spearman correlation coefficient refrains from making assumptions about data distribution or homoscedasticity. Instead of relying on the actual values, the Spearman correlation is based on ranking the data points for both variables. Moreover, the Spearman correlation is robust against outliers because it relies on ranking rather than actual values, making it less susceptible to the influence of extreme data points. The determination of the optimal number of ICs in the ICA-by-Blocks method heavily relies on

visual interpretation of a signal-correlation plot. However, this approach is inherently subjective and potentially time-consuming, particularly when the plot shows multiple significant drops at similar levels. In contrast, the proposed CW_ICA method automates this process by quantitatively defining what constitutes a "significant drop", thereby eliminating the need for manual interpretation or inspection.

Additionally, the ICA-by-Blocks method lacks repetitions, which increases the risk of obtaining results by chance. To avoid the risk of biased random splitting due to specific row distribution, we propose repeating the entire process — randomization, equitable division of mixed signals into two blocks, and the subsequent steps — multiple times, denoted as Rep. This systematic repetition within the CW_ICA methodology ensures a more robust and reliable determination of the optimal number of ICs.

3.3.3 Illustrative Example

To provide a clear explanation of the CW_ICA procedure, we consider a dataset comprising $p = 20$ mixed signals, each with a length of $n = 500$, which are generated by linear combination of 5 source signals ($q_{opt} = 5$). Our objective is to determine the optimal number of ICs, q_{opt} .

CW_ICA starts with randomly dividing the 20 mixed signals into two blocks: \mathbf{B}_1 and \mathbf{B}_2 , with the dimension of each block being $\frac{p}{2} \times n$ (*i. e.*, 10×500). With the maximum number of ICs being 10 ($A_{max} = 10$), we perform ICA on each block using a range of preset IC numbers ($q = 2, \dots, 10$). For instance, when $q = 8$, after applying ICA to both \mathbf{B}_1 and \mathbf{B}_2 , we obtain two sets of ICs, one from each block, resulting in a combined matrix of $2q \times n$ (*i. e.*, 16×500). Then we compute the Spearman correlation coefficient between each pair of ICs and the correlation matrix is denoted as \mathbf{P} with dimension being $2q \times 2q$ (*i. e.*, 16×16). \mathbf{P} is composed of four distinct blocks with each block being a $q \times q$ (*i. e.*, 8×8) matrix. The diagonal blocks are indeed identity matrices because they are the correlation between ICs from the same block, which are orthogonal. The two symmetric off-diagonal blocks, which depict correlations between the ICs in \mathbf{B}_1 and \mathbf{B}_2 , contains the same information. therefore, we only need to consider one off-diagonal matrix, \mathbf{P}' , as shown in Figure. 3.2B,

For each column of \mathbf{P}' , we first record the highest absolute correlation value ($\rho_1, \rho_2, \dots, \rho_8$), which indicates the strength of correlation between each IC in \mathbf{B}_1 and each IC in \mathbf{B}_2 . Then, the

quantitative measurement of this model is $R_8 = \min_{1 \leq i \leq 8} \{\rho_i\}$. If R_8 is close to 1, it suggests that all extracted ICs in \mathbf{B}_1 have strong correlations with ICs from \mathbf{B}_2 . This indicates that the optimal number of ICs is greater than or equal to 8. Conversely, if R_8 is close to 0, it implies that some extracted ICs from \mathbf{B}_1 are not correlated with any ICs from \mathbf{B}_2 . These redundant ICs indicate that the optimal number of ICs should be less than 8. In Fig. 2B, we observe that R_8 is close to 0, which indicates that q_{opt} is less than 8 in this iteration. Moreover, we also show the situation when q is 5, which is indeed the true number of signals, in Fig 2A. R_5 is close to 1, suggesting that the q_{opt} is greater than or equal to 5.

To determine the q_{opt} , we look at R_2, \dots, R_{10} and locate the significant drop. Specifically, we calculate the first order differences between these values. For instance, we compute $D_6 = R_6 - R_5$, if D_6 is found to be the smallest value among all the differences, we conclude that the optimal number of ICs in the first iteration is $q_{opt,1} = 5$. Furthermore, we generate a signal-correlation plot to visually examine if there is a significant drop at the 5th point. A significant drop in the plot suggests $q_{opt,1}$.

In the simulation, we repeat the whole process 10 times, i.e., $Rep = 10$, and record the optimal number of ICs from each iteration as $q_{opt,i}, i = 1, \dots, 10$. Based on previous observations, the performance of CW_ICA, regardless of combining with various ICA methods, become stable in 10 repetitions. Nevertheless, it is essential to adjust the number of repetitions based on the number of input signals to avoid excessive computation time while maintaining accurate outcomes. By examining the 10 signal-correlation lines overlaid on the plot (Figure. 3.2C), we consistently observe the significant drop occurring at the 5th point from most iterations. Based on this frequent occurrence, we confidently conclude that the optimal number of ICs is $q_{opt} = 5$.

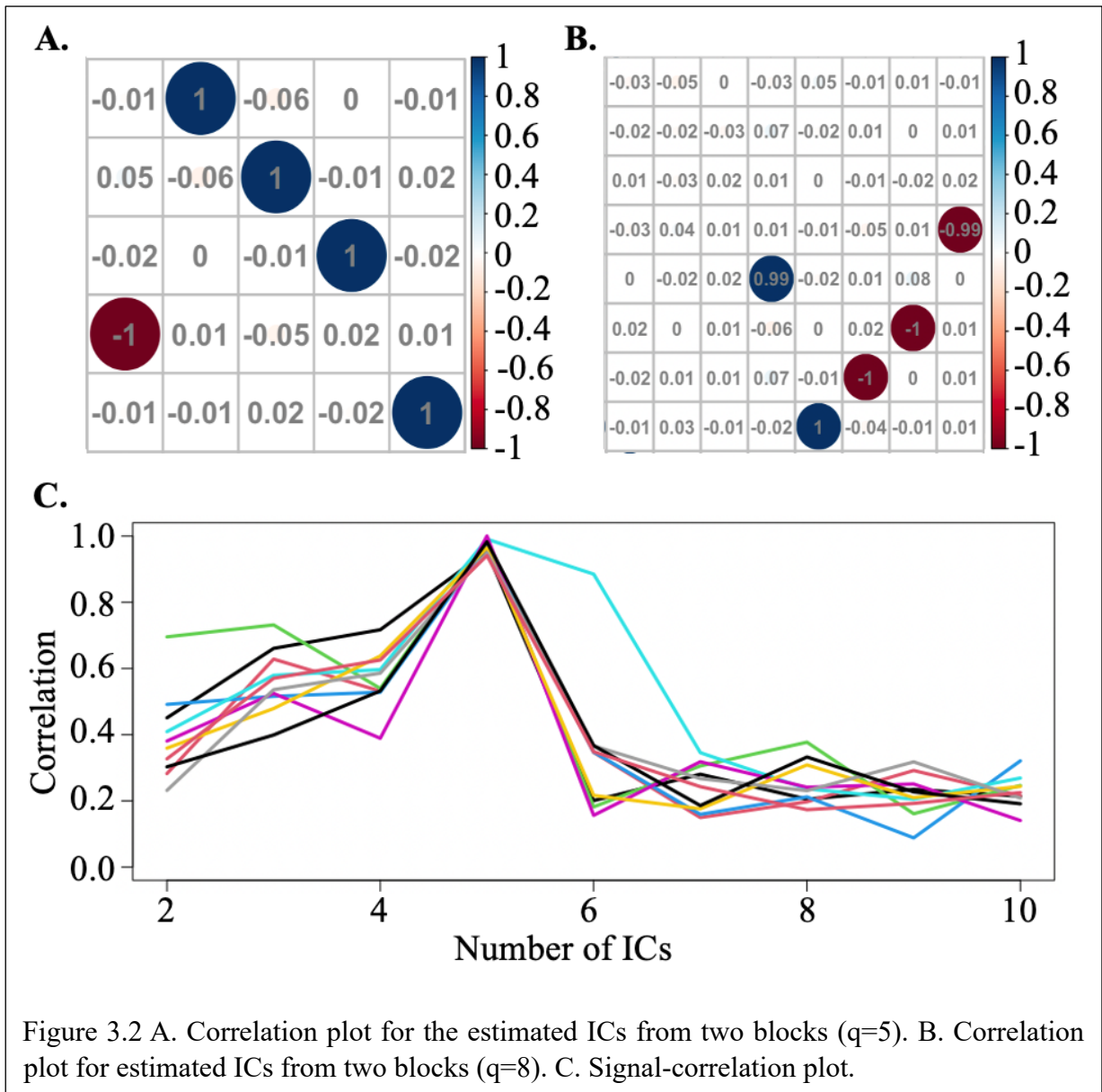


Figure 3.2 A. Correlation plot for the estimated ICs from two blocks ($q=5$). B. Correlation plot for estimated ICs from two blocks ($q=8$). C. Signal-correlation plot.

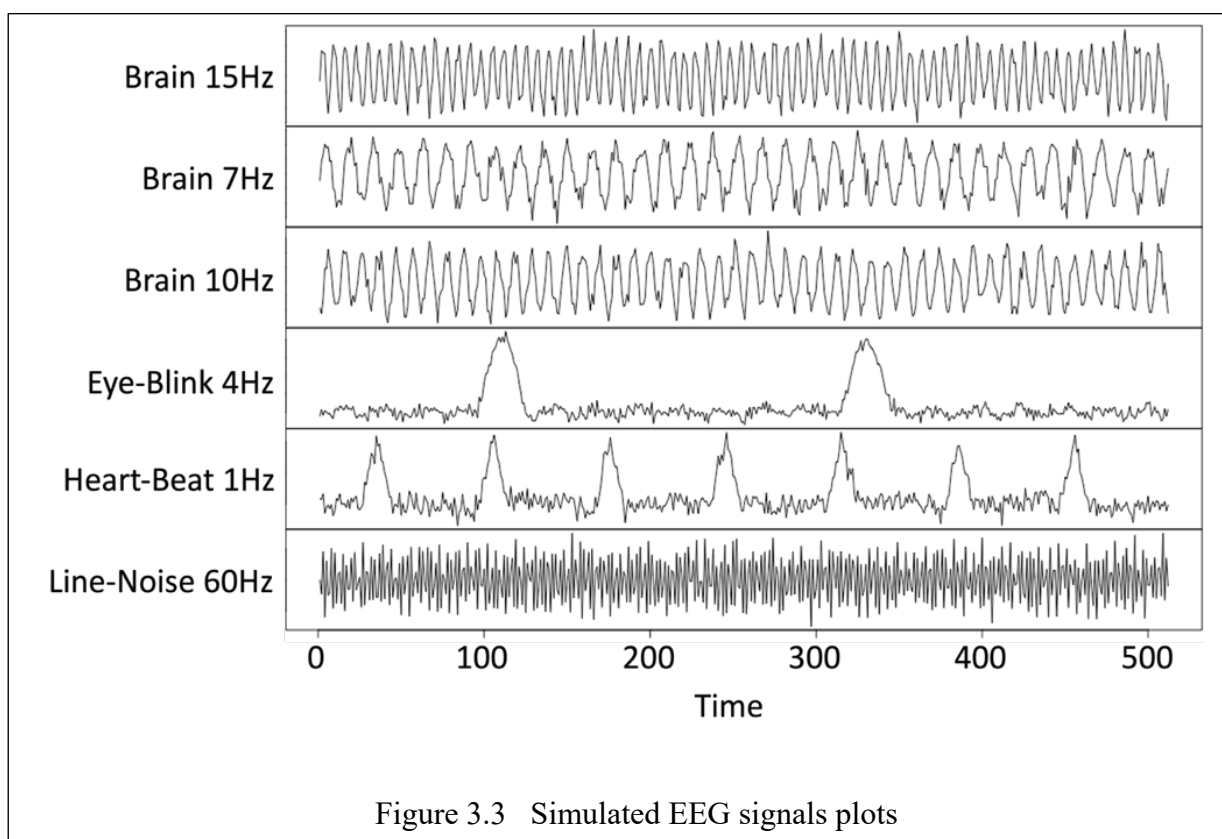
3.4 Simulation

Artifacts in the collected scalp EEG signals are inevitable and can affect the subsequent analysis of brain activity. To address this issue, ICA techniques are widely utilized to “clean” scalp EEG signals by filtering out artifacts (e.g., eye movements, cardiac activity, muscle activity, etc.) from brain signals. If too few ICs are used, the resulting brain signal may still contain artifacts, reducing the effectiveness of artifact removal. On the other hand, using an excessive number of ICs can lead to over-separation of the brain signal, potentially causing the loss of important features and information. Therefore, determining the accurate number of

source signals is crucial when applying ICA to scalp EEG signals. In this section, we carry out simulations and implement the proposed method on real EEG data to assess the proposed CW_ICA.

3.4.1 Simulation

Firstly, we assess the performance of the proposed method by applying CW_ICA in conjunction with different ICA methods on simulated EEG signal data. Since the true number of source signals is known in simulation data, the accuracy of CW_ICA along with three determination methods, the DW, ICA-by-Blocks and ICA_corr_y methods, will be compared. We select three widely used ICA methods, namely FastICA, Infomax, and JADE, to combine with each determination method. However, it is important to note that JADE have convergence issues if the preset number of ICs is greater than the number of source signals. Therefore, JADE is only performed on the real data, where the true number of signals is unknown.



3.4.1.1 Simulated Data Generation

According to characteristics of EEG components[91], eight analog EEG source signals (i.e., true $q = 8$), which consist of both periodic and non-periodic signals as shown in Figure 3.3, are simulated. Then, $p = 30$ mixed signals, denoted as \mathbf{X} , are generated by linearly combining the source signals as $\mathbf{X} = \mathbf{A}\mathbf{S}$, where \mathbf{A} is a randomly generated mixing matrix whose elements are normally distributed.

3.4.1.2 Impact of Correlation Coefficients on Determination Methods

To assess the impact of different correlation coefficients on two determination methods, specifically ICA-by-Blocks and CW_ICA, we conduct this simulation study. Noting that the true optimal number of ICs is $q_{opt} = 8$ in this simulation.

We employ both CW_ICA and ICA-by-block coupled with Pearson and Spearman correlation to determine the optimal number of IC, q_{opt} . Figure 3.4 shows the signal-correlation plots. For ICA-by-Blocks, when it is coupled with Spearman correlation (Figure 3.4 A2 and A4), the estimated q_{opt} is 8, which is the same as the true q . However, when coupled with Pearson correlation (Figure 3.4 A1 and A3), ICA-by-block determines q_{opt} being 10, which leads to over-decomposition. This discrepancy suggests that the choice of correlation coefficient greatly influences the determination accuracy of ICA-by-Blocks. On the other hand, CW_ICA, clearly exhibits a sharp drop at the true number of ICs (*i.e.* $q_{opt} = 8$), which indicates that CW_ICA outperforms ICA-by-Blocks regardless of the type of correlation coefficient employed (Figure 3.4 B1-B4). Moreover, the result obtained by the Spearman correlation-based CW_ICA provides even much more compelling evidence for accurately identifying the true q_{opt} (Figure 3.4 B2 and B4). This is because Spearman correlation coefficient captures monotonic relationships that exist among ICs, providing enhanced performance in determining the optimal number of ICs.

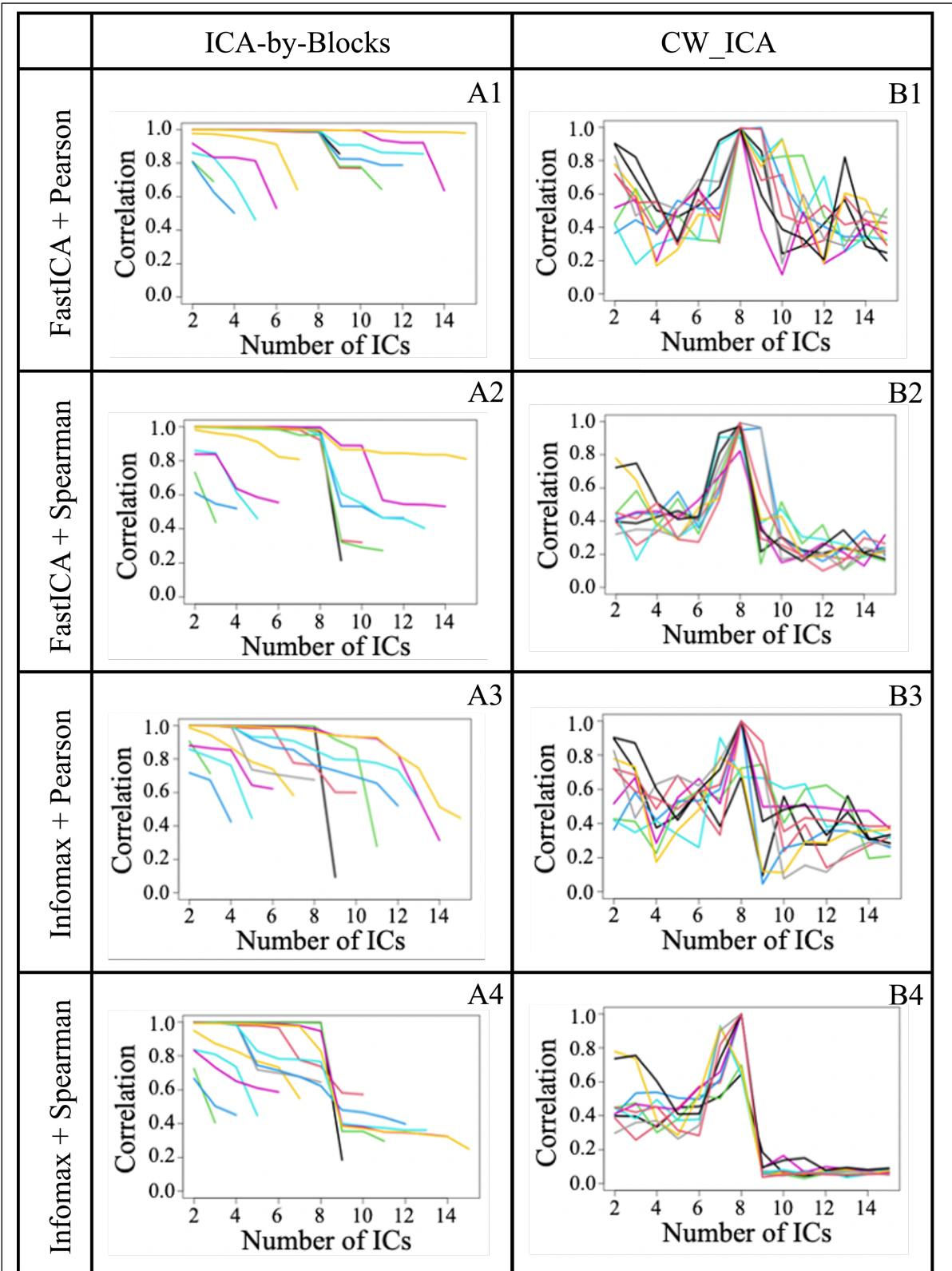


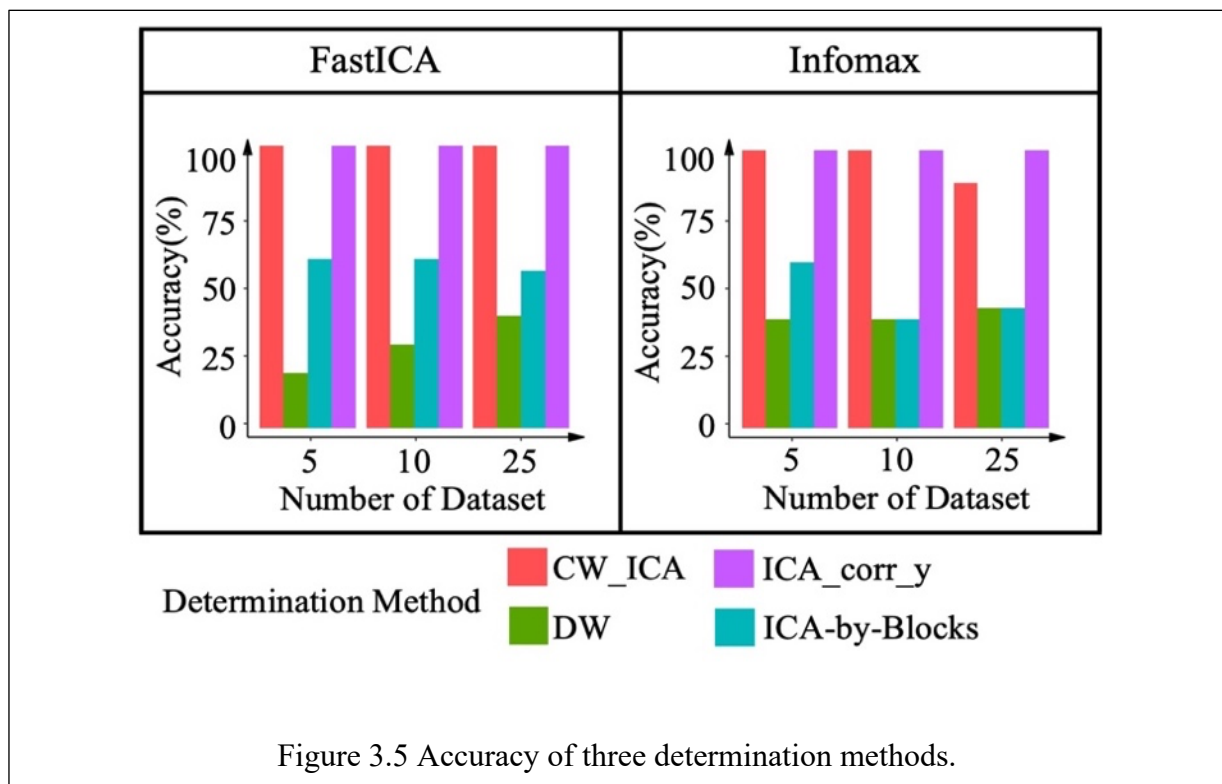
Figure 3.4 Signal-Correlation plots for CW_ICA (A) and ICA-by-Blocks (B) methods.

3.4.1.3 Accuracy

To show how accurately the proposed CW_ICA method detects the optimal number of the ICs, we compare our method with the ICA-by-Blocks and DW determination methods. Similarly, we randomly generate $p = 30$ mixed signals \mathbf{X} by linear combination of $q = 6$ simulated EEG component signals. Multiple mixed-signal datasets, \mathbf{X} , are obtained by repeatedly and randomly generating the mixing matrix, \mathbf{A} , while keeping the original source signal, \mathbf{S} , unchanged. Accuracy in this context refers to the percentage of simulation runs that correctly identify the optimal number of ICs.

$$Accuracy = \frac{m}{N} \times 100\% \quad (3.6)$$

where m is the number of simulations runs that correctly identify the optimal number of ICs and N denotes the total number of simulations runs. To estimate the optimal number of ICs, all three determination methods (i.e., CW_ICA, ICA-by-Blocks and DW) are combined with FastICA and Infomax, respectively. We only compare the accuracy of three methods for limited number of simulations runs (i.e., $N = 5, 10, 25$), due to the fact that the DW and ICA-by-Blocks methods are time consuming since they require graphs to identify q_{opt} .



As shown in Figure 3.5, CW_ICA provides 100% accuracy combined with either FastICA or Infomax. The accuracy of ICA-by-Blocks combined with FastICA is also 100%, which is much higher than the result when combined with Infomax. DW performed worst, especially when combined with FastICA.

3.4.1.4 Robustness

To further understand the robustness of existing and proposed determination methods on datasets with different characteristics, we generated datasets with varying numbers of mixed signals, signal lengths, signal-to-noise ratio, and frequency ranges, respectively. A total of three methods for determining the number of ICs are compared with the combination of FastICA and Infomax.

Figure 3.7 illustrates the estimated number of the source signals, q_{opt} , obtained by four methods as we vary the levels of mixed signals, signal lengths, signal-to-noise ratio, and frequency ranges. First, we conclude that DW coupled with FastICA is not suitable, since the Figure 3.7A1-A4 show that the results obtained using the DW criteria is inconsistent as the signal parameters change. Second, we observe that if ICA-by-Blocks is used in conjunction with Infomax, it generates inconsistent results in certain cases, for instance as the number of mixed signals increases, or the signal length changes (i.e., Figure 3.7B1, B3). Additionally, the results obtained by ICA_corr_y show variations with changes in the characteristics of mixed signals (i.e., Figure 3.7B3, B4). Furthermore, when combined with FastICA, ICA_corr_y produces incorrect results under certain conditions (i.e., Figure 3.7A2). In comparison, CW_ICA shows more consistent results regardless the characteristic of mixed signals and the ICA methods.

3.4.2 Scalp EEG Data Application

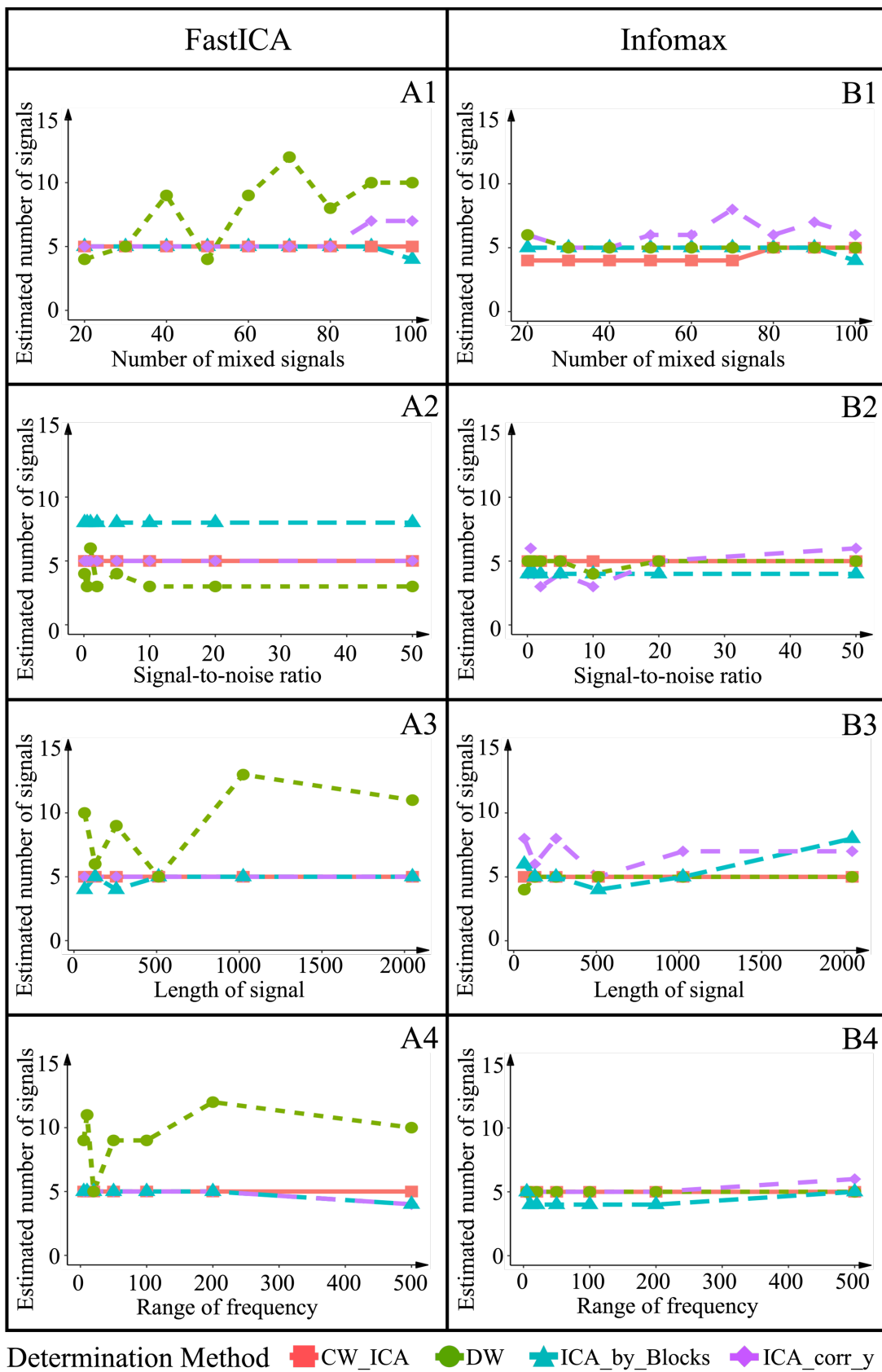


Figure 3.6 Estimated number of source signals from the varying mixed signals.

Artifacts in the collected scalp EEG signals are inevitable, which affects the subsequent analysis of brain activity. ICA techniques are widely utilized in the scalp EEG signals to filter out artifacts (e.g., eye movements, cardiac activity, muscle activity, etc.) and extract brain signals. However, too few ICs results in a contaminated brain signal that still contains artifacts, while too many ICs causes the brain signal to be over separated, losing some relevant features. Therefore, it is important to determine the number of source signals when applying ICA on scalp EEG signals. In this chapter, we applied our data to the raw scalp EEG dataset, introduced in Chapter 1.1.

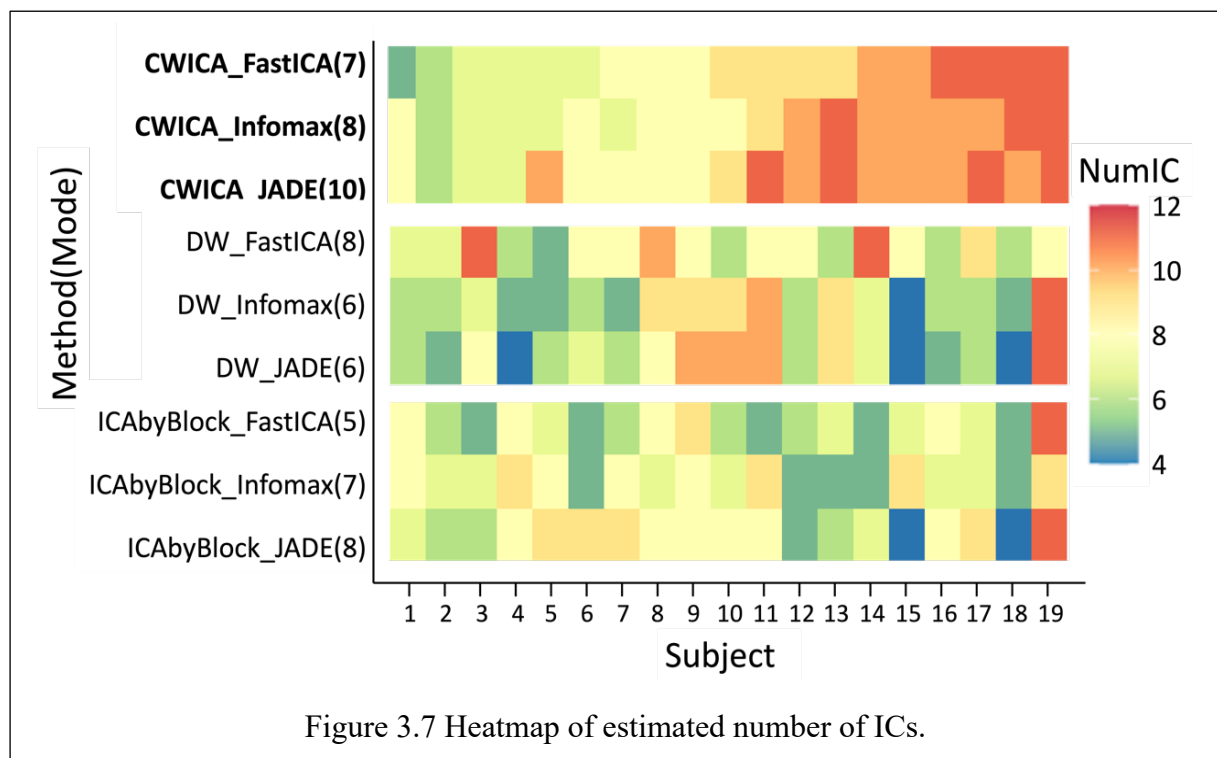
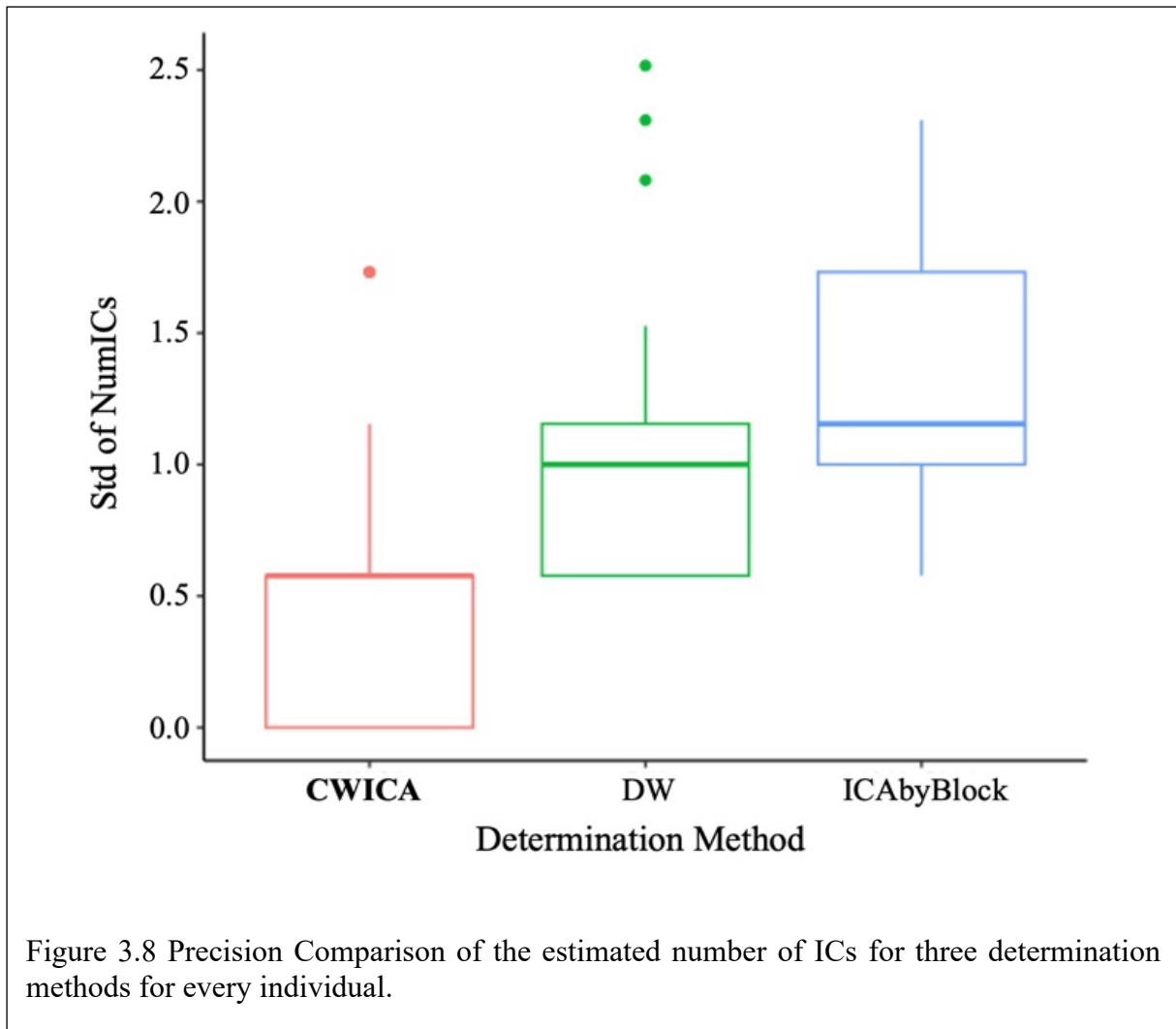


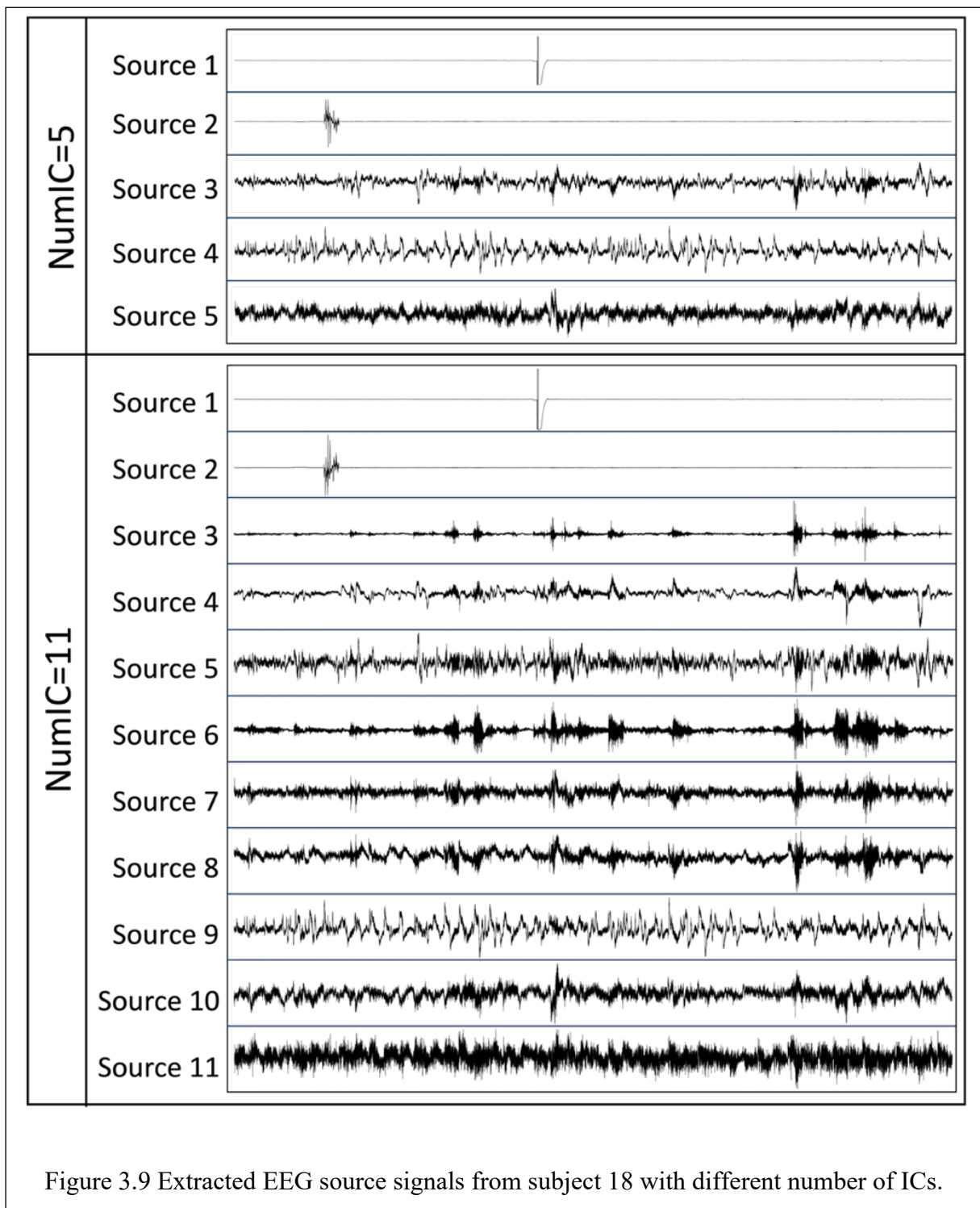
Figure 3.7 Heatmap of estimated number of ICs.

To validate the effectiveness and robustness of the proposed CW_ICA, we consider three selection methods for determining the optimal number of ICs (i.e., CW_ICA, DW, ICABlock), and each is coupled with three ICA methods (Fast ICA, Infomax, and JADE). We also consider determining the optimal number of ICs at two levels: subject-wise and channel-wise. First, the optimal number of ICs is determined for each subject respectively, with the dimension of mixed signals \mathbf{X} being 2828×48 (n) by 64 (p). The optimal ICs obtained using nine methods over 19 subjects are summarized in the heatmap (Figure 3.7). In this plot, each entry denotes the estimated number of ICs, the redder the larger value is. Each column records the estimated number of ICs obtained by 9 methods for one subject. For each method, the



number in the parenthesis summarizes the mode of IC number across all subjects. For each subject, the resulting ICs numbers selected by CW_ICA are more consistent across three ICA methods, compared to DW and ICABlock. The variation of the determined IC number, quantified by the variance, is further summarized in Figure 3.8. Overall, CW_ICA shows the minimal variation when combining with different ICA methods, which indicates the consistency of CW_ICA. Besides, the optimal IC number obtained by CW_ICA for all subjects is around 7-10 while DW and ICABlock give smaller IC number, around 5-8. In real data, there is no ground truth, therefore, we further investigate the corresponding ICs obtained using the optimal IC number determined by three selection algorithms. Take subject 18 as an example. The optimal IC number obtained by CW_ICA is 11 while the number determined by the other two selection methods is 5. Therefore, we preset the number of IC being 5 and 11 respectively when applying ICA on subject 18 and summarize the obtained ICs in Figure 3.9. With number of ICs being 5, some channel noise can be separated (e.g., IC 1 and 2). However, the rest three

ICs are still mixed signals. With IC number being 11, sources are much better separated, for instance, channel noise (IC 1 and 2) and muscle (IC 6 and 7).



Besides subject-wise analyses, we also consider channel-wise determination of the optimal IC number. This approach accounts for the possibility that the number of source signals may

vary across different brain regions. For example, channels located near the eyes may have a higher number of sources compared to other regions. To determine the channel-wise optimal IC number we implement the nine algorithms on the mixed signals \mathbf{X} , with dimensions of 2828 (n) by 48 (p), for each subject and each channel. The average of obtained optimal IC number across all 19 participants is calculated and summarized in the heatmap given in Figure 3.10. In this plot, each entry denotes the estimated number of ICs, the redder the larger value is. Each column describes the estimated number of ICs by 9 methods at the same electrode.

When examining the standard deviation (Std) of the determined IC numbers in Figure 3.11, it is evident that the CW_ICA method exhibits more consistent outcomes when combined with different ICA methods, even though the optimal IC number may vary across different electrodes.

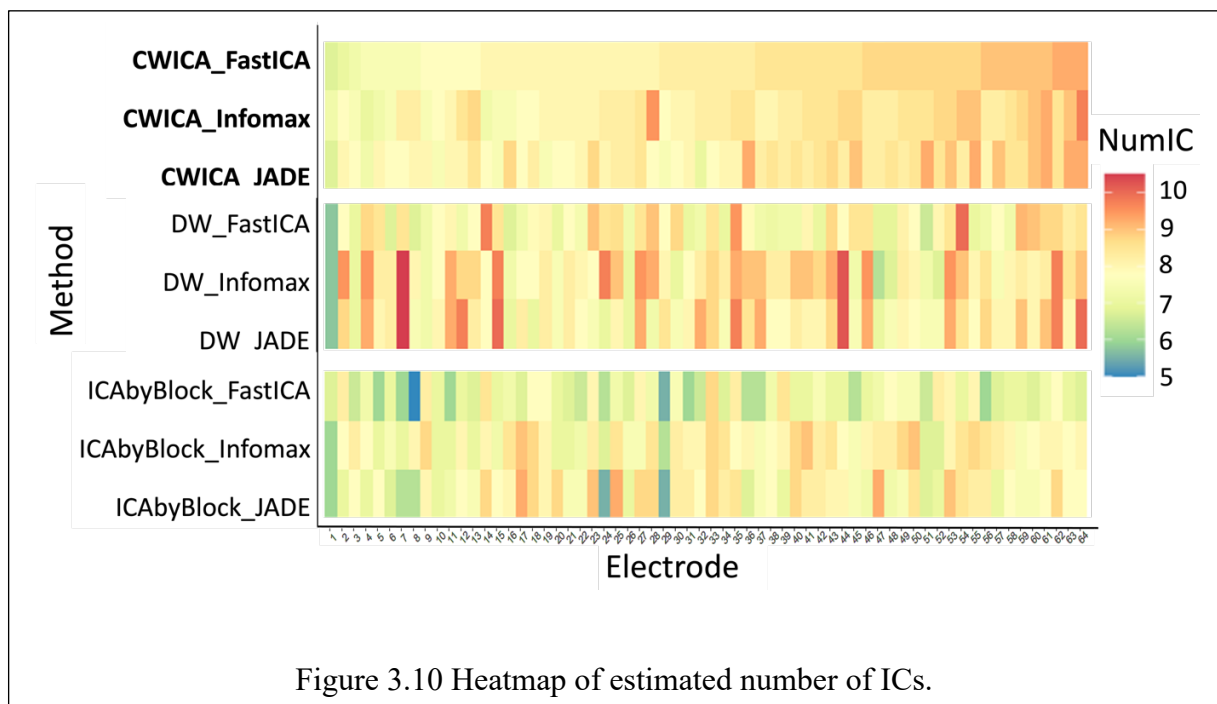
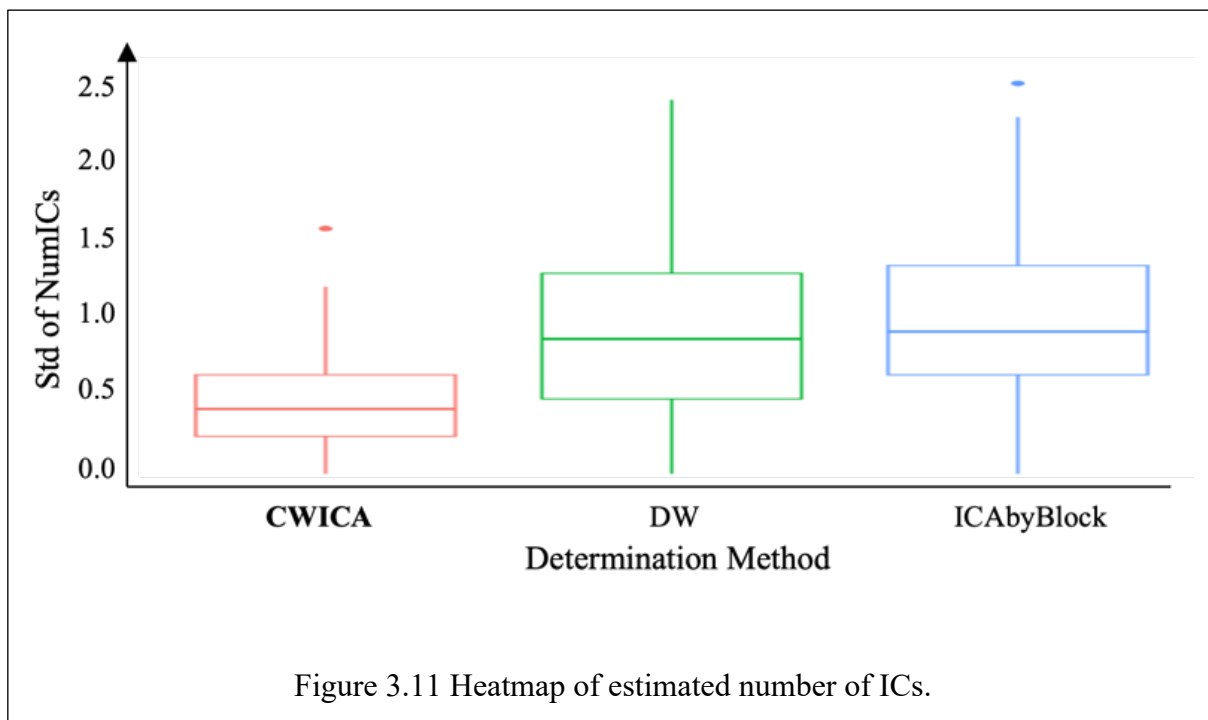
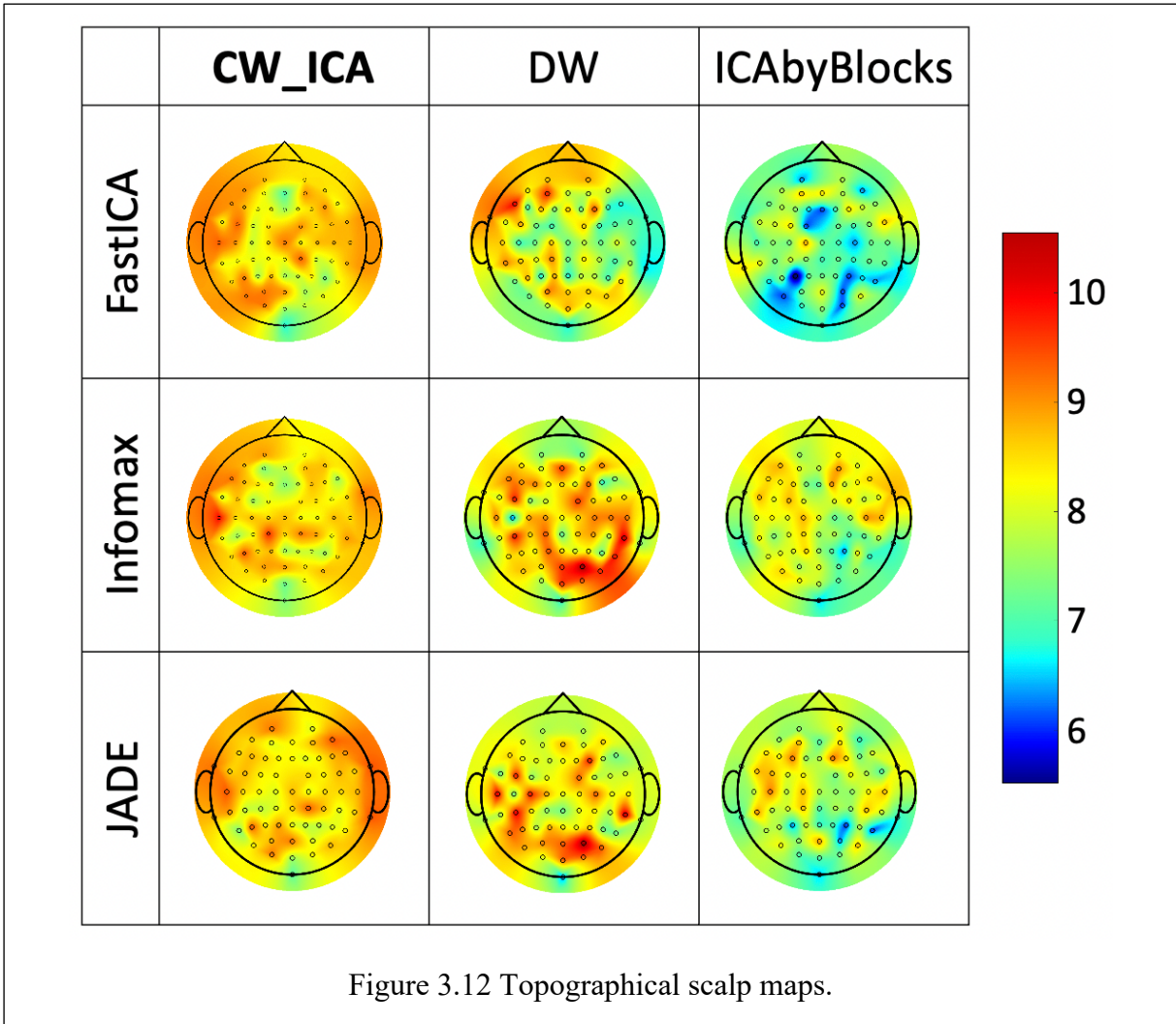


Figure 3.10 Heatmap of estimated number of ICs.



Topographical scalp maps (Figure 3.12) represent the channel-wise optimal number of ICs for all 19 participants and provide valuable insights into the distribution of the optimal IC numbers across different brain regions. Overall, the optimal IC number obtained by the CW_ICA method is relatively lower in the frontal cortex, indicating potentially less interference from physical artifacts. Whereas a larger number of ICs may be required to effectively separate the brain signals from the artifacts in prefrontal and temporal cortex. This pattern can be attributed to the fact that the prefrontal and temporal cortex regions are near facial muscles and signals are more susceptible to physical artifacts, such as muscle movements and eye blinking. In contrast, the other two determination methods, particularly ICA-by-Blocks, do not show significant variations in the optimal IC numbers across different electrodes.



3.5 Discussion

In this study, we propose a robust method, CW_ICA, to determine the optimal number of ICs. Although there have been many studies on determining the number of ICs, the ability of combination with different ICA methods and the comparison of their performance on signals with different natures have not been discussed. The current methods for determining the number of ICs are mainly used only in conjunction with JADE[112], and applied to the real data from some specific fields. There is no information about whether they can be combined with other simpler or widely used ICA methods, i.e., FastICA, Infomax, etc., and the robustness of these methods. Therefore, to fill in this missing information and address the shortcomings

of existing determination methods, we propose CW_ICA, which can be combined with multiple ICA methods and automatically determine the optimal number of ICs.

We summarize the significant advantages of the proposed CW_ICA method over existing approaches here. First, its computational cost is significantly lower than the existing methods since it eliminates the need for source signal-related information. Moreover, it relies on a quantitative measurement instead of visual identification. Secondly, the CW_ICA method simplifies the process of determining the optimal number of ICs compared to the ICA-by-Blocks method. Rather than determining the number of blocks in ICA-by-Blocks, CW_ICA divides mixed signals into only two blocks. This simplification eliminates potential challenges in selecting the maximum number of ICs when dealing with a large number of blocks. In addition to the aforementioned advantages, CW_ICA offers several other notable benefits. Firstly, CW_ICA extracts only one value from each column of the off-diagonal matrix, whereas ICA-by-Blocks preserves all values in the off-diagonal matrix. This streamlined approach significantly reduces complexity and computational overhead while maintaining the accuracy and reliability of determining the optimal number of ICs. Secondly, CW_ICA can be coupled with multiple ICA methods, consistently yielding reliable results. Researchers can focus solely on selecting an appropriate ICA method based on the properties of the mixed signal, without concerns about the compatibility of the determination method and the ICA method. Thirdly, CW_ICA is a robust method because it uses rank-based correlation instead of Pearson correlation coefficient as used in ICA-by-Blocks. The rank-based correlation measures the relationship between ICs from different blocks based on ranks, avoiding reliance on assumptions and generating more robust results. Finally, CW_ICA automatically determines the optimal number of ICs. Unlike ICA-by-Blocks and DW, which require visual identification from plots, CW_ICA quantifies the identification process. This quantification allows for automated determination, eliminating the time-consuming manual analysis required by the other methods.

To better illustrate the advantages of CW_ICA, we apply it and some existing determination methods on both simulated signals and collected raw EEG signals, respectively. First, we compare the measurements of two correlation coefficients (Pearson and Spearman correlation coefficient) using ICA-by-Blocks and CW_ICA. Results show that CW_ICA with Spearman clearly display a significant drop at 6, which is the true number of ICs (Figure 3.4). Second, we check the accuracy of CW_ICA by applying it to multiple datasets and comparing it with ICA-by-Blocks and DW. Results imply that only CW_ICA keeps 100% accuracy rate, the

performance of the other two methods varies when they are combined with different ICA methods (Figure 3.5). Next, to verify the robustness of CW_ICA, it is applied to multiple sets of mixed signals with different properties (i.e., length, quantity, signal-to-noise ratio, and range of frequency), and compared with the results obtained by ICA-by-Blocks and DW. Results indicate that the correct number of ICs is identified by CW_ICA and not affected by changes of mixed signals (Figure 3.6). Further, we compare 9 combinations of determination methods (CW_ICA, ICA-by-Blocks, DW) and ICA methods (FastICA, Infomax, JADE) to determine the optimal number of ICs in raw EEG signals (Figure 3.7, Figure 3.8, Figure 3.10 and Figure 3.11). Among these combinations, only the proposed CW_ICA provides the same number of ICs across each electrode, demonstrating its compatibility with different ICA algorithms and consistent determination capability. It is worth noting that our methods are adaptable and can be applied to other datasets for the determination of the number of source signals as well.

Although the robustness of the proposed CW_ICA is proved by comparison with other methods under different conditions, the measurement difference between models with consecutive number of ICs in CW_ICA is not as significant as before, when the size of the signal matrix becomes larger (length or number of mixed signals increases). Besides, the computing complexity increases as the number of mixed signals and signal length increase, since it is measured by the correlation matrix. To tackle these issues, in the future work, the proposed method will be extended to functional methods to reduce the dimensionality of the data and obtain more robust and obvious results. Besides, we only discuss the classical ICA methods here, we will also consider performing functional ICA with the proposed CW_ICA method in next chapter.

3.6 Conclusion

The proposed CW_ICA method addresses limitations in current determination methods by introducing a quantitative measurement and a block splitting approach to reduce computational complexity. By focusing on the smallest column-wise maximum absolute value, CW_ICA offers a versatile solution that can be seamlessly integrated with various ICA methods. Moreover, it leverages the robustness of the Spearman correlation coefficient, which leads to reliable and consistent results in determining the optimal number of ICs automatically. To evaluate the performance of CW_ICA, it is compared with existing determination methods in

combination with multiple ICA methods using extensive simulated data and real raw EEG signals. In conclusion, the proposed CW_ICA method offers a versatile and robust approach for automatically determining the number of ICs in signal analysis. Its compatibility with multiple ICA methods, reduction in computational complexity, utilization of Spearman correlation coefficient, and strong performance in comparative evaluations make it a valuable tool for researchers in various fields.

Chapter 4

Robust Functional Independent Component Analysis for Functional Data

4.1 Introduction

Functional Independent Component Analysis (fICA) holds a unique advantage over classical ICA methods in that it is specifically tailored for applications where the temporal dimension is critical, such as in neuroimaging. While classical ICA seeks to find statistically independent components in multi-dimensional data, fICA takes this a step further by considering context-specific patterns of activity. In domains like functional Magnetic Resonance Imaging (fMRI) and Electroencephalography (EEG), where the timing of neural responses is essential, fICA excels by extracting components that represent their corresponding temporal dynamics. This consideration of time continuity enhances the interpretability and relevance of the extracted components, making fICA a powerful tool for understanding complex data with inherent temporal dependencies.

While there are several fICA methods available, the majority of them merely concentrate on the classification based on the extracted ICs. Mehta and Gray [117] firstly performed ICA on functional observations estimated by fPCA by minimizing entropy. D. Peña et al. [118] defined functional kurtosis operator to identify outliers and cluster structures, but classification, implementation and theoretical properties under mixtures of Gaussian processes rather than a functional version of ICA. Li et al. [119] defined functional Fourth-Order Blind Identification (fFOBI), using an estimation procedure stemmed from the finite Karhunen-Loève (K-L) expansion, orthogonal expansion is optimal in the least-squared error sense. Virta et al. [120] extended two ICA methods, FOBI and JADE, to vector-valued functional data. Lastly, in a more recent work by Vidal et al. [121], they proposed a functional independent component analysis, pFICA, based on the spectral decomposition of the kurtosis operator of a smoothed principal component expansion. However, a critical gap in the existing literature is the lack of

emphasis on evaluating the recovering of the extracted ICs in comparison to the true ICs. Additionally, there is a noticeable absence of discussions regarding the robustness of these methods. The need for a robust fICA method becomes particularly evident when considering signal preprocessing. In practical scenarios, such as the analysis of neuroscience data, the presence of noise is virtually unavoidable, e.g., sample outliers, bump outliers, peak outliers, etc. In this study, we focus on sample outliers, that are individual EEG recordings that exhibit patterns significantly different from the expected or other trials, potentially indicating the presence of noise, artifacts, or rare events in neural activity. These noise signals can significantly impact the subsequent steps of signal analysis.

In this study, we propose a novel and robust functional Independent Component Analysis method, named rFICA. Unlike current approaches solely focused on IC classification, rFICA addresses the multifaceted challenges of accurately assessing the extracted ICs concerning the ground truth ICs while maintaining resilience to noise interference. To evaluate the robustness of rFICA, we conduct a comparative analysis, contrasting the performance of classical fICA and rFICA when applied to mixed signals subjected to varying levels of contamination. This assessment incorporates a representative type of outliers—sample outliers—to simulate real-world scenarios encountered in datasets. Additionally, we extend the evaluation to real scalp EEG signals by comparing the proposed robust fICA method (i.e., rFICA) with existing fICA techniques. Notably, rFICA not only showcases the robustness but also demonstrates superior performance, yielding higher classification rates based on the extracted ICs. This groundbreaking development highlights the potential of rFICA to improve the quality of EEG signal analysis and interpretation, paving the way for more precise and efficient advances in neuroscience applications.

The paper's contributions can be summarized as follows:

1. **Development of robust functional Independent Component Analysis (rFICA):** We introduce a novel and robust method for Functional Independent Component Analysis (i.e., rFICA). Our approach incorporates Kendall's function and ROBPCA into the framework of fICA, enhancing its robustness and applicability.
2. **Algorithm for extracting functional ICs from contaminated mixed signals:** We present a comprehensive algorithm designed to efficiently extract functional Independent Components (ICs) from datasets containing a mixture of various data types. This algorithm facilitates the extraction of meaningful insights from complex datasets.

3. **Robustness validation through simulated and real data:** We validate the robustness of rFICA by applying it to both simulated datasets with sample outliers and real EEG datasets. This empirical evaluation demonstrates the effectiveness and reliability of our method in real-world scenarios.

The structure of Chapter 4 is organized as follows. We briefly introduce fICA method and provide a literature review of existing fICA methods in Section 4.1. In Section 4.2, we present the rFICA algorithm. Section 4.3 features a simulation study that assesses the performance of our proposed rFICA method. In Section 4.4, we apply this method to analyze a real-world dataset, demonstrating its practical applicability. In Section 4.5, we delve into discussions by summarizing our key findings, exploring their implications, and proposing potential directions for future research. Finally, in Section 4.6, we provide the concluding remarks for this chapter.

4.2 Proposed Algorithm

4.2.1 Algorithm Development

ICA methods typically involve a crucial step of standardizing observed data by applying the inverse square root of the covariance matrix. This standardization process effectively eliminates linear dependencies and normalizes variance across different dimensions. However, when dealing with infinite-dimensional spaces, a challenge arises due to the non-invertibility of covariance operators, resulting in an ill-posed problem.

To tackle this challenge, we refer to the definition of independence within the functional domain, as introduced by Gutch and Theis [122]. This definition states that a functional random variable demonstrates independent components when the coordinates, derived after projection onto a specified orthonormal basis, behave as independent variables. Thus, the aim of fICA is to find a linear operator Γ , such that for a truncated orthonormal basis $\{\phi_1(t), \dots, \phi_N(t)\} \in L^2(T)$, the variables $\langle \Gamma X, \phi_j \rangle$ are mutually independent.

Suppose that we are given a dataset X , comprising p signals, each with a signal length of T . Firstly, signals are expressed as linear combination of N basis functions $X(t) = A\phi(t)$, where $A = (a_{ij}) \in \mathbb{R}^{p \times N}$, $\phi(t) = (\phi_1(t), \dots, \phi_N(t))^T$. Then, following the ICA pre-processing steps, we standardize the approximated curves by defining the whitening operator as

$$\Psi\{X(t)\} = \tilde{X}(t) = A_W \phi(t) \quad (4.1)$$

where $A_W = AW$, $W = U\Lambda^{-1}U^T$, $\Sigma_A = \frac{1}{p}(AA^T) = U^T\Lambda$, Λ is diagonal matrix that contains eigenvalues of A , U contains eigenvectors corresponding to the eigenvalues of A .

Next, we employ robust fPCA approaches inspired by ROBPCA [123] and Kendall's τ function [50] to estimate sample curves in terms of robust functional principal components based on sample covariance Mercer decomposition. Let us consider the B-spline basis expansion of the eigenfunctions of covariance of $\tilde{X}(t)$, $\gamma(t) = \phi(t)^T b$, where $b = (b_1, \dots, b_N)^T$. Let $G = (\langle \phi_i, \phi_{i'} \rangle)$, $i, i' = 1, \dots, N$ denotes the covariance of basis functions. Since B-spline basis functions are non-orthonormal with respect to the usual L^2 geometry, we can apply Cholesky factorization of the form $G = LL^T$ to find a non-singular matrix that allows us to operate in terms of the B-spline geometrical structure induced into \mathbb{R}^N . Then, finding the coefficients of $(b_1, \dots, b_N)^T$ corresponds to solve the eigenvalue problem

$$L^{-1}G\Sigma_{A_W}(L^{-1}G)^T e_j = \lambda_j e_j \quad (4.2)$$

where $e_j = L^T b_j$, leading to a set of orthonormal functions. The j^{th} PC is then given by $z_j = AG(L^{-1})^T e_j$, and the problem is reduced to the multivariate PCA of the matrix $C = A_W G(L^{-1})^T$ in \mathbb{R}^N (More detailed information can be found in [121]). In this study, we utilize two robust PCA algorithms that are particularly effective in handling sample outliers. The ROBPCA method combines ideas of both projection pursuit and robust covariance estimation based on Minimum Covariance Determinant (MCD) method. Based on the ROBPCA algorithm, we introduce specific modifications to effectively handle sample outliers. Specifically, we apply the ROBPCA method to the matrix C for the case of sample outliers. In addition to utilizing the ROBPCA method, we also incorporate Kendall's τ function [50] into our approach to construct functional PCs. This integration leverages rank information and draws inspiration from both the Kendall's τ correlation coefficient and the spatial sign covariance function proposed by Gervini [49], enhancing the robustness of our methods (More detailed information can be found from Zhong et al. [50]).

After estimating the coefficients $(b_1, \dots, b_K)^T$ and principal component scores Z , let's proceed by assuming that the principal component is truncated at the q -term, i.e., the number of functional ICs. Then, the column vector of sample curves is given by $\tilde{X}^q(t) = Z^q \beta(t)$, where $Z^q \in \mathbb{R}^{p \times q}$ is the matrix that contains first q PC scores with respect to the basis of principal component functions $\beta(t) = (\beta_1(t), \dots, \beta_q(t))^T$. Let $D_{Z^q} = \text{diag}(Z^q Z^{qT})$, the kurtosis operator of $\tilde{X}^q(t)$ can be expressed as

$$\mathcal{K}_{\tilde{X}^q}(\beta(t)^T h) = \frac{1}{p} (Z^{qT} D_{Z^q} Z^q h)^T \beta(t) \quad (4.3)$$

The eigen-analysis of this kurtosis operator leads to the diagonalization of the kurtosis matrix of the principal components Z^q

$$\Sigma_{4,Z^q} v_j = u_j v_j, j = 1, \dots, q \quad (4.4)$$

where $\Sigma_{4,Z^q} \in \mathbb{R}^{q \times q}$ is defined as $\Sigma_{4,Z^q} = \frac{1}{p} \sum_{i=1}^p \|z_i^q\|^2 z_i^q z_i^{qT} = \frac{1}{p} Z^{qT} D_{Z^q} Z^q$.

Thus, the robust function ICA of $X(t) \in L^2(T)$ can be obtained from the classical ICA of $Z^q \in \mathbb{R}^q$. Then, the extracted independent components can be expressed as $C_j = Z^q v_j$, and the operator Γ is

$$\Gamma(\tilde{X}^q) = \beta^T U^T \Sigma_{Z^q}^{-1/2} z_i^q \quad (4.5)$$

where $U \in \mathbb{R}^{q \times q}$ is the matrix of eigenvectors of the kurtosis matrix Σ_{4,Z^q} .

The algorithm of rFICA is summarized as below:

Algorithm rFICA

Input: Observed signals \mathbf{X} , number of ICs q , robust term $\mathbf{R}(\mathbf{rob}, \tau)$

Output: Extracted ICs $\hat{\mathbf{S}}$

Step 1: Functional data estimation $\mathbf{X}(t) = \mathbf{A}\phi(t)$

Step 2: Functional ICA pre-processing $\Psi\{\mathbf{X}(t)\} = \tilde{\mathbf{X}}(t) = \mathbf{A}_w \phi(t)$

Step 3: Construct robust fPCA $\tilde{\mathbf{X}}^q(t) = \mathbf{Z}^q \beta(t)$

If $R == \mathbf{rob}$ then perform fPCA with ROBPCA on $\mathbf{C} = \mathbf{A}_w \mathbf{G}(\mathbf{L}^{-1})^T$

Else if $R == \tau$ then perform fPCA with Kendall's τ function on $\tilde{\mathbf{X}}(\mathbf{t})$

Step 4: Eigen-analysis of kurtosis operator $\Sigma_{4, \mathbf{Z}^q} \mathbf{v}_j = \mathbf{u}_j \mathbf{v}_j, j = 1, \dots, q$

Step 5: Independent components extraction $\mathcal{C}_j = \mathbf{Z}^q \mathbf{v}_j, j = 1, \dots, q$

Return $\hat{\mathbf{S}} = \Gamma(\tilde{\mathbf{X}}^q) = \boldsymbol{\beta}^T \mathbf{U}^T \Sigma_{\mathbf{Z}^q}^{-1/2} \mathbf{z}_i^q$

4.2.2 Validation

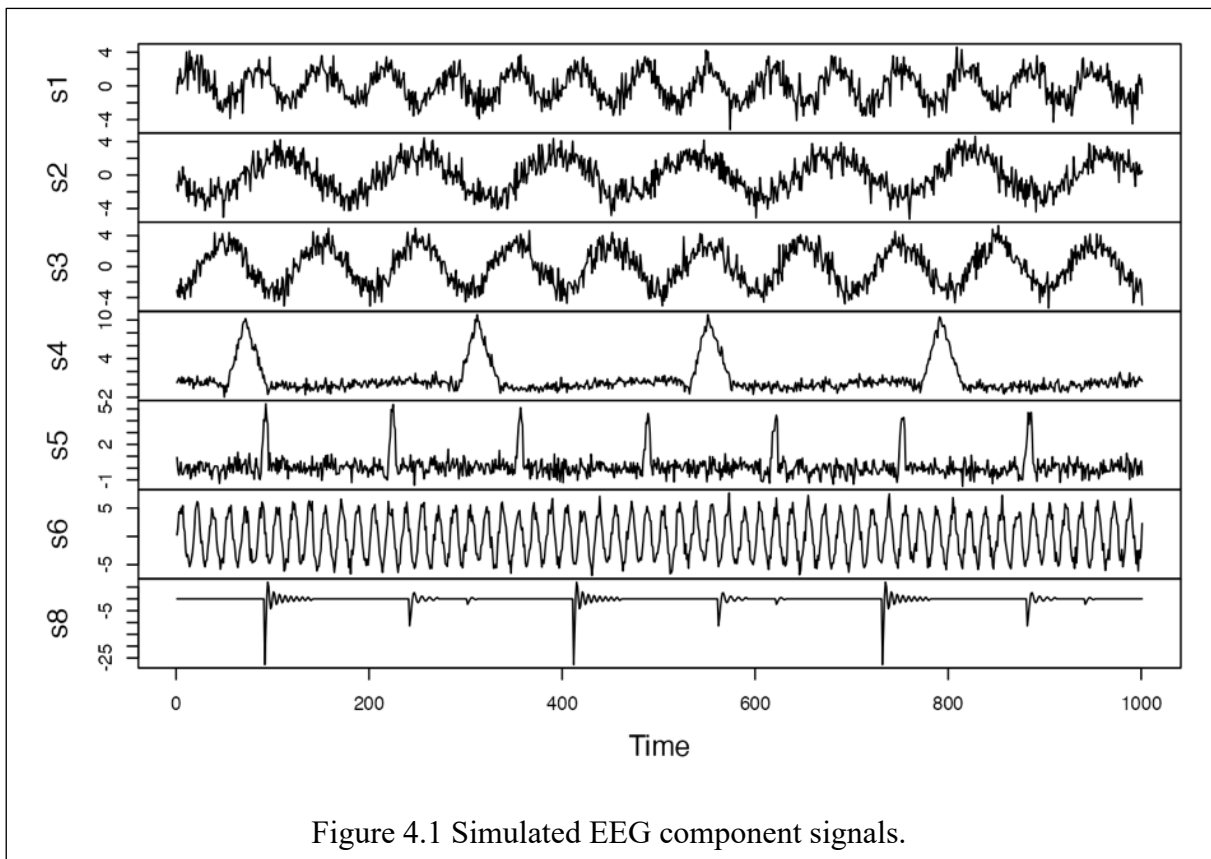
The ROBPCA method, as introduced by Hubert et al. [123], seamlessly integrates the principles of projection pursuit and robust covariance estimation, primarily utilizing the MCD method, which is rooted in the pursuit of an h -subset with the smallest determinant for the classical covariance matrix (More detailed information can be found by Hubert et al. [123], [124]. Specifically, the key step of ROBPCA revolves around identifying $h (< p)$ data points that demonstrate the least outlying characteristics. Typically, h is set to approximately $0.75n$, or it can be specified by the user. Then, we robustly estimate the covariance matrix for the mean-centered matrix obtained in the second step, by employing the MCD estimator. Similar to classical PCA method, ROBPCA also assumes that the data is approximately normally distributed. For case of sample outliers, we perform ROBPCA on the matrix C , to detect the data points that are far away from others and remove it.

The robustness of Kendall's τ function stems from its utilization of rank information. The key point of this method is the construction of the Kendall's τ covariance, which excels in handling sample outliers owing to its reliance on rank-based statistics. Specifically, Kendall's τ serves as a non-parametric measure of association, assessing the correspondence between the ranks of two variables without assuming any specific data distribution. This attribute renders it less sensitive to extreme values and outliers, particularly when contrasted with parametric methods tied to specific data distributions. Additionally, Kendall's τ considers tied values in the data, where outliers typically influence only a limited number of tied data points, thereby reducing their overall impact on the rank correlation. Therefore, the integration of Kendall's τ function into fPCA offers an effective approach for addressing sample outliers, especially when they exhibit similar patterns.

4.3 Simulation

4.3.1 Simulated Data Generation

Aligned with the ground truth of EEG components, source signals are generated with a sampling frequency of 1000 Hz and an epoch duration of 1 second. This simulation yields a total of $p=30$ EEG signals, each crafted through a linear combination of $q=7$ distinct source signals. These source signals encompass various components, including brain rhythms at 7, 10, and 15 Hz, an eye-related component at 4 Hz, a heart-related component at 1 Hz, a component attributed to line noise at 60 Hz, and a muscle-related component (Figure 4.1).



In order to examine the robustness of proposed rFICA method, we introduce additional contaminations into simulated EEG signals, i.e., sample outliers (Figure 4.2). Sample outliers are generated based on gaussian noise with varying parameter setting. Detailed information about the contamination data can be found in Table 4.1 below.

Sample Outliers	Magnitude	Proportion
Contamination Level	1, 5, 10, 20	3%, 14%, 25%, 40%, 50%

Table 4.1 Contamination Parameter setting.

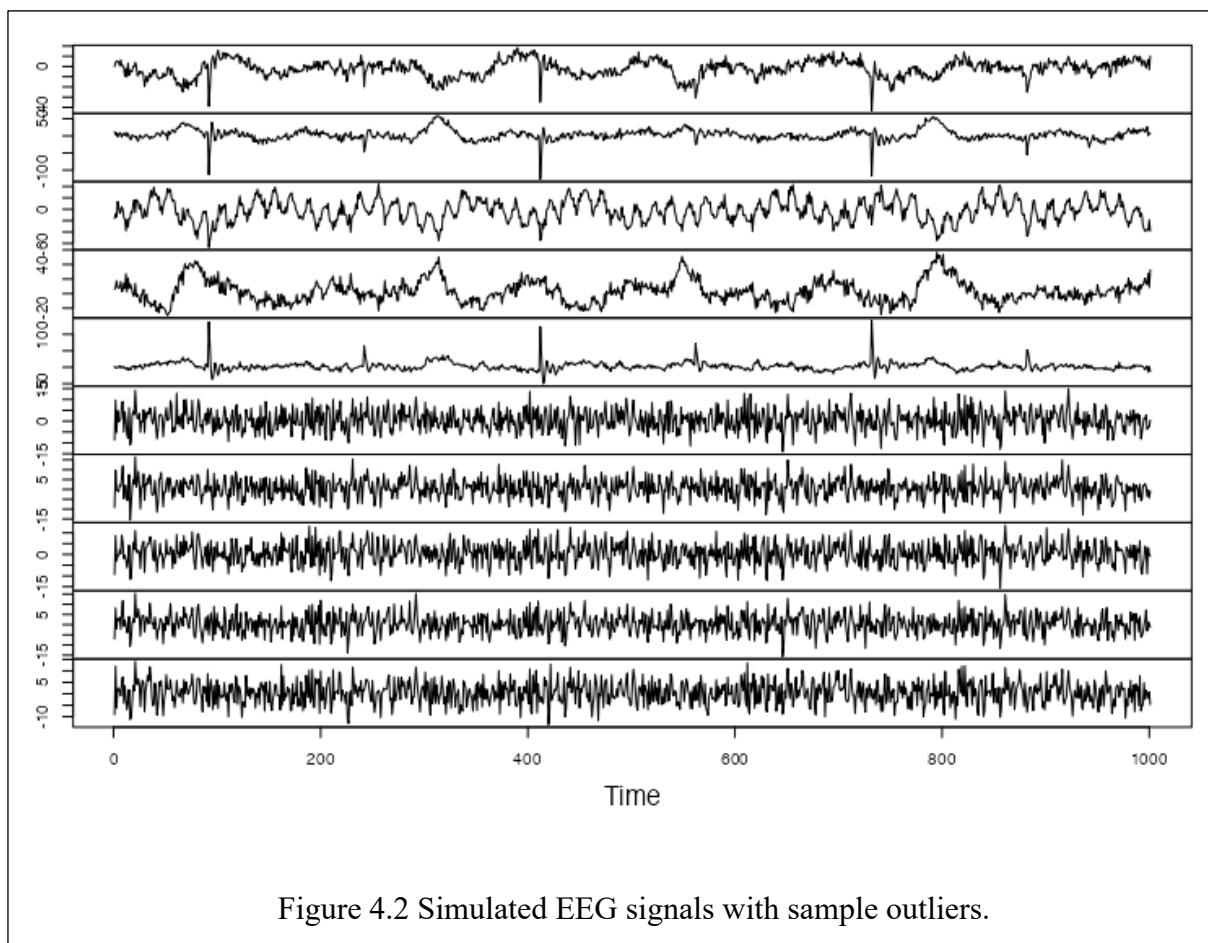


Figure 4.2 Simulated EEG signals with sample outliers.

4.3.2 Simulation Performance Metrics

The ideal extracted ICs are expected to encompass both temporal and frequency information. Consequently, the performance of fICA methods will be evaluated in both the time and frequency domains. In the time domain, the Cross-Correlation Function (CCF) is used to measure the similarity between signals. Let s_{jt} and \hat{s}_{it} denotes the t^{th} point of i^{th} true and estimated signal, respectively.

$$CCF_{ij} = \sum_{t=1}^T (\hat{s}_{i(t+\tau)} \cdot \text{conj}(s_{jt}), i, j = 1, \dots, p \quad (4.6)$$

Following parameter tuning experiments, the delay parameter, τ , is set as 10 in this study, allowing for up to a 10-time-point delay between two signals.

To assess the frequency information of the extracted ICs, we employ the Fast Fourier Transformation (FFT) on extracted ICs and measure the similarity by Cosine Similarity (COS). Let f_j and \hat{f}_i denotes the spectrum vector of i^{th} true and estimated signal, respectively.

$$COS_{ij} = \frac{\hat{f}_i \cdot f_j}{\|\hat{f}_i\| \cdot \|f_j\|}, i, j = 1, \dots, p \quad (4.7)$$

While the Mean Square Error (MSE) is commonly employed as a metric, it has a limitation in that it does not consider the possibility of time shifts between true and estimated ICs, which can occasionally be unavoidable. Besides that, another option for capturing the frequency information of signals is the Euclidean Distance (EUC). Nevertheless, the COS, which falls within a range of -1 to 1, is generally preferred over EUC. This preference arises from COS's insensitivity to the magnitude of frequency, rendering it a more robust choice.

4.3.3 Simulation result analysis

Our analysis involves the comparison of two non-robust fICA techniques (i.e., pFICA, fFOBI) with newly proposed robust fICA method that incorporates three robust fPCA methods tailored to data with varying contamination levels. Our expectation is that under uncontaminated conditions, all fICA methods will yield similar results. Furthermore, we hypothesize that the robust fICA method will excel over other methods when applied to datasets containing contamination.

We introduce sample outliers as additional mixed signals, simulating them using Gaussian signals. To thoroughly evaluate the robustness of the proposed method, we systematically vary the parameters associated with these sample outliers. Specifically, we adjust the magnitude of

sample outliers, ranging from 1 to 20. Additionally, we explore different scenarios by varying the number of sample outliers from 1 to 30. The result of each case can be found below (Figures 4.3-4.4, Tables 4.2-4.5).

Case 1: Varying magnitude of sample outliers

From Figure 4.3 and Tables 4.2-4.3, we observe that the proposed rFICA that utilizes Kendall's τ function and ROBPCA outperforms than others when mixed signals contain sample outliers, regardless of the magnitude of outliers.

CCF	fFOBI	pFICA	rFICA (τ)	rFICA (rob)
X	0.7400	0.7269	0.7298	0.7202
Xs(1)	0.6439	0.5499	0.6883	0.6514
Xs(5)	0.6428	0.5499	0.6839	0.6505
Xs(10)	0.6422	0.5495	0.6833	0.6491
Xs(20)	0.6411	0.5490	0.6825	0.6469

Table 4.2 Performance of fICA methods based on CCF metric for varying magnitude of sample outliers (i.e., 1, 5, 10, 20).

COS	fFOBI	pFICA	rFICA (τ)	rFICA (rob)
X	0.7851	0.7649	0.7766	0.8046
Xs(1)	0.7321	0.6251	0.7541	0.7658
Xs(5)	0.7328	0.6250	0.7495	0.7651
Xs(10)	0.7332	0.6248	0.7497	0.7649
Xs(20)	0.7336	0.6245	0.7490	0.7627

Table 4.3 Performance of fICA methods based on COS metric for varying magnitude of sample outliers (i.e., 1, 5, 10, 20).

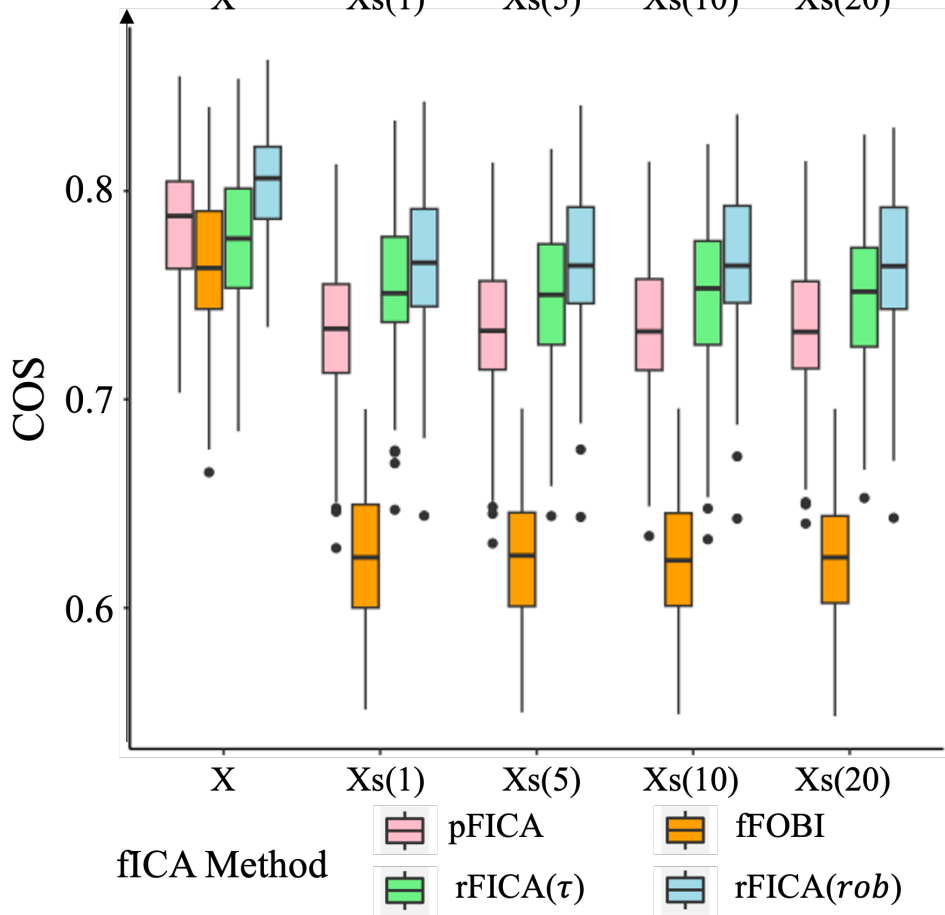
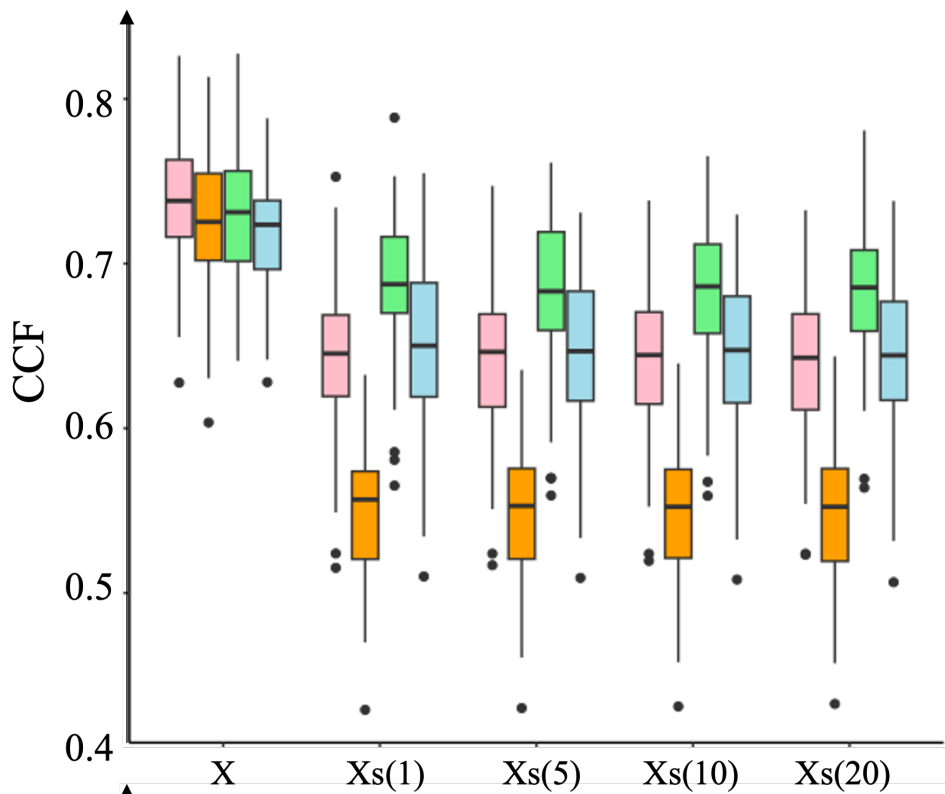


Figure 4.3 Side-by-side boxplot of CCF and COS values for robust and non-robust fICA methods with different with varying magnitude of sample outliers.

Case 2: Varying contamination levels for sample outliers

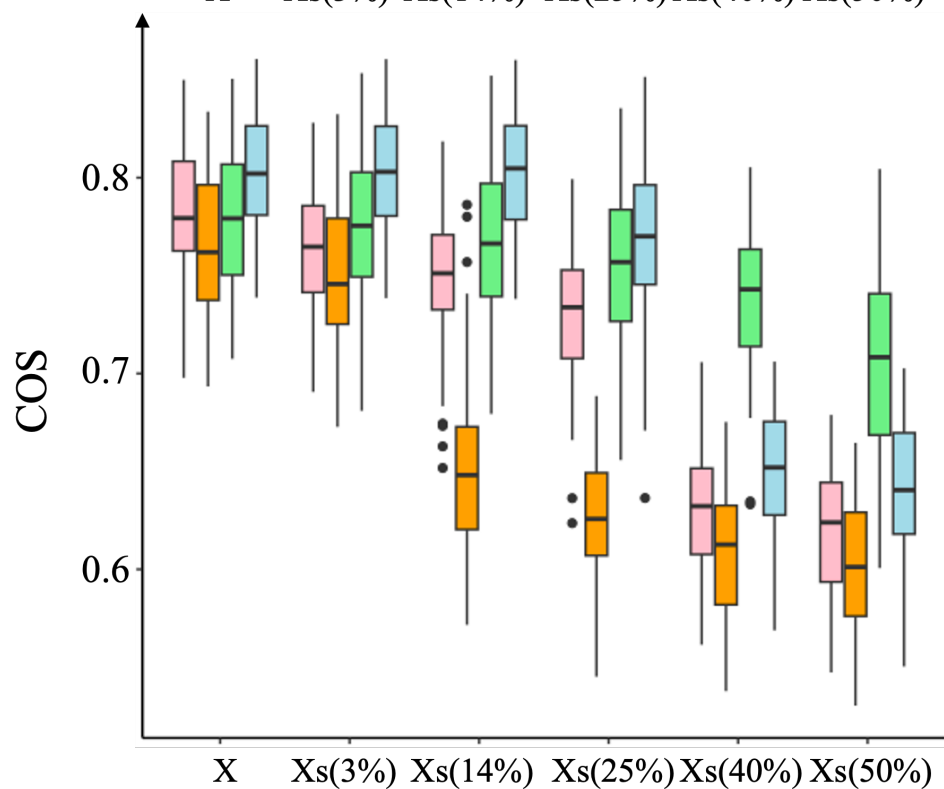
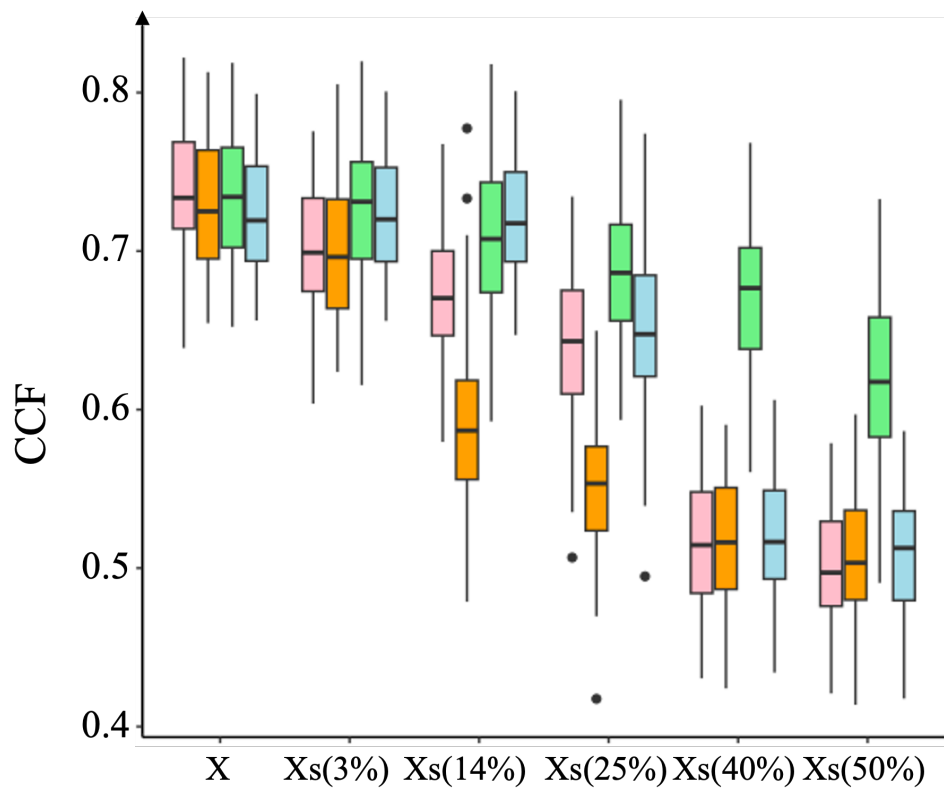
From Figure 4.4 and Tables 4.4-5, we find that both Cross-Correlation Function (CCF) and Cosine Similarity (COS) measurements exhibit a decrease as the number of sample outliers increases. In the presence of mixed signals containing sample outliers, the proposed rFICA method consistently outperforms other methods. Specifically, the rFICA(τ) demonstrates superior performance, irrespective of the number of outliers.

CCF	fFOBI	pFICA	rFICA (τ)	rFICA (rob)
X	0.7373	0.7281	0.7338	0.7215
Xs(3%)	0.6999	0.6995	0.7259	0.7212
Xs(14%)	0.6748	0.5883	0.7081	0.7204
Xs(25%)	0.6399	0.5513	0.6877	0.6495
Xs(40%)	0.5151	0.5167	0.6702	0.5201
Xs(50%)	0.5024	0.5074	0.6196	0.5100

Table 4.4 Performance of fICA methods based on CCF metric for varying contamination levels of sample outliers (i.e., 3%, 14%, 25%, 40%, 50%).

CCF	fFOBI	pFICA	rFICA (τ)	rFICA (rob)
X	0.7816	0.7639	0.7786	0.8042
Xs(3%)	0.7633	0.7492	0.7755	0.8039
Xs(14%)	0.7499	0.6503	0.7665	0.8032
Xs(25%)	0.7313	0.6250	0.7549	0.7671
Xs(40%)	0.6300	0.6085	0.7387	0.6509
Xs(50%)	0.6198	0.6010	0.7022	0.6424

Table 4.5 Performance of fICA methods based on COS metric for varying contamination levels of sample outliers (i.e., 3%, 14%, 25%, 40%, 50%).



fICA Method





 pFICA	 fFOBI
 rFICA(τ)	 rFICA(<i>rob</i>)

Figure 4.4 Side-by-side boxplot of CCF and COS values for robust and non-robust fICA methods with different with varying contamination level of sample outliers.

In summary, irrespective of the magnitude and number of sample outliers, the proposed rFICA method, which incorporates Kendall's τ function and ROBPCA, consistently outperforms other techniques when dealing with mixed signals containing sample outliers. Notably, the rFICA(τ) consistently demonstrates superior performance, whereas the rFICA(rob) excels in extracting more accurate frequency.

4.4 Implementation of the rFICA to Scalp EEG Data

In this section, we apply the proposed rFICA method to a real EEG dataset, assessing its performance through the classification results of the extracted IC scores. Furthermore, we conduct a comparative analysis, contrasting our findings with those obtained using established fICA methods, specifically pFICA and FFOBI.

In the scalp EEG dataset analyzed here, the researcher recruited a total of 19 adults (7 females, 12 males) from the University of Arizona and recorded their scalp EEG signals. Participants navigated in the virtual reality and were asked to monitor the distances travelled: short (100 virtual meters) vs long (200 virtual meters) distances. Each task was repeated 24 trials, and each trial lasted 5.656 seconds. Participants walked freely on an omnidirectional treadmill while wearing a wireless scalp EEG cap. The sampling rate was 500 Hz. Details of the experiment design can be found in [69], [92].

Using the extracted ICs, we conduct random forest classification that incorporated both frequency and IC score information. When comparing the results obtained from non-robust and robust fICA methods, we observe that the classification outcomes from rFICA are significantly better than those from non-robust methods (p-value (IC score) = 0.0117, p-value (Frequency) = 0.0002). In Figure 4.5, we present the highest classification results of 19 individuals achieved by non-robust and robust fICA methods, respectively.

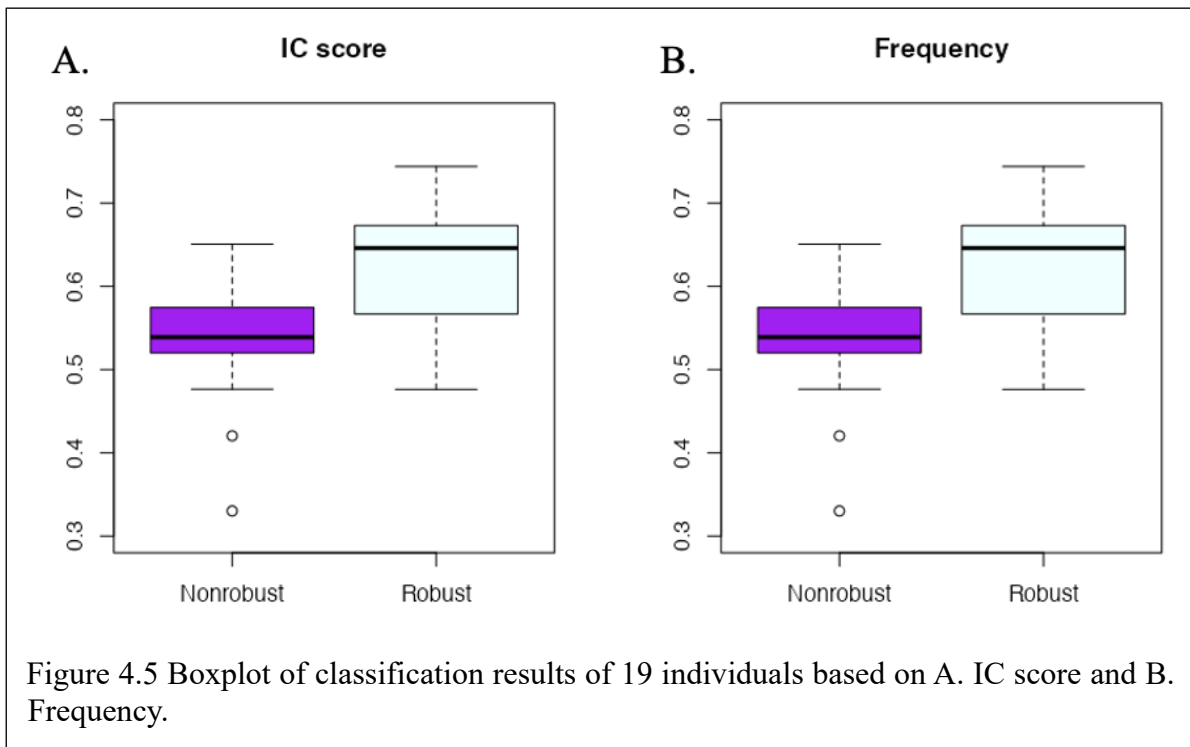
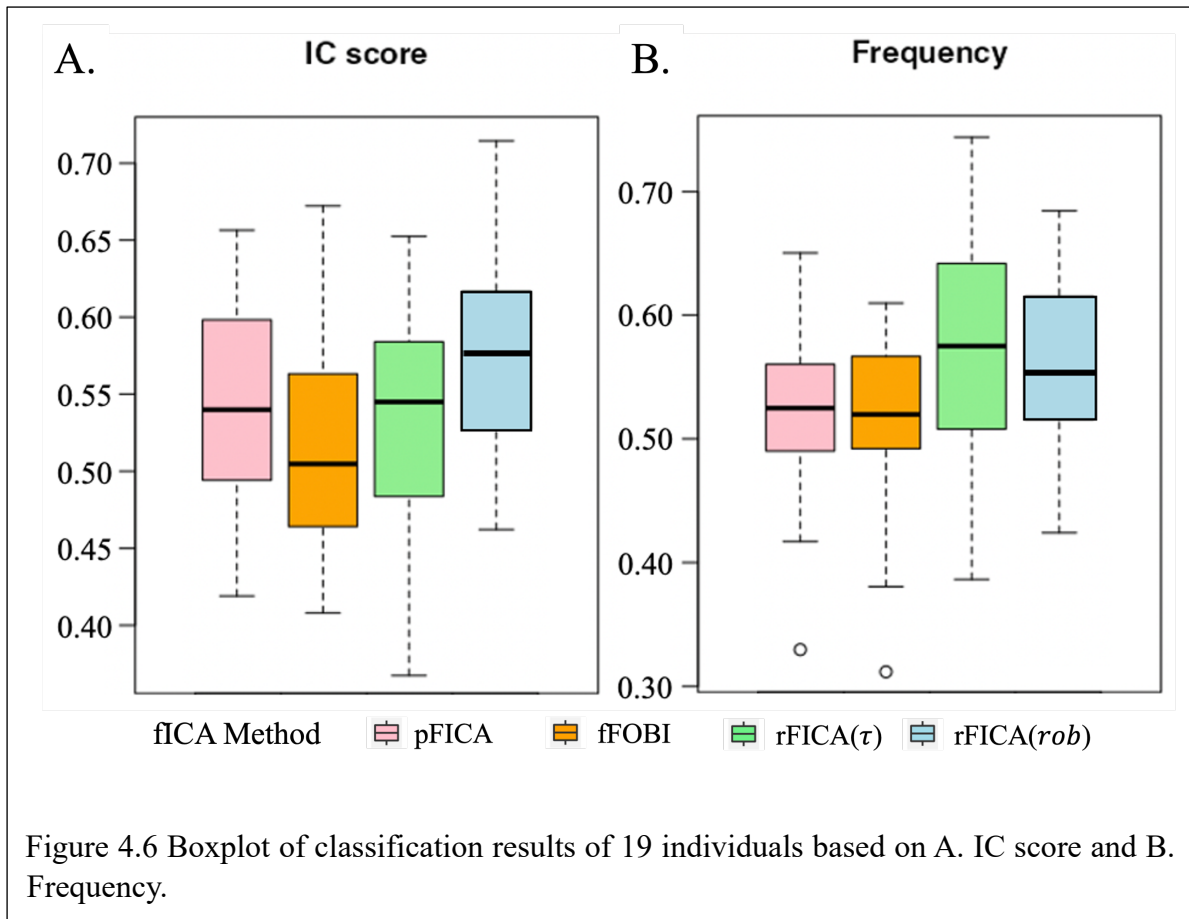


Figure 4.5 Boxplot of classification results of 19 individuals based on A. IC score and B. Frequency.

Furthermore, as illustrated in Figure 4.6, we observe that $\text{rfICA}(\text{rob})$ performs best from perspective of IC score, whereas $\text{rfICA}(\tau)$ performs best from perspective of frequency, which indicates the existence of sample outliers.

In conclusion, our findings confirm that rFICA outperforms in classification, particularly in its capacity to detect both sample outliers. This finding is consistent with the notion that EEG signals often encompass various types of outliers, including mixed outliers that may pose visual identification challenges. Nevertheless, it is essential to recognize that individual subject variations may impact the selection of the most appropriate robust technique for implementing the rFICA method. Consequently, we advise a meticulous assessment of parameter tuning and the adjustment of robust techniques based on the unique signal characteristics inherent to each subject.



4.5 Discussion

The necessity for a robust functional Independent Component Analysis method becomes apparent when considering signal preprocessing. In practical scenarios, the presence of noise is virtually inevitable, and this noise can significantly impact subsequent signal analysis steps. While recent scientific developments have introduced several functional independent component analysis methods, most of them primarily focus on classification based on the extracted ICs. There is a noticeable gap in discussions regarding the robustness of these methods and their ability to accurately recover extracted ICs compared to true ICs. To bridge this gap and address the limitations of existing determination methods, we present a novel and robust fICA method named rFICA that incorporates two types of robust parameters.

The key highlights of the proposed rFICA method are as follows. Firstly, rFICA tackles the complex task of precisely assessing the extracted ICs concerning the ground truth ICs. Secondly, by incorporating ROBPCA and Kendall's τ function, rFICA maintains resilience to

noise interference, particularly in case involving sample outliers, which represent one of common types of outliers encountered in real-world scenarios. To better illustrate the advantages of rFICA, we apply it alongside some existing fICA methods to both simulated mixed signals subjected to varying contamination levels and collected raw EEG signals. This comprehensive evaluation demonstrates the effectiveness of rFICA in addressing the challenges posed by noisy data and highlights its potential in real-world applications, especially in EEG data analysis.

To assess the robustness of rFICA, we initiate our analysis by applying it to mixed signals containing sample outliers, varying in terms of magnitude and quantity. In this scenario, we compare rFICA with pFICA and fFOBI. The results affirm the superior performance of rFICA, particularly the variant that incorporates Kendall's τ function and ROBPCA, when confronted with mixed signals featuring sample outliers. Furthermore, we have implemented the proposed rFICA method on a real EEG dataset and evaluated its performance by analyzing the classification results of the extracted IC scores. We also compare our method with pFICA and fFOBI. The superior classification results demonstrate the robustness of our approach, especially when applied to raw scale EEG data that contains outliers.

While the robustness of our proposed rFICA has been validated through comparisons with other methods in scenarios involving sample outliers, there remain avenues for further exploration. Specifically, our current work has primarily focused on one type of contaminations, but there are other prevalent sources of noise that warrant investigation, such as channel noise. Channel noise encompasses correlated noise that uniformly affects all observations, posing a distinct challenge. Channel noise is not easily distinguishable from the genuine sources. This presents a critical issue since ICA assumes the statistical independence of sources. When sources are mixed with correlated noise like channel noise, it violates this independence assumption, resulting in dependencies between the sources. These dependencies present a significant hurdle for ICA methods in accurately separating the sources.

Hence, our future endeavors should include an exploration of methodologies to address the unique complexities introduced by correlated noise, such as channel noise, in order to further enhance the applicability and robustness of rFICA.

4.6 Conclusion

Many of the current functional ICA methods do not prioritize the recovery of extracted ICs. Hence, there is a significant need and practical value in developing a robust functional independent component analysis method. In this research, we introduce rFICA, which leverages ROBPCA and Kendall's τ function to extract ICs from mixed signals, even when they are contaminated, taking into account the temporal continuity of the signals. Through our experiments on simulated data and real raw EEG signals, we can confidently conclude that rFICA is a robust algorithm for the recovery of independent components. It exhibits remarkable resilience to contaminations, making it a valuable tool in the field of functional data analysis.

Chapter 5

General Conclusion

In this dissertation, we have advanced statistical methodologies for the interpretable classification of neuroscience data by leveraging both time and frequency information, as well as took into account the temporal continuity of signal data. Furthermore, we introduced an efficient approach for the automated identification of the optimal number of independent components, applicable across multiple ICA methods. Additionally, we have developed a resilient functional independent component analysis method that remains robust in the presence of outliers. This chapter offers a concise summary of these three projects and outlines potential avenues for future research endeavors.

5.1 Summary

In Chapter 2, we introduce an innovative three-stage algorithm that offers a structured approach to the analysis of EEG data using Functional Data Analysis (FDA) techniques. Our methodology starts by applying FDA methods to EEG signals transformed into the time-frequency domain, enabling the extraction of rich information from the EEG signals. While this transformation enhances information capture, it also raises the data's dimensionality, making direct application of FDA techniques impractical. To address this challenge, we introduce the second stage of our algorithm, incorporating functional testing methods for initial dimension reduction, which involves feature selection. This ensures that the number of features used for modeling remains manageable within the context of functional modeling techniques. Through the application of this algorithm to Scalp EEG data, we demonstrate its effectiveness in achieving interpretable classification. Our findings indicate that frontal delta-theta and parietal delta-theta oscillations are more strongly associated with distance judgment compared to other oscillatory components.

In Chapter 3, we introduce the CW_ICA method, which simplifies computation complexity by dividing the data into two blocks and recording the smallest column-wise maximum value.

This method not only offers compatibility with multiple ICA techniques but also capitalizes on the robustness of the Spearman correlation coefficient. It automatically identifies the optimal number of Independent Components (ICs) through a quantitative measurement. We conducted a comparative analysis of CW_ICA against other existing approaches, when used in combination with multiple ICA methods. Our assessments involved simulated data and real raw EEG signals. Based on the performance of these determination methods, we conclude that CW_ICA stands out as a robust algorithm for the automated determination of the number of ICs. It consistently yields results that can be effectively applied across various ICA methods.

In Chapter 4, we introduce rFICA, an innovative method that combines the principles of ROBPCA with the utility of Kendall's τ function to extract functional ICs from mixed signals. What sets rFICA apart is its remarkable ability to perform this extraction even when the signals are contaminated, and it does so while accounting for the temporal continuity of these signals. Through a series of rigorous experiments that encompass simulated data as well as real raw EEG signals, we have gained profound insights into the capabilities of rFICA. It is clear from our findings that rFICA is not just a robust algorithm for the recovery of independent components; it is also highly resilient to contaminations. This resilience makes it an invaluable tool for signal analysis, especially in fields where data quality may be compromised, such as biomedical signal processing and environmental monitoring. The ability of rFICA to work under challenging conditions and still produce meaningful, interpretable results makes it a noteworthy contribution to the field of functional data analysis. Its capacity to uncover hidden patterns and information within complex and noisy data holds great promise for a wide range of applications, ranging from neuroscience and medical diagnostics to environmental science and beyond.

5.2 Future Work

Building upon the implementation of the interpretable classification algorithm outlined in Chapter 1, we have introduced innovative techniques focused on optimizing information extraction during the signal preprocessing stage. This effort is aimed at augmenting the efficacy of neuroscience data analysis. However, it's imperative to recognize that our current research primarily emphasizes the temporal and frequency domains, inadvertently sidelining the substantial spatial dimension. As we look to the future, our research endeavors will venture

into a promising domain: the development of three-dimensional functional data analysis and modeling. This comprehensive framework will span the dimensions of time, space, and frequency, affording us the capability to capture and account for the intricate spatial correlations among electrodes within the data. This holistic vantage point is expected to significantly enhance the precision and depth of our analyses, ushering in novel insights within the field of neuroscience.

Furthermore, Chapter 3 raises important practical considerations that warrant further exploration. One such issue pertains to the diminishing distinction in measurement between models with consecutive numbers of Independent Components (ICs) within CW_ICA, particularly as the dimensions of the signal matrix expand, either in length or by an increase in the number of mixed signals. Moreover, there is an observable uptick in computational complexity, in tandem with the growth of mixed signals and signal length, as it is directly measured by the correlation matrix. To address these challenges, we will propose an extension of functional CW_ICA. This extension aims to mitigate data dimensionality while simultaneously enhancing the robustness and clarity of the results. By doing so, we intend to overcome the limitations associated with the increasing data size and pave the way for more effective and efficient signal analysis.

Our investigation into robust functional independent component analysis in Chapter 4 has not only yielded valuable insights but also uncovered new research avenues. Our current focus has been primarily on one type of contaminations, yet there are other prevalent sources of noise that merit exploration, with channel noise being a notable example. Channel noise, characterized by correlated noise that uniformly impacts all observations, presents a distinctive challenge. Therefore, our future research should encompass the development of methodologies designed to tackle the specific complexities introduced by correlated noise, particularly channel noise. This pursuit aims to broaden the scope of applicability and enhance the overall robustness of rFICA.

References

- [1] S. Siuly, Y. Li, and Y. Zhang, “Electroencephalogram (EEG) and Its Background,” in *EEG Signal Analysis and Classification: Techniques and Applications*, S. Siuly, Y. Li, and Y. Zhang, Eds., in Health Information Science. , Cham: Springer International Publishing, 2016, pp. 3–21. doi: 10.1007/978-3-319-47653-7_1.
- [2] “What is an Electroencephalogram (EEG) Test?” Accessed: Oct. 27, 2023. [Online]. Available: <https://www.simplypsychology.org/what-is-an-eeg.html>
- [3] Y. Petrov, “Harmony: EEG/MEG Linear Inverse Source Reconstruction in the Anatomical Basis of Spherical Harmonics,” *PLOS ONE*, vol. 7, no. 10, p. e44439, Oct. 2012, doi: 10.1371/journal.pone.0044439.
- [4] A. Campbell *et al.*, “NeuroPhone: brain-mobile phone interface using a wireless EEG headset,” in *MobiHeld '10*, 2010. doi: 10.1145/1851322.1851326.
- [5] H. Maki, T. Toda, S. Sakti, G. Neubig, and S. Nakamura, “EEG signal enhancement using multi-channel wiener filter with a spatial correlation prior,” in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2015, pp. 2639–2643. doi: 10.1109/ICASSP.2015.7178449.
- [6] S. Jirayucharoensak and P. Israsena, “Automatic removal of EEG artifacts using ICA and Lifting Wavelet Transform,” in *2013 International Computer Science and Engineering Conference (ICSEC)*, Sep. 2013, pp. 136–139. doi: 10.1109/ICSEC.2013.6694767.
- [7] C. Vidaurre, N. Krämer, B. Blankertz, and A. Schlögl, “Time Domain Parameters as a feature for EEG-based Brain–Computer Interfaces,” *Neural Networks*, vol. 22, no. 9, pp. 1313–1319, Nov. 2009, doi: 10.1016/j.neunet.2009.07.020.
- [8] J. S. Richman and J. R. Moorman, “Physiological time-series analysis using approximate entropy and sample entropy,” *Am J Physiol Heart Circ Physiol*, vol. 278, no. 6, pp. H2039–2049, Jun. 2000, doi: 10.1152/ajpheart.2000.278.6.H2039.
- [9] B. Qian and K. Rasheed, “HURST EXPONENT AND FINANCIAL MARKET PREDICTABILITY”.
- [10] J.-H. Kang, C. H. Lee, and S.-P. Kim, “EEG feature selection and the use of Lyapunov exponents for EEG-based biometrics,” in *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, Feb. 2016, pp. 228–231. doi: 10.1109/BHI.2016.7455876.
- [11] S. Chatterjee, S. Pratiher, and R. Bose, “Multifractal detrended fluctuation analysis based novel feature extraction technique for automated detection of focal and non-focal electroencephalogram signals,” *IET Science, Measurement & Technology*, vol. 11, no. 8, pp. 1014–1021, Jul. 2017, doi: 10.1049/iet-smt.2017.0117.
- [12] “Genetic Algorithms in EEG Feature Selection for the Classification of Movements of the Left and Right Hand | SpringerLink.” Accessed: Oct. 10, 2023. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-00969-8_57
- [13] “A Primer on the Discrete Fourier Transform: American Journal of EEG Technology: Vol 34, No 4.” Accessed: Oct. 10, 2023. [Online]. Available: <https://www.tandfonline.com/doi/abs/10.1080/00029238.1994.11080490>
- [14] “EEG Dataset Reduction and Feature Extraction Using Discrete Cosine Transform | IEEE Conference Publication | IEEE Xplore.” Accessed: Oct. 10, 2023. [Online]. Available: <https://ieeexplore.ieee.org/document/6410152>
- [15] T. Ouyang and H.-T. Lu, “Vigilance Analysis Based on Continuous Wavelet Transform of EEG Signals,” in *2010 International Conference on Biomedical Engineering and Computer Science*, Apr. 2010, pp. 1–4. doi: 10.1109/ICBECS.2010.5462289.

- [16] “Feature Extraction and Classification of ECG Signal Using Neuro-Wavelet Approach - ProQuest.” Accessed: Oct. 10, 2023. [Online]. Available: <https://www.proquest.com/openview/e7a8b612a1dfe8123f10ea31fed5ab0f/1?pq-origsite=gscholar&cbl=2037360>
- [17] “(PDF) Statistical Wavelet Features, PCA, and SVM Based Approach for EEG Signals Classification.” Accessed: Oct. 10, 2023. [Online]. Available: https://www.researchgate.net/publication/284593790_Statistical_Wavelet_Features_PCA_and_SVM_Based_Approach_for_EEG_Signals_Classification
- [18] “Classification of EEG signals using time domain features | IEEE Conference Publication | IEEE Xplore.” Accessed: Oct. 10, 2023. [Online]. Available: <https://ieeexplore.ieee.org/document/7130354>
- [19] J. Wang, Z. Feng, and N. Lu, “Feature extraction by common spatial pattern in frequency domain for motor imagery tasks classification,” in *2017 29th Chinese Control And Decision Conference (CCDC)*, May 2017, pp. 5883–5888. doi: 10.1109/CCDC.2017.7978220.
- [20] “Comparative Study of Feature Subset Selection Methods for Dimensionality Reduction on Scientific Data | IEEE Conference Publication | IEEE Xplore.” Accessed: Oct. 10, 2023. [Online]. Available: <https://ieeexplore.ieee.org/document/7544805>
- [21] “A Comparison of Three Classification Algorithms for Handwritten Digit Recognition | IEEE Conference Publication | IEEE Xplore.” Accessed: Oct. 10, 2023. [Online]. Available: <https://ieeexplore.ieee.org/document/8723702>
- [22] A. S. Eesa, Z. Orman, and A. M. A. Brifcani, “A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems,” *Expert Systems with Applications*, vol. 42, no. 5, pp. 2670–2679, Apr. 2015, doi: 10.1016/j.eswa.2014.11.009.
- [23] M. S. A. Daweri, S. Abdullah, and K. A. Z. Ariffin, “A Migration-Based Cuttlefish Algorithm With Short-Term Memory for Optimization Problems,” *IEEE Access*, vol. 8, pp. 70270–70292, 2020, doi: 10.1109/ACCESS.2020.2986509.
- [24] “Overview and comparative study of dimensionality reduction techniques for high dimensional data - ScienceDirect.” Accessed: Oct. 10, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S156625351930377X?casa_token=tkqn5Dg3Y9IAAAAA:s880YP-cpaa5xlpV_A5_6MIBGRXk8dQb2NiZ0n4NC2abZ_LlWzs9OzZEGxYLkjsIs01Q8SHL3w
- [25] “A Review of Dimensionality Reduction Techniques for Efficient Computation - ScienceDirect.” Accessed: Oct. 10, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050920300879>
- [26] “(PDF) A new feature selection model based on ID3 and bees algorithm for intrusion detection system.” Accessed: Oct. 10, 2023. [Online]. Available: https://www.researchgate.net/publication/276371477_A_new_feature_selection_model_based_on_ID3_and_bees_algorithm_for_intrusion_detection_system
- [27] U. M. Khaire and R. Dhanalakshmi, “Stability of feature selection algorithm: A review,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 4, pp. 1060–1073, Apr. 2022, doi: 10.1016/j.jksuci.2019.06.012.
- [28] S. Visalakshi and V. Radha, “A literature review of feature selection techniques and applications: Review of feature selection in data mining,” in *2014 IEEE International Conference on Computational Intelligence and Computing Research*, 2014, pp. 1–6. doi: 10.1109/ICCIC.2014.7238499.

- [29] M. Verleysen and D. François, “The Curse of Dimensionality in Data Mining and Time Series Prediction,” in *Computational Intelligence and Bioinspired Systems*, J. Cabestany, A. Prieto, and F. Sandoval, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2005, pp. 758–770. doi: 10.1007/11494669_93.
- [30] Y. Leung and Y. Hung, “A multiple-filter-multiple-wrapper approach to gene selection and microarray data classification,” *IEEE/ACM Trans Comput Biol Bioinform*, vol. 7, no. 1, pp. 108–117, 2010, doi: 10.1109/TCBB.2008.46.
- [31] C. Lazar *et al.*, “A survey on filter techniques for feature selection in gene expression microarray analysis,” *IEEE/ACM Trans Comput Biol Bioinform*, vol. 9, no. 4, pp. 1106–1119, 2012, doi: 10.1109/TCBB.2012.33.
- [32] M. R. Mahmood and A. M. Abdulazeez, “A Comparative Study of a New Hand Recognition Model Based on Line of Features and Other Techniques,” in *Recent Trends in Information and Communication Technology*, F. Saeed, N. Gazem, S. Patnaik, A. S. Saed Balaid, and F. Mohammed, Eds., in Lecture Notes on Data Engineering and Communications Technologies. Cham: Springer International Publishing, 2018, pp. 420–432. doi: 10.1007/978-3-319-59427-9_45.
- [33] “Classification of EEG Data using k-Nearest Neighbor approach for Concealed Information Test | Request PDF.” Accessed: Oct. 10, 2023. [Online]. Available: https://www.researchgate.net/publication/329051407_Classification_of_EEG_Data_using_k-Nearest_Neighbor_approach_for_Concealed_Information_Test
- [34] E. Neto, F. Biessmann, H. Aurlien, H. Nordby, and T. Eichele, “Regularized Linear Discriminant Analysis of EEG Features in Dementia Patients,” *Front Aging Neurosci*, vol. 8, p. 273, 2016, doi: 10.3389/fnagi.2016.00273.
- [35] M. Arvaneh, C. Guan, K. K. Ang, and H. C. Quek, “EEG Channel Selection using Decision Tree in Brain-Computer Interface”.
- [36] “A study of the Naive Bayes classifier for analyzing imaginary movement EEG signals using the Periodogram as spectral estimator | IEEE Conference Publication | IEEE Xplore.” Accessed: Oct. 10, 2023. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/6487514>
- [37] J. Ramsay, “Functional Data Analysis,” in *Encyclopedia of Statistics in Behavioral Science*, American Cancer Society, 2005. doi: <https://doi.org/10.1002/0470013192.bsa239>.
- [38] P. Craven and G. Wahba, “Smoothing noisy data with spline functions,” *Numer. Math.*, vol. 31, no. 4, pp. 377–403, Dec. 1978, doi: 10.1007/BF01404567.
- [39] *Nonparametric Functional Data Analysis*. in Springer Series in Statistics. Springer New York, 2006. doi: 10.1007/0-387-36620-2.
- [40] J. A. Rice and C. O. Wu, “Nonparametric mixed effects models for unequally sampled noisy curves,” *Biometrics*, vol. 57, no. 1, pp. 253–259, Mar. 2001, doi: 10.1111/j.0006-341x.2001.00253.x.
- [41] F. Yao, H.-G. Müller, and J.-L. Wang, “Functional linear regression analysis for longitudinal data,” *The Annals of Statistics*, vol. 33, no. 6, pp. 2873–2903, Dec. 2005, doi: 10.1214/009053605000000660.
- [42] J. S. Morris and R. J. Carroll, “Wavelet-based functional mixed models,” *J R Stat Soc Series B Stat Methodol*, vol. 68, no. 2, pp. 179–199, Apr. 2006, doi: 10.1111/j.1467-9868.2006.00539.x.
- [43] G. Cao, L. Yang, and D. Todem, “Simultaneous Inference For The Mean Function Based on Dense Functional Data,” *J Nonparametr Stat*, vol. 24, no. 2, pp. 359–377, Jun. 2012, doi: 10.1080/10485252.2011.638071.

- [44] “Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference - ScienceDirect.” Accessed: Oct. 21, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0047259X82900884>
- [45] P. Hall, H.-G. Müller, and J.-L. Wang, “Properties of principal component methods for functional and longitudinal data analysis,” *The Annals of Statistics*, vol. 34, no. 3, pp. 1493–1517, Jun. 2006, doi: 10.1214/009053606000000272.
- [46] P. Sawant, N. Billor, and H. Shin, “Functional outlier detection with robust functional principal component analysis,” *Comput Stat*, vol. 27, no. 1, pp. 83–102, Mar. 2012, doi: 10.1007/s00180-011-0239-3.
- [47] J. L. Bali, G. Boente, D. E. Tyler, and J.-L. Wang, “Robust functional principal components: A projection-pursuit approach,” *The Annals of Statistics*, vol. 39, no. 6, pp. 2852–2882, Dec. 2011, doi: 10.1214/11-AOS923.
- [48] “Robust Nonparametric Estimation for Functional Data | SpringerLink.” Accessed: Oct. 21, 2023. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-7908-2062-1_18
- [49] D. Gervini, “Robust Functional Estimation Using the Median and Spherical Principal Components,” *Biometrika*, vol. 95, no. 3, pp. 587–600, 2008.
- [50] R. Zhong, S. Liu, H. Li, and J. Zhang, “Robust functional principal component analysis for non-Gaussian longitudinal data,” *Journal of Multivariate Analysis*, vol. 189, p. 104864, May 2022, doi: 10.1016/j.jmva.2021.104864.
- [51] J. O. Ramsay and C. J. Dalzell, “Some Tools for Functional Data Analysis,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 53, no. 3, pp. 539–572, 1991.
- [52] T. Hastie and C. Mallows, “[A Statistical View of Some Chemometrics Regression Tools]: Discussion,” *Technometrics*, vol. 35, no. 2, pp. 140–143, 1993, doi: 10.2307/1269658.
- [53] C. Beckmann, C. Mackay, N. Filippini, and S. Smith, “Group comparison of resting-state fMRI data using multi-subject ICA and dual regression,” *NeuroImage*, vol. 47, p. S148, Jul. 2009, doi: 10.1016/S1053-8119(09)71511-3.
- [54] T. Adali, M. Anderson, and G.-S. Fu, “Diversity in Independent Component and Vector Analyses: Identifiability, algorithms, and applications in medical imaging,” *Signal Processing Magazine, IEEE*, vol. 31, pp. 18–33, May 2014, doi: 10.1109/MSP.2014.2300511.
- [55] I. Rejer and P. Gorski, “Benefits of ICA in the Case of a Few Channel EEG,” *Annu Int Conf IEEE Eng Med Biol Soc*, vol. 2015, pp. 7434–7437, 2015, doi: 10.1109/EMBC.2015.7320110.
- [56] M. Boiret, D. N. Rutledge, N. Gorretta, Y.-M. Ginot, and J.-M. Roger, “Application of independent component analysis on Raman images of a pharmaceutical drug product: pure spectra determination and spatial distribution of constituents,” *J Pharm Biomed Anal*, vol. 90, pp. 78–84, Mar. 2014, doi: 10.1016/j.jpba.2013.11.025.
- [57] B. Debrus *et al.*, “Application of new methodologies based on design of experiments, independent component analysis and design space for robust optimization in liquid chromatography,” *Analytica Chimica Acta*, vol. 691, no. 1, pp. 33–42, Apr. 2011, doi: 10.1016/j.aca.2011.02.035.
- [58] A. Hyvarinen, “Fast ICA for noisy data using Gaussian moments,” in *1999 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 1999, pp. 57–61 vol.5. doi: 10.1109/ISCAS.1999.777510.

- [59] D. N. Rutledge and D. Jouan-Rimbaud Bouveresse, “Independent Components Analysis with the JADE algorithm,” *TrAC Trends in Analytical Chemistry*, vol. 50, pp. 22–32, Oct. 2013, doi: 10.1016/j.trac.2013.03.013.
- [60] L. N. O. Moreno, M. A. A. Arce, and J. G. Lamont, “Implementation of Infomax ICA Algorithm for Blind Source Separation,” in *2008 Electronics, Robotics and Automotive Mechanics Conference (CERMA '08)*, Sep. 2008, pp. 447–451. doi: 10.1109/CERMA.2008.37.
- [61] M. E. Davies and C. J. James, “Source separation using single channel ICA,” *Signal Processing*, vol. 87, no. 8, pp. 1819–1832, Aug. 2007, doi: 10.1016/j.sigpro.2007.01.011.
- [62] I. Rejer and P. Górski, “MAICA: an ICA-based method for source separation in a low-channel EEG recording,” *J. Neural Eng.*, vol. 16, no. 5, p. 056025, Sep. 2019, doi: 10.1088/1741-2552/ab36db.
- [63] B. Mijovic, M. De Vos, I. Gligorijevic, and S. Van Huffel, “Combining EMD with ICA for extracting independent sources from single channel and two-channel data,” *Annu Int Conf IEEE Eng Med Biol Soc*, vol. 2010, pp. 5387–5390, 2010, doi: 10.1109/IEMBS.2010.5626482.
- [64] A. Hyvärinen, “The Fixed-Point Algorithm and Maximum Likelihood Estimation for Independent Component Analysis,” *Neural Processing Letters*, vol. 10, no. 1, pp. 1–5, Aug. 1999, doi: 10.1023/A:1018647011077.
- [65] B. Hjorth, “EEG analysis based on time domain properties,” *Electroencephalography and Clinical Neurophysiology*, vol. 29, no. 3, pp. 306–310, 1970, doi: [https://doi.org/10.1016/0013-4694\(70\)90143-4](https://doi.org/10.1016/0013-4694(70)90143-4).
- [66] U. R. Acharya, V. S. Subbhuraam, S. Goutham, R. Martis, and J. Suri, “Automated EEG analysis of epilepsy: A review,” *Knowledge-Based Systems*, vol. 45, pp. 147–165, Jun. 2013, doi: 10.1016/j.knosys.2013.02.014.
- [67] P. Hosseini, H. Soltanian-Zadeh, K. Elisevich, and D. Pompili, “Cloud-based Deep Learning of Big EEG Data for Epileptic Seizure Prediction,” Dec. 2016. doi: 10.1109/GlobalSIP.2016.7906022.
- [68] F. Abd El-Samie, T. Alotaiby, S. Alshebeili, T. Alshawi, and I. Ahmad, “EEG seizure detection and prediction algorithms: A survey,” *EURASIP Journal on Advances in Signal Processing*, vol. 2014, Dec. 2014, doi: 10.1186/1687-6180-2014-183.
- [69] J. Zheng *et al.*, “time-frequency analysis of scalp EEG with Hilbert-Huang transform and deep learning,” *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2021, doi: 10.1109/JBHI.2021.3110267.
- [70] M. Golmohammadi, A. H. Harati Nejad Torbati, S. Lopez de Diego, I. Obeid, and J. Picone, “Automatic Analysis of EEGs Using Big Data and Hybrid Deep Learning Architectures,” *Frontiers in Human Neuroscience*, vol. 13, p. 76, 2019, doi: 10.3389/fnhum.2019.00076.
- [71] J. Zheng, F. Hsieh, and L. Ge, “A Data-Driven Approach to Predict and Classify Epileptic Seizures from Brain-Wide Calcium Imaging Video Data,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 17, no. 6, pp. 1858–1870, 2020, doi: 10.1109/TCBB.2019.2895077.
- [72] J. Zheng, M. Liang, A. Ekstrom, L. Ge, W. Yu, and F. Hsieh, “On association study of scalp eeg data channels under different circumstances,” in *International Conference on Wireless Algorithms, Systems, and Applications*, Springer, 2018, pp. 683–695.
- [73] T. J. Sullivan, S. R. Deiss, T.-P. Jung, and G. Cauwenberghs, “A brain-machine interface using dry-contact, low-noise EEG sensors,” in *2008 IEEE International*

- Symposium on Circuits and Systems*, 2008, pp. 1986–1989. doi: 10.1109/ISCAS.2008.4541835.
- [74] Y. Zhang, C. Wang, F. Wu, K. Huang, L. Yang, and L. Ji, “Prediction of working memory ability based on EEG by functional data analysis,” *Journal of Neuroscience Methods*, vol. 333, p. 108552, Dec. 2019, doi: 10.1016/j.jneumeth.2019.108552.
- [75] S. K. Kendrick, Q. Zheng, N. C. Garbett, and G. N. Brock, “Application and interpretation of functional data analysis techniques to differential scanning calorimetry data from lupus patients,” *PLOS ONE*, vol. 12, no. 11, p. e0186232, Nov. 2017, doi: 10.1371/journal.pone.0186232.
- [76] K. Hasenstab *et al.*, “A multi-dimensional functional principal components analysis of EEG data,” *Biometrics*, vol. 73, Jan. 2017, doi: 10.1111/biom.12635.
- [77] C. Happ and S. Greven, “Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains,” *Journal of the American Statistical Association*, vol. 113, no. 522, pp. 649–659, Feb. 2018, doi: 10.1080/01621459.2016.1273115.
- [78] X. Wang, B. Nan, J. Zhu, and R. Koeppe, “Regularized 3D functional regression for brain image data via Haar wavelets,” *The Annals of Applied Statistics*, vol. 8, no. 2, pp. 1045–1064, 2014, doi: 10.1214/14-AOAS736.
- [79] X. Wang, B. Nan, J. Zhu, R. Koeppe, and K. Frey, “Classification of ADNI PET images via regularized 3D functional data analysis,” *Biostatistics & Epidemiology*, vol. 1, no. 1, pp. 3–19, 2017, doi: 10.1080/24709360.2017.1280213.
- [80] P. Shangguan, T. Qiu, T. Liu, S. Zou, Z. Liu, and S. Zhang, “Feature extraction of EEG signals based on functional data analysis and its application to recognition of driver fatigue state,” *Physiological Measurement*, vol. 41, no. 12, p. 125004, Jan. 2021, doi: 10.1088/1361-6579/abc66e.
- [81] “Feature extraction of EEG signals based on functional data analysis and its application to recognition of driver fatigue state - PubMed.” Accessed: Nov. 01, 2023. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/33126235/>
- [82] S. Mousavi and H. Sørensen, “Functional logistic regression: a comparison of three methods,” *Journal of Statistical Computation and Simulation*, vol. 88, pp. 1–19, Oct. 2017, doi: 10.1080/00949655.2017.1386664.
- [83] M. Yuan and Y. Lin, “Model selection and estimation in regression with grouped variables,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.
- [84] B. J. Roach and D. H. Mathalon, “Event-Related EEG Time-Frequency Analysis: An Overview of Measures and An Analysis of Early Gamma Band Phase Locking in Schizophrenia,” *Schizophrenia Bulletin*, vol. 34, no. 5, pp. 907–926, Aug. 2008.
- [85] J. S. Lee, *Aspects of functional data inference and its applications*. Rice University, 2006.
- [86] A. Pini and S. Vantini, “The interval testing procedure: A general framework for inference in functional data analysis,” *Biometrics*, vol. 72, p. n/a-n/a, Jan. 2016, doi: 10.1111/biom.12476.
- [87] J.-T. Zhang and X. Liang, “One-way ANOVA for functional data via globalizing the pointwise F-test,” *Scandinavian Journal of Statistics*, vol. 41, no. 1, pp. 51–71, 2014.
- [88] J. Ramsay and B. W. Silverman, “Functional data analysis (Springer series in statistics),” 1997.
- [89] M. Escabias, A. Aguilera, and M. Valderrama, “Principal component estimation of functional logistic regression: discussion of two different approaches,” *Journal of Nonparametric Statistics*, vol. 16, no. 3–4, pp. 365–384, 2004.

- [90] L. Meier, S. Van De Geer, and P. Bühlmann, “The group lasso for logistic regression,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 70, no. 1, pp. 53–71, 2008.
- [91] L. Pion-Tonachini, K. Kreutz-Delgado, and S. Makeig, “ICLabel: An automated electroencephalographic independent component classifier, dataset, and website,” *NeuroImage*, vol. 198, pp. 181–197, Sep. 2019, doi: 10.1016/j.neuroimage.2019.05.026.
- [92] M. Liang, J. Zheng, E. Isham, and A. Ekstrom, “Common and distinct roles of frontal midline theta and occipital alpha oscillations in coding temporal intervals and spatial distances,” *bioRxiv*, p. 2020.08.05.237677, Jun. 2021, doi: 10.1101/2020.08.05.237677.
- [93] H. Li, X. Mao, and L. Chen, “An Emotion Classification Method Based on Energy Entropy of Principal Component,” *Journal of Physics: Conference Series*, vol. 1487, p. 012002, Mar. 2020, doi: 10.1088/1742-6596/1487/1/012002.
- [94] W. L. McFarland, H. Teitelbaum, and E. K. Hedges, “Relationship between hippocampal theta activity and running speed in the rat.,” *Journal of comparative and physiological psychology*, vol. 88, no. 1, p. 324, 1975.
- [95] C. H. Vanderwolf, “Hippocampal electrical activity and voluntary movement in the rat,” *Electroencephalography and clinical neurophysiology*, vol. 26, no. 4, pp. 407–418, 1969.
- [96] J. Jacobs, “Hippocampal theta oscillations are slower in humans than in rodents: implications for models of spatial navigation and memory,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, no. 1635, p. 20130304, 2014.
- [97] L. M. Sangalli, P. Secchi, S. Vantini, and A. Veneziani, “Efficient estimation of three-dimensional curves and their derivatives by free-knot regression splines, applied to the analysis of inner carotid artery centrelines,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 58, no. 3, pp. 285–306, 2009, doi: <https://doi.org/10.1111/j.1467-9876.2008.00653.x>.
- [98] P. K. R. Boppidi, V. J. Louis, A. Subramaniam, R. K. Tripathy, S. Banerjee, and S. Kundu, “Implementation of fast ICA using memristor crossbar arrays for blind image source separations,” *IET Circuits, Devices & Systems*, vol. 14, no. 4, pp. 484–489, 2020, doi: 10.1049/iet-cds.2019.0420.
- [99] G. Sahonero-Alvarez and H. Calderon, “A Comparison of SOBI, FastICA, JADE and Infomax Algorithms,” 2017.
- [100] N. Ito, R. Ikeshita, H. Sawada, and T. Nakatani, “A Joint Diagonalization Based Efficient Approach to Underdetermined Blind Audio Source Separation Using the Multichannel Wiener Filter,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 1950–1965, 2021, doi: 10.1109/TASLP.2021.3079815.
- [101] Y. Hao, L. Song, M. Wang, L. Cui, and H. Wang, “Underdetermined Source Separation of Bearing Faults Based on Optimized Intrinsic Characteristic-Scale Decomposition and Local Non-Negative Matrix Factorization,” *IEEE Access*, vol. 7, pp. 11427–11435, 2019, doi: 10.1109/ACCESS.2019.2892559.
- [102] A. Hyvärinen and E. Oja, “Independent component analysis: algorithms and applications,” *Neural Networks*, vol. 13, no. 4–5, pp. 411–430, Jun. 2000, doi: 10.1016/S0893-6080(00)00026-5.
- [103] G. Sahonero-Alvarez and H. Calderon, “A Comparison of SOBI, FastICA, JADE and Infomax Algorithms,” p. 6, 2017.

- [104] J. F. Cardoso and A. Souloumiac, “Blind beamforming for non-gaussian signals,” *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, no. 6, pp. 362–370, Dec. 1993, doi: 10.1049/ip-f-2.1993.0054.
- [105] S. Pandey, N. Billor, and A. Turkmen, “The Effect of Outliers in Independent Component Analysis,” *American Journal of Mathematical and Management Sciences*, vol. 28, no. 3–4, pp. 399–418, Feb. 2008, doi: 10.1080/01966324.2008.10737735.
- [106] S. Amari, A. Cichocki, and H. Yang, “A New Learning Algorithm for Blind Signal Separation,” in *Advances in Neural Information Processing Systems*, MIT Press, 1995. Accessed: Sep. 26, 2022. [Online]. Available: <https://proceedings.neurips.cc/paper/1995/hash/e19347e1c3ca0c0b97de5fb3b690855a-Abstract.html>
- [107] P. Ablin, J.-F. Cardoso, and A. Gramfort, “Faster Independent Component Analysis by Preconditioning With Hessian Approximations,” *IEEE Trans. Signal Process.*, vol. 66, no. 15, pp. 4040–4049, Aug. 2018, doi: 10.1109/TSP.2018.2844203.
- [108] J. Montoya-Martínez, J.-F. Cardoso, and A. Gramfort, “Caveats with stochastic gradient and maximum likelihood based ICA for EEG,” in *Latent Variable Analysis, Independent Component Analysis LVA-ICA International Conference*, Grenoble, France, Feb. 2017. Accessed: Sep. 26, 2022. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-01451432>
- [109] G. Wang, W. Cai, and X. Shao, “A primary study on resolution of overlapping GC-MS signal using mean-field approach independent component analysis,” *Chemometrics and Intelligent Laboratory Systems*, vol. 82, no. 1, pp. 137–144, May 2006, doi: 10.1016/j.chemolab.2005.05.009.
- [110] Y. B. Monakhova, S. A. Astakhov, A. Kraskov, and S. P. Mushtakova, “Independent components in spectroscopic analysis of complex mixtures,” *Chemometrics and Intelligent Laboratory Systems*, vol. 103, no. 2, pp. 108–115, Oct. 2010, doi: 10.1016/j.chemolab.2010.05.023.
- [111] A. Kassouf, D. Jouan-Rimbaud Bouveresse, and D. N. Rutledge, “Determination of the optimal number of components in independent components analysis,” *Talanta*, vol. 179, pp. 538–545, Mar. 2018, doi: 10.1016/j.talanta.2017.11.051.
- [112] D. Jouan-Rimbaud Bouveresse, A. Moya-González, F. Ammari, and D. N. Rutledge, “Two novel methods for the determination of the number of components in independent components analysis models,” *Chemometrics and Intelligent Laboratory Systems*, vol. 112, pp. 24–32, Mar. 2012, doi: 10.1016/j.chemolab.2011.12.005.
- [113] F. Bach and M. Jordan, “Finding Clusters In Independent Component Analysis,” *4th International Workshop on Independent Component Analysis and blind Signal Separation, ICA2003*, May 2003.
- [114] J. Durbin and G. S. Watson, “Testing for Serial Correlation in Least Squares Regression: I,” *Biometrika*, vol. 37, no. 3/4, pp. 409–428, 1950, doi: 10.2307/2332391.
- [115] M. P. Gómez-Carracedo, J. M. Andrade, D. N. Rutledge, and N. M. Faber, “Selecting the optimum number of partial least squares components for the calibration of attenuated total reflectance-mid-infrared spectra of undesigned kerosene samples,” *Anal Chim Acta*, vol. 585, no. 2, pp. 253–265, Mar. 2007, doi: 10.1016/j.aca.2006.12.036.
- [116] H. F. Kaiser, “An index of factorial simplicity,” *Psychometrika*, vol. 39, no. 1, pp. 31–36, Mar. 1974, doi: 10.1007/BF02291575.
- [117] N. Mehta and A. Gray, “FuncICA for Time Series Pattern Discovery,” in *Proceedings of the 2009 SIAM International Conference on Data Mining*, Society for Industrial and Applied Mathematics, Apr. 2009, pp. 73–84. doi: 10.1137/1.9781611972795.7.

- [118] “Independent components techniques based on kurtosis for functional data analysis.” Accessed: Mar. 24, 2023. [Online]. Available: <https://e-archivo.uc3m.es/handle/10016/18868>
- [119] B. Li, G. V. Bever, H. Oja, and R. Sabolova, “Functional independent component analysis : an extension of fourth-order blind identification.” Accessed: Jun. 23, 2022. [Online]. Available: <https://www.semanticscholar.org/paper/Functional-independent-component-analysis-%3A-an-of-Li-Bever/be161c7433beeafb0e72023bf4721f3983632117>
- [120] J. Virta, B. Li, K. Nordhausen, and H. Oja, “Independent component analysis for multivariate functional data,” *Journal of Multivariate Analysis*, vol. 176, p. 104568, Mar. 2020, doi: 10.1016/j.jmva.2019.104568.
- [121] M. Vidal, M. Rosso, and A. M. Aguilera, “Bi-Smoothed Functional Independent Component Analysis for EEG Artifact Removal,” *Mathematics*, vol. 9, no. 11, Art. no. 11, Jan. 2021, doi: 10.3390/math9111243.
- [122] H. Gutch and F. Theis, *To Infinity and Beyond: On ICA over Hilbert Spaces*. 2012, p. 187. doi: 10.1007/978-3-642-28551-6_23.
- [123] M. Hubert, P. J. Rousseeuw, and K. Vanden Branden, “ROBPCA: A New Approach to Robust Principal Component Analysis,” *Technometrics*, vol. 47, no. 1, pp. 64–79, Feb. 2005, doi: 10.1198/004017004000000563.
- [124] P. J. Rousseeuw and K. V. Driessen, “A Fast Algorithm for the Minimum Covariance Determinant Estimator,” *Technometrics*, vol. 41, no. 3, pp. 212–223, Aug. 1999, doi: 10.1080/00401706.1999.10485670.
- [125] J. Zheng *et al.*, “Time-Frequency Analysis of Scalp EEG With Hilbert-Huang Transform and Deep Learning,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 4, pp. 1549–1559, Apr. 2022, doi: 10.1109/JBHI.2021.3110267.
- [126] M. Liang, J. Zheng, E. Isham, and A. Ekstrom, “Common and Distinct Roles of Frontal Midline Theta and Occipital Alpha Oscillations in Coding Temporal Intervals and Spatial Distances,” *Journal of Cognitive Neuroscience*, vol. 33, no. 11, pp. 2311–2327, Oct. 2021, doi: 10.1162/jocn_a_01765.