New methods for discovering hidden dependence and for assessing the possible influence of unobserved variables

by

Yeo Jung Park

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Statistics)
in The University of Michigan
2013

Doctoral Committee:

Professor Kerby A. Shedden, Chair
Associate Professor Timothy D. Johnson
Professor Vijay N. Nair
Assistant Professor Long Nguyen

To my family and Tresbien

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

ABSTRACT

New methods for discovering hidden dependence and for assessing the possible influence of unobserved variables

by

Yeo Jung Park

Advisor: Kerby Shedden

The biological interpretation of neuroimaging data often depends on changes in the dependence structure between locations in the brain. A major challenge in neuroscience is uncovering the relationship between consciousness and brain activity. Electroencephalography (EEG) recordings made on human subjects who are given anesthesia for surgery provide an opportunity to directly study this relationship. The main focus in this area has been on changes in the connectivity between brain regions that occur as the consciousness state changes. Connectivity can be assessed in terms of the statistical dependence between EEG measurements from different recording sites on the scalp.

In this thesis, we consider two approaches for capturing changes in the dependence structure among several time series. We first consider the possibility that dependence between two series may be localized to a specific frequency band, and hence cannot be uncovered using global measures dependence. We propose methods to characterize the frequency-specific dependence in such data. We then consider the possibility that the dependence between two series can be revealed by applying a local transformation.

We optimize over a class of such transformations to maximize a simple association measure, leading to a new measure of dependence for serially observed data. These two new methods are used to analyze a data that consists of multi-channel EEG recordings of multiple subjects under several consciousness states.

Another question that arises in analyzing complex biological data sets is whether there exists an unobserved variable responsible for all apparent relationships between a given set of observed variables and the outcome. In the last chapter, we propose an approach to understanding under what circumstances a single unmeasured variable could explain the entire observed relationship between an outcome and several observed predictors. The unobservable regression of interest is characterized in terms of three quantities: the distribution of the unobserved covariate, the effect size of the unobserved covariate, and the net dependence between the unobserved and the observed covariates. We derive an explicit functional relationship among these quantities, and how this in turn can be used to learn about possible alternative explanations for an observed multiple regression relationship.

CHAPTER I

Introduction

Neuroscientists who study consciousness are interested in identifying and characterizing the neural activity patterns associated with different conscious experiences (*Rees et al.* (2002)). General anesthesia, combined with brain imaging technology, can be used as a tool for addressing such questions (*Alkire and Miller* (2005); *Hudetz* (2012)). There are number of hypotheses about the brain state under anesthesia-induced unconsciousness (*Nallasamy and Tsao* (2011)). *Alkire et al.* (2008) propose that anesthesia-induced unconsciousness is the result of disrupted functional connectivity among certain brain regions. Figure 1.1 is a simple illustration of the hypotheses. Fp1, Fp2, F3, F4 indicates the front part of the brain and P3, P4 are the back part of the brain. As can be seen, the hypothesis states that when a person is under anesthesia-induced unconscious, the connectivity between front and back part of brain are reduced, while within front and back region connectivity are remained. Therefore, the problem of characterizing neural activity for both resting awake state and anesthesia state becomes to show how the neural connectivity among brain regions changes between two settings.

In statistics, there is a large literature on methods for analysis of brain imaging data. In this thesis, we focus on the analysis of brain connectivity. In this area, functional Magnetic Resonance Imaging (fMRI) is widely used. Usually, there are

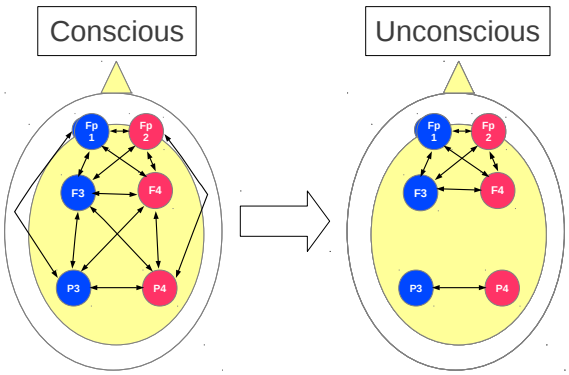


Figure 1.1: A simple illustration of the hypotheses on the brain connectivity under consciousness and anesthesia-induced unconsciousness states. Nodes represent the regions of the brain, and edges are drawn if two regions are connected. For simplicity, we only show 6 regions that we will be considering in our data analysis.

two kinds of connectivity that require different approaches: undirected and directed. In the fMRI literature, the undirected and directed connectivity are also known as functional and effective connectivity, respectively (*Friston* (1994)). For undirected connectivity analysis, methods such as Principal Components Analysis (PCA) and Independent Components Analysis (ICA) are mainly used. For directed connectivity analysis, Structural Equation Modeling (SEM), Dynamic Causal Modeling (DCM), and Granger causality are commonly used. We refer to *Lindquist* (2008); *Friston* (2011) for a more detailed review.

There are a growing number of studies of consciousness that utilize general anesthesia. Three regions of the brain - thalamus, frontoparietal and posterior cingulate cortex[1] - are mainly studied in terms of activation and connectivity in anesthesia-induced unconsciousness (*Nallasamy and Tsao* (2011); *Hudetz* (2012)). We focus on frontoparietal connectivity: the connection between the front and back part of the brain.

It is commonly understood that the brain is spatially organized, so the different

[1]The surface of the brain is divided by four sections: front, middle, back and bottom. Each sections are called the frontal lobe, the parietal lobe, the occipital lobe, and the temporal lobe respectively. Thalamus is located at the center of inner brain and posterior cingulate cortex is back part of the inner brain near thalamus. See *Alkire et al.* (2008) for the picture.

regions have different functions. Specifically, *Crick and Koch* (2003) proposed a hypothesis that the front and back regions of the brain have their own synthetic processes, yet with active interaction with each other. This suggests that connectivity within regions may play a different role than connectivity between regions. *Peltier et al.* (2005) used fMRI and showed that under anesthesia, connectivity between motor-related regions and others are lost. Their approach used a correlation map between seed region (motor cortex) and all other voxels. Each voxel contains a single series that is an average of preprocessed multiple time series in that voxel. *Schrouff et al.* (2011) also used fMRI and applied an information-theory based approach to assess connectivity. They observed that the frontoparietal connectivity reduces under anesthesia.

The data sets considered in this thesis are multi-channel EEG recordings. All data sets contain multiple subjects whose EEG signals are collected from 6 brain locations under several different conscious states: Awake, Induction and Anesthesia. Details of the data are given in Chapter II.

There is a literature that uses multi-channel EEG data to characterize the neural connectivity pattern in several different consciousness states. *Ku et al.* (2011) used an information-theory based method, called Symbolic Transfer Entropy, which identifies changes in information flow among brain regions. This study showed that there was a decreased amount of information exchange from the parietal to frontal regions under anesthesia, and no significant change of the flow from the frontal to parietal areas. *Cimenser et al.* (2011) also explored the connectivity pattern by applying a method based on the spectrogram and global coherence[2]. In this study, they found that under anesthesia certain frequency bands showed distinctive changes of the power in the occipital and frontal lobes. These studies suggest that the connectivity pattern changes are seen particularly between the frontal and parietal regions (*Alkire and*

[2]Spectrum is a frequency domain measure that utilize the Fourier transformation, and Global coherence is a summary measure of the spectrogram matrix.

Miller (2005)).

This thesis develops two approaches to understanding connectivity, and connectivity changes in EEG measurements of the brain activity. The focus is on identifying connections that are mostly invisible to simple descriptive measures of association. Chapter II provides an approach that utilize scale space decomposition of series. This approach is motivated from the possibility that the hidden association pattern appears in certain temporal scale. In Chapter III, we propose a local transformation approach that reveals hidden association. This approach is motivated from the possibility that we can reveal hidden association by capturing some local characteristics of two series using localized transformation. In addition, Chapter IV presents a new method for understanding the role of unobserved variables in regression analysis. This technique is motivated by issues that arisen genetic analysis, but analogous issues arise in neuroscience application.

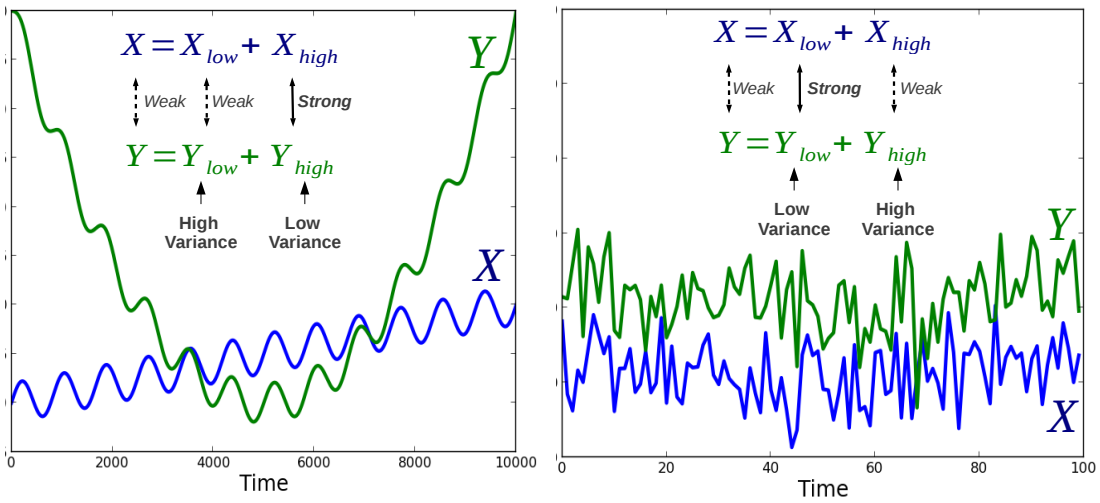
CHAPTER II

A correlation decomposition approach to connectivity analysis for multivariate time series with application to a study of anesthesia-induced unconsciousness

2.1 Introduction

In this chapter, we consider an approach that explores the changes in neural connectivity under different conditions using multi-channel EEG. We view connectivity as being any type of statistical dependency between two series. Therefore, the problem can be defined as exploring how the association patterns among multiple time series change under several conditions. The approach developed in this chapter locally decomposes the time series into frequency bands, to identify relationships that are hidden in the raw data.

The approach is motivated by the observation that simple association measures such as the correlation coefficient have limited ability to capture relationships in sequentially observed data. As an illustration, we can consider that a series arises as a sum from two frequency bands: low and high. Figure 2.1 (a) shows two series whose correlation is strong in the high frequency band and weak in the low frequency



(a) High frequency shared scheme

(b) Low frequency shared scheme

Figure 2.1: Illustrative examples of two series with hidden correlation in certain frequency band.

band. If the variation in the high frequency band is dominated by variance at other frequencies, the Pearson correlation between raw series would not reveal such hidden association. The opposite case is also possible. Figure 2.1 (b) shows the case that strong correlation is in the high frequency band but its variation is small.

Classical frequency domain analysis based on Fourier methods has been studied and used extensively across a broad range of fields such as neuroscience and economics (*Platt and Damon* (1975); *Iacobucci* (2003); *Cimenser et al.* (2011)). In particular, in the neural connectivity analysis, the coherence measure (analogous to the correlation coefficient for frequency domain) is widely used. For multivariate time series, *Brillinger* (1975) developed a method that measures association between two sets of multiple time series. This method is an extension of *Hotelling* (1936)'s Canonical Correlation Analysis to time series using the frequency domain.

Fourier transformation (FT) captures periodic behavior that remains synchronized over the entire length of the data series. Unless the series is almost perfectly stationary, this is unlikely to effectively capture the important features of the data. In order to accommodate non-stationary time series while using a Fourier-like approach,

Short-time FT (STFT) has been proposed (*Allen* (1977)). This method applies FT to local windows of a series. However, STFT has limitations due to rigid size of the windows regardless of the frequency level. That is, there is a trade off in resolutions of frequency and time according to the window size; using a narrow window to capture high frequency gives good frequency resolution but poor time resolution and using a wide window has the opposite effect.

Wavelet approaches have become a more popular means to decompose variation in time series that are not strictly stationary. The wavelet transformation uses orthogonal, compactly supported basis functions to resolve effects that occur at specific temporal scales, while allowing these effects to evolve freely over time. We will briefly review wavelet decompositions in Section 2.2. We also refer to *Abramovich et al.* (2000) for a review of wavelet analysis in statistical applications.

In this chapter, we introduce a procedure that uses a scale space transformation, such as the wavelet transformation. This procedure aims to identify dependencies between two series that are localized to a relatively narrow frequency band. Figure 2.2 shows the schematic of our approach. As can be seen, given two series from location 1 and 2, each series can be decomposed into different temporal scale (high to low). Then, we look at the associations between two decomposed series with the same temporal scale. This approach is interpretable, efficient to compute, and enables effective visualization.

The rest of the chapter is organized as follows. In Section 2.2, we discuss our method in detail including review of wavelet transformation. Section 2.3 provides a simulation study that assesses the effect of series length on detecting the hidden association. Section 2.4 discuss a real data application using EEG data with three different anesthetic drugs. Section 2.5 provides concluding remarks.

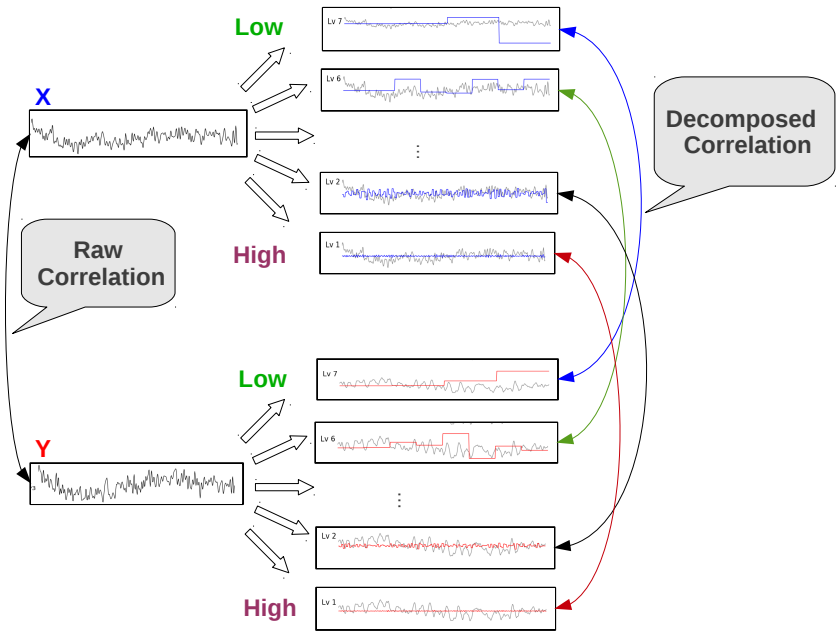


Figure 2.2: Schematic of our approach. The pairs of series that are used to compute an association measure are connected with the arrowed line. Each color denotes the each pair of association.

2.2 Correlation decomposition using wavelet transformation

In this section, we provide details of the approach. The approach can be viewed as consisting of two steps. First, apply a scale space transformation of series as depicted in Figure 2.2. Second, we explore the dependence structure between the series, within each level of the scale space decomposition. This exploration takes the form of visualizations as well as more formal statistical analysis. The scale space decomposition we consider in this analysis is wavelet transformation. We begin this section by providing a review of wavelet transformation.

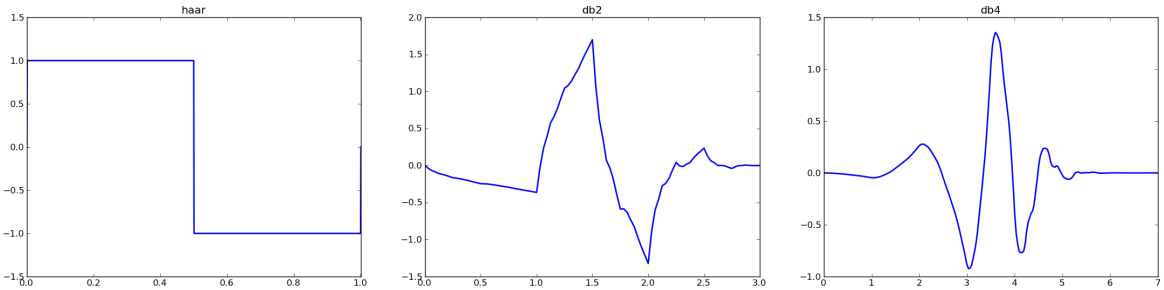
2.2.1 Discrete Wavelet Transformation

Overview:

The wavelet transformation (WT) is well known for its localization in both time and frequency. This dual localization is achieved by taking an inner-product of an original

series and a small wave shaped basis function. This basis function can have various spans according to frequency bands. WT has two different formats: continuous and discrete. Continuous wavelet transform (CWT) is defined over the entire real line of time, whereas discrete wavelet transform (DWT) is defined over discretized time points (integer values). In the analysis, we use DWT because of its orthogonality of basis function and computational efficiency. Note that DWT differs from a discrete approximation of CWT, which is a practical implementation of CWT. For details of CWT, see *Samar et al.* (1999) and *Addison* (2005).

DWT can have many different forms of basis function. The Daubechies wavelet family is one of the most popular ones that are used in a wide range of fields. The shape of the basis function is provided by a mother wavelet (or wavelet function). Figure 2.3 shows selected examples of the mother wavelet from the Daubechies family. A subclass of Daubechies family is classified by its vanishing moment, which is half of the number of coefficients used to construct mother wavelet function. For example, in the wavelet examples in Figure 2.3, (a) has vanishing moment 1, (b) has 2 and (c) has 4. In the analysis, we use Daubechies wavelet with order 4 based on the study by *Adeli et al.* (2003).



(a) Haar **(b)** Daubechies Order 2 **(c)** Daubechies Order 4

Figure 2.3: Various mother wavelets from Daubechies wavelet family.

Applying DWT to original series produces two kinds of coefficients: wavelet (or detail) and scaling coefficient. Wavelet coefficients are the product of a wavelet basis and the series, and scaling coefficients are product of a scale function and the series.

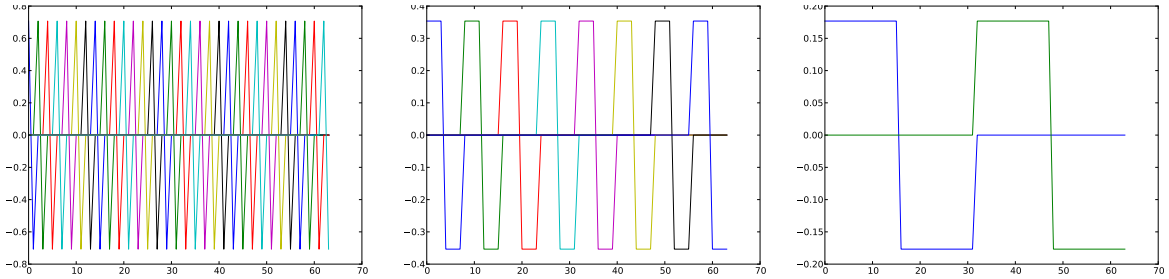
Usually, wavelet coefficients are considered to be series that passed a high-pass (or bandpass) filter, and scaling coefficients are series that passed a low-pass filter. As an illustration, in the Haar wavelet case, at a time point t, wavelet coefficients are differences of the local average of signals over a certain time interval before and after t. The scaling coefficients are just an average of the series over the entire time. Other families of DWT basis function work in similar way as the Haar wavelet, but have more complicated structure in order to have flexibility in the shape.

Decomposition of a series into different frequency bands is achieved by applying multi-resolution DWT. Multi-resolution DWT is an iterative application of rescaled mother wavelet according to frequency bands; the mother wavelet in a basis function repeats in narrower span for higher frequencies and wider span for lower frequencies. Figure 2.4 shows the shape of basis function of Haar wavelet in different frequencies. This way of constructing a basis achieves localization in both time and frequency scales that naturally leads to a good resolution in both scales simultaneously, and the decomposed series completely reconstruct the original series without having redundancy. As DWT is based on a dyadic subsampling scheme, the length of series cut into half after going thorough each level of transformation. The computational efficiency is achieved by an iterative algorithm, the so-called Pyramid algorithm[1].

Wavelet analysis of variance:

A beneficial consequence of applying orthogonal wavelet basis is that it also decomposes the variance and covariance into different frequency bands. Correlation coefficients at each wavelet level can be calculated using these decomposed variances and covariances. We review how DWT decomposes the variance and covariance of the series by frequency bands, and introduce resulting wavelet correlation coefficient. We borrowed some notations from *Lindsay et al.* (1996).

[1]This algorithm applies wavelet and scale function on the scale coefficients from the previous level in multi-resolution DWT repeatedly (*Lindsay et al.* (1996)). A good review of the algorithm can be found in Chapter 4, *Percival and Walden* (2000)



(a) High frequency **(b)** Medium frequency **(c)** Low frequency

Figure 2.4: Illustration of different wavelet basis for multi-resolution DWT (Haar wavelet). Mother wavelet is rescaled and shifted in order to capture both frequency and time information.

Multi-resolution DWT provides alternative representation of a series Y with T time points as the following way:

$$Y = \sum_{j=1}^{L} D_j \psi_j + \bar{y}\overrightarrow{1}, \tag{2.2.1}$$

where D_j is a vector of wavelet coefficients at level j (contain $2^{-j}N$ coefficients), ψ_j is a wavelet basis at level j, and $\overrightarrow{1} = [1, \cdots, 1]'$. As the basis are orthonormal, the inner product of ψ_i and ψ_j is 0 if $i \neq j$ and 1 if $i = j$.

The alternative representation (2.2.1), which is a linear combination of wavelet and scaling coefficients, implies that the variance of a series Y can be also decomposed. Let $D = [D_1', D_2', \cdots, D_L', \bar{y}]$ and $\Psi = [\psi_1, \psi_2, \cdots, \psi_L, \overrightarrow{1}]$, where D_j is a vector of wavelet coefficients ($2^{-j}T$ elements) and ψ_j is a $T \times 2^{-j}T$ wavelet basis matrix. Hence, (2.2.1) can be expressed as $Y = \Psi' D$. Also, as DWT are constructed for the wavelet coefficients to have zero mean, sum of squared elements in a wavelet coefficient vector at level j is equivalent to the variance of wavelet coefficients at level j, i.e. $\widehat{Var}(D_{y,j}) = \sum_{k=1}^{2^{-j}T} D_{y,jk}^2$. Then the sample variance of Y is decomposed by variances of wavelet coefficients D as following:

$$\hat{\sigma}^2 \equiv \frac{Y'Y}{N} - \overline{Y}^2 = \frac{D'\Psi\Psi'D}{N} - \overline{Y}^2 = \frac{D'D}{N} - \overline{Y}^2 = \frac{1}{N}\sum_{j=1}^{L}\left(\sum_{k=1}^{2^{-j}N} D_{y,jk}^2\right), \quad (2.2.2)$$

where $D_{y,jk}$ is kth wavelet coefficient of series Y at level j. Similarly, the sample covariance between two series X and Y can be derived as following:

$$\hat{\sigma}_{xy} \equiv \frac{1}{N}(\sum_{t=0}^{N-1}(X_t - \bar{X})(Y_t - \bar{Y}) = \frac{1}{N}\sum_{j=1}^{L}\left(\sum_{k=1}^{2^{-j}N} D_{x,jk}D_{y,jk}\right). \quad (2.2.3)$$

Using (2.2.2) and (2.2.3), we can define wavelet sample correlations at each wavelet level j. The raw correlation between X and Y is not decomposed the same way variance and covariance did, that is, variances (or covariances) of each level approximately sum up to the total variance (or covariance). The definition of wavelet sample correlation at level j is as following:

$$\hat{\rho}_{xy,j} = \frac{\hat{\sigma}_{xy,j}}{\hat{\sigma}_{x,j}\hat{\sigma}_{y,j}}, \quad (2.2.4)$$

where $\hat{\sigma}_{xy,j}$ is a sample covariance of X and Y at level j, and $\hat{\sigma}_{x,j}$, $\hat{\sigma}_{y,j}$ are sample variance of X and Y at level j respectively.

2.2.2 Visualizing the association change pattern with the transformed series

In the previous subsection, we discussed how to decompose each time series and calculate the wavelet correlation coefficients from those decomposed series. In this subsection, we elaborate on the post-analysis that provide a summary measure of the association change for visualization. Recall, we have n number of subjects, and each subject has a $p \times p$ correlation matrix. After applying a multi-resolution DWT, we have a total of L frequency levels (for $L \geq 1$). Then each subject have L number of

$p \times p$ correlation coefficient matrix per experimental condition. Given the number of resulting $p \times p$ correlation matrices is $2 \times n \times L$, we take the difference of the squared correlation matrices between two conditions, as our interest is in the magnitude of association change. Hence, the number of matrices is now reduced to $n \times L$. Each element of $p \times p$ correlation difference matrix will be averaged over subjects to have L number of adjacency matrix.

Figure 2.5 shows the schematic of the method producing an adjacency matrix for each wavelet level. In order to achieve some sparsity for the adjacency matrix at each level, we conduct one sample t-test using n differences for each element of the difference matrix. This tests whether the mean difference is 0 or not for that particular location pair at the level l. We consider multiple comparison correction for each wavelet level. In our data analysis, we apply Bonferroni corrected significance level. The final adjacency matrix at level l will have nonzero value if the t-test result is significant, and zero otherwise. Now we have L number of thresholded correlation difference matrices, and we can draw a network plot of the connectivity change at each wavelet level based on these matrices.

2.3 Simulation study

In this section, we provide a simulation study that assesses the effect of series length on detecting the frequency band where the association between two series is hidden. Generating two series with a hidden association is done using the multi-resolution wavelet representation of a series:

$$\begin{aligned} X &= \sum_{j=1}^{L} D_j^x \psi_j + \bar{x}\overrightarrow{1} \\ Y &= \sum_{j=1}^{L} D_j^y \psi_j + \bar{y}\overrightarrow{1}. \end{aligned}$$

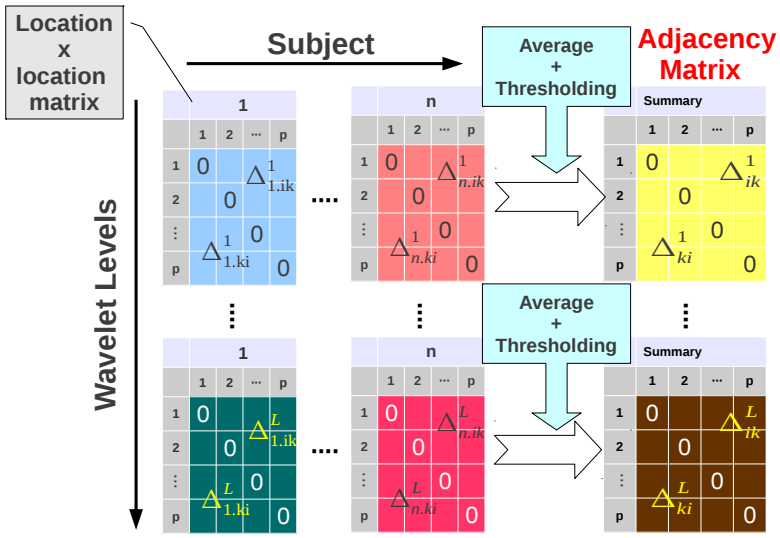


Figure 2.5: Schematic of producing adjacency matrices for each wavelet level j. Δ_{ik} indicate the difference of squared correlation from condition 1 to condition 2, i.e. $\Delta^j_{ik} = \hat{\rho}^2_{ik,j,1} - \hat{\rho}^2_{ik,j,2}$, for $j = 1, \cdots, L$ and $i, k \in 1, \cdots, p$. At each location pair, one sample t-test with 90% confidence level is applied on $\Delta^j_{1.ij}, \cdots, \Delta^j n.ik$ to threshold the connectivity differences .

We generate two sets of wavelet coefficients D^x_j and D^y_j for each level $j = 1, 2, \ldots L$ to have certain correlation. The L is determined by total length of the series T. Two sets of coefficients with correlation r are generated by

$$D^x_j, \epsilon \sim N(0, 1)$$
$$D^y_j = rD^x_j + \sqrt{1 - r^2}\epsilon.$$

The characteristics of wavelet basis are shared across the different wavelet families. Therefore, for simplicity we use Haar wavelet without loss of generality. The wavelet level j indicates the width of the Haar basis. The width of basis function is equivalent to the number of points in a series that is used to calculate coefficients. Specifically, the Haar basis function of level 1 uses 2 non-overlapping neighboring time points to calculate the coefficients. Hence, the length of resulting coefficients is $T/2$. Similarly, at the jth level, 2^j number of non-overlapping points are used to calculate correlation

and the number of coefficient is $T/2^j$. The lower level indicate higher frequency band.

In the simulation, we use series constructed from Haar wavelet with lengths in power of 2. As an illustration, if we want to generate series with length 2^5, we need to generate 5 levels of wavelet coefficients each with length 2^j for $j = 0, 1, 2, 3, 4$, and a scale coefficient (mean of the series). The considered series lengths are $2^{10}, 2^{11}, 2^{12}, 2^{13}$, and 2^{14}. We set level 3 (basis width 8) to have highest correlation (0.5) for each generated time series. Other levels have correlation between 0.2 and 0.4. We also only consider the wavelet level with enough series length to have reasonable correlation coefficient. The series should contain at least 20 time points. Therefore, the considered levels for each length are up to 6.

The simulation is repeated 500 times. At each iteration, we generate two series, decompose the series using Haar wavelet, and calculate correlation coefficients for each wavelet level. Figure 2.6 shows the resulting distribution of argument of the maximum (argmax) among the correlations from 6 wavelet levels. Regardless of the series length, the true level 3 is detected most frequently. When series length increases, we see that the detected argmax wavelet level converges to the true level.

2.4 Application to EEG data

In this section, we discuss the structure of the real data (EEG recordings from surgical subjects who are anesthetized with three different anesthetics) and our analysis results.

2.4.1 Data description

Our data set is composed of EEG recordings (sampling rate 256 Hz) of surgical subjects whose unconscious states are induced by three different anesthetic agents: Propofol[2], Sevoflurane and Ketamine. The recordings were taken under five experi-

[2]This data was previously analyzed by *Ku et al.* (2011).

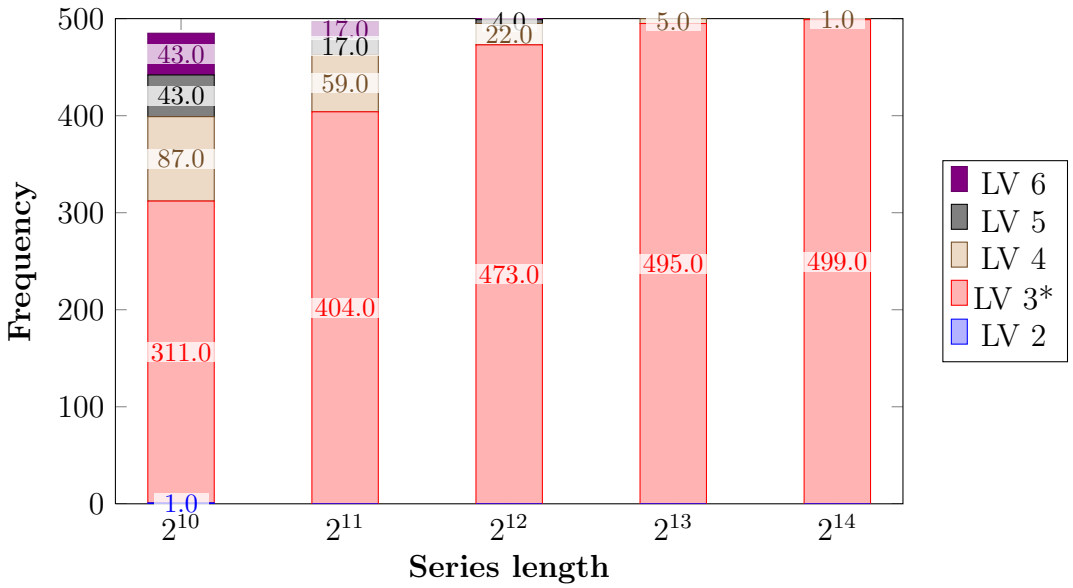


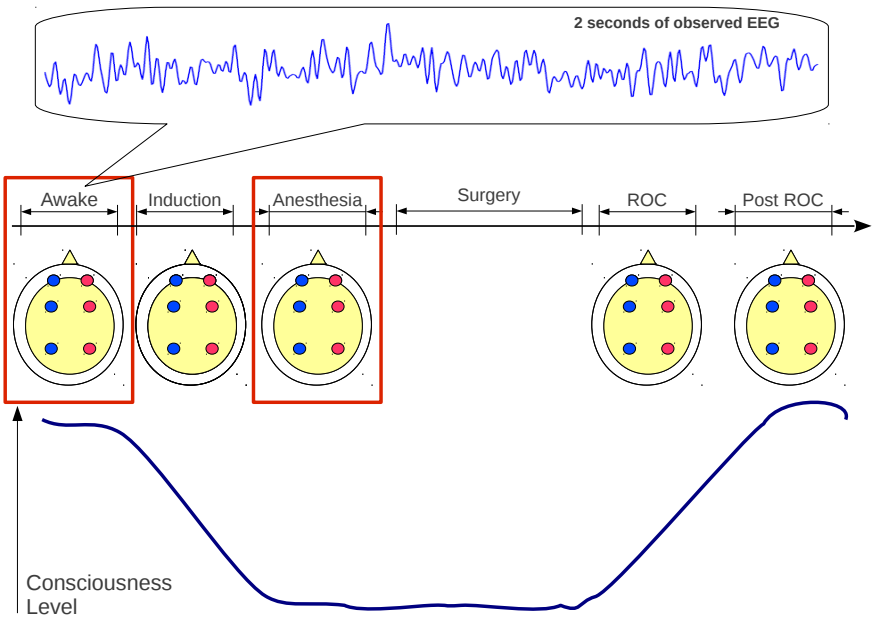
Figure 2.6: Distribution of the argmax from the simulation for different series length. True wavelet level that contain higher correlation is 3. All simulations are repeated 500 times.

mental conditions[3] for Propofol and Sevoflurane, and three conditions[4] for Ketamine. As there are various surgeries conducted on the subjects, EEG recordings from ROC state shows huge individual variation. In order to obtain better contrast, we use two states of consciousness, which are Resting wakefulness (baseline) and Under anesthesia (anesthesia). A schematic of the experiment is shown in Figure 2.7 (a). Each recording (per subject and state for all anesthetics) is from 6 locations in the brain according to 10-20 international system: 4 frontal and 2 parietal regions (purple circles in Figure 2.7 (b)). For Propofol and Sevoflurane, 5 minutes of EEG were originally recorded at each state. We use a manually selected minute long segment in order to minimize artificial effects. Ketamine data are 5 minutes long for baseline state and 3 minutes long for anesthesia state, and we used all the recordings. Table 2.1 shows a summary of the data structure.

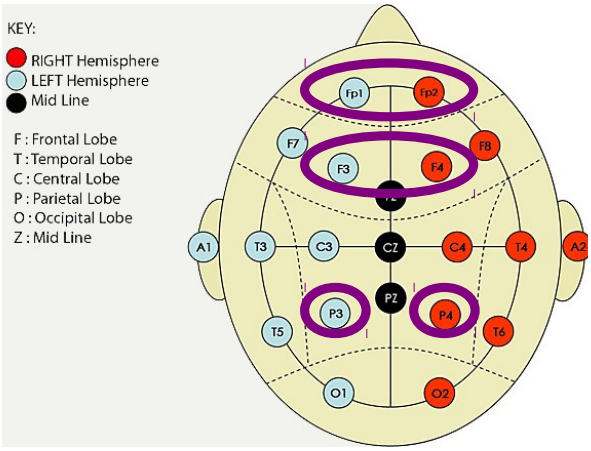
EEG are considered to be a composition of several oscillators with various frequen-

[3]Resting Wakefulness, Anesthesia induction, Under anesthesia, Return of consciousness (ROC), and Post ROC.

[4]Resting wakefulness, Anesthesia induction and Under anesthesia



(a) Schematic of the experiment



(b) 10-20 International System

Figure 2.7: Experiment description and brain location map. (a) In the analysis, we use two consciousness states (red boxes). (b) Purple circles are locations used in the analysis. The picture is reprinted from *Jenkins et al.* (2009).

cies. Therefore, the frequency domain analysis is widely used in order to classify the sequence into different frequency components (*Nunez and Srinivasan* (2006)). These classified frequency ranges were named with Greek letters and the corresponding tasks are extensively studied (*Herrmann et al.* (2005); *Dietrich and Kanso* (2010)). The decomposition levels from multi-resolution DWT can be approximated to this clas-

Table 2.1: Summary of the collected data

Anesthetics	Propofol	Sevoflurane	Ketamine
No. of Subject	9	9	29
No. of Location	6 (4 frontal, 2 parietal)	6	6
No. of Condition	5	5	3
Length	5 min. (all conditions)	5 min. (all conditions)	5 min.(Baseline) 2 min.(Induction) 3 min.(Anesthesia)

sical frequency range (*Bassett et al.* (2006)). Table 2.2 is a list of the wavelet levels and its approximation to the classical frequency bands with related tasks.

Table 2.2: Approximated Wavelet decomposition levels to the classical EEG frequency bands with sampling rate 256 Hz.

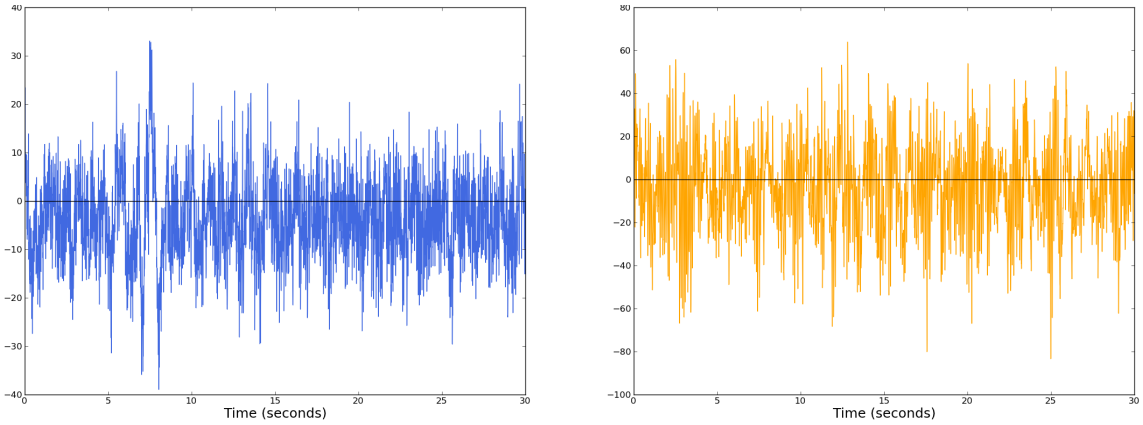
Wavelet Level	Frequency (Hz)	Name	Associated task
1	32 - 64	Gamma (γ)	Higher brain function such as binding of a perceptual information
2	16 - 32	Beta (β)	Alert, active thinking and motor actions
3	8 - 16	Alpha (α)	Awake, relaxed and large number of cognitive process
4	4 - 8	Theta (θ)	Working memory
5, 6, 7	0.5 - 4	Delta (δ)	Deep sleep

* In this table, we only include the wavelet levels that correspond to the classical frequency bands. We rearranged the wavelet levels to match with the classical frequency bands.

2.4.2 Results

Characteristic of EEG:

EEG is an noninvasive method that measures electrical signal generated by up to a billion of neurons' post-synaptic potentials on the human scalp (*Nunez and Srinivasan*



(a) Condition 1: Baseline **(b)** Condition 2: Anesthesia

Figure 2.8: 30 seconds of raw EEG data in two different condition for the same person and location.

(2006)). Due to this complicated relationship among neurons, generally raw EEG does not show any distinctive patterns for a certain experimental condition (*Rampil* (1998)). Figure 2.8 shows EEG series in two different states. The plot of raw EEG suggests that though it is hard to find any distinctive pattern in the mean structure, it is possible that the variance of EEG increases in the anesthesia state. Figure 2.9 shows mean, variance and covariance of EEG at all recorded states. On average, the mean of EEG recordings are almost constant over different conditions regardless of the anesthetic. Variances, on the other hand, tend to increase when subjects are under anesthesia for all anesthetics; strong increasing tendency shown in Propofol and Sevoflurane, and weaker tendency for Ketamine. Covariances also show a similar pattern as variances, but have more variety by the anesthetic agents. This suggests that the neural activities are reflected in variance structure of the EEG rather than the mean.

Before we get into the analysis with transformed series, we explore the association pattern reflected from the raw EEG series. To obtain this overview of connectivity patterns among brain regions, we made a scatter plot matrix of the raw EEG from six brain channels for each subject and condition (see Figure 2.10 for a selected subject

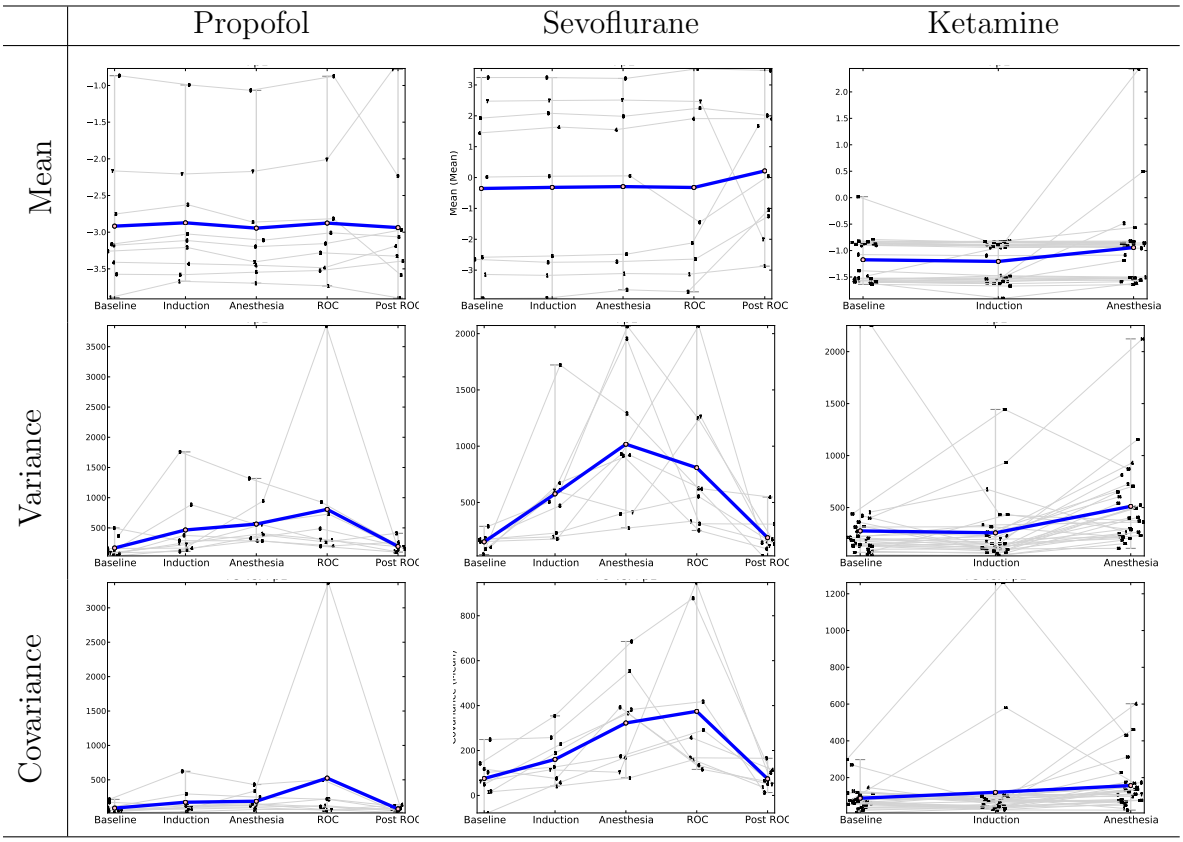


Figure 2.9: Three descriptive statistics (each row) plot in different treatment conditions of three different anesthetic agents (each column) for selected brain locations. The selected locations are 1) Fp1 for mean and variance, and 2) Fp1 and P3 pair for covariance. Gray lines represent the changes of each subject. Blue lines represent average of the descriptive statistics over all subjects.

and condition). Regardless of the anesthetic agents, the relationship between EEG series from two brain regions are fairly linear and show some heterogeneity. This linear pattern suggests that using the Pearson correlation coefficient may provide some useful summary that shows the alternation in the connectivity pattern under two conditions. However, the provided information would be limited as it does not fully characterize the temporal inter-relationship.

Wavelet analysis result:

In multi-resolution DWT, the total possible wavelet decomposition levels depend on the length of the EEG series. The resulting maximum wavelet level by applying multi-resolution DWT is 11 for a minute long EEG (Propofol and Sevoflurane case) and 12 for 2$\sim$3 minutes long EEG (Ketamine case)[5]. At each level, the length of transformed series (i.e. wavelet coefficients) become approximately half of the previous wavelet level's series length. Our results only include the levels that correspond to the classical frequency bands (see Table 2.2).

In order to investigate whether there is a dominant frequency scale, wavelet variances and covariances are calculated according to (2.2.2) and (2.2.3) respectively. Figure 2.11 shows the variances and covariances for each wavelet level for selected locations[6]. The increasing variances and covariances under anesthesia is observed at most of the wavelet levels. In both conditions (baseline and anesthesia), the maximum of decomposed variance happens at wavelet level 4 (or approximately α range) for Propofol and Sevoflurane and wavelet level 7 (δ range) for Ketamine. Covariances over multiple wavelet levels do not show any consistent pattern for all the anesthetics (see the second column of Figure 2.11). Overall, the most noticeable frequency bands are level 4 and 7, which corresponds to α and δ range respectively.

[5]As the sampling rate is 256 Hz, the length of a minute long series is $256 * 60 = 15360$, and two or three minutes long series have length 30720 and 46080, respectively.

[6]Other regions have similar pattern as the selected locations

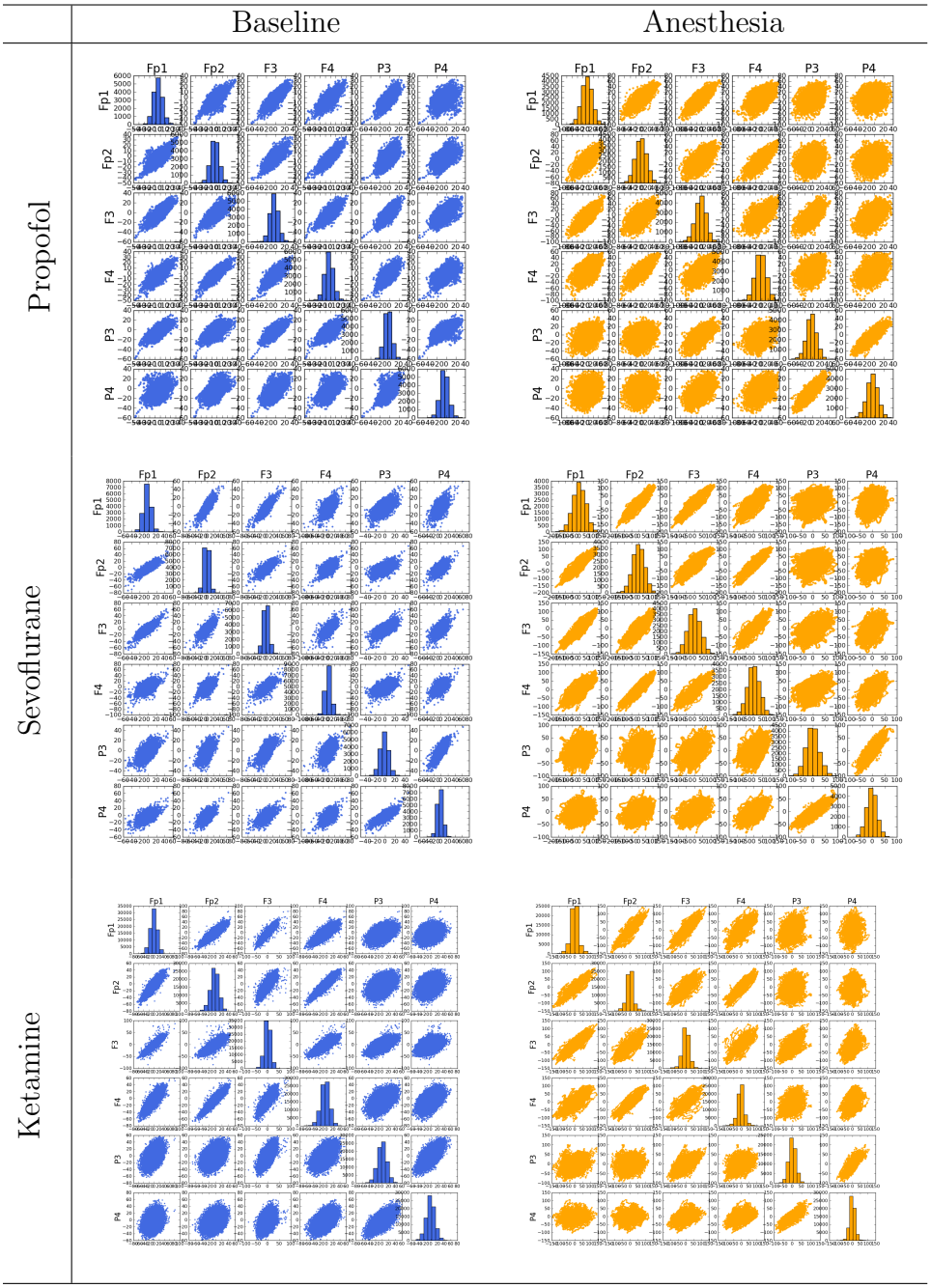


Figure 2.10: Scatter plot matrix of raw EEG from six brain locations for selected subjects.

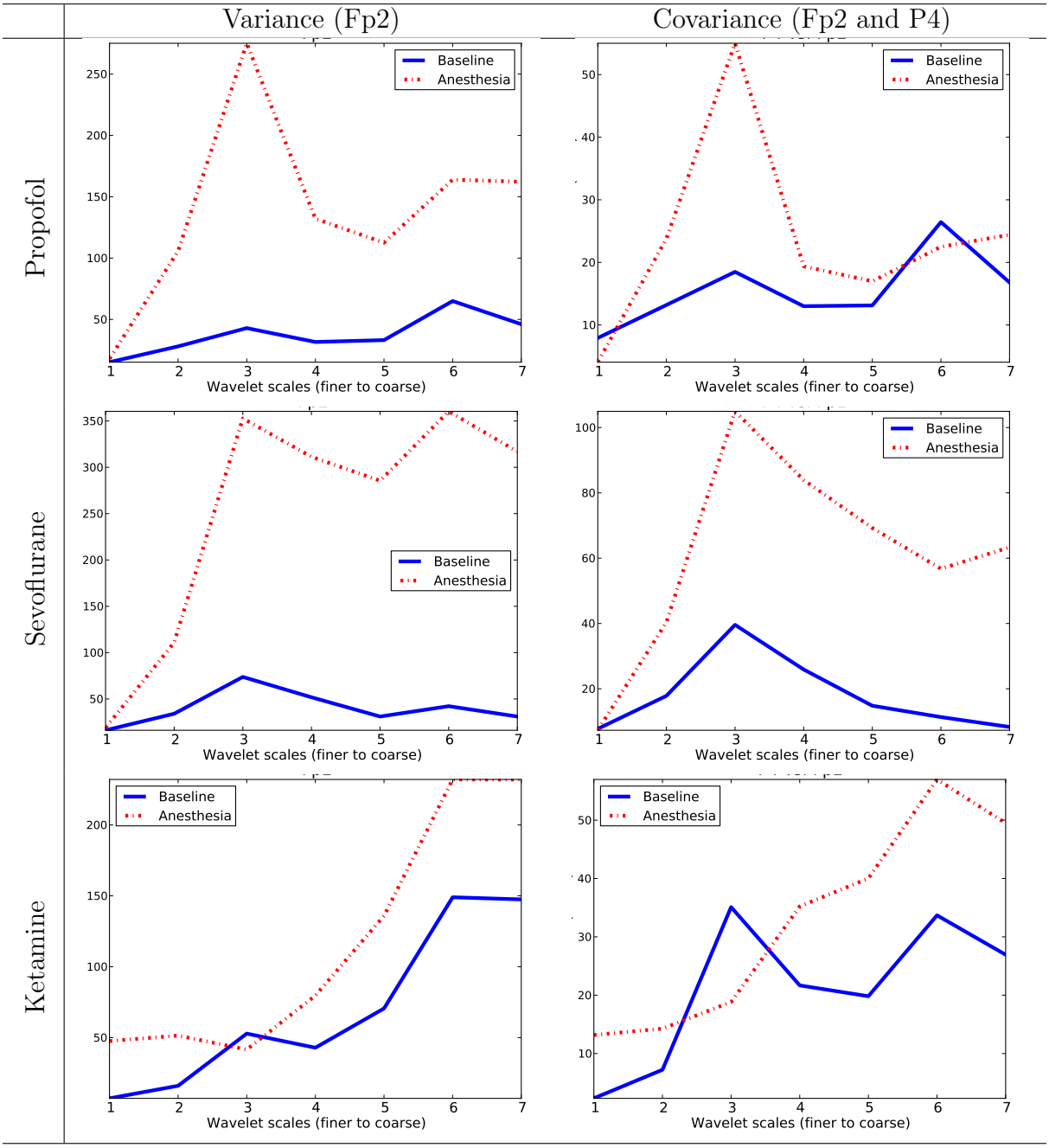


Figure 2.11: Average of wavelet variance and covariance over all subjects for selected locations: 1) Fp2 for the variance, 2) Fp2 and P4 for the covariance.

Connectivity pattern analysis:

As described in Section 2.2, at each wavelet level, we obtain an adjacency matrix where each element is the squared wavelet correlation difference between two conditions that passed a threshold. With these adjacency matrices, we draw network plots with 6 nodes that are specific regions of the brain. Two nodes are connected if there is significant change in association pattern between two locations. As a threshold, we use two different significance levels: nominal (0.05) and Bonferroni correction (0.05/15). Edges have two kinds, increasing and decreasing, depending on the direction of the correlation change after anesthesia. Based on the previous studies, connectivity changes among frontal and parietal regions (longer distance relationship) are particularly of interest[7]. Figure 2.12 shows the resulting network plots at each wavelet level.

We first focus on the association pattern between any frontal and parietal regions. In the δ and θ range, we see strong decreased connectivity under anesthesia between most of the pairs from frontal and parietal for all three anesthetics. In the α range, the association change pattern differs by anesthetics. The Propofol and Ketamine cases show similar patterns (decrease connectivity between any frontal and parietal), whereas Sevoflurane only shows a decreased change in any pairs between (Fp1, Fp2) and (P3, P4).

More distinctive changes in association pattern across different anesthetics occur in the β and γ range. In Propofol case, both β and γ range show the similar change pattern as that of α and δ range: decreased connectivity between any pairs from (Fp1, Fp2) to (P3, P4). Sevoflurane case does not show significant alternation between any frontal and parietal regions in these ranges. Ketamine, on the other hand, shows increasing connectivity between (Fp1, F3) and P3 in γ range and no significant change

[7]Previous studies consistently showed that there is disconnection between frontal and parietal regions especially for Propofol and Sevoflurane case (Ketamine case is still under study, *Alkire and Miller* (2005)).

in β range. This increasing trend is interesting because Ketamine is well known as a dissociative agent, that is it causes massively increased brain activity in order to arise unconsciousness, unlike Propofol or Sevoflurane that suppress the brain activity (*Alkire and Miller* (2005)).

The connectivity alternation pattern among any frontal regions vary according to both anesthetic agents and wavelet levels. In δ range, selective pairs of location in frontal region show increased connectivity for Propofol and Sevoflurane case. For Ketamine, there are increased connectivities in all possible location pairs among the frontal regions. In θ range, none of the anesthetics show significant change. In contrast, α range show decreased connectivity between frontal regions for Propofol and Ketamine case. The high frequency bands (γ and β ranges), Propofol case show decreased connectivity pattern among few pairs, and Sevoflurane show decresed (γ) and increased (β) connectivity pattern for a pair, respectively.

The connectivity change between two parietal regions are mostly non-significant. Few exceptions are γ range (decreased) and α range (increased) for Propofol, and δ range (decreased) for Ketamine.

2.5 Discussion

In this chapter, we have presented a correlation-based approach that explores how the association pattern among multiple time series alters under different conditions. The proposed approach utilizes wavelet analysis to allow the association pattern changes in different frequency bands to be identified. Wavelet correlation coefficients, which are calculated from each decomposed level, are used to show the connectivity change pattern at different frequency bands. Our method is applied to EEG data to analyze neural connectivity changes from a conscious state to general anesthesia-induced unconscious state. Our results suggest that there is different connectivity alteration pattern according to frequency band. Among three anesthetic

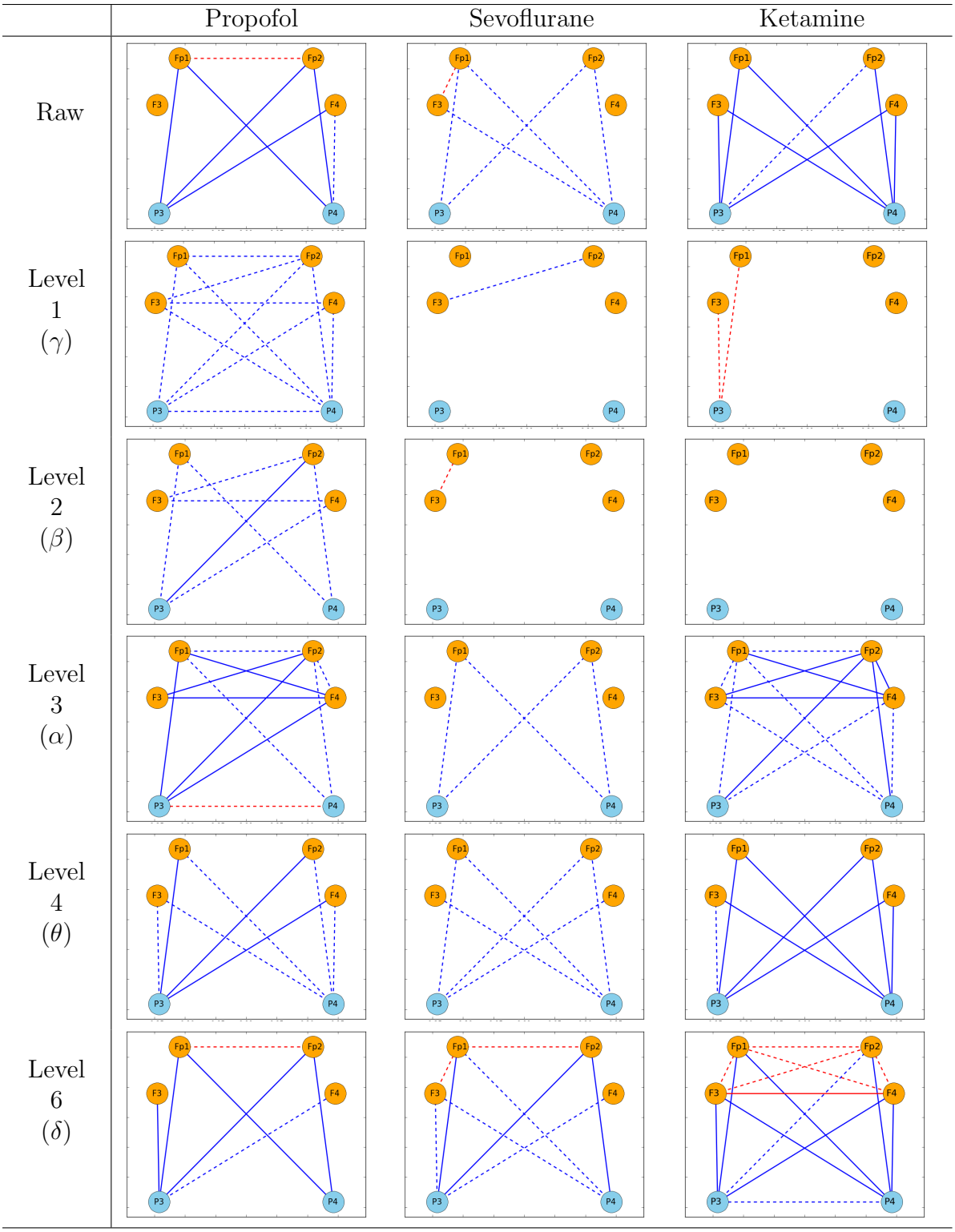


Figure 2.12: Network plot of the squared correlation differences, i.e. Anesthesia $Corr^2$ - Baseline $Corr^2$. Red and blue lines indicate that increased and decreased connectivity under anesthesia respectively. Solid lines are for location pairs whose p-value is lower than a Bonferonni significance level (0.05/15), and dashed lines are the pairs whose p-value is lower than a nominal significance level (0.05). Orange nodes indicate frontal locations and light blue nodes denote parietal locations of the brain.

agents, for Propofol and Sevoflurane case, we had consistent results with the previous literature. We observed that the association change pattern for the Ketamine case had some discrepancy from the other two agents. Though there is little existing literature using Ketamine agent, it is known that Ketamine have different anesthesia mechanism from the other two agents.

In conducting a discrete wavelet analysis, there are numerous possible families and subfamilies of wavelet functions that we can use. There are relatively few literature on how to select a wavelet family. Daubechies family is commonly used one in wavelet analysis (*Lindsay et al.* (1996)). Our choice of the wavelet family (Daubechies order 4) is based on *Adeli et al.* (2003), which focused on selecting a wavelet family in terms of the detection of epileptic transition in EEG data. Though our results showed meaningful discretion of association change pattern in different frequency bands, we can further investigate the sensitivity of the analysis to choice of wavelet for other wavelet family such as coiflets, symlets, and biorthogonal.

A possible extension of our method is using a modified version of the discrete wavelet transformation (DWT) that has potential to provide better resolution of the series. This modified DWT[8] uses highly redundant non-orthogonal bases unlike DWT. This transformation method has advantages such as 1) meaningful alignment between the transformed and original series (so called zero-phase filter), and 2) obtaining more efficient estimator for decomposed variance and covariance than DWT (*Percival and Mofjeld* (1997)). To apply this approach, we need to address how to deal with boundary issues.

This correlation-based framework provides the basis for a sensitivity analysis on the connectivity change, that is, assessing the effect of a signal from unobserved brain regions on the observed connectivity. The idea of this sensitivity analysis is based on

[8]The modified version of DWT is called by many different names: maximal overlap DWT (MODWT), undecimated DWT, stationary DWT, translation-invariant DWT, and time-invariant DWT (*Percival and Mofjeld* (1997)).

the hypothesis that $\mathbb{E}(Y|X,U) = \mathbb{E}(Y|U)$, where Y and X are observed signals and U is an unobserved signal. This approach has a potential to answer an interesting scientific question that has arisen by neuroscientists, e.g. "Is the observed connectivity solely resulting from the neural connectivity between two brain regions?".

CHAPTER III

A local transformation approach to revealing hidden dependence in multivariate time series

3.1 Introduction

Suppose we are interested in assessing the associations between pairs of series in a collection of time series. As an illustration, consider the two series shown in Figure 3.1 (a). We could calculate the Pearson correlation coefficient between the two series, ignoring the serial structure. As seen in the scatter plot Figure 3.1 (c), the Pearson correlation does not reveal much of a relationship. However, it is possible that a hidden relationship remains to be discovered, if we take account of the serial structure of the time series. One way to proceed is to recalculate the Pearson correlation after applying local transformations to the two series. A simple local transformation is to smooth the series over a bandwidth of w points. Figure 3.1 (b) shows the Pearson correlation after applying the smoothing transformation with different bandwidths. As can be seen in Figure 3.1 (d), the correlation increases as the two series are smoothed, up to a point where the correlation is maximized, and then the correlation decreases as the two series are further smoothed. Building on this idea, our goal will be to find a local transformation that can reveal such hidden associations among time series.

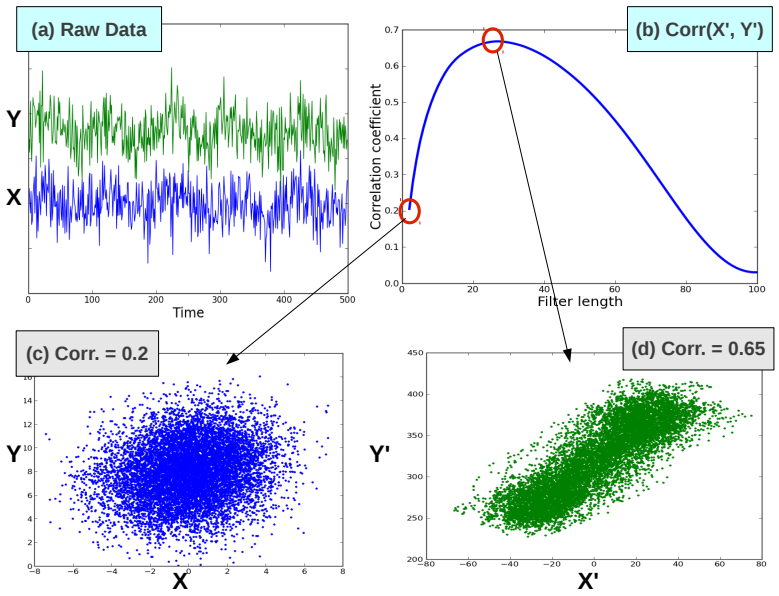


Figure 3.1: An illustrative example of the motivation. (a) Simulated two series X, Y (b) Correlation between transformed series X' and Y' using smoothing transformation with different bandwidth (c) Scatter plot between the raw series (d) Scatter plot between the transformed series

A transformation is one of the standard techniques in statistics that is used to facilitate the discovery of the relationships among variables. A familiar example of a transformation is the use of a log transform to remove certain types of mean/variance relationships. In regression analysis, the Box-Cox power transformations allow optimal transformations to be adaptively identified from data. More recently, the alternating conditional expectation (ACE) algorithm (*Breiman and Friedman* (1985)) finds coordinate transformations that give a best fit between the response and predictor variables. For the simple regression setting, the ACE algorithm estimates the maximal correlation. Note that the maximal correlation is one of dependence measure defined as

$$\arg\max_{f,g} Corr(f(X), g(Y)),$$

where f and g are real-valued functions for random variables X and Y, respectively. These transformation techniques are mainly developed for non-sequential data.

In this chapter, we propose methods that construct local transformations that

reveal some underlying association structure among the time series. In particular, we focus on revealing the hidden correlation, that is maximizing correlations between the transformed series. In order to handle the temporal structure, we consider local transformations that have the following form:

$$\begin{aligned} \widetilde{X}_t &= f(X_t, X_{t-1}, \ldots, X_{t-w}) \\ \widetilde{Y}_t &= g(Y_t, Y_{t-1}, \ldots, Y_{t-w}), \end{aligned}$$

where T is the total length of the series X, w is arbitrary time point less than T, and f, g are arbitrary functions. The proposed method adaptively finds the transformation that maximizes the correlation between $\tilde{X}$ and $\tilde{Y}$. In this analysis, we use a linear transformation.

The rest of this chapter is organized as follows. In Section 3.2, the characteristics of a filter structure are discussed. In Section 3.3, 3.4 and 3.5, we discuss our method in detail including relevant literature reviews. Then in Section 3.6, we provide simulation study that investigates the performance of our transformation method. Section 3.7 is about a real data application using EEG data from the Ketamine study. Lastly, Section 3.8 provides concluding remarks.

3.2 Linear time-invariant transformation

One widely studied class of transformation for time series is the linear time-invariant (LTI) transformation or linear filter. The LTI transformation is used in many fields. One of the main purposes of using the linear filter is noise reduction. For example, in biomedical engineering, decomposing electrocardiogram (ECG) to detect heartbeat uses extensive signal processing techniques utilizing linear filters (*Afonso et al.* (1999)). In image processing, the filter is used to sharpen the edges of an image (*Glasbey and Horgan* (1995)). Another example can be found in economics

for estimating business cycle for macroeconomics data (*Baxter and King* (1995)).

The filter produces an output series that highlights certain characteristics of an input series and suppresses others. A linear filter can be represented as the convolution of an input series $x(t)$ and specified coefficients $b(t)$:

$$\tilde{x}_t = \sum_{u=-\infty}^{\infty} b_u x_{t-u}, \quad t = 0m \pm 1, \ldots, \tag{3.2.1}$$

where b_u's are real number satisfying $\sum_{u=-\infty}^{\infty} b_u^2 < \infty$. The coefficient b_u that determines the filter is also called the impulse response function (*Koopmans* (1974)).

We are particularly interested in the finite-time finite filter coefficients in discrete time, also known as the finite-time digital filter or Finite Impulse Response (FIR) filter. The output series from this filter can be expressed as the following:

$$\begin{aligned} \tilde{X}(t) &= \sum_{u=0}^{T-1} \phi_1(t-u)X(u), \text{ for } t = 0, \pm 1, \cdots, T \\ \tilde{Y}(t) &= \sum_{u=0}^{T-1} \phi_2(t-u)Y(u), \text{ for } t = 0, \pm 1, \cdots, T, \end{aligned} \tag{3.2.2}$$

where $\phi_1 \in \mathbb{R}^{d_1}$ and $\phi_2 \in \mathbb{R}^{d_2}$ are finite filter coefficients for X and Y separately.

There are numerous literature on constructing linear filters. The non-adaptive method uses prespecified filter coefficients to reveal certain characteristic of an input series. For example, we can construct a filter that only allows the certain frequency band of an input series to pass through, so the output series contains the signal from the certain frequency band. The low-pass filter only passes low frequencies, the high-pass filter passes high frequencies, and the band-pass filter passes user specified frequency band of an input series. See Chapter 4 of *Koopmans* (1974) for details of these filter constructions. These types of filter are extensively used in image processing applications.

The adaptive way of constructing a filter involves some type of optimization

procedure to find its coefficients. Some examples of this kind of filter are Wiener-Kolmogorov filter, Kalman filter, and the matched filter. These approaches view an input signal as a mixture of signal and noise. The Wiener filter constructs the filter coefficients based on the minimum mean square of the error signal. The Kalman filter constantly updates the filter coefficients constructed from this least mean squares algorithm. The Kalman filter can handle nonstationary input series, whereas the Wiener filter only handles stationary input series. See Chapter 7 of *Hayes* (1996) for more details. The matched filter uses different optimization criteria. This filter constructs the coefficients that minimizes signal-to-noise ratio of a series (*Turin* (1960)). Note that the introduced examples are only considering one input series.

Our approach is an adaptive way of constructing the optimal local transformation that maximizes correlation between two series. Using the expression in (3.2.2), we can rewrite our objective as the following:

$$\text{Construct the filter } \phi_1 \text{ and } \phi_2 \text{ that maximizes } Corr(\tilde{X}, \tilde{Y}).$$

The filters ϕ_1 and ϕ_2 can be either different or the same. The construction of the these two cases are illustrated in following two sections. Furthermore, we extend the filter construction to multiple time series.

The length of the filter can be determined using external knowledge. As our focus is on the high frequency time series, the length of the filter ϕ_i should be long enough to capture some local characteristics of the input series. The high-dimensional filter would cause computational issue as well as identifiability. One way to avoid these issues is to construct the filter based on the B-spline basis. This way we reduce the dimension of unknown parameters from length of filter to the number of necessary knots for the spline basis. For example, if we want to apply a filter with length 50 with 5 knots, the number of unknown parameters is 5 instead of 50.

Any order of B-spline can be used, but for simplicity we use the linear B-spline of order 2. Also, we include a delta function in our basis in order to consider the raw series as a possible resulting filter. Figure 3.2 shows the shape of the bases used in our analysis.

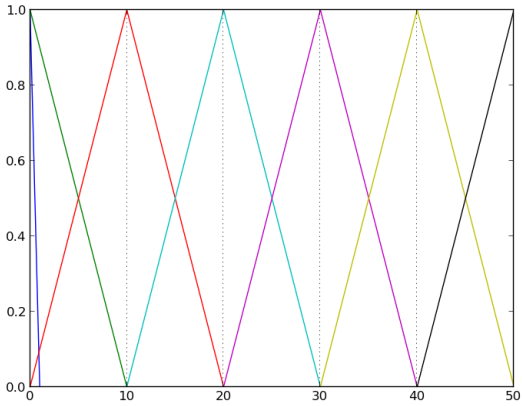


Figure 3.2: B-spline basis of order 2 with delta function (Blue).

3.3 Constructing series specific filter for a pair of series

In this section, we describe the method finding the optimal local transformations for two series. Here, we let each series have its own local transformation. Consider the following objective function:

$$\arg\max_{\phi_1,\phi_2\in\mathbb{R}^d} Corr(\phi_1 * X, \phi_2 * Y), \tag{3.3.1}$$

where * is a convolution operator and $d \in \mathbb{Z}_+$. The construction of two local transformations ϕ_1, ϕ_2 for each series X and Y can be achieved by using the Canonical Correlation Analysis (CCA). We begin this section by introducing CCA.

3.3.1 Canonical correlation analysis (CCA)

CCA is used to measure the linear association between two sets of variables. Since it was proposed by *Hotelling* (1936), CCA has been widely used in many fields such

as computer science, engineering, psychology, and economics. In particular, CCA has been extensively applied to image analysis (*Friman et al.* (2001); *Hardoon et al.* (2004); *Donner et al.* (2006); *Jing et al.* (2011)). In the literature, CCA is often presented as a dimension reduction technique such as Principal Components Analysis (PCA), Independent Component Analysis (ICA), among others. The main difference between CCA and PCA dimension reduction methods is that CCA finds a projection for a different basis for each set of the variables, whereas PCA considers only one basis for one set of variables. Specifically, CCA finds new coordinates for each set to maximize the correlation between the projection of the two sets of variables.

Let X and Y be vectors of random variables with p and q dimensions, respectively. Then we have mean vectors $\mu_X \in \mathbb{R}^p$ and $\mu_Y \in \mathbb{R}^q$, covariance matrices $\Sigma_{XX} \in \mathbb{R}^{p\times p}$, $\Sigma_{YY} \in \mathbb{R}^{q\times q}$ and $\Sigma_{XY} \in \mathbb{R}^{p\times q}$. Consider the linear combination of X and Y:

$$\tilde{X} = \alpha' X \text{ and } \tilde{Y} = \beta' Y.$$

Then the correlation between $\tilde{X}$ and $\tilde{Y}$ is the following:

$$Corr(\tilde{X}, \tilde{Y}) = \frac{\alpha' \Sigma_{XY} \beta}{(\alpha' \Sigma_{XX} \alpha)^{1/2} (\beta' \Sigma_{YY} \beta)^{1/2}} \tag{3.3.2}$$

CCA searches projections α and β that maximize (3.3.2). This problem is equivalent to the following:

$$\begin{aligned} &\arg\max_{\alpha,\beta} \alpha' \Sigma_{XY} \beta \\ &\text{subject to } \alpha' \Sigma_{XX} \alpha = \beta' \Sigma_{YY} \beta = 1. \end{aligned} \tag{3.3.3}$$

Finding a linear combination that maximizes (3.3.3) can be solved analytically.

The solutions are obtained from the following eigenvalue equations:

$$\Sigma_{XX}^{-1/2}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1/2}\alpha = \lambda\alpha$$
$$\Sigma_{YY}^{-1/2}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1/2}\beta = \lambda\beta. \tag{3.3.4}$$

The eigenvalue equation (3.3.4) can be solved by applying the singular vector decomposition (SVD) to the correlation matrix $\Sigma_{XX}^{-1/2}\Sigma_{XY}\Sigma_{YY}^{-1/2}$. The resulting r non-zero eigenvalues $\lambda_1 > \lambda_2 > \cdots, \lambda_r$ for $r = \min(p, q)$ correspond to the maximized correlation. We use the eigenvectors α and β corresponding to the first eigenvector as our linear filter. Sample covariance matrices for X, Y and XY will be calculated in the following way:

$$\hat{\Sigma}_{XX} = \frac{\sum_{i=1}^{n} X_i X_i'}{n},\ \hat{\Sigma}_{YY} = \frac{\sum_{i=1}^{n} Y_i Y_i'}{n},\ \hat{\Sigma}_{XY} = \frac{\sum_{i=1}^{n} X_i Y_i'}{n}, \tag{3.3.5}$$

where n is the sample size and mean vectors μ_x, μ_y are assumed to be zero. See Chapter 10 of *Johnson and Wichern* (2002) and Chapter 10.2 of *Brillinger* (1975) for details.

3.3.2 Our approach

In this subsection, we will show the formulation of our problem in terms of CCA. Given two series, the convolution of filter and a series can be represented as a matrix product. This matrices contain the time lagged series of each series respectively. If we view each matrix as a set of variables, the problem is finding optimal filters that maximizes correlation, and this can be solved using CCA framework. To be specific, let X, Y be infinite time series and l is the filter length. Then we can construct a collection of lagged series

$$X_m = [X_t, X_{t+1}, \cdots, X_{t+l}],\ Y_m = [Y_t, Y_{t+1}, \cdots, Y_{t+l}],$$

for any t. Using these restructured data matrices X_m, Y_m, we can express the convolution representation (3.2.2) as a multiplication between this data matrix and a filter coefficient vector. With the reconstructed matrix, we have the following mean and covariance:

$$\begin{pmatrix} X_m \\ Y_m \end{pmatrix} \sim \left(\begin{pmatrix} \mu \\ \nu \end{pmatrix}, \begin{pmatrix} \Sigma_{X_m X_m} & \Sigma_{X_m Y_m} \\ \Sigma_{Y_m X_m} & \Sigma_{Y_m Y_m} \end{pmatrix} \right),$$

where $\Sigma_{X_m X_m}$, $\Sigma_{Y_m Y_m}$ are autocovariance matrices and $\Sigma_{X_m Y_m}$, $\Sigma_{Y_m X_m}$ are cross-covariance matrices. The definition of autocovariance and cross-covariance for (i, j)th element are

$$\begin{aligned} [\Sigma_{X_m X_m}]_{ij} &= E[(X_i - \mu_i^X)(X_j - \mu_j^X)] \\ [\Sigma_{Y_m Y_m}]_{ij} &= E[(Y_i - \mu_i^Y)(Y_j - \mu_j^Y)] \\ [\Sigma_{X_m Y_m}]_{ij} &= E[(X_i - \mu_i^X)(Y_j - \mu_j^Y)], \end{aligned}$$

where μ_i^X and μ_i^Y are mean of X and Y at time point i. If both series were stationary, we have $\mu_i^X = \mu_j^X = \ldots = \mu^X$ and $\mu_i^Y = \mu_j^Y = \ldots = \mu^Y$ for all time points i, j.

The linear filter can be found by applying SVD on the correlation matrix

$$\Sigma_{X_m X_m}^{-1/2} \Sigma_{X_m Y_m} \Sigma_{Y_m Y_m}^{1/2}. \tag{3.3.6}$$

Let F_x, F_y be the resulting linear filter from SVD (right and left eigenvector respectively). It is well known that the filters F_x, F_y are the same if the correlation matrix (3.3.6) is symmetric. Hence, it is possible that we have the same resulting filters F_x, F_y. This result depends on the level of symmetry of $\Sigma_{X_m Y_m}$, since $\Sigma_{X_m X_m}$ and $\Sigma_{Y_m Y_m}$ are symmetric.

In classical CCA setting, it is not necessarily expected to have a symmetric covariance matrix between two sets of variables. However, in our setting, it is possible especially we deal with stationary time series. We will explore what type of time series may have a symmetric cross-covariance matrix using a vector autoregressive (VAR) process with two time series.

Suppose the following VAR(1) model:

$$Z_t = \nu + B_1 Z_{t-1} + u_t,\ t = 0, \pm 1, \pm 2, \ldots, \tag{3.3.7}$$

where $Z_t = (X_t, Y_t)'$ is 2×1 random vector, $\nu = (\nu_X, \nu_Y)$ is 2×1 fixed intercept vector, B_1 is (2×2) fixed coefficient matrix, and u_t is the white noise with $E(u_t) = 0$, $E(u_t u_t') = \Sigma_u$, $E(u_t u_s') = 0$ for $t \neq s$. In order to derive the cross-covariance expression, we first represent the series by an infinite sum of white noise

$$Z_t = \mu + \sum_{i=0}^{\infty} B_1^i u_{t-i},\ t = 0, \pm 1, \pm 2, \ldots,$$

where $\mu \equiv (I_2 - B_1)^{-1}\nu$ and B_1^i denote that B_1 is multiplied i times. The mean vector is $E(Z_t) = \mu$ for all t, and the cross-covariance with h time lag is

$$Cov(Z_t, Z_{t-h}) = \begin{bmatrix} Cov(X_t, X_{t-h}) & Cov(X_t, Y_{t-h}) \\ Cov(Y_t, X_{t-h}) & Cov(Y_t, Y_{t-h}) \end{bmatrix}. \tag{3.3.8}$$

See *Lütkepohl* (2005) for more details.

Symmetry of the cross-correlation can be checked by comparing the (1, 2)th and (2, 1)th elements of the cross-covariance matrix (3.3.8). Using (3.3.7), this cross-covariance matrix can be expressed as the following:

$$\begin{aligned} Cov(Z_t, Z_{t-h}) &\equiv E(Z_t - \mu)(Z_{t-h} - \mu)' \\ &= \lim_{T\to\infty} \sum_{i=0}^{T} \sum_{j=0}^{T} B_1^i E(u_{t-i} u_{t-h-j}')(B_1^j)' \\ &= \lim_{T\to\infty} \sum_{i=0}^{T} B_1^i E(u_{t-i} u_{t-h}') + B_1^i E(u_{t-i} u_{t-h+1}') B_1' + \cdots \\ &= \lim_{T\to\infty} B_1^h E(u_{t-h} u_{t-h}') + B_1^{h+1} E(u_{t-h+1} u_{t-h+1}') B_1' + \cdots \\ &= \lim_{T\to\infty} \sum_{i=0}^{T} B_1^{h+i} \Sigma_u (B_1^i)' = \sum_{i=0}^{\infty} B_1^{h+i} \Sigma_u (B_1^i)'. \end{aligned} \tag{3.3.9}$$

Here, Σ_u is a symmetry matrix. It is hard to control the symmetry of the cross-covariance matrix as the matrix would be more affected by the values in Σ_u than B_1. Therefore, we expect that the resulting linear filters for X and Y would have very similar shape in VAR(1) model case.

We close this subsection by introducing the reconstruction of time series with finite time. Consider a series X with T time points and a filter coefficient vector b with length l. Without considering boundary effects, the elements of the series from a convolution between X and b is the following:

$$\begin{aligned} t = l \quad & b_l x_1 + b_{l-1} x_2 + \cdots + b_1 x_l \\ t = l+1 \quad & b_l x_2 + b_2 x_3 + \cdots + b_1 x_{l+1} \\ & \vdots \\ t = T \quad & b_l x_{T-l+1} + b_2 x_{T-l+1} + b_1 x_T. \end{aligned} \tag{3.3.10}$$

This convoluted series (3.3.10) can be expressed by matrix multiplication $X_m b$, where $b = [b_l, b_2, \cdots, b_1]'$ and X_m is a restructured matrix from a series X as the following:

$$X_m = \begin{pmatrix} x_1 & x_2 & \cdots & x_l \\ x_2 & x_3 & \cdots & x_{(l+1)} \\ \vdots & \vdots & & \vdots \\ x_{(T-l+1)} & x_{(T-l+2)} & \cdots & x_T \end{pmatrix}_{(T-l+1)\times l}. \tag{3.3.11}$$

Note that the each column of X_m shows the lagged time series starting from the full series. Similarly, we can reconstruct another series Y and the reconstructed matrix is denoted as Y_m.

The sample covariance matrices for (3.3.4) can be computed from these restruc-

tured matrices X_m, Y_m and $X_m Y_m$ as the following:

$$\begin{aligned}
\hat{C}_{XX} &= \frac{1}{T-l+1}\begin{pmatrix} \sum_{k=1}^{T-l+1} x_k x_k & \sum_{k=1}^{T-l+1} x_k x_{k+1} & \cdots & \sum_{k=1}^{T-l+1} x_k x_{k+l-1} \\ \sum_{k=2}^{T-l+2} x_k x_{k-1} & \sum_{k=2}^{T-l+2} x_k x_k & \cdots & \sum_{k=2}^{T-l+2} x_k x_{k+l} \\ \vdots & \vdots & & \vdots \\ \sum_{k=l}^{T} x_k x_{k-l+1} & \sum_{k=l}^{T} x_k x_{k-l+2} & \cdots & \sum_{k=l}^{T} x_k x_k \end{pmatrix}_{l\times l} \\
\hat{C}_{YY} &= \frac{1}{T-l+1}\begin{pmatrix} \sum_{k=1}^{T-l+1} y_k y_k & \sum_{k=1}^{T-l+1} y_k y_{k+1} & \cdots & \sum_{k=1}^{T-l+1} y_k y_{k+l-1} \\ \sum_{k=2}^{T-l+2} y_k y_{k-1} & \sum_{k=2}^{T-l+2} y_k y_k & \cdots & \sum_{k=2}^{T-l+2} y_k y_{k+l} \\ \vdots & \vdots & & \vdots \\ \sum_{k=l}^{T} y_k y_{k-l+1} & \sum_{k=l}^{T} y_k y_{k-l+2} & \cdots & \sum_{k=l}^{T} y_k y_k \end{pmatrix}_{l\times l} \\
\hat{C}_{XY} &= \frac{1}{T-l+1}\begin{pmatrix} \sum_{k=1}^{T-l+1} x_k y_k & \sum_{k=1}^{T-l+1} x_k y_{k+1} & \cdots & \sum_{k=1}^{T-l+1} x_k y_{k+l-1} \\ \sum_{k=2}^{T-l+2} x_k y_{k-1} & \sum_{k=2}^{T-l+2} x_k y_k & \cdots & \sum_{k=2}^{T-l+2} x_k y_{k+l} \\ \vdots & \vdots & & \vdots \\ \sum_{k=l}^{T} x_k y_{k-l+1} & \sum_{k=l}^{T} x_k x_{k-l+2} & \cdots & \sum_{k=l}^{T} x_k y_k \end{pmatrix}_{l\times l}
\end{aligned} \tag{3.3.12}$$

Then using these sample covariance $\hat{C}_{xx}, \hat{C}_{yy}, \hat{C}_{xy}$, the linear filter ϕ_1, ϕ_2 are constructed from (3.3.4).

3.3.3 Brillinger's CCA extension to time series

Brillinger (1975) extended Hotelling's CCA to multivariate time series. Brillinger's CCA is interested in finding filters that measure the association between two sets of multivariate time series. Consider two vector-valued stationary time series $X_t \in \mathbb{R}^{d_1}$ and $Y_t \in \mathbb{R}^{d_2}$ for $t = 0, \pm 1, \ldots$. The transformed series are as the following:

$$\begin{aligned}
\tilde{X}_t &= \textstyle\sum_u \phi_{t-u} X_u \\
\tilde{Y}_t &= \mu + \textstyle\sum_u \psi_{t-u} \tilde{X}_u,
\end{aligned}$$

where $\{\phi_t\}$ is $q \times d_1$ filter, $\{\psi_t\}$ is $d_2 \times q$ filter and μ is a d_2 vector of constants.

Brillinger approached this problem as the least squares set up. That is, the filter $\{\phi_t\}$ and $\{\psi_t\}$ are constructed to minimize

$$\mathbb{E}[(Y_t - \tilde{Y}_t)'(Y_t - \tilde{Y}_t)]. \tag{3.3.13}$$

It can be shown that this approach leads to the same as (3.3.3). See the proof of Theorem 10.2 in *Brillinger* (1975) for more details.

Given spectral density matrices[1] $f_{XX}(\lambda)$, $f_{YX}(\lambda)$, $f_{XY}(\lambda)$ and $f_{YY}(\lambda)$ at the frequency λ, the optimum filters $\{\phi_t\}$ and $\{\psi_t\}$ are constructed using eigenvectors $V_k(\lambda)$ of the matrix

$$f_{YX}(\lambda)f_{XX}^{-1}(\lambda)f_{XY}(\lambda).$$

The resulting filters are the following:

$$\begin{aligned} \phi_t &= \tfrac{1}{2\pi}\textstyle\int_0^{2\pi} B(\lambda)\exp(itu)d\lambda \\ \psi_t &= \tfrac{1}{2\pi}\textstyle\int_0^{2\pi} C(\lambda)\exp(itu)d\lambda, \end{aligned}$$

where

$$\begin{aligned} B(\lambda) &= \begin{bmatrix} \overline{V_1(\lambda)'} \\ \vdots \\ \overline{V_q(\lambda)'} \end{bmatrix} f_{YX}(\lambda)f_{XX}(\lambda)^{-1} \\ C(\lambda) &= [V_1(\lambda), \ldots, V_q(\lambda)]\,. \end{aligned}$$

The main difference in Brillinger's approach from ours is in the construction of the covariance matrix. Brillinger's method focuses on the dimension reduction of multiple time series, so the filter is constructed for the whole series. In other words, the covariance matrices or spectral densities are constructed using cross-covariance functions between two different series. On the other hand, our filter is designed

[1]The spectral density matrix is the Fourier transformation of the cross-covariance matrix. To be specific, the off diagonal elements of the spectral density of f_{XX} are cross-spectrum of series X_i, X_j for $i \neq j$, and the diagonal elements are power spectrum of X_i. Other spectral density matrices are constructed similar way.

to capture localized structure, so the covariance matrix contains cross-covariance function between two different time points in a series.

3.4 Constructing pairwise common filter

In this section, we discuss the method that constructs a filter $\phi \in \mathbb{R}^d$ that maximizes correlation between two time series. For given two series X and Y, the objective function is the following:

$$\arg\max_{\phi \in A} Corr(\phi * X, \phi * Y), \tag{3.4.1}$$

where $A \subset \mathbb{R}^d$ for $d \in \mathbb{Z}_+$.

Optimization algorithm

We propose a heuristic algorithm that searches for the optimum filter coefficients. Our algorithm adopts the idea of random search algorithms[2]. A random search algorithm is a sequential optimization algorithm that moves the feasible space by a randomly selected direction. This algorithm is particularly used for optimization problems with non-convex objective function (*Schumer and Steiglitz* (1968); *White* (1971); *Zabinsky et al.* (2010)). The new candidate point in each iteration is determined by two components: direction and step size. For step size determination, we consider the fixed step size random search proposed by **?**.

Our algorithm generates a direction vector at each iteration from the standard normal distribution with unit length. Mainly our algorithm moves a unit length, if there is no improvement we decrease the length by half until either the improvement is made or the length reaches near 0. Our algorithm is following the structure of Improving Hit-and-Run (IHR) algorithm proposed by *Zabinsky et al.* (1993) with a

[2]This algorithm can also be called as a Monte Carlo method or a stochastic algorithm (*Zabinsky et al.* (2010))

modification in the step when improvement is not made. Pseudo code is given in III.2.

Algorithm III.1 *Conv.Corr*: A function that construct a filter and calculate correlation coefficient between transformed series

Require: Two time series X, Y, a B-spline basis B, a starting point $S \in \mathbb{R}^d$

1: Construct a filter $\phi_k = BS$.
2: Compute convolutions: $\tilde{X} = \phi_k * X$, $\tilde{Y} = \phi_k * Y$
3: **return** $Corr(\tilde{X}, \tilde{Y})$.

Algorithm III.2 Heuristic Up Hill Algorithm

Require: Time series X, Y and a B-spline matrix B.

1: Initialize $S_k \in \mathbb{R}^d$.
2: **repeat**
3: Compute a transformed correlation: $C_k = Conv.Corr(X, Y, B, S_k)$.
4: Generate a direction r from the standard normal distribution.
5: **if** $Conv.Corr(X, Y, S_k + r, B) > C_k$ **then**
6: Set $S_{k+1} = S_k + r$.
7: **else**
8: Set $j = 1$.
9: **while** $Conv.Corr(X, Y, S_k + r/2^j) < C_k$ **and** $j < J$ **do**
10: Set $r_j = r/2^j$, compute $Conv.Corr$ and increment j
11: **if** $C_{kj} > C_k$ **then**
12: Set $S_{k+1} = S_k + r$.
13: **else**
14: Set $S_{k+1} = S_k$.
15: **end if**
16: **end while**
17: **end if**
18: **until**
19: Stopping rule met.

In addition, we use the n-dimensional spherical coordinate system with a unit radius. The conversion to Cartesian coordinates can be done by the formula derived by *Blumenson* (1960). This coordinate conversion allow us to have one reduced dimension for ϕ.

Visualizing the surface

In order to explore the resulting correlation surface, we developed a visualization tool

that explores the surface around a local maximum point. We use Principle Component Analysis (PCA) to set a direction for the exploration. As a start, we repeatedly apply the Heuristic up hill algorithm for multiple starting points. PCA is applied on the resulting optimum coefficients. Let M be the optimized filter coefficients at the maximum correlation from all the iterations. Then we compute correlation coefficient from transformed series with filter coefficients from

$$M + a_1 PC1 + a_2 PC2,$$

where a_1, a_2 are grid points over certain range. A contour plot can be made using the correlation values along the a_1 and a_2. This contour plot can be made for different PC combinations.

3.5 Constructing an optimal filter for multiple series

In this section, we discuss constructing a filter that reveals hidden association structure in multiple time series. In order to extend to the multiple series, we use a summary measure of the correlations from all possible pairs from the multiple time series. This approach is analogous to the idea of the Projection pursuit algorithm. Projection pursuit algorithm is designed to find a lower dimensional projection that reveals some interesting structure of the original high dimensional data. This projection is constructed through optimizing over a certain projection index. The projection index are defined to construct a "useful" projection, so it can be different for each applications (*Friedman and Tukey* (1974)).

The linear filter we try to construct has the similar purpose as the projection pursuit algorithm: reveal a hidden structure. Hence, we refer this approach, constructing a filter for multiple series, as Projection Pursuit (PP) approach. In this study we consider two projection indexes: mean and variance of transformed correlations.

Suppose that X is a collection of multiple time series. Following is the objective function using the mean projection index:

$$\arg\max_{\phi \in A} \frac{\sum_{(k1,k2)\in\mathcal{K}} Corr(\phi * X_{k1}, \phi * X_{k2})}{|\mathcal{K}|}, \tag{3.5.1}$$

where $\mathcal{K}$ is the set of indexes for all possible pairs in X, $|S|$ is the cardinality of a set S, and $A \subset \mathbb{R}^d$ for $d \in \mathbb{Z}_+$. Similarly, we can obtain the objective function with the variance index. Then we use the heuristic algorithm (III.2) to find a filter ϕ.

3.6 Simulation Study

In this simulation study, we explore the characteristics of the linear filter constructed from our methods. We consider four scenarios to generate time series, that carries unique characteristics. The simulation scenarios are as following:

1. *Difference*: Two series that one is the difference of the other

$$\begin{aligned} X &\sim \text{Normal}(0,1) \\ Y_t &= X_t - X_{t-1}, \text{ for } t = 2, 3, \ldots \end{aligned}$$

2. *Time Delay*: Two series generated from VAR(1) model.

$$\begin{aligned} E_t &\sim \text{Multivariate Normal}(0, \Sigma_E) \\ X_t &= 0.2X_{t-1} + 2Y_{t-1} + E_t^x \\ Y_t &= 0.2X_{t-1} + 0.5Y_{t-1} + E_t^y \end{aligned}$$

3. *Low frequency shared (LFS)*: Two sinusoid that share the same structure, but have independent errors.

$$\begin{aligned} X &= \sin(2\pi ft) + 0.5E_x \\ Y &= \sin(2\pi ft) + 0.5E_y, \end{aligned}$$

where $f = 0.01$Hz and E_x, E_y are standard normal errors.

4. *High frequency shared (HFS)*: Two sinusoid that have distinctive structure (different frequency and phase), but share the same error.

$$\begin{aligned} X &= \sin(2\pi \cdot ft) + 0.3 * E \\ Y &= \sin(2\pi \cdot 3ft + \pi/4) + 0.3 * E, \end{aligned}$$

where $f = 0.1$Hz and E is standard normal error.

All simulated series have length of 10000.

The first scenario, a series Y is the difference of values between consecutive time points of another series X. The scenarios *Low Frequency Shared* and *High Frequency Shared*, we view the series consists of two frequency component. For illustrative purpose, we let the series in both scenarios share one of the frequency components. To be specific, the smoothing case share the low frequency component, and the contrasting case share the high frequency component. Similarly, the series in time delay case are constructed to have high emphasis on the effect of Y_{t-1} to X_t. Then in all of the scenarios, we expect that the well constructed linear filter will give a perfect correlation. Figure 3.3 shows the simulated series according to the scenarios and the scatter plots.

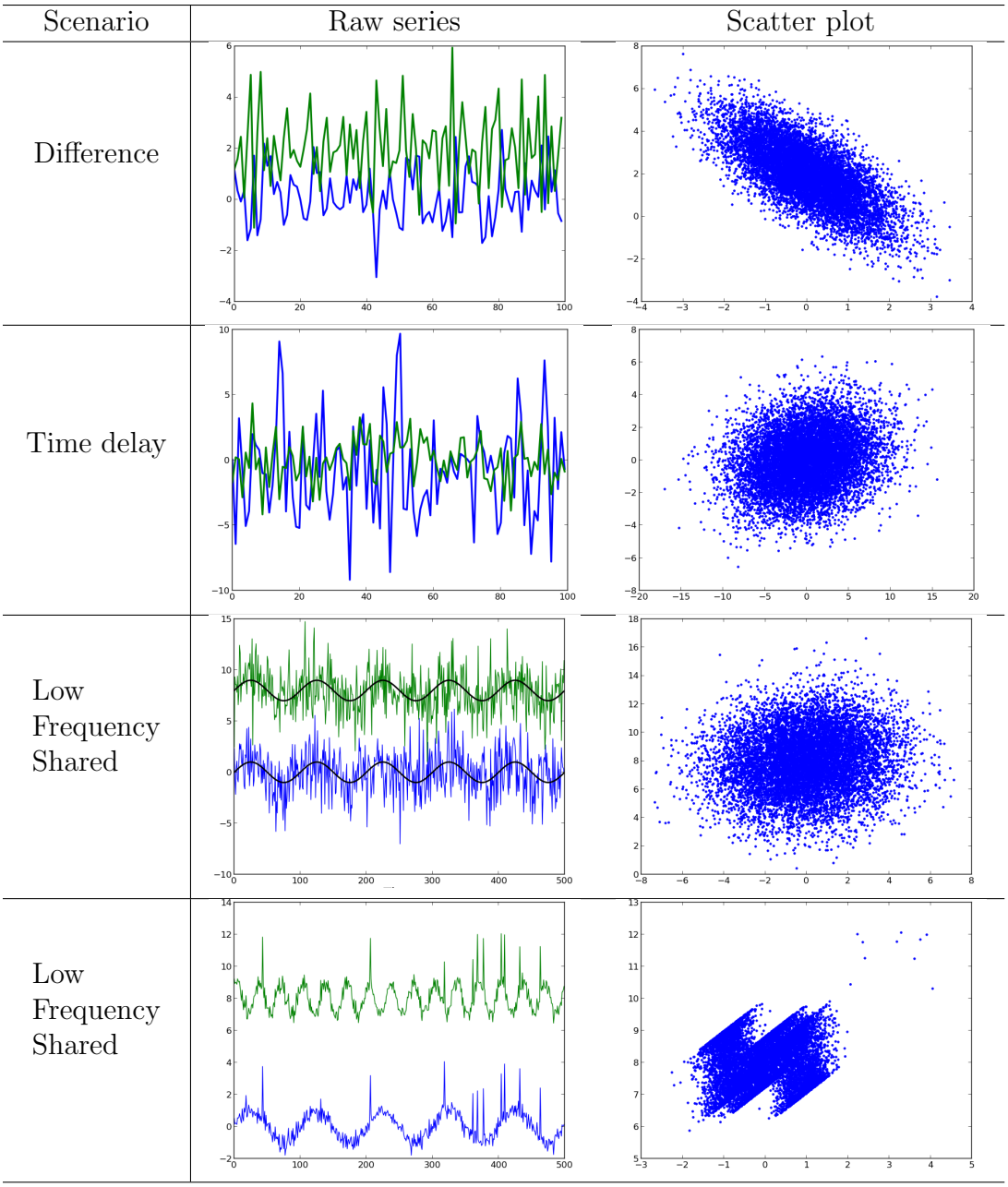


Figure 3.3: Simulated series for each scenario and the scatter plot for those two series. For raw series plot, we show first 100 or 500 time points of the series.

3.6.1 Series-specific filter

In this subsection, we show the simulation results of using the series specific filters. In order to compare the smoothness of the filters, we used two kinds of basis function: 1) B-spline of order 2 (linear) with 5 knots, and 2) identity matrix. The identity basis

function is equivalent to impose no restriction on smoothness of the filter. We refer each basis function as B-spline and Free, respectively.

Table 3.1 shows the correlation between two raw series and between two output series. The correlation between output series increased quite dramatically in all scenario but the Time delay. There are very little difference in these optimized correlation between B-spline and Free basis functions.

Table 3.1: Correlation coefficient computed from the raw series and output series from the series-specific filters.

Scenario	Difference	Time delay	LFS	HFS
Raw	-0.707	0.214	0.117	0.153
B-spline	1.000	0.291	0.793	1.0
Free	1.000	0.341	0.788	1.000

Figure 3.4 shows the resulting series-specific filters for each four scenarios. In the Difference and the Time delay scenarios, not only we see fairly different shape of filters for both series, but also using B-splines and Free basis functions. The filters for each series in the Low Frequency Shared and High Frequency Shared scenarios are almost the same regardless of the basis functions.

3.6.1.1 Model selection

In this subsection, we explore the filters in different length. We focus on the constrained filter case. The evaluation is on the optimized correlation coefficient and over-fitting. In the first part of evaluation, the effect of filter length, considered filter lengths are 5, 25, 50, 100, 500, 1000 and 3000, which are 0.05%, 0.25%, 0.5%, 1%, 5%, 10% and 30% of the filter respectively. Due to an issue with covariance structure, being non-positive definite, for high dimensional filters, we only consider simulated series from Time Delay, High Frequency Shared (HFS) and Low Frequency Shared (LFS) scenarios. Also, for the high dimensional case (length over 50), we use B-splines basis for the filters. Number of knots for the B-splines basis are set to be either differ

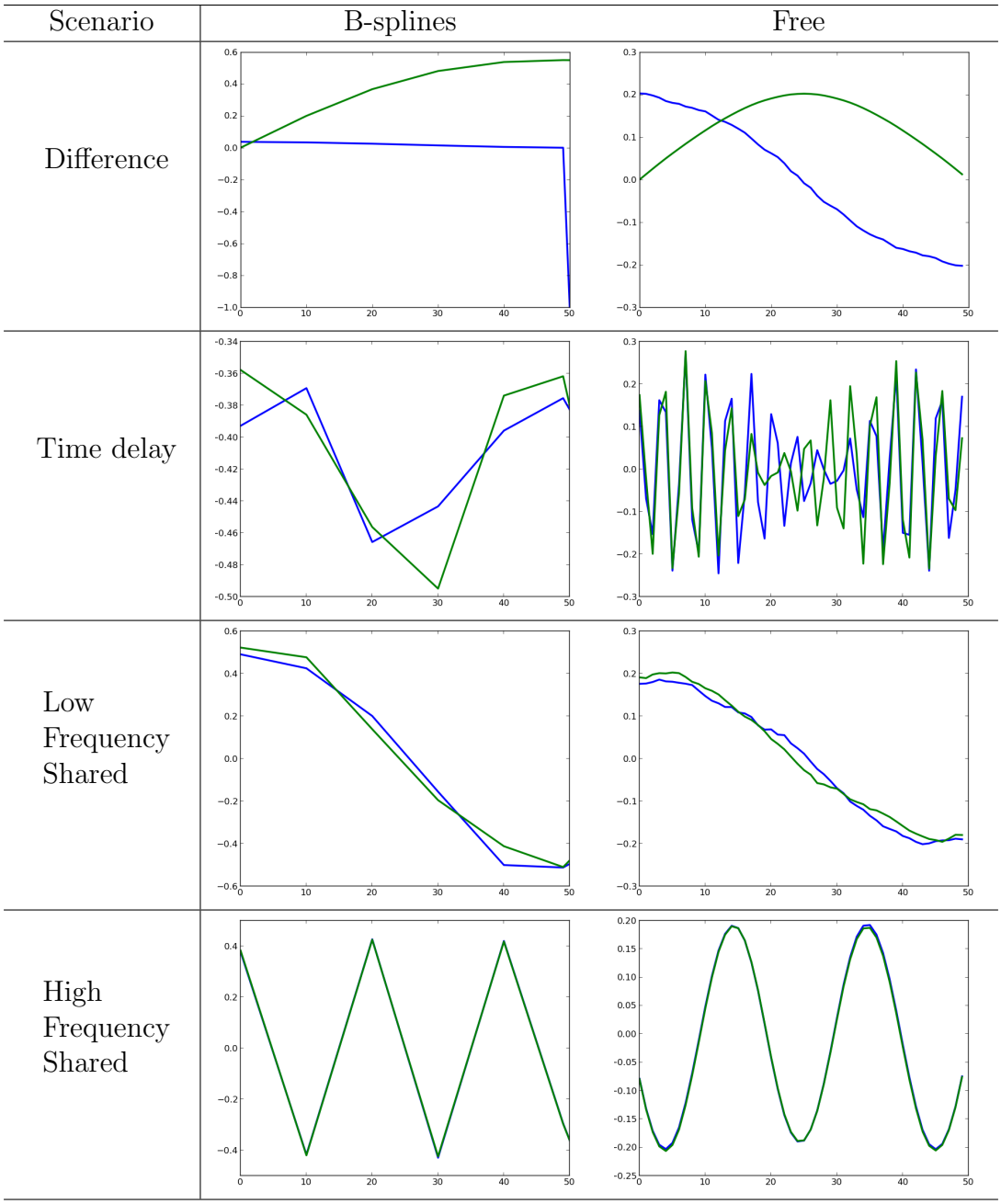


Figure 3.4: Constructed filters for two series from each simulation scenario. The first column shows the filter constructed using B-spline basis with order 2, and the second column shows the free parametrized filter. Blue line is the resulting filter for series X, and green line is the filter for series Y.

by the filter length and a fixed number. In the different know case, the knot size is determined to have a 10 time points between knots. For example, there are 5 knots for the filter length 50 and 100 knots for the filter length 1000. For the fixed knot

case, we use 5 knots.

Figure 3.5 shows the correlation between output series for the three scenarios. For each scenario, we plotted two knot cases. The solid line shows the resulting correlation for the different knot case, and the dashed line shows the result for the fixed knot case. Both the differed and fixed knot cases will have the same resulting correlations for the filter length 5, 25 and 50. This is because the B-splines basis is not used for the length 5 and 25, and there is the same number of knots for the length 50. In the Time Delay scenario, increasing filter length leads to the increased optimum correlation regardless of knot size. However, if you have more knots for the basis, the optimum correlation tend to be higher. On the other hand, though the change is very small. The optimized correlation has a decreasing tendency as the filter length increases in the HFS scenario regardless of knot size. If the knot size increases, it seems that the noticeable drops in the correlation happen in the shorter filter length. Lastly, the LFS scenario shows two different pattern.

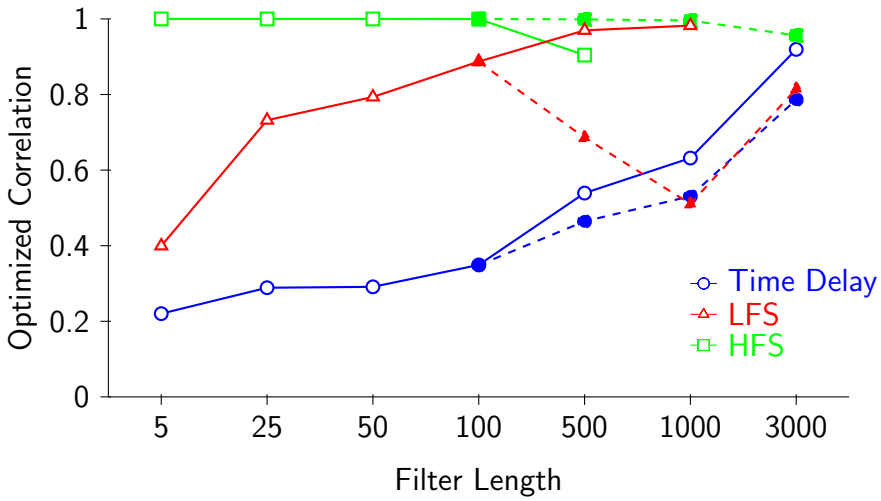


Figure 3.5: Optimized correlation result for different filter length. The solid lines with white filled marks are the results from the changing number of knots by filter length. The dashed lines with color filled marks are the results from the case where the number of knots is fixed to be 5 regardless of filter length. The optimized correlation for the length 5, 25, and 50 are the same for both cases.

The overfitting of the constructed filter is assessed via cross validation analysis.

For cross validation, we cut the series into half and construct a filter with the first half of the series (training). Then the constructed filter is applied to the second half of the series (testing). As an overfitting measure, we use the absolute difference between the two optimized correlations, training and testing series. Since we are using half of the series, the covariance structure issue mentioned in the length study case happens for shorter filter length. Hence, the simulation is performed for the filter length up to 500. Similar as the previous length study, we consider the two cases of knot size: differ by filter length and fixed to be 5.

Table 3.2 shows the mean absolute difference (MAD) from 100 repeats for each scenario and filter length. In most cases, the magnitude of absolute difference is somewhat negligible, especially in the High Frequency Shared scenario. The Time Delay scenario showed the largest MAD and it increases as the filter length increase.

Table 3.2: Mean absolute difference between optimized correlation from training and test sample.

Filter Length	Time delay	LFS	HFS
5	0.02087	0.02300	1.18E-07
25	0.05553	0.02547	1.10E-06
50	0.05445	0.02676	3.43E-06
100	0.08004	0.01969	1.32E-03
500	-	0.01226	-

3.6.2 Pairwise common filter

In this subsection, we explore the performance of pairwise common filter, or the same filter for two series. B-splines of order 2 basis with 5 knots is used to construct the filter, in order to reduce the dimension of the filter. Also, we set the length of filter to be 50.

We first check the stability of our heuristic algorithm as well as the surface of the filtered correlations. For given two simulated series, we repeat the optimization procedure using the Heuristic up-hill algorithm for 100 different starting points. The

results of those 100 optimized filter and correlations are shown in Figure 3.6.

The left column of Figure 3.6 shows resulting the filters that gives optimized correlation near the maximum. As can be seen, the resulting filters are clustered into 2 groups[3] The filters in each cluster are mirror image to each other for most of the scenarios. The filters in Difference scenario shows few exceptions. Though we don't have nearly perfect two clusters in the Difference scenario, the filters at least share the similar pattern. The right column of Figure 3.6, shows the distribution of optimized correlation from 100 repeats. Both Time delay and Smoothing scenarios reached the same optimal correlation value from any starting point. For Difference and Contrasting scenarios, 70% of the starting points reached the similar optimized correlation value.

Table 3.3 shows the optimized correlation coefficient from the pairwise common filter. The pairwise common filter generally gives the lower optimized correlation than the series-specific filter. In particular, the Difference scenario showed the biggest gap in the optimized correlations between series-specific and pairwise common filter. This result is somewhat expected as the Difference scenario simulation is constructed to have two different filter for each series. For other scenarios, the optimized correlations from the common filter are similar to the optimized correlation from the B-spline series-specific filter.

Table 3.3: Correlation coefficient computed from the raw series and output series from the pairwise common filter.

Scenario	Difference	Time delay	LFS	HFS
Raw	-0.707	0.214	0.117	0.153
Optimum	0.001	0.292	0.719	0.878

Surfaces of the filtered correlations around the maximum filter are shown in Figure 3.7. The bottom part of the figure is contour plots of shows the correlation coefficients at each principal components (PC) coordinates. In each plot, we set the

[3]We applied k-means clustering using filter coefficients.

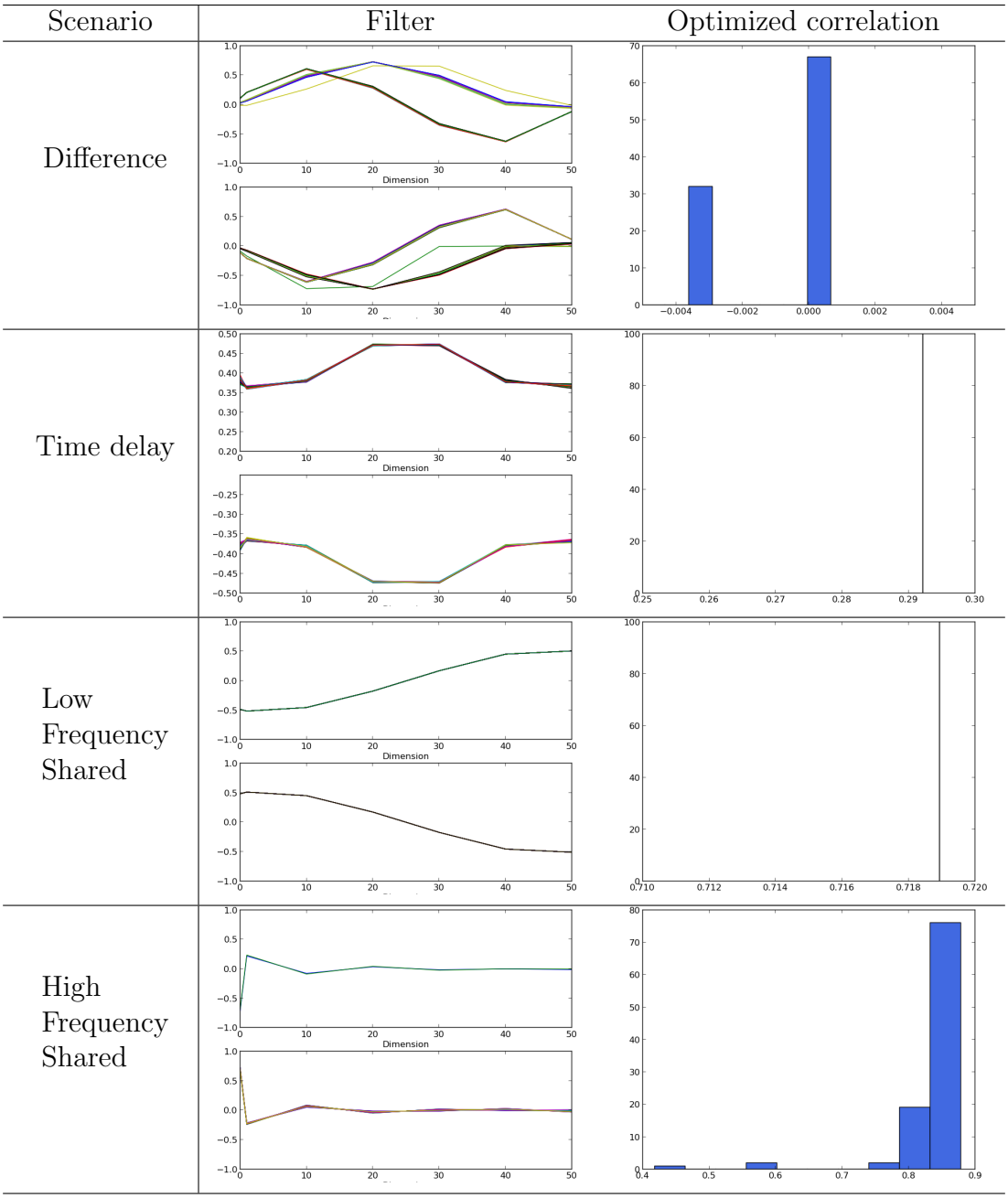


Figure 3.6: Resulting filters for each scenario. Each simulation were repeated for 100 different starting points. First column shows clustering analysis of the resulting filters. Second column shows histograms of the resulting correlation between output series.

(0,0) coordinate to be the filter that gives the maximized correlation value. We can see the multiple modes in the three scenarios - Difference, Time delay and Smoothing, and the area of modes are quite large. The Contrasting scenario, on the other hand,

have very narrow region that gives maximized correlation.

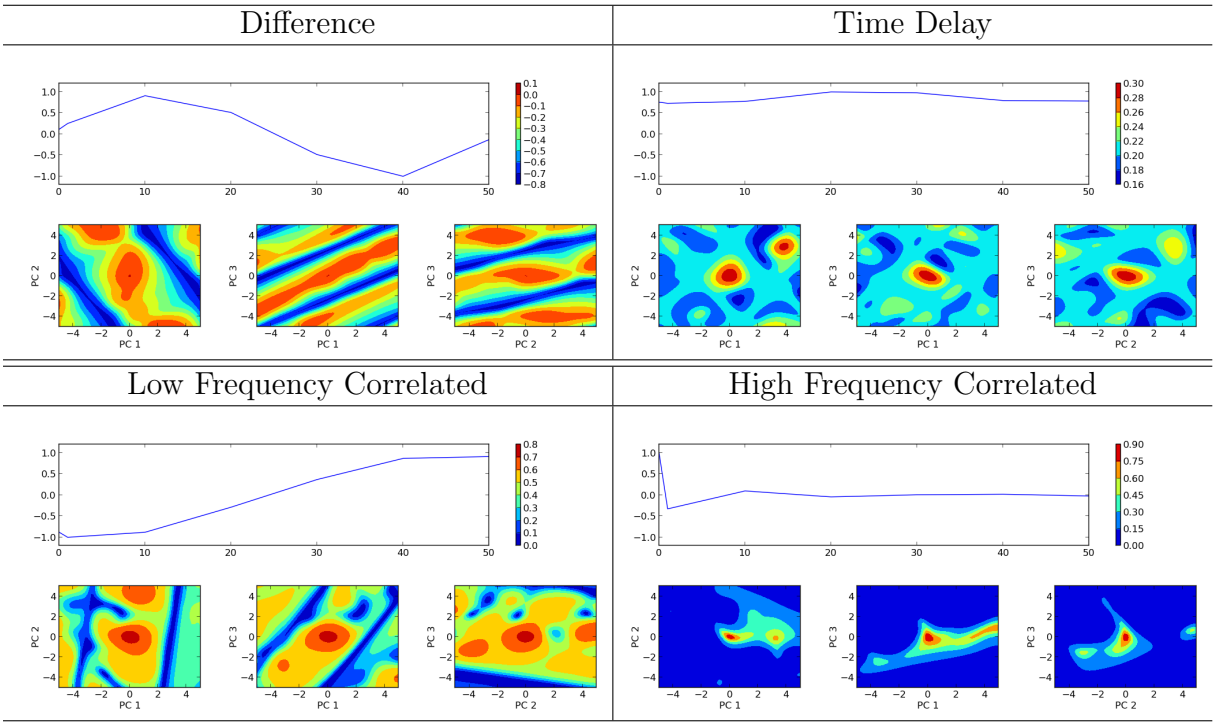


Figure 3.7: Resulting filters and the response surfaces using the pairwise same filter approach. Results from the four simulation studies are shown. The length of filter is set to be 50 with 5 knots.

3.6.3 Multiple series with same filter

In this simulation study, we explore the performance of the projection pursuit approach that constructs a linear filter for multiple time series. We generate six time series with a hidden connectivity structure. There are two groups of series that are highly correlated each other within each group, and independent across the group. Table 3.4 shows the summary of the simulation.

Figure 3.8 shows the correlation between all possible pairs from raw data series. most of these correlations are very small (up to 2∼3 decimal points) due to large errors. From this plot, the group structure of the series is not seen that clear. If we highlight the pairs with the largest correlation, we see spurious relationships between X_2, X_3 and Y_1. Still the interpretation of the connectivity is somewhat meaningless

Table 3.4: Summary of the simulation

Group	1	2
No. of series	4	2
Formula	$\sin(2\pi ft) + 5\epsilon$	$\sin(2\pi \cdot 3ft + \pi/4) + 5\epsilon$
Notation	X_1, X_2, X_3, X_4	Y_1, Y_2

* $f = 0.01$Hz, $t = 1, \ldots, 10000$ and $\epsilon \sim N(0, 1)$.

since the magnitudes of correlation are very small.

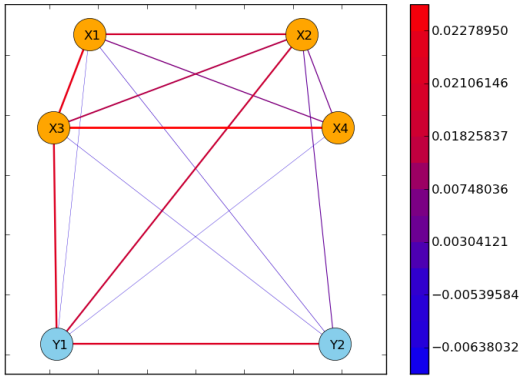


Figure 3.8: Correlation coefficients of possible pairs from the simulated data.

We consider two filter lengths 50 and 200 with alternating knot size 5 and 20, respectively. The filter is constructed from the Heuristic up-hill algorithm using mean and variance projection indexes. Figure 3.9 shows the optimized correlation among the six series. With filter length 50, we see that the correlation among X_1, X_2, X_3, X_4 are increased to at least 0.3, and correlations between any X_i and Y_j are less than 0.1. The optimized correlation recovered the relationships among Xs, but the relationship between Y_1 and Y_2 was not recovered. Note that, the mean and variance index gave very similar results. Similarly, with filter length 200, we reveal the highly correlated relationships among Xs. The magnitude of the optimized correlations are much higher than that of the filter length 50.

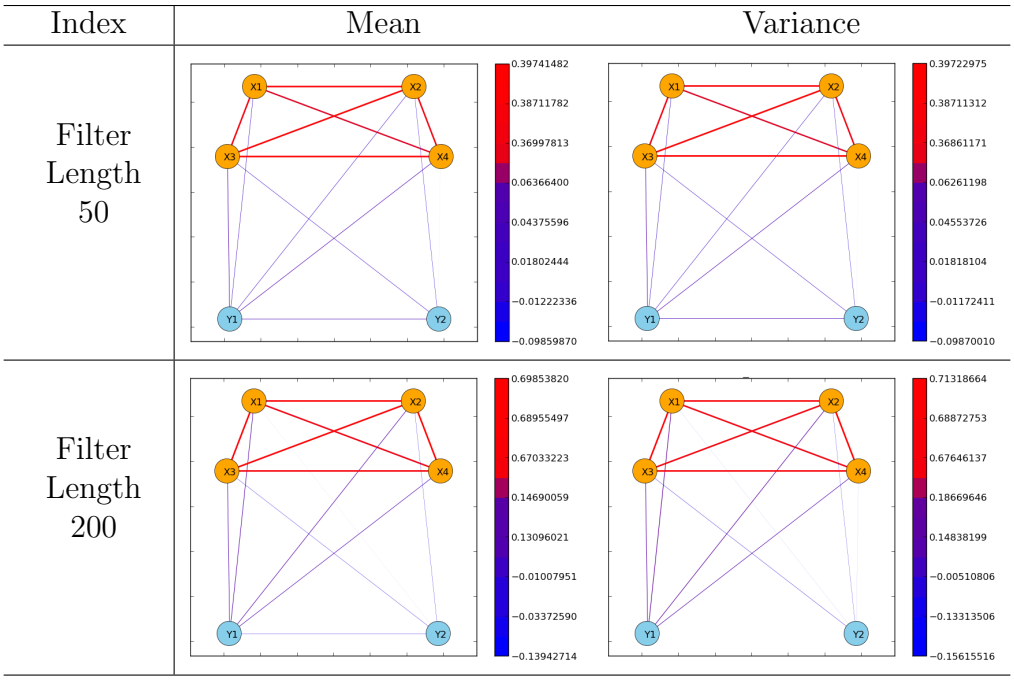


Figure 3.9: Optimized correlation from the projection pursuit approach with simulated data. Mean and variance indexes are shown for filter length 50 and 200 with knot size 5 and 20, respectively.

Table 3.5: Mean and Standard deviation of the resulting correlations for each projection indexes.

Index	Mean	Variance
Raw	0.008 (0.012)	
FL*= 50	0.151 (0.189)	0.142 (0.182)
FL*= 200	0.280 (0.326)	0.276 (0.340)

*FL = Filter Length

3.7 Application to EEG data

In this section, we illustrate the use of optimal local transformations with data from the Ketamine study. This data set contains EEG recordings of 29 subjects. For illustration, we use the pre-treatment (awake) portion of the data. For each subject, 5 minute duration EEG recordings were made in 6 locations in the brain (4 frontal and 2 parietal). Our goal is to assess whether optimal local transformations can be used to aid in understanding the neural connectivity in this baseline resting state. Throughout the analysis, the time interval for the local transformation, or the filter

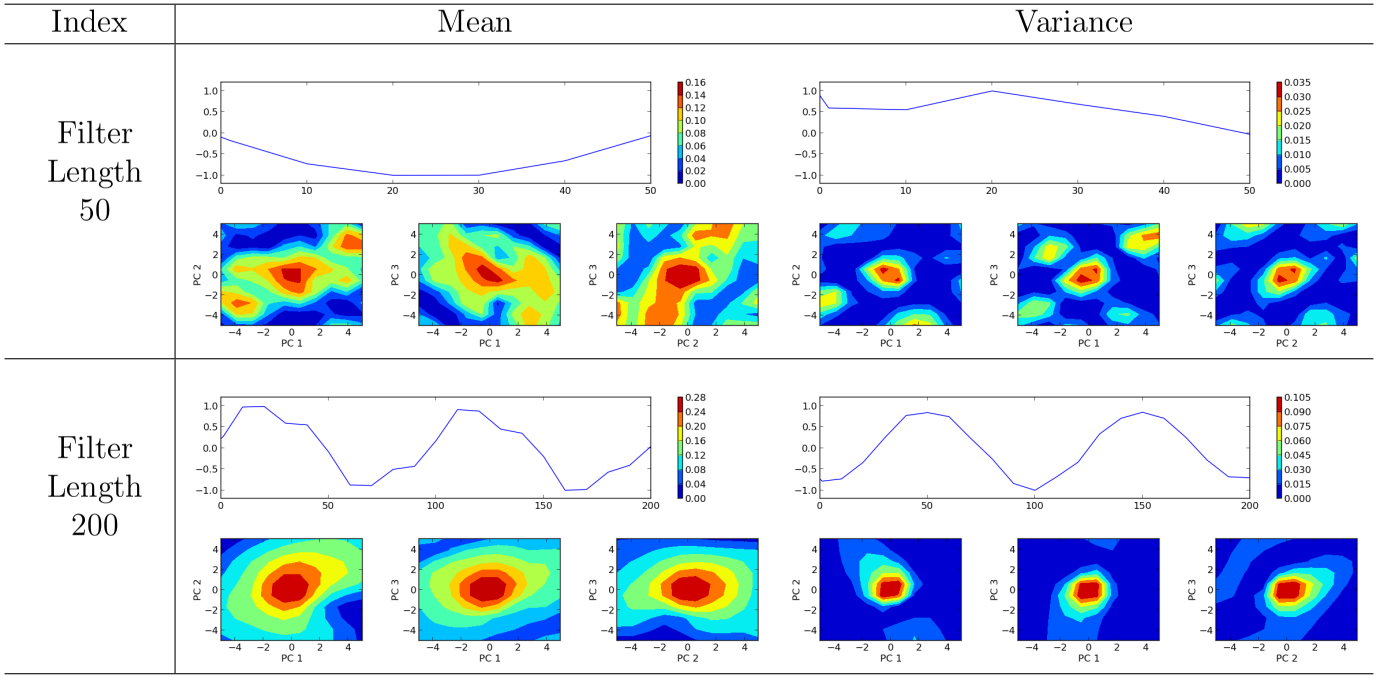


Figure 3.10: Resulting filters (upper part of each plot) and the filtered correlation surfaces (bottom part of each plot) from the the projection pursuit approach with simulated data. Mean and variance indexes are shown for filter length 50 and 200 with knot size 5 and 20, respectively.

length, is set to be 50. This length is chosen based on the scientific reason that the length matches with the time that a person responds to a visual stimulus[4].

The analysis is conducted in the following ways: 1) using position-specific local transformations, 2) using a a common filter for each position pair, 3) using a common transformation for multiple position pairs. For simplicity, in each analysis we show the results for 1) each subject for a fixed location pair (subject-wise), and 2) each of the possible location pairs for selected subjects (location-wise).

3.7.1 Pairwise analysis with position-specific transformations

Given a location pair, transformations are constructed for each location. For example, if we are interested in the location pair Fp1 and P3, two transformation filters are constructed for Fp1 and P3, respectively. As described in Section 3.3.2, CCA is used to construct filters. To explore the effect of the number of knots (we consider equal space between knots), we consider two basis functions for the filter: B-splines with 5 knots and free. The free basis function is equivalent to have a basis with 50 knots[5].

First, we discuss the subject-wise analysis results. Figure 3.11 shows the correlations from raw and filtered series constructed with the two basis. Most of the subjects have similar filtered correlation for both B-spline and Free basis functions. The subjects whose correlation showed more than a 0.1 difference between Free and B-spline basis are 4, 17, 19, 20, 22, 23 and 24. This suggests that the filter could be quite smooth.

Figure 3.12 shows the resulting filters from the B-splines and Free basis functions for selected subjects. Subject 1 and 13 have very small difference in the filtered

[4]It is well known that 200 millisecond is the time that a visual stimulus travel through retina, thalamus and parietal to frontal area. The sampling rate of our data set is 256 Hz, which has 256 points in a second. So using 50 filter length means the filter covers 1/5 second or 200 milliseconds.

[5]We also used knot sizes 10 and 25, but the both of filtered correlations were very similar to that of 5 knots.

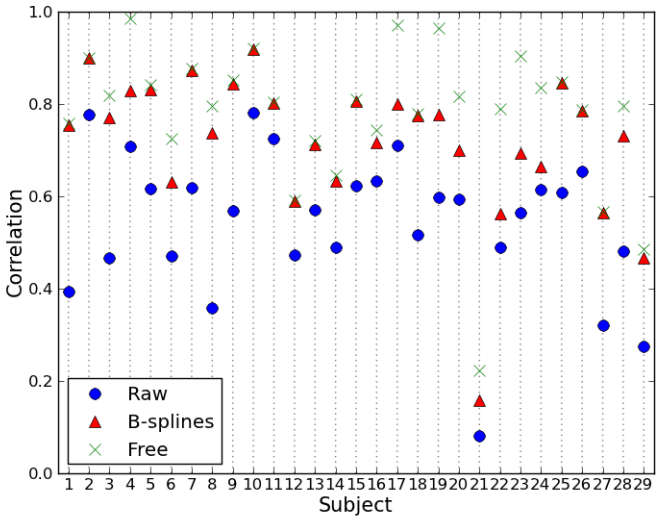


Figure 3.11: Subject-wise correlations between EEG series from Fp1 and P3 locations. Raw and filters series are shown together. The filtered series are from B-spline and Free position-specific filters.

correlation between the two basis. Subjects 4, 17 and 23 are the ones whose optimized correlation increased over 0.1, and the optimized correlations of Free filter are near 1. Interestingly, the subjects with similar filtered correlations from two basis functions tend to share the similar filter shape for Free basis as seen in subjects 1 and 13. Others share the shape of subjects 4, 17 and 23. Also, this distinction is more apparent in the Free basis filter than B-spline filter.

The constructed filters have some nice interpretation. For example, the Free filter from subject 4 and 17 take differences of the neighboring points. Subject 4's filters are almost the same for each location, whereas subject 17's filters have a time lag between the locations. Fig 3.13 shows subject 4's filtered series for the location pair Fp1 and P3 (up to about 2 seconds), and scatter plot of those two series. The plot shows that the filtered series are indeed have strong linear relationship.

Now, let's look at the results from location-wise analysis. For simplicity, we show the results for subjects 1 and 4 using B-spline basis filter. Both Free and B-spline filters give very similar results. With the EEG data, it is known that the signals from the nearby locations synchronize better the signals from further locations. Therefore,

using the optimized connectivity to assess the connectivity pattern may not be as effective. Instead, we use differences of the correlations of locations pairs between raw and transformed series. For example, let series from Fp1 be X and P3 be Y, then we calculate

$$Corr(\hat{\phi} * X, \hat{\phi} * Y) - Corr(X, Y), \tag{3.7.1}$$

where $\hat{\phi}$ is the constructed common filter for X and Y, and * is the convolution operator. As the optimized correlation should be larger than the raw correlation, all the differences are expected to be positive. This calculation is repeated for each possible pairs of 6 locations (15 location pairs).

Figure 3.14 shows the resulting differences for each position pair in subjects 1 and 4. In both subjects, the most noticeable differences occur at the pairs Fp1-P3, Fp1-P4, Fp2-P3, and Fp2-P4. This result is coherent with the scientific hypothesis that the connectivity between frontal and parietal regions are active when a person is conscious.

3.7.2 Analysis with pairwise common local transformation

In this analysis, we restrict the optimal local transformations to be the same for a pair of location. Allowing different transformation for each series may give the largest increase of the optimized correlation. However, most of the subjects or the location pairs had very similar filter for each series. Hence, it may also be scientifically meaningful to assume that there exists some shared behavior in a certain location pairs.

In the analysis, the filter is constructed for each pair of location. The heuristic algorithm (III.2) introduced in Section 3.4 is used to construct the transformation. Note that we let the transformation to be different for the signal from same location when the signal is paired with different locations. For example, the constructed filters may differ for the pair Fp1-P3 and Fp1-Fp1 though the same Fp1 series is involved.

Similar as the position-specific filter analysis, the results are shown in subject-wise and location-wise ways.

In the subject-wise transformation analysis, we construct filters for each subject using the location pair Fp1-P3. Figure 3.15 shows the optimized correlation from the common filter. For the comparison, the resulting correlation from the B-spline position-specific filter and raw series are also plotted. As can be seen, the optimized correlations for each subjects with common filters are slightly lower than that of the position-specific filters. The average increase of the optimized correlation from these common filters across the subjects is 0.158 and the standard deviation 0.088. The most prominent change is seen in subjects 1, 3 and 8 with the increase 0.294, 0.298 and 0.354 respectively.

Now, we discuss the results from the location-wise analysis. For simplicity, we show the results from subject 1 and 4. All possible pair of 6 locations are considered. Note that the analysis is still done for each pair. Similar as the location-wise analysis with the position-specific transformation, we look at the correlation change before and after the transformation. Recall, we use the differences in order to avoid observing dominant correlation pattern among the nearby location pairs. Figure 3.16 shows the resulting differences between raw and transformed series pair. The pattern of the differences is very similar to the result of position-specific transformation analysis: both subjects show prominent increase in the frontal-parietal connectivity.

Post-hoc Analysis

Suppose that we are interested in clustering analysis for either subjects or location pairs based on the filters. This clustering analysis can be tricky if we were to use the filter directly, due to the identifiability issue. Instead, we create a measure that can be used to compute the similarity between clustering objects. The basic idea is that if the common filters from two different location pairs differ, the resulting filtered correlation from each others filter would be significantly lower than using their own

filter. Therefore, we calculate the filtered correlation using the other pair's filter, and call it Cross-matched correlation. The idea is that, if certain subjects or location pairs share a similar pattern of the filter, the distribution of their cross-matched correlation would be similar.

This Cross-matched correlation can be computed for both subject and location-wise analysis. In the subject-wise analysis, the Cross-matched correlation can be computed for each subject. Hence, in our data, there will be 28 filtered correlation values applying the subject i's filter to others. By repeating this calculation for all 29 subjects, we obtain 29×29 matrix with the diagonal be the filtered correlation using subject i's own filter. Here, the each row indicates where the filter is from, and the each column indicates that subject whom the filter is applied to. For example, if we apply a filter from the subject 1 to subject 3, the resulting correlation will be at (1,3) coordinate of the matrix. Similarly, we can compute this Cross-matched correlation for the location-wise analysis, where each row indicates the location pair that create the filter and each column indicates the location pair that the filter is applied to.

We first apply hierarchical clustering with single linkage to explore the grouping pattern, then apply the k-means to cluster the subjects. Both of the clustering analysis give very similar results. For the hierarchical clustering, we define the similarity matrix as the following:

$$D_{ij} = \sum_{k=1}^{n} (Corr(\hat{\phi}_i * X_k, \hat{\phi}_i * Y_k) - Corr(\hat{\phi}_j * X_k, \hat{\phi}_j * Y_k))^2,$$

where $\hat{\phi}_i$ is the common filter constructed for series pair X_i and Y_i, n is the number of clustering objects, and $*$ is the convolution operator. For the k-means, we use the Euclidean distance of the cross-matched correlation between two rows (subjects or location pairs). In the between subject analysis, the subjects who share similar filter would be clustered together. Similarly, in the location-wise analysis, the location

pairs that share the similar filter are clustered together.

Figure 3.17 shows the clustering result for the between subject analysis. From the dendrogram, we see that most of the subjects are grouped together except subjects 21 and 27. Though subject 27 is not abnormal in terms of its raw and filtered correlation value, subject 21 have significantly lower raw and filtered correlation compare to that of other subjects. This analysis suggests that we can consider a common filter that accounts for all the subjects at a certain location pair.

Figure 3.18 shows the clustering results for the within subject analysis. These clustering methods are applied to subject 1 and 4 separately. In both subject, we can see 2 clusters from the dendrogram. Interestingly, the location pairs in each clusters are the same for both subjects. Table 3.6 shows the location pairs for the each cluster. One cluster contains the pairs only within frontal or parietal regions, and another cluster contains the pairs across any frontal and parietal regions. The shapes of filters in each cluster also show some coherent patterns.

Table 3.6: List of location pairs for each cluster from K-means clustering analysis. Subject 1 and 4 have the same clustering results, and the cluster number is arbitrary.

Cluster	Location pair
1	Fp1-Fp2, Fp1-F3, Fp1-F4, Fp2-F3, Fp2-F4, F3-F4, P3-P4
2	Fp1-P3, Fp1-P4, Fp2-P3, Fp2-P4, F3-P3, F3-P4, F4-P3, F4-P4

3.7.3 Multiple series with the same filter

In this subsection, we analyze the data by constructing a filter over multiple series using projection pursuit approach. Many different kind of projection index, or the summary function of all possible pairwise correlation among the multiple series, can be used. In this analysis, we use mean and variance indexes. Similar as the previous analysis using position-specific transformation and pairwise common transformation, we consider subject-wise and location-wise level analysis. In the subject-wise analysis,

we construct a common filter across the subject for a location pair Fp1 and P3. Similarly, in the location-wise analysis, we construct a common filter across all the possible pairs among the 6 locations for a subject.

Figure 3.19 shows the optimized correlation from the subject-wise analysis. As the filter is constructed to increase the mean correlation of all subjects, not all of the subject's correlation increase after the transformation. For the subjects whose transformed correlation increased, the magnitude is slightly lower than the transformed correlation from the position-specific filter. Also, for decreased subjects, there is a subject who has noticeable decrease: subject 27. Recall that this subject was in the other cluster along with subject 21 whose correlations are significantly lower than others. The change pattern of the transformed correlation are very similar in the two indexes.

In the location-wise analysis, we look at the correlation change before and after the transformation. Figure 3.20 shows the resulting differences for each index and subject. As can be seen, the mean and variance index showed quite different pattern in both subjects. In the mean index case, we see the increased connectivity between frontal and parietal regions after the transformation, which is consistent with the previous analysis. However, with the variance index, we see the decreased connectivity between frontal and parietal regions after the transformation. Note that the filter used in this analysis is different from the one we constructed across the subject.

Lastly, we show the resulting filters and their response surfaces for 6 different projection pursuit approach: two indexes for subject-wise and location-wise analysis. See Figure 3.21. The resulting filters have fairly different shape across the analysis cases. Also, most of the case have very narrow regions of maximum filtered correlation.

3.7.4 Summary

As a summary, we provide mean and sample standard deviation of optimized correlation from each method (Table 3.7). The position-specific approach have the largest optimal correlation on average. However, the optimal correlation from pairwise common filter is also very similar to that of position-specific filter. Both method have almost the same standard deviation. The projection pursuit approach have less increase in average optimal correlation for both projection indexes. However, both measure have larger standard deviation than that of pairwise methods.

Table 3.7: Summary of the filtered correlation from subject-wise analysis for all the methods.

	Raw	Position-specific	Common	PP Mean	PP Variance
Mean	0.544	0.720	0.702	0.630	0.613
SD	0.155	0.153	0.151	0.213	0.230

3.8 Discussion

In this chapter, we developed local transformation methods that reveal hidden association among multivariate time series. The local transformation take the form of convolution between filter coefficients and series, or linear filter. The filter coefficients are constructed to maximize the correlation between output series. We considered several different ways of constructing the coefficients: series-specific, pairwise common, and common transformation for multiple series. In the construction of series-specific transformation, we used Canonical correlation analysis approach. In the construction of common filter for a pair or multiple series, we developed the Heuristic up-hill algorithm that finds the filter coefficients for optimum correlation.

The proposed approach can be easily changed to reveal different aspect of the relationships. For example, in the neuronal connectivity application, we can find the filter that maximizes differences of the correlation between EEG series collected under

different conditions. Also, the method can be extended using different measures of association such as information measures, or different transformation method. All this can be done by just adjusting the objective function.

It is more likely to have non-convex objective function by imposing minimal restrictions on choosing one. This arises the question of identifiability and uniqueness of the constructed transformation. Though the local transformations constructed in our simulation study and data analysis did not show severe problem regarding the identifiability (giving quite small group of patterns), this issue is still remained to be further investigated.

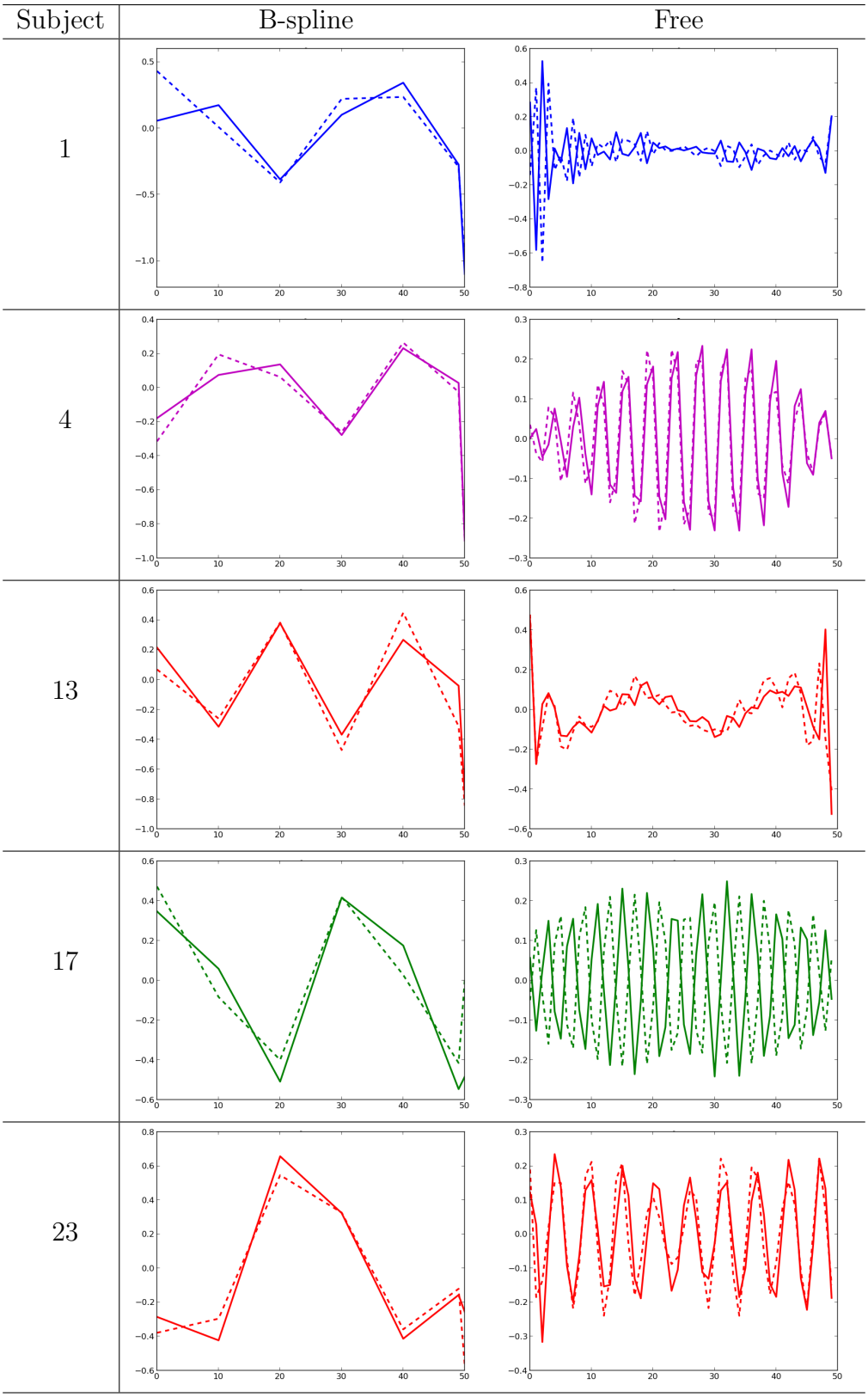
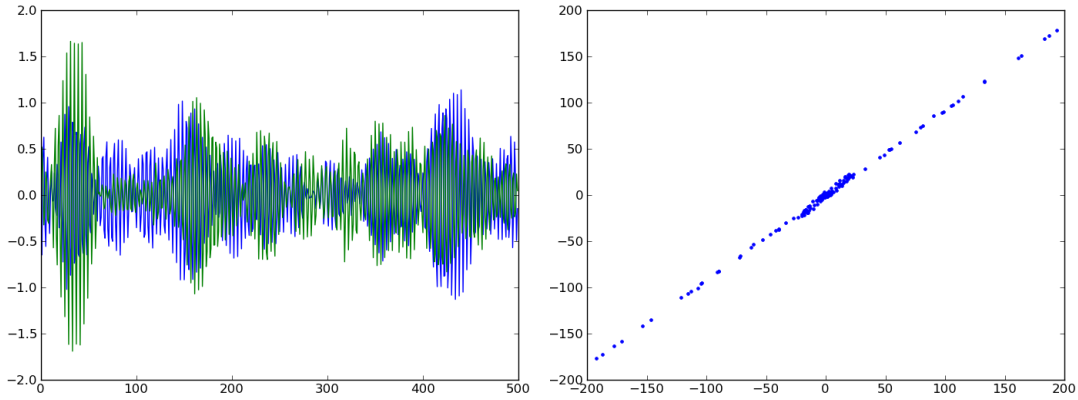
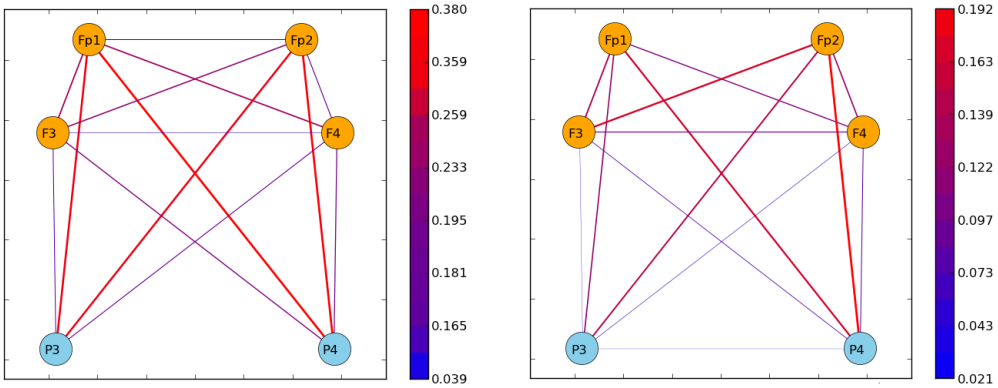


Figure 3.12: Constructed filters from CCA for the selected subjects The first column shows the filter constructed using B-spline basis with order 2, and the second column shows the free parametrized filter. Solid line is the resulting filter for EEG series from Fp1, and dashed line is the filter for series from P3.



(a) Filtered series **(b)** Scatter plot

Figure 3.13: Subject 4's filtered series for each location and its scatter plot. Blue is the filtered series for Fp1 and green is for P3. The length of original series is 77632, and we only show first 500 time points. For the filter construction, the Free basis function is used.



(a) Subject 1 **(b)** Subject 4

Figure 3.14: Location-wise correlation differences between raw and filtered series (Filtered - Raw) for selected subjects. The position-specific filter is used.

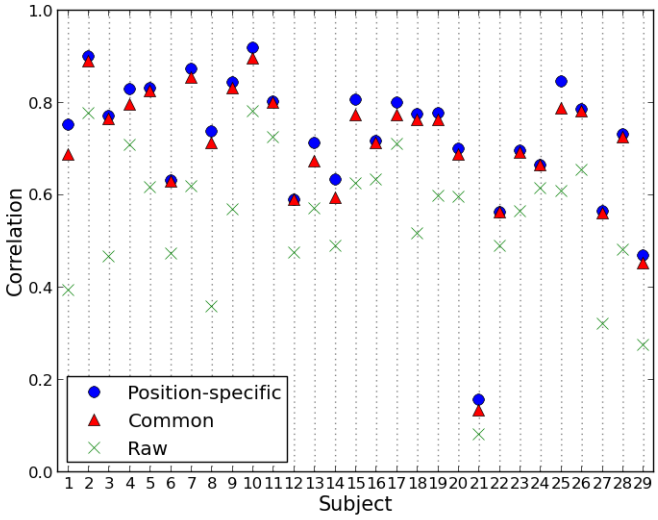
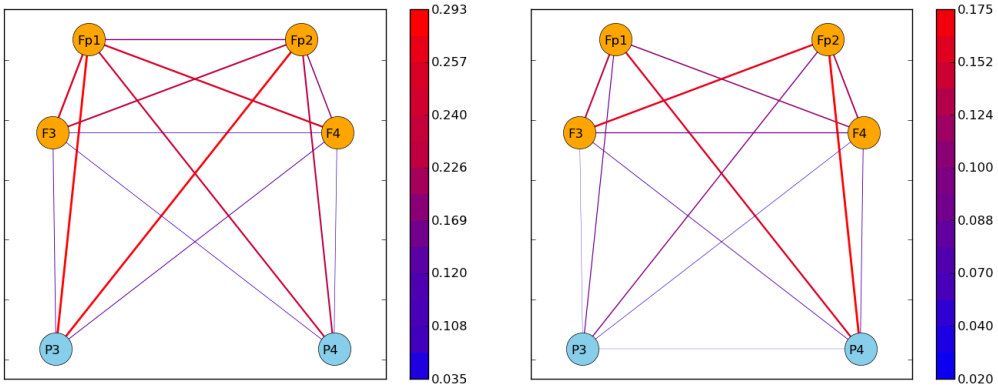


Figure 3.15: Subject-wise correlations between EEG series from Fp1 and P3 locations. Raw and filters series are shown together. The filtered series are from 1) the pairwise common filter and 2) B-spline position-specific filter



(a) Subject 1 **(b)** Subject 4

Figure 3.16: Location-wise correlation differences between raw and filtered series (Filtered - Raw) for selected subjects. The pairwise common filter is used.

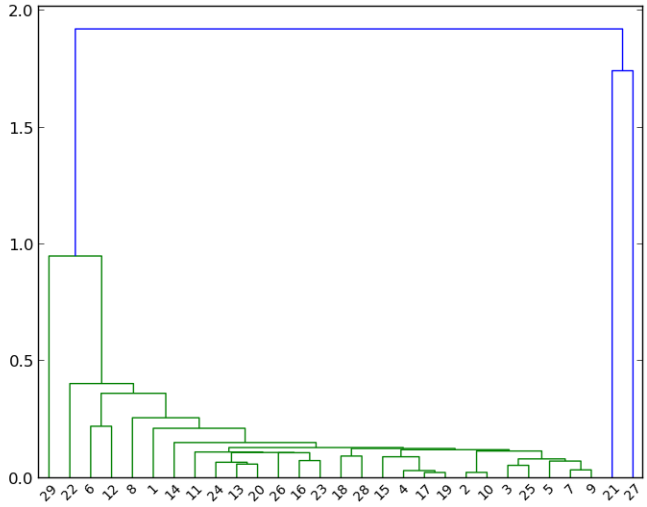


Figure 3.17: Dendrogram from the hierarchical clustering result on the subjects. The numbers on the x-axis denote the each subject.

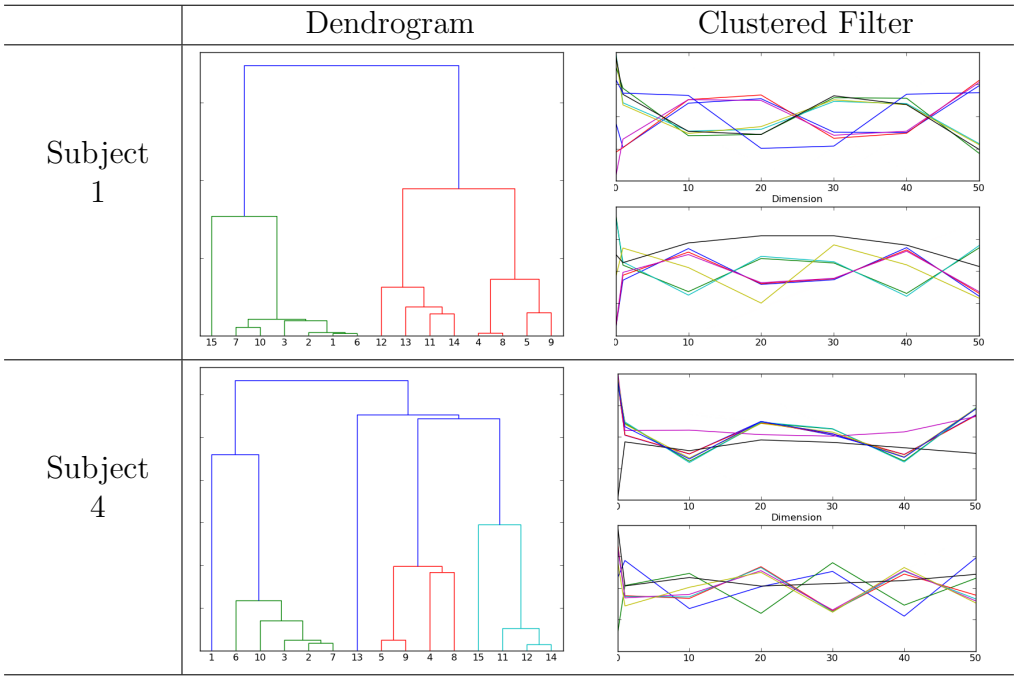
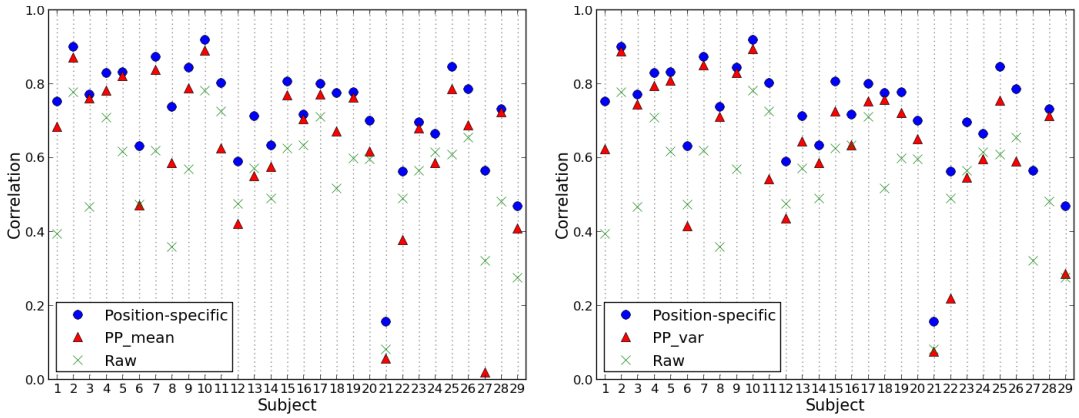


Figure 3.18: Clustering analysis results for the location-wise level analysis. The first column is dendrograms from the hierarchical clustering analysis, and the second column shows the filter for each cluster from k-means. The labels in the dendrograms indicate the location pairs.



(a) Mean **(b)** Variance

Figure 3.19: Subject-wise correlations between EEG series from Fp1 and P3 locations. Raw and filtered series are shown together. The filtered series are from 1) the projection pursuit approach and 2) B-spline CCA approach.

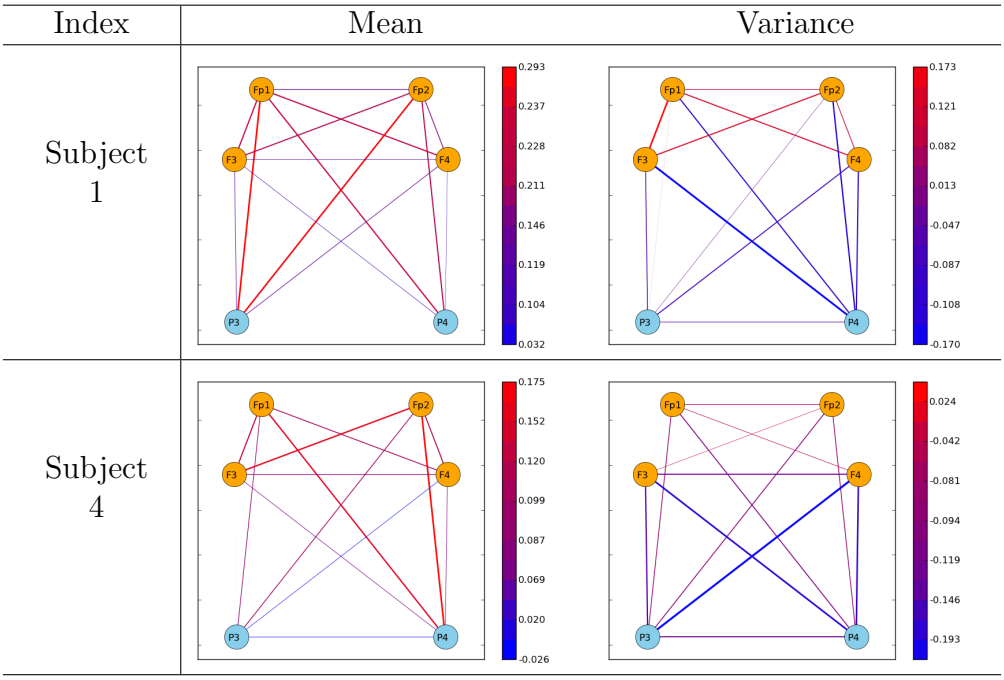


Figure 3.20: Location-wise correlation differences between raw and filtered series (Filtered - Raw) for selected subjects. The projection pursuit filter is used.

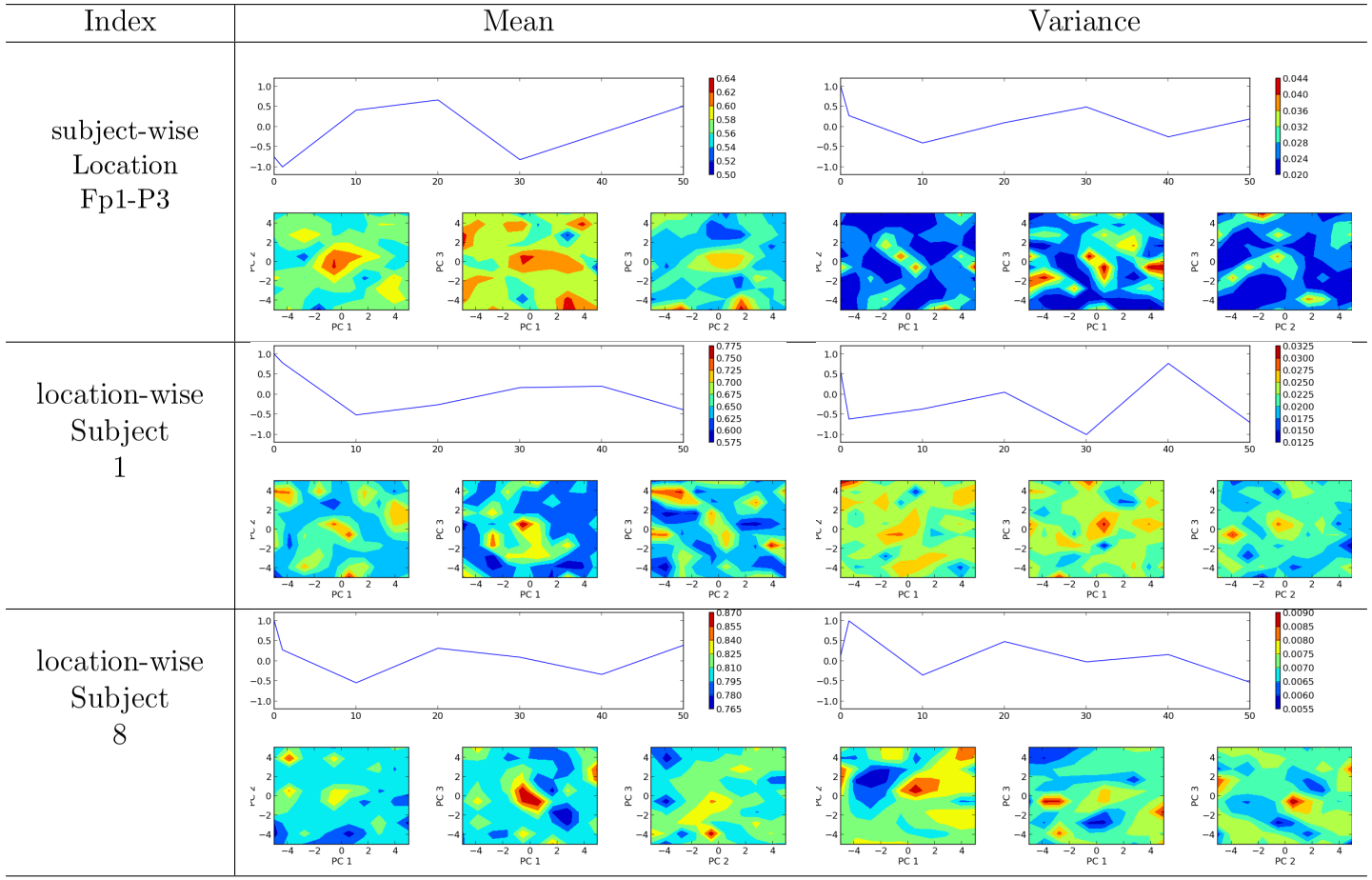


Figure 3.21: Resulting filters from subject-wise and location-wise analysis. The second column is from the mean projection index and the second column is from the variance index.

CHAPTER IV

Assessing the impact of a single unobserved covariate on the estimated effects of multiple observed covariates in regression analysis

4.1 Introduction

Multiple regression analysis is used in many scientific fields to aid in understanding the relationship between measured covariates and an outcome. However, in practice some important covariates that influence an outcome may not be observed, and this complicates the interpretation of fitted regression models. In this project we consider problems relating to questions about what can be learned regarding the possible effects of unobserved variables in a regression analysis. We start by providing several examples of scientific problems where regression analysis is usefully applied, and where there are likely to be important variables that cannot be measured and included in the regression.

- *Neuroscience*

  Suppose a researcher measures neural activity in several parts of the brain, in order to understand its relation to the perception of pain. If a multiple regression analysis indicates that neural activity in multiple brain regions contributes to

the outcome, one may interpret this as an indication that pain perception is influenced by several distinct neural systems. However, as activity levels from all parts of the brain cannot be measured in practice, it is possible that neural activity in a single unmeasured brain region is the only important variable. If this is the case, the results of our multiple regression analysis are qualitatively misleading, in that not only are the relevant brain regions misidentified, but we also erroneously conclude that multiple distinct regions, rather than one region, contribute to the perception of pain.

- *Genetics*

  A geneticist wants to perform multiple regression analysis to understand the association between an observable trait and genetic variables. In current practice, only a subset of possible genetic variables are directly measured. Due to linkage disequilibrium, the unmeasured genetic variables will be correlated with the measured genetic variables. Suppose we perform a multiple regression analysis using the measured genetic data, in a situation where only a single unmeasured genetic variable is truly important. The effect of the unmeasured variable will be transferred to one or more of the measured variables in the regression analysis. This may lead to the incorrect suggestion that "allelic heterogeneity"[1] is present, when in fact the genetic effect is quite simple.

- *Nutrition*

  A doctor wants to study how dietary intake of food components such as fat or protein is related to a health outcome. Nutrient intake is usually assessed by administering food-frequency questionnaires. Foods consist of numerous components, and not all foods are present on the questionnaire. Therefore, it is possible that a single eating behavior underlies the risk for a given health outcome, even when several behaviors, analyzed as incomplete data, seem to

[1] The presence of multiple functional variants within a linked region.

play major roles.

There are numerous models that deal with unobserved variables. One model is the latent variable model. The latent variable model assumes that there exist latent (or hidden) variables that explains the joint distribution of the observed variable usually with a reduced number of variables. For example, factor analysis (one of the simplest form of latent model) is based on modeling observed variable by a linear function of latent variables (*Bishop* (1999)). Another form is sensitivity analysis proposed by *Rosenbaum and Rubin* (1983). In this approach, the unobserved variable is considered to be a hidden bias that is not controlled by adjusting observed variables (*Rosenbaum* (2002)). Their main interest is to measure the effect of this hidden bias where it alters the significant effect of the treatment. Hence, this method is based on modeling the odds ratio with an unobserved variable and derive the boundary of the p-value with a related hypothesis test.

Our work has a similar goal as Rosenbaum style of sensitivity analysis, in the sense that the unobserved variable has its own effect beside the portion explained by observed variables. In all the scientific problem examples above, the key issue is to understand the effect of an unobserved covariate U. Such a question can only be partially answered in practice. As a special case, we will consider the situation where a single unobserved variable U contains all the information in X about Y, i.e., $P(Y|X,U) = P(Y|U)$. Note that U must be correlated with X. We focus initially on linear models, so that the effect size of U can be defined as the coefficient θ where $E(Y|U) = \theta U$. Here θ cannot be estimated as U is unobserved. But it will be seen that in this setting, as long as X and Y are correlated, the effect size of U must be nonzero and should be somehow related to the effect sizes of the Xs, which are estimable.

One special situation is the case when U is a function of X, e.g. $U = \delta' X$. In this case, U and X may contain the same information about Y, but it is not possible

for U to contain information about Y that is not in X. However, there is some situation that U may explain additional variance in Y, beyond that explained by X. For example, take the conventional way of expressing the linear model,

$$Y = \beta' X + \epsilon.$$

Given any decomposition of the error $\epsilon = \eta + \tilde{\epsilon}$, where $\tilde{\epsilon}$ is uncorrelated with $\beta' X + \eta$. If we set $U = \beta' X + \eta$, then U is an unobserved variable that explains a greater fraction of the response variance than X does. Based on this example, it is clear that we should only hope to place a lower bound on the effect size of U. It is also interesting to consider how weakly U and X can be related, with U still capturing all the information in X about Y.

The rest of this report is organized as follows. In section 2, we elaborate on how to address the problems we discussed: finding lower bound for effect size of U and relationship between U and X. In Section 3, we will show an illustrative example via simulation study based on an application to genetics. In the last section, we will discuss future work.

4.2 Problem formulation

In this section we introduce several quantities that can be used to learn about the influence of an unmeasured covariate. Our goal at this point is to establish what can be learned about the population level relationship between Y and U in a situation where we have complete information about the population level relationship between Y and X. Here we do not address the further complicating factor that characteristics of the joint distribution of Y and X must be estimated.

4.2.1 Bounds on moments between observed and unobserved predictors

We have n subjects with continuous responses $Y_1, \cdots, Y_n$. For each subject i, we observe p independent variables $X_i = (X_{i1}, \cdots, X_{ip})$. Suppose that there is an unobserved variable U which carries all the information in X about Y, that is

$$E(Y|X_1, \cdots, X_p, U) = E(Y|U). \tag{4.2.1}$$

The assumption (4.2.1) implies that when both X, U are present in the regression model, the coefficients of X will be zero. Also, we assume a linear relationship between U and Y, i.e.,

$$E(Y|U) = \alpha + \theta U. \tag{4.2.2}$$

As it is not possible to recover the subject level of U_i values, we focus on a much lower dimensional quantity $X'U$ – the collection of the inner products between each observed component X_j and the unobserved variable U for $j = 0, \cdots, p$ (we include intercept in the design matrix X).

While we cannot recover the exact value of $X_j'U$ for $j = 0, \cdots, p$, we demonstrate that this term can be bounded in a couple of ways using available information. First, as $(I - X(X'X)^{-1}X')$ is a projection matrix, the semi-definiteness of the projection matrix gives the first inequality

$$U'X(X'X)^{-1}X'U \leq U'U. \tag{I1}$$

Moreover, by the Cauchy-Schwartz inequality, we get another inequality

$$\left| \frac{X'U}{n} - \bar{X}\bar{U} \right| \preceq SD(X)SD(U), \tag{I2}$$

where $SD(X)$ is a vector of standard deviation of $X_j (j = 0, \cdots, p)$, $SD(U)$ is a

standard deviation of U (scalar) and $\preceq$ denotes element-wise inequality. To be specific, the inequality (I2) is derived from the fact that the absolute value of a sample correlation between X_j and U cannot exceed 1 for all $j = 0, \cdots, p$.

The terms containing U in (I1)-(I2) can be substituted with the terms that can be calculated from the observed data. Suppose that (4.2.1) and (4.2.2) hold. Under the linearity assumption between Y and U as in (4.2.2), the assumption (4.2.1) is equivalent to the following:

$$Y = \theta U + \epsilon \text{ and } E(\epsilon|U, X) = 0. \tag{4.2.3}$$

The above assumption (4.2.3) allows us to have the following expected normal equation when both X, U are observed:

$$\begin{pmatrix} X'EY \\ U'EY \end{pmatrix}_{(p+2)\times 1} = \begin{pmatrix} X'X & X'U \\ U'X & U'U \end{pmatrix} \begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \\ \theta \end{pmatrix}, \tag{4.2.4}$$

where $EY = E(Y|X, U) = E(Y|U) \in \mathbb{R}^n$. Since U is unknown, (4.2.4) has $p + 2$ equations and $p + 6$ scalar unknowns which are: α, θ, $U'U$, $\sum_{i=1}^{N} U_i$, $U'EY$ and $X'U$. Here, the unknown quantities are mostly the first or second moments except α and θ, which are the intercept and the effect size of U respectively. By expanding (4.2.4) we get the following equations :

$$X'EY = \alpha S_x + \theta X'U \tag{4.2.5}$$

$$U'EY = \alpha U. + \theta U'U, \tag{4.2.6}$$

where $S_x = (n, X_{1\cdot}, \cdots, X_{p\cdot})'$ is the first column of $X'X$ and $U_\cdot = \sum_{i=1}^{N} U_i$. Then rearranging (4.2.5) yields:

$$X'U = \frac{X'EY - \alpha S_x}{\theta}. \tag{4.2.7}$$

We are now able to express the key moment vector $X'U$ in terms of just two unknowns (α and θ) instead of $p+1$ unknowns ($U_\cdot, X_1'U, \cdots, X_p'U$), which can be quite high dimensional.

Moreover, the terms $U_\cdot/n$, $U'U/n$ can be replaced with $E(U)$ and $E(U^2)$ by using external information regarding the distribution of U. As illustrations, if we consider a Bernoulli distribution with parameter $p_u = P(U = 1)$, then we automatically have following first and second moments expressed by p_u:

$$E(U^2) = p_u, \ E(U) = p_u.$$

When the distribution of U can be expressed in terms of one unknown parameter, we are able to obtain fairly explicit formula for θ, the effect size of U . If U comes from a more general distribution family, we can only make very weak statements about these quantities unless we have some idea about the behavior of $Var(U)$. Therefore, in this paper, we consider distributions such that $Var(U)$ depend on only one parameter such as p_u. Also, we use $E(U), Var(U)$ rather than expressing these terms in a particular form under a certain distribution,

4.2.2 Combining all the known information

So far, we have established inequality constraints (I1)-(I2) involving the key unknown quantity $X'U$. Also, we have expressed $X'U$ explicitly in terms of the two unknowns α and θ. In this subsection, we will express the inequality (I1)-(I2) in terms of p_u and θ. These two unknown values p_u and θ are of primary interest and both have practical interpretation: θ is the effect size of the unobserved variable U

with respect to Y, and p_u determines the first two moments of U.

First, by plugging (4.2.7) into (I1)-(I2), the inequality relationship become the following inequality constraints (I1′)-(I2′) with three unknown parameters p_u, θ and α:

$$(X'EY - \alpha S_x)'(X'X)^{-1}(X'EY - \alpha S_x) \leq nE(U^2)\theta^2 \tag{I1′}$$

$$\left| \frac{X'EY - \alpha S_x}{\theta} - \bar{X}E(U) \right| \preceq SD(X)SD(U). \tag{I2′}$$

Here p_u plays a role in $E(U)$, $E(U^2)$ and $Var(U)$, which may have different form according to the distribution of U. Note that the value of α can be computed from the first row of inequalities in (I2′) as the first column of X is $(1, \cdots, 1)'$. Hence, the standard deviation of the first column of X is 0, and therefore

$$\alpha = \overline{EY} - \theta\bar{U}. \tag{4.2.8}$$

From (I1′)-(I2′), which are explicit algebraic functions of θ and p_u, we can find a set of (θ, p_u) values that satisfy both constraints. However, it is not possible to analytically express the range of these values. Thus we adopt a numerical approach exploring the range of possible (θ, p_u) values. The lower bound for θ found by the numerical approach will be denoted as θ_{LB}.

We first use a grid-search over $\theta \in [-3, 3], p_u \in [0.5, 1]$ range. Fig. 4.1 shows the region for possible pairs of (θ, p_u) found through this grid-search (the black colored area). The plot shows that any value of θ greater than a lower bound θ_{LB} satisfies (I1′)-(I2′). Hence, instead of searching over all possible grid for (θ, p_u), we can apply a bisection-like root finding algorithm to find corresponding θ_{LB} for a given p_u. We consider only positive part in the algorithm, as there is the symmetry pattern in possible (θ, p_u) values area. An interesting point revealed from this numerical search is that (I1′) provides the most tight bound for θ, that is once (I1′) is satisfied (I2′) does not impose any additional restriction other than deciding α value. Below is a

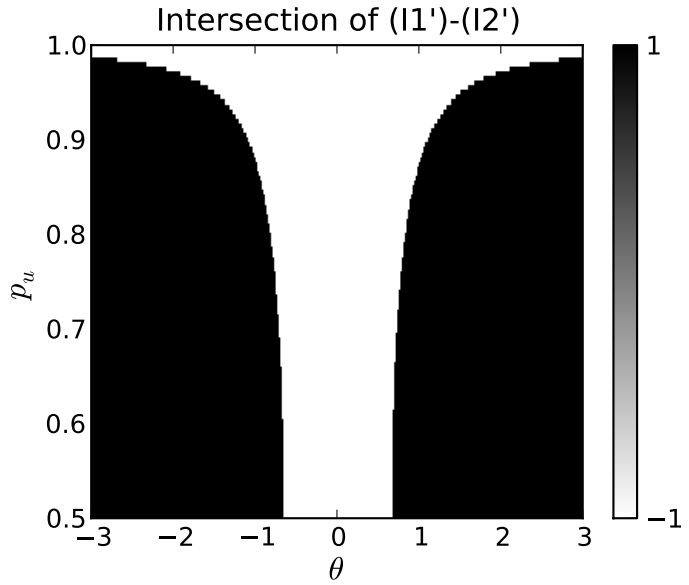


Figure 4.1: A collection of possible (θ, p_u) pairs satisfying (I1′) - (I2′) from the numerical example in Section 4.4. We performed this numerical grid search under different scenarios of simulated data and all gave the same pattern for the possible pairs. We will discuss about simulating different kind of data structure in depth in Section 4.4. Here considered range of θ is $[-3, 3]$ and by symmetry $p_u \geq 0.5$ are considered. The plot has only two values such that -1(white) for pair values not satisfying at least one of the constraints and 1(black) for pair values satisfying all constraints.

root(θ_{LB})-seeking algorithm that was used in the numerical study.

4.2.3 Relationship between X and U

If an unobserved variable U contains all the information in X about Y, then U must be related with X in some degree. Intuitively, if X can explain a good portion of variance in Y, then a weak dependence between X and U will cause the effect size of U to be large. However, as U is not observed, we cannot estimate this relationship between U and X with given data. Nevertheless, there are some cases in practice such that possible level of the relationship between observed variable X and an unobserved variable U can be assumed. For example, in genetic analysis we know that any genetic variable U must have a certain minimal level of dependence with the measured X variables based on the density of markers. Hence, one can utilize the

Algorithm IV.1 Bisection like algorithm that finds θ_{LB}

For fixed value p_u, consider an initial value $\theta_0 = [0, t]$.
if $f(t) = -1$, where

$$f(x) = \begin{cases} 1 & \text{if both (I1}') \text{ and (I2}') \text{ are satisfied} \\ -1 & \text{otherwise} \end{cases}$$

then
  Increase the value t until $f(t) = 1$.
  $\theta_0 \leftarrow [0, t^*]$
end if
Set $\theta = \theta_0$.
while Absolute value of the difference in $\theta > \epsilon$ **do**
  Take mean of the values in θ, and denote it as s.
  Compute $f(s)$.
  if $f(s) == 1$ **then**
    $\theta \leftarrow [0, s]$
  else
    $\theta \leftarrow [s, t]$
  end if
end while
Return s as θ_{LB} for the fixed p_u.

external information to bound the possible level of the relationship between X and U.

In this subsection, we establish an equation relating the level of dependency between X and U with the effect size θ and the variance of U. As we are considering multiple Xs, we quantify the net dependence between U and X as a sample $R^2(\hat{R}^2_{u|x})$ of a hypothetical regression of U on X:

$$\hat{R}^2_{u|x} = \frac{\| \hat{U} - \bar{U} \|^2}{\| U - \bar{U} \|^2} = \frac{\hat{U}'\hat{U}/n - \bar{U}^2}{U'U/n - \bar{U}^2}, \tag{4.2.9}$$

where $\hat{U}$ is a projection of U onto X space, i.e. $\hat{U} = X(X'X)^{-1}X'U$. The $\hat{R}^2_{u|x}$ in (4.2.9) is a ratio of sample variances of $\hat{U}$ to that of U, and this allows us to express the net dependence as a function of θ and p_u. Once a functional relationship among $\hat{R}^2_{u|x}$, θ and p_u is established, we can explore the surface of possible influences of an

unobserved variable U.

In order to establish this relationship, we need to substitute the terms containing U in (4.2.9). As $\hat{U}$ contains the key unknown moment $X'U$, we can substitute this term with the following expression by using (4.2.7) in Subsection 4.2.2. As a result, we have

$$\hat{R}^2_{u|x} = \frac{\left((X'EY - \alpha S_x)'(X'X)^{-1}(X'EY - \alpha S_x)/n\theta^2 - \bar{U}^2\right)}{U'U/n - \bar{U}^2}. \tag{4.2.10}$$

Furthermore, by using distribution information of U, we have following functional relationship among the three parameters θ, $R^2_{u|x}$ and P_u:

$$R^2_{u|x} \approx \frac{(A/\theta^2 - E(U)^2)}{Var(U)}, \tag{4.2.11}$$

where $A = (X'EY - \alpha S_x)'(X'X)^{-1}(X'EY - \alpha S_x)/n$. The left-hand side $R^2_{u|x}$ cannot be calculated from the data, whereas the right-hand side can be. Therefore, we do not have an exact equation as (4.2.10).

The relationship (4.2.11) can be solved for θ by plugging in $\alpha = \overline{EY} - \theta E(U)$:

$$\theta^2_p \approx \left(\frac{EY'P_xEY}{n} - \overline{EY}^2\right)\left(\frac{1}{R^2_{u|x}Var(U)}\right). \tag{4.2.12}$$

We denote the functional relationship (4.2.12) as θ_p to indicate that this relationship contains population value such as EY. The relationship (4.2.12) will be called as the first approximation and it consists of two parts: data and parameters. That is (4.2.12) can be expressed as

$$\theta_p = h(EY)g(R^2_{u|x},\ P_u).$$

This population functional relationship can have a sample version by using Y

instead of EY. We will call this sample version of relationship as the second approximation. Let a sample version of A be $A' = (X'Y - \alpha S_x)'(X'X)^{-1}(X'Y - \alpha S_x)/n$. If we expand A', we get the following:

$$\begin{aligned} A' = & \ (X'Y - \alpha S_x)'(X'X)^{-1}(X'Y - \alpha S_x)/n \\ = & \ (X'(Y - EY + EY) - \alpha S_x)'(X'X)^{-1}(X'(Y - EY + EY) - \alpha S_x)/n \\ = & \ (X'(Y - EY) + X'EY - \alpha S_x)'(X'X)^{-1}(X'(Y - EY) + X'EY - \alpha S_x)/n \\ = & \ A + \epsilon' P_x \epsilon/n + 2\theta\epsilon' P_x U/n - 2\bar{\epsilon}(\bar{Y} - \theta\bar{U}). \end{aligned} \tag{4.2.13}$$

Then the second approximation can be expressed as

$$\hat{R}^2_{u|x} + \frac{2\theta\epsilon' P_x U/n - 2\bar{\epsilon}(\bar{Y} - \theta\bar{U}) + \epsilon' P_x \epsilon/n}{\theta^2 \left(U'U/n - \bar{U}^2\right)}, \tag{4.2.14}$$

which has some discrepancy from the $\hat{R}^2_{u|x}$ in (4.2.10).

Now, we define a sample version of the functional relationship among three parameters as following:

$$\theta_s^2 = \left(\frac{Y' P_x Y}{n} - \overline{Y}^2\right)\left(\frac{1}{R^2_{u|x} Var(U)}\right). \tag{4.2.15}$$

Similar as θ_p in (4.2.12), the sample version θ_s can be divided into data and parameter parts

$$\theta_s = h(Y) g(R^2_{u|x}, P_u).$$

Note that $h(Y)$ is a sample version of $h(EY)$, which can be denoted as $\hat{h}(EY)$.

4.3 Toy example

In this section, we provide a toy example to illustrate our method. We consider normally distributed observed, unobserved and response variables with several differ-

ent correlation structures. Simulated data are generated following way.

1. Observed variable:

   $X_j = \sqrt{1-r_x}Z_j + \sqrt{r_x}e$, where $Z_j, e \sim N(0,1)$ and $Corr(X_j, X_{\tilde{j}})^2 = r_x$ for any $j \neq \tilde{j}$.

2. Unobserved variable:

   $U = \sqrt{1-r_u}Z_u + \sqrt{r_u}e$, where $Corr(X_j, U)^2 = r_u$ for $\forall j$

3. Response variable:

   $Y = \theta U + \sqrt{c_2}\epsilon_u$, where $c_2 = \theta^2 Var(U)\frac{1-Corr(Y,U)}{Corr(Y,U)}$ for a given $Corr(Y,U)$ and $\epsilon_u \sim N(0,1)$

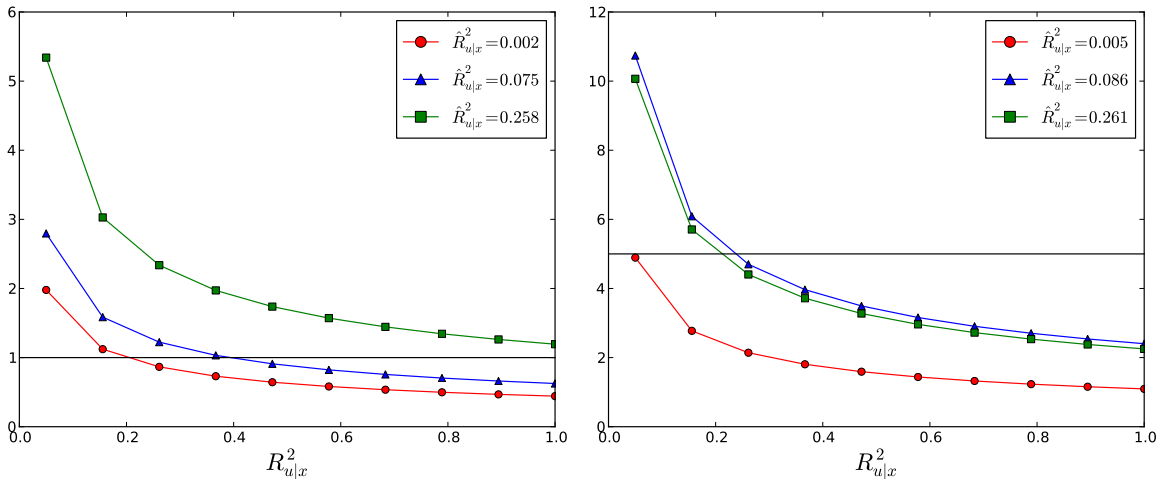
For simplicity, variances of all X_i and U are set to be 1.

The simulation study has two purposes: one to demonstrate the utility of the equation (4.3.1), another to assess possible factors that may affect the performance of θ_s. With our simulation setting, the functional relationship (4.2.15) derived in Section 4.2.3 contains only two unknown parameters which are θ and $R^2_{u|x}$:

$$\theta_s^2 = \left(\frac{Y'P_xY}{n} - \overline{Y}^2\right)\left(\frac{1}{R^2_{u|x}}\right) \tag{4.3.1}$$

Using the equation (4.3.1), we can produce a plot of the relationship between the two parameters. In order to offer a practical interpretation, the simulation is set to reflect reality. Also, θ_s will be infinite when $R^2_{u|x} = 0$, so we consider a range of possible $R^2_{u|x}$ from 0.05 to 1.

The plots for the relationship between $R^2_{u|x}$ and θ_s are shown in Fig. 4.2. As can be seen, θ_s is monotonically decreasing as $R^2_{u|x}$ increases. This decreasing relationship is expected regardless of the data structure, such as level of net dependency between X and U or true θ value. Intuitively, this decreasing relationship implies that if U is explained well by X, the effect size of U will be small, or the role of U is very



(a) $\theta = 1$ **(b)** $\theta = 5$

Figure 4.2: This plot show the relationship between θ_s and $R^2_{u|x}$ derived in (4.3.1). $\hat{R}^2_{u|x}$ in the legend box denotes sample regression R^2 calculated from simulated U and X and $R^2_{u|x}$ in x-axis is a symbol varying from 0.05 to 1. Correlation between X_1 and X_2 for both plot is set to 0.1 and 1, 5 is used for true θ value.

weak. Yet we are not discussing the effect of correlation structure of the data on the parameter relationship. This plot is also useful to gain better ideas about possible range of the effect size of U, especially if one can use external information to restrict the possible range of the X and U relationship. Though the parameter relationship differs for different true θ, there is no distinct tendency in the relationship for different correlation structures between X and U.

Next, we assess the performance of θ_s under different combinations of several levels of possible covariates that may affect the performance such as correlation structure of data and number of observed covariates. For the correlation structure of data, we consider different levels of correlations 1) between X and U, 2) between Y and U, 3) among X. We also consider 2 and 10 number of observed variable X. In this simulation study, the performance we are interested in is the absolute magnitude of discrepancy of θ_s from the true value θ, denoted as |Bias| in the table. Note that our bias is not the same as conventional statistical definition of bias. Also, we fix

$\hat{R}^2_{u|x}$ with the value calculated from the simulated data X and U rather than using possible range of $R^2_{u|x}$ as in (4.3.1).

The simulation result with 1000 repetition is shown in Table 4.1. As can be seen, the average of absolute bias has largest value when the dependence 1) X & U and 2) Y & U are weak. This result can be explained from the second approximation formula (4.2.14), which shows where the discrepancy coming from. To be specific, the error ϵ will be larger if U explains less portion of the variation in Y. If X has nearly no relationship with U then it is natural that the bias gets larger. Though the dependence 1) is weak, if 2) gets stronger the bias reduces significantly. Hence, in order to have reasonable bias we need moderate dependence level in either 1) or 2). It is interesting to note that higher multicollinearity (we call dependence 3) here) helps to reduce the bias. It seems that this is because adequate amount of multicollinearity increases $\hat{R}^2_{u|x}$.

4.4 Numerical example: Genetic mapping

In this section, we provide a simulation study based on a genetics application. Genetic mapping aims to locate a gene that is associated with a certain quantitative trait (*Altshuler et al.* (2008)). In such association studies, usually genetic markers called single-nucleotide polymorphism (SNP) are used as explanatory variables. SNP is a gene's location on a chromosome where allelic differences occur among individuals. SNPs on the same chromosome are supposed to be in linkage disequilibrium (LD), i.e. correlated with each other if their positions are close. Details about density of SNPs and LD can be found in *Reich et al.* (2001).

4.4.1 Parametrize U for biallelic SNPs

Under a targeted region, we use a multiple regression analysis to understand the association between a quantitative trait $Y = (Y_1, \cdots, Y_n)$ and the causal DNA variant.

Table 4.1: Simulation results. Here R^2s are sample regression R^2 calculated from simulated X, U and Y. The second row of the table stands for three levels of generated relationship between U and X – 1) 'NN' for nearly none, 2) 'M' for moderate and 3) 'S' for severe relationship. k is the number of observed covariate X.

k = 2		θ_s			$\lvert Bias \rvert$			$R^2_{u\lvert x}$			$R^2_{y\lvert x}$		
$Corr(Y,U)$	$Corr(X_i,X_j)$	NN	M	S	NN	M	S	NN	M	S	NN	M	S
0.1	0.1	4.679	1.661	1.334	3.741	0.858	0.552	0.020	0.107	0.191	0.020	0.029	0.039
	0.7	4.471	1.123	1.041	3.533	0.366	0.261	0.021	0.418	0.826	0.020	0.061	0.097
0.5	0.1	1.846	1.073	1.014	1.013	0.276	0.189	0.022	0.109	0.194	0.022	0.064	0.106
	0.7	1.901	1.002	1.000	1.079	0.137	0.100	0.019	0.418	0.826	0.021	0.217	0.418
k = 10													
0.1	0.1	3.243	1.900	1.568	2.243	0.902	0.573	0.104	0.334	0.568	0.100	0.124	0.146
	0.7	3.349	1.608	1.345	2.349	0.615	0.368	0.100	0.532	0.962	0.103	0.143	0.182
0.5	0.1	1.429	1.139	1.071	0.448	0.182	0.123	0.102	0.339	0.570	0.099	0.223	0.332
	0.7	1.431	1.085	1.032	0.447	0.140	0.095	0.102	0.532	0.962	0.101	0.319	0.525

For each subject, indexed by $i = 1, \cdots, n$, p number of SNPs $(X_{ij}, j = 1, \cdots, p)$ are observed. Suppose that the causal SNP that carries all the signals reflected in X about Y is untyped and denote this variant as U. Each SNP in X and U have values in $\{-1, 0, 1\}$, where 1 stands for homozygote[2] with allele whose frequency is greater than 0.5, -1 stands for homozygote with allele whose frequency is less than 0.5 and 0 stands for heterozygote[3]. As noted above, within the targeted region, SNPs are in LD, which guarantees nonzero correlation between observed and unobserved SNPs.

In genetics, each SNP is generally assumed to satisfy Hardy-Weinberg equilibrium (HWE), an important assumption that assures independence of individuals. Under HWE, U is distributed according to the following one parameter family of distributions:

$$Pr(U = 1) = p_u^2,\ Pr(U = -1) = (1 - p_u)^2 \text{ and } Pr(U = 0) = 2p_u(1 - p_u),$$

where p_u is the major allele frequency (MAF)[4] for U. Hence, we get the following expressions for the mean and the variance of U:

$$E(U) = 2p_u - 1,\ Var(U) = 2p_u(1 - p_u). \tag{4.4.1}$$

Hence, we get following approximated functional relationship with three unknown parameters:

$$\theta_s^2 = \left(\frac{Y'P_xY}{n} - \overline{Y}^2\right)\left(\frac{1}{R_{u|x}^2 2P_u(1 - P_u)}\right). \tag{4.4.2}$$

4.4.2 Simulation study set up

The simulation study consists of two parts as the toy example in Section 4.3. However, in order to generate discrete variables for X and U, we use a probit type of

[2] A SNP whose genotype consists of two same allele. e.g. AA or BB
[3] A SNP whose genotype consists two different allele. e.g. AB
[4] Allele frequency > 0.5

method that first generates continuous random variables then classifies the generated random numbers into several categories. Specifically, we first sample X and U from multivariate normal distribution with mean 0 and standard deviation Σ (the same method as toy example). Then for a given p, MAF for a SNP, use $(1-p)^2$ and p^2 quartiles of standard normal as thresholds to classify the generated continuous variables into three classes $\{-1, 0, 1\}$. MAF p can differ for the SNPs in X and U, but we use 0.5 for all SNPs in the simulation.

We consider several different data structures reflecting potential covariates that may affect the performance of the approximated functional relationship (4.4.2). The potential covariates are the following dependence structures: 1) between X and U, 2) between Y and U, 3) among X. We generate X and U from multivariate normal distribution with the following exchangeable covariance matrix that has two different covariance value; one among $X(\rho_x)$ and another between X and $U(\rho_u)$, i.e.

$$Cov([U, X]) = \begin{pmatrix} 1 & \rho_u & \cdots & \rho_u \\ \rho_u & 1 & \cdots & \rho_x \\ \vdots & \vdots & \ddots & \vdots \\ \rho_u & \rho_x & \cdots & 1 \end{pmatrix}. \tag{4.4.3}$$

Entries of the covariance matrix (4.4.3) are the same as correlation as we set variance of X and U to be 1. To make sure that we control the dependence structure of simulated covariates as categorical variables, we calculate average correlation among X and between each X and U to select the data set with a certain correlation structure after classifying the continuous random variables into three classes.

In the simulations, three different levels of dependence structures 3) among X are considered by choosing data set with average correlation $\{0.1, 0.4, 0.7\}$. Similarly, we use three levels of dependency 1) X & U, which are denoted as 'NN' for nearly no, 'M' for moderate and 'S' for severe relationship in resulting tables. Once we generate

U, the quantitative trait Y is generated as

$$Y = \theta U + r\epsilon,$$

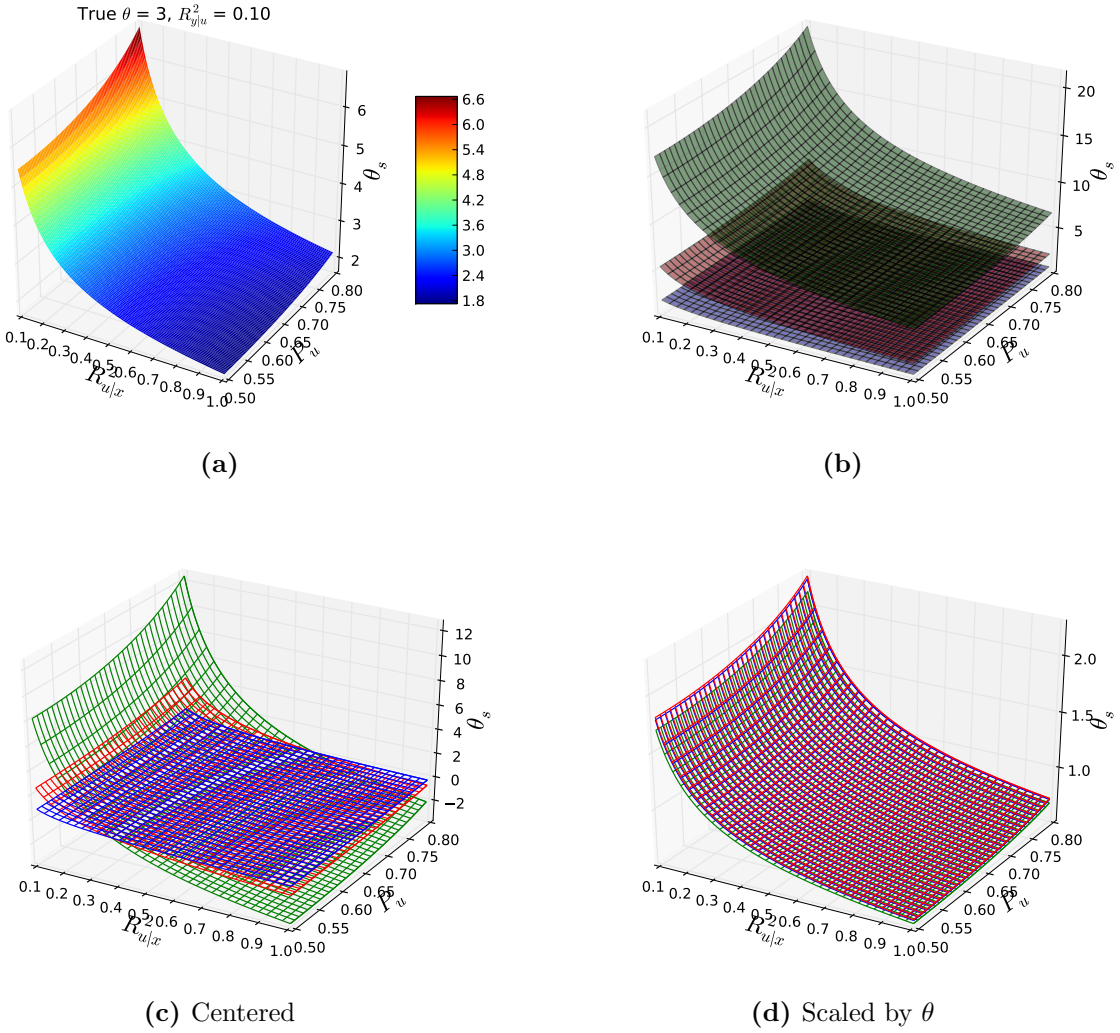
where ϵ is standard normal error and r is a scaling factor for U and Y, which controls the level of dependence 2) Y & U. The squared correlation between Y and U is $\frac{\theta^2 Var(U)}{\theta^2 Var(U)+r^2}$. We consider 10 observed SNPs that all have MAF 0.5 as well as U. The considered true effect size θ of U are [0.5, 1, 3, 5, 10] with sample size 250. Also, we set squared correlation between U and Y to be 0.1 as a single gene may account for 10 to 15% of the response variance (*Allen et al.* (2010)),

4.4.3 Relationship between X and U

In the simulation study, the equation for the functional relationship in (4.2.15) contains three unknown parameters: θ, $R^2_{u|x}$, P_u. Hence, we get a surface of feasible θ_s values for given $R^2_{u|x}$ and P_u (Fig. 4.3). For better visualization, we consider the range [0.1, 1] and [0.5, 0.8] for $R^2_{u|x}$ and P_u respectively as either $R^2_{u|x}$ is close to 0 or P_u is close to 1, θ_s value shoots up.

Analogous to Fig. 4.2, 4.3 (a) shows the tendency that the value of θ_s increases when $R^2_{u|x}$ gets small and P_u gets larger (upper right corner of the plot). The effect of $R^2_{u|x}$ is consistent with toy example in Section 2 and possible reason is discussed in the section. About the effect of P_u, this can be explained as when P_u increases $Var(U)$ in this simulation will decrease and the reduced variation in U causes effect size to become larger in order to account the portion of variation explained by X.

Particularly, in a genetic mapping effort, typically it is known that all genetic variants would have a certain minimal dependence level with the measured variants. The value of this minimal level of dependence would be determined by the density of the measured genetic markers. Thus, if we know that any unmeasured genetic



(a) **(b)**

(c) Centered **(d)** Scaled by θ

Figure 4.3: Resulting 3-dimensional plot for functional relationship with one time simulation (no repetition). Both $R^2_{y|u}$ and $Corr(X_i, X_j)$ are set to be 0.1. (a) Surface plot when a true θ is 3. (b) Surface plot for three different θ values. Bottom blue is when $\theta = 1$, middle red is when $\theta = 3$ and upper green is when $\theta = 10$. The same values are used for dependence structures as (a). (c) Centered plot of (b) by each mean. (d) Scaled plot of (b) by true θ.

variable would need to have at least 40% of its variance explained by the measured genetic variables, then we would know that the effect size of the unmeasured variable need to be in between 2.5 and 3 according to P_u. However, if it was around 10%, then the unmeasured genetic factor would need to have an effect size about 5 to 6 according to P_u value.

Fig. 4.3 (b),(c),(d) shows how the shape of functional relationship responds to different values of θ. As can be seen in plot (b), the surface moves up along the axis of θ_s as the true θ value increases. in order to compare the surface shape according to different θ values, the plot (b) is centered and scaled version by its mean and true θ respectively (plot (c) and (d)). It seems that from plot (b) the shape of functional relationship are not so different, since we seek θ_s through the fixed grid of $R^2_{u|x}$ and P_u. However, plots in (c) and (d) shows that the shape actually differ correspondent to the each data set.

Performance of the functional relationship (4.2.15) is assessed in terms of bias and variance using simulated $\hat{R}^2_{u|x}$ and P_u. Here the bias is calculated as $(\theta_s - \theta)$ and we take average of squared bias as the bias can be either positive or negative according to (4.2.14). Additionally, we divide this average squared bias by true θ^2 to scale out the effect of magnitude of true θ value. Table 4.2 shows the simulation result with different θ values, different correlation structure between X and U and different approximations. The first approximation, which plug in $E(U)$ and $Var(U)$ instead of sample values $U./n$ and $U'U/n$ in (4.2.15), has nearly 0 bias across all the combinations. We can see that most of the bias comes from the second approximation that is plugging in Y instead of EY from the third approximation that is combination of the first and the second approximation.

Table 4.2: Bias and variance result of 100 repeated simulation. The correlation between Y and U is set to be 0.1 and the average of correlations among X is set to be 0.1 as a nearly orthogonal case.

		Mean of squared bias/ θ^2			Standard deviation		
		Sample mean of $Corr(X, U)$					
	θ	0.1	0.4	0.7	0.1	0.4	0.7
First Approximation	0.5	0.00005	0.00006	0.00009	0.00352	0.00339	0.00318
	1	0.00004	0.00005	0.00007	0.00664	0.00695	0.00536
	3	0.00005	0.00006	0.00008	0.02078	0.02176	0.01880
	5	0.00005	0.00005	0.00008	0.03654	0.03475	0.03101
	10	0.00003	0.00005	0.00009	0.05827	0.06765	0.05671
Second Approximation	0.5	7.08625	0.00538	0.00306	0.60169	0.03561	0.02739
	1	5.90011	0.00646	0.00214	0.90290	0.07736	0.04466
	3	6.37932	0.00692	0.00195	3.60179	0.24068	0.12660
	5	6.66666	0.00650	0.00201	5.34683	0.37874	0.22416
	10	8.16516	0.00556	0.00289	14.47273	0.70376	0.50179
Third Approximation	0.5	7.07827	0.00525	0.00306	0.60119	0.03551	0.02767
	1	5.88771	0.00634	0.00203	0.90033	0.07710	0.04471
	3	6.40138	0.00675	0.00178	3.61006	0.23988	0.12508
	5	6.65841	0.00657	0.00207	5.35292	0.38401	0.22546
	10	8.15888	0.00559	0.00262	14.47224	0.71394	0.49832

4.5 Discussion

In this chapter, we proposed a method that assesses the effect of unobserved variable in multiple linear regression analysis. As an assessment, we derived an explicit formula that represents a functional relationship among three parameters: variance of an unobserved variable, relationship between the observed and an unobserved variable, and relationship between the unobserved variable and an outcome. Using this functional relationship we can explore the possible range of the relationship between an unobserved variable and the outcome.

This approach has a major limitation in representing true existence of U. That is, the functional relationship curve would be the same regardless of the existence of an unobserved variable that accounts all the observed regression relationship. One possible solution to this is using different measure for the relationship between an unobserved variable and observed predictors. We may not obtain an explicit formula

by introducing another measure of the relationship between observed covariates and an unobserved variable. However, it would be more interesting if the functional relationship can incorporate some underlying effect of this unobserved variable.

We close this chapter by providing future works.

4.5.1 Future work

Extensions of the method

Lastly, we would like to consider possible extensions of the method in the following ways:

1. Time series: This method can be extended to time series combined with the correlation methods proposed in previous chapters.

2. Multiple unmeasured variable U: U_1, U_2 with (P_1, θ_1), (P_2, θ_2)

   Adding one more unobserved variant would increases the combination of possible situations exponentially. Therefore, to address this problem, we may need to consider some additional constraints regarding relationship between Us.

3. Additional constraint using higher moments.

   As we saw limitation of inequality constraints (I1)-(I2), we may consider constructing additional constraints which are helpful to get more precise bounds for the effect size of unobserved variable. One natural way is using higher moments, since it is possible we may still obtain simple expression including θ and p_u.

Sensitivity to the modeling framework

As we assume a linear relationship between Y and U, we may ask how robust our results are to the case when there is a nonlinear relationship. As we work with moments of X and U, the non-linearity assumption on the relationship between X

and U should not affect our results. We seek to explore the effect of nonlinearity via simulations involving such models.

Mapping

We can derive lower and upper bounds for the correlation between each observed covariate and U. Particularly for genetics application, this has a practical interpretation regarding to the location of the trait-causing gene. The correlation between each X and U can be expressed as following:

$$Corr(X_i, U) = \frac{Cov(X_i, Y)}{\theta SD(X_i) SD(U)}$$

We can bound the squared correlation between X_i and U by using the fact that $Var(Y) > Var(U)$ and the inequality constraint (**??**):

$$\frac{Corr(X_i, Y)^2}{\theta^2} \leq Corr(X_i, U)^2 \leq \frac{Cov(X_i, Y)^2}{Var(\hat{Y}_x) Var(X)} \tag{4.5.1}$$

However, as the lower bound in (4.5.1) depends on θ, it is still ambiguous how this inequality would behave when $|\theta| \leq 1$. Thus, we need to investigate further the bounds when $|\theta| \leq 1$. Alternatively, we may consider another upper bound not depending on θ.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abramovich, F., T. C. Bailey, and T. Sapatinas (2000), Wavelet analysis and its statistical applications, *Journal of the Royal Statistical Society. Series D (The Statistician)*, *49*(1), pp. 1–29.

Addison, P. S. (2005), Wavelet transforms and the ECG: a review, *Physiological Measurement*, *26*(5), R155.

Adeli, H., Z. Zhou, and N. Dadmehr (2003), Analysis of EEG records in an epileptic patient using wavelet transform, *Journal of Neuroscience Methods*, *123*(1), 69 – 87, doi:10.1016/S0165-0270(02)00340-0.

Afonso, V., W. J. Tompkins, T. Nguyen, and S. Luo (1999), ECG beat detection using filter banks, *Biomedical Engineering, IEEE Transactions on*, *46*(2), 192–202, doi:10.1109/10.740882.

Alkire, M. T., and J. Miller (2005), General anesthesia and the neural correlates of consciousness, in *The Boundaries of Consciousness: Neurobiology and Neuropathology*, *Progress in Brain Research*, vol. 150, edited by S. Laureys, pp. 229 – 597, Elsevier, doi:10.1016/S0079-6123(05)50017-7.

Alkire, M. T., A. G. Hudetz, and G. Tononi (2008), Consciousness and anesthesia, *Science*, *322*(5903), 876–880, doi:10.1126/science.1149213.

Allen, H. L., K. Estrada, G. Lettre, S. I. Berndt, et al. (2010), Hundreds of variants clustered in genomic loci and biological pathways affect human height, *Nature*, *467*(7317), 832–838, doi:10.1038/nature09410.

Allen, J. (1977), Short term spectral analysis, synthesis, and modification by discrete fourier transform, *Acoustics, Speech and Signal Processing, IEEE Transactions on*, *25*(3), 235–238, doi:10.1109/TASSP.1977.1162950.

Altshuler, D., M. J. Daly, and E. S. Lander (2008), Genetic mapping in human disease, *Science*, *322*(5903), 881 –888, doi:10.1126/science.1156409.

Bassett, D. S., A. Meyer-Lindenberg, S. Achard, T. Duke, and E. Bullmore (2006), Adaptive reconfiguration of fractal small-world human brain functional networks, *Proceedings of the National Academy of Sciences*, *103*(51), 19,518–19,523, doi:10.1073/pnas.0606005103.

Baxter, M., and R. G. King (1995), Measuring business cycles approximate band-pass filters for economic time series, *Working Paper 5022*, National Bureau of Economic Research.

Bishop, C. M. (1999), Latent variable models, in *Learning in Graphical Models*, pp. 371–403, MIT Press.

Blumenson, L. E. (1960), A derivation of n-dimensional spherical coordinates, *The American Mathematical Monthly*, *67*(1), pp. 63–66.

Breiman, L., and J. H. Friedman (1985), Estimating optimal transformations for multiple regression and correlation, *Journal of the American Statistical Association*, *80*(391), 580–598, doi:10.2307/2288473.

Brillinger, D. R. (1975), *Time series: data analysis and theory*, International series in decision processes, Holt, Rinehart, and Winston, New York, v.

Cimenser, A., P. L. Purdon, E. T. Pierce, J. L. Walsh, A. F. Salazar-Gomez, P. G. Harrell, C. Tavares-Stoeckel, K. Habeeb, and E. N. Brown (2011), Tracking brain states under general anesthesia by using global coherence analysis, *Proceedings of the National Academy of Sciences*, doi:10.1073/pnas.1017041108.

Crick, F., and C. Koch (2003), A framework for consciousness, *Nat Neurosci*, *6*(2), 119–126, doi:10.1038/nn0203-119.

Dietrich, A., and R. Kanso (2010), A review of EEG, ERP, and neuroimaging studies of creativity and insight., *Psychological Bulletin*, *136*(5), 822 – 848.

Donner, R., M. Reiter, G. Langs, P. Peloschek, and H. Bischof (2006), Fast active appearance model search using canonical correlation analysis, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *28*(10), 1690–1694, doi:10.1109/TPAMI.2006.206.

Friedman, J., and J. Tukey (1974), A projection pursuit algorithm for exploratory data analysis, *Computers, IEEE Transactions on*, *C-23*(9), 881–890, doi:10.1109/T-C.1974.224051.

Friman, O., M. Borga, P. Lundberg, and H. Knutsson (2001), Canonical correlation as a tool in functional MRI data analysis, in *Proceedings of the SSAB Symposium on Image Analysis*.

Friston, K. J. (1994), Functional and effective connectivity in neuroimaging: A synthesis, *Human Brain Mapping*, *2*(1-2), 56–78, doi:10.1002/hbm.460020107.

Friston, K. J. (2011), Functional and effective connectivity: A review, *Brain Connectivity*, *1*, 13–36.

Glasbey, C. A., and G. W. Horgan (1995), *Image analysis for the biological sciences*, John Wiley & Sons, Inc., New York, NY, USA.

Hardoon, D. R., S. Szedmak, and J. Shawe-Taylor (2004), Canonical correlation analysis: An overview with application to learning methods, *Neural Computation*, *16*(12), 2639–2664.

Hayes, M. H. (1996), *Statistical digital signal processing and modeling*, xv, 608 p.– pp., Wiley, New York.

Herrmann, C. S., M. Grigutsch, and N. A. Busch (2005), EEG oscillations and wavelet analysis, in *Event-related potentials: A methods handbook*, pp. pp. 229–259, MIT Press, Cambridge, MA, editor: Handy, Todd C.

Hotelling, H. (1936), Relations between two sets of variates, *Biometrika*, *28*(3-4), 321–377, doi:10.1093/biomet/28.3-4.321.

Hudetz, A. G. (2012), General anesthesia and human brain connectivity, *Brain Connectivity*, *2*(6), 291–302.

Iacobucci, A. (2003), Spectral analysis for economic time series, *Documents de Travail de l'OFCE 2003-07*, Observatoire Francais des Conjonctures Economiques (OFCE).

Jenkins, S., R. Brown, and N. Rutterford (2009), Comparing thermographic, EEG, and subjective measures of affective experience during simulated product interactions, *International Journal of Design; Vol 3, No 2 (2009)*.

Jing, X., S. Li, C. Lan, D. Zhang, J. Yang, and Q. Liu (2011), Color image canonical correlation analysis for face feature extraction and recognition, *Signal Processing*, *91*(8), 2132 – 2140, doi:10.1016/j.sigpro.2011.02.016.

Johnson, R. A., and D. W. Wichern (2002), *Applied multivariate statistical analysis (Fifth Ed.)*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Koopmans, L. H. (1974), *The spectral analysis of time series*, xiv, 366 p.– pp., Academic Press, New York.

Ku, S.-W., U. Lee, G.-J. Noh, I.-G. Jun, and G. A. Mashour (2011), Preferential inhibition of frontal-to-parietal feedback connectivity is a neurophysiologic correlate of general anesthesia in surgical patients, *PLoS ONE*, *6*(10), e25,155, doi:10.1371/journal.pone.0025155.

Lindquist, M. A. (2008), The statistical analysis of fmri data, *Statistical science*, *23*.

Lindsay, R., D. Percival, and D. Rothrock (1996), The discrete wavelet transform and the scale analysis of the surface properties of sea ice, *Geoscience and Remote Sensing, IEEE Transactions on*, *34*(3), 771 –787, doi:10.1109/36.499782.

Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer-Verlag GmbH, Berlin Heidelberg.

Nallasamy, N., and D. Y. Tsao (2011), Functional connectivity in the brain: Effects of anesthesia, *The Neuroscientist*, *17*, 94–106, doi:10.1177/1073858410374126.

Nunez, P. L., and R. Srinivasan (2006), *Electric fields of the brain: the neurophysics of EEG*, xvi, 611 p.– pp., Oxford University Press, Oxford ; New York.

Peltier, S. J., C. Kerssens, S. B. Hamann, P. S. Sebel, M. Byas-Smith, and X. Hu (2005), Functional connectivity changes with concentration of sevoflurane anesthesia, *NeuroReport*, *16*(3).

Percival, D. B., and H. O. Mofjeld (1997), Analysis of subtidal coastal sea level fluctuations using wavelets, *Journal of the American Statistical Association*, *92*(439), 868–880, doi:10.1080/01621459.1997.10474042.

Percival, D. B., and A. T. Walden (2000), *Wavelet Methods for Time Series Analysis (Cambridge Series in Statistical and Probabilistic Mathematics)*, Cambridge University Press.

Platt, T., and K. L. Damon (1975), Spectral analysis in ecology, *Annual Review of Ecology and Systematics*, *6*, pp. 189–210.

Rampil, I. J. (1998), A primer for EEG signal processing in anesthesia, *Anesthesiology*, *89*(4), –.

Rees, G., G. Kreiman, and C. Koch (2002), Neural correlates of consciousness in humans, *Nat Rev Neurosci*, *3*(4), 261–270.

Reich, D. E., et al. (2001), Linkage disequilibrium in the human genome, *Nature*, *411*(6834), 199–204, doi:10.1038/35075590.

Rosenbaum, P. R. (2002), *Observational studies*, xiv, 375 p.– pp., Springer, New York.

Rosenbaum, P. R., and D. B. Rubin (1983), Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome, *Journal of the Royal Statistical Society. Series B (Methodological)*, *45*(2), 212–218.

Samar, V. J., A. Bopardikar, R. Rao, and K. Swartz (1999), Wavelet analysis of neuroelectric waveforms: A conceptual tutorial, *Brain and Language*, *66*(1), 7 – 60, doi:10.1006/brln.1998.2024.

Schrouff, J., et al. (2011), Brain functional integration decreases during propofol-induced loss of consciousness, *NeuroImage*, *57*(1), 198 – 205, doi:10.1016/j.neuroimage.2011.04.020.

Schumer, M., and K. Steiglitz (1968), Adaptive step size random search, *Automatic Control, IEEE Transactions on*, *13*(3), 270 – 276, doi:10.1109/TAC.1968.1098903.

Turin, G. (1960), An introduction to matched filters, *Information Theory, IRE Transactions on*, *6*(3), 311–329, doi:10.1109/TIT.1960.1057571.

White, R. (1971), A survey of random methods for parameter optimization, *SIMULATION*, *17*(5), 197–205, doi:10.1177/003754977101700504.

Zabinsky, Z. B., R. L. Smith, J. F. Mcdonald, H. Edwin, and D. E. Kaufman (1993), Improving hit-and-run for global optimization, *J. Global Optim*, *3*, 171–192.

Zabinsky, Z. B., J. J. Cochran, L. A. Cox, P. Keskinocak, J. P. Kharoufeh, and J. C. Smith (2010), *Random Search Algorithms*, John Wiley & Sons, Inc., doi: 10.1002/9780470400531.eorms0704.