Evaluating the PCA-based Eigenconnectivity Approach to Extract Dynamic Functional Brain Connectivity Patterns

Xueying Zhang (s1605747)

Thesis advisors: Dr. Tom F. Wilderjans and Dr. Wouter D. Weeda

MASTER THESIS

Defended on Month Day, Year

Specialization: Statistical Science

Universiteit Leiden

Statistical Science

Leiden University Medical Center • Mathematical Institute • Institute of Psychology
Wageningen University • VU Medical Center
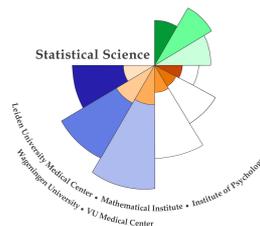
**STATISTICAL SCIENCE
FOR THE LIFE AND BEHAVIOURAL SCIENCES**

Abstract

Functional connectivity (FC) is an important metric to characterize brain mechanisms. Assessment of resting-state FC is a popular tool for studying brain disease mechanisms. Correlations between functional magnetic resonance imaging (fMRI) blood-oxygenation-level-dependent (BOLD) time courses in different brain regions can measure FC which has revealed a meaningful organization of spontaneous fluctuations in the brain during rest. Therefore, in most studies, the presence of temporal and spatial dynamics of FC are usually measured by the correlation coefficients between the fMRI signals of several brain regions. However, recent research has shown that FC is not stationarity. That is, FC dynamically changes over time reflecting additional and rich information about brain organization. In 2013, Leonardi et al. proposed a new approach which was based on principal component analysis (PCA) to reveal hidden patterns of coherent FC dynamics across multiple subjects.

This thesis evaluates this new approach in a simulation study. Moreover, also a framework to test the new approach is proposed. The simulation study showed advantages and disadvantages of the new approach. The results of the simulation study showed that the new approach can extract the most important dynamic connectivity features underlying fMRI data. It can retrieve time-varying connectivity between dynamic brain regions during rest effectively. The new approach identified connections with similar fluctuations, and gave an efficient linear representation, but only sensitive to linear relations between connectivity pairs, and it yielded robust results in restricted conditions. Finally, some recommendations for researchers using this method to study dynamic brain functional brain connectivity at rest are provided.

Contents

Chapter 1

Introduction

In the brain, when people engage in different activities, the activations of neurons will fluctuate constantly, from simple tasks such as controlling one's hand to pick up something to complex cognitive activities like understanding dialogues [1]. The brain also has many specialized parts. Therefore, activities involving sight, hearing, touch, language, memory, etc. have different patterns of neuronal activity [1]. Even when people rest quietly with their eyes closed, the brain is still very active, and this mode of action in a resting state is thought to reveal specific regional networks that often act together. Resting state is a state in which the brain does not perform specific cognitive tasks, maintains a quiet, relaxed, awake state, and is the most basic and most essential state of the complex state of the brain. Resting-state networks have always existed in healthy subjects, with different stages of consciousness and between different species.

Functional Magnetic Resonance Imaging (fMRI) is a safe technique for measuring and mapping non-invasively brain activity [1]. It has been used in many studies to better understand how healthy brains work. In recent years, there has been an explosive growth in the number of neuroimaging studies performed using fMRI [2]. For example, many important functional networks in the brain were discovered, including the default mode network (DMN), visual neural networks, auditory networks, emotional networks, etc.; the connection strength of DMN in resting state is obviously greater than the task state, which reduces its activity during attention-demanding tasks [3]. These networks with coherent activity during rest are consistent across subjects and

are very similar to the functional organization of brain-evoked responses [3]. DMN is the core area of the structure of the brain functional research [4]. Furthermore, these research results have been widely applied to disease research, such as to study the differences in resting state functional connectivity between patients with depression and healthy subjects; in this regard, substantial differences were found in the emotional networks, visual cortical areas and cerebellum [5].

Spontaneous fluctuations of the fMRI blood-oxygen-level-dependent (BOLD) signal are not random but temporally coherent between distinct brain regions [3]. Coherent BOLD activity persists during sleep and in anaesthetized monkeys, suggesting that it reflects a fundamental property of the brain's functional organization [6, 7]. Coherent BOLD activity, known as "functional connectivity" (FC), is crucial for elucidating how neurons and neural networks process information [8, 9]. It refers to the pattern of statistical dependencies between different units within the nervous system. FC is modulated by learning, cognitive and affective states and also shows some spontaneous patterns that are consistent [8, 10, 11]. FC measurement of fMRI BOLD correlations between different brain regions has revealed meaningful organization of spontaneously fluctuations in the resting brain [3]. Connectivity refers to links between functionally interacting brain regions. FC is usually measured during resting-state fMRI and is typically analyzed in terms of correlation, coherence, and spatial grouping based on temporal similarities [9].

Research on resting-state FC is mostly based on stability state analysis, that is, assuming the functional connectivity network is stable over time. However, more and more evidence shows the non-stationary of FC in resting-state. Indeed, FC is not constant over time but changes over time. Several recent studies have since used the term dynamic FC to specifically refer to fluctuating connections during rest; stationary FC refers to connectivity estimates under the assumption of temporal stationarity [3]. When there is no temporal stationarity assumed, whole-brain and multi-subject FC studies become very complicated. So far, FC analysis mainly used sliding time-window technology to study its dynamic nature [12].

Although recent studies have shown that resting-state brain activity has substantial dynamic characteristics that are time-varying, there is a lack of an effective method for estimating time-varying FCs; this is mainly due to the difficulty in selecting a suitable window to locate the time-

varying correlation coefficient (TVCC) [13]. Pearson product-moment correlation coefficient is a measure of connectivity between two variables and it is also a fundamental measure of FC between the time-series of different brain regions. The correlation coefficient is not dependent on time for stationary signals. However, converging evidence has shown that the FC measured by correlation shows substantial fluctuations over time. The time-varying correlation coefficient (TVCC) is estimated by a sliding window approach which is a simple but useful approach, and it employs a window to localize the time-varying information [14]. It is very important for the performance of the estimator to select an appropriate window size in order to avoid excessive bias and to reduce the estimation variance [15]. To estimate slowly time-varying correlations, a long time window is needed to average out the additive noise, and for fast time-varying correlations, a small window captures dynamic features preferably. When applying the sliding window method to study dynamic FC, a short window can help detect more transitions, but due to the small number of samples used, the variability can be large. On the other hand, a long window may reduce the variability of the FC estimate and thus have a higher risk of smoothing out meaningful transitions (i.e., changes over time in functional connectivity). The current sliding window analysis of the dynamic FC depends on the empirically selected fixed window size, ranging from 30 seconds to 60 seconds [16]. Therefore, it is still an open problem to estimate dynamic FC because of the need of an appropriate window selection.

In this regard, Leonardi et al. (2013) proposed a data-driven adaptive window estimation approach based on principal component analysis (PCA) to reveal hidden patterns of coherent FC dynamics -which are called eigenconnectivities- across multiple subjects [3]. Principal component analysis (PCA) is applied to find intrinsic structure in the data that here represent patterns of coherent FC dynamics [3]. PCA is a dimension reduction statistical method. It not only reduces the dimension of high-dimensional data but also importantly eliminates noise through dimension reduction and discovers patterns in the data. Specifically, the method of Leonardi et al. (2013) [3] used sliding time window to estimate the whole brain dynamic FC of the regional- averaged BOLD activity. The PCA is then used to determine the FC patterns that reflects the meaningful pattern in FC fluctuations. They utilize multiple whole brain FCs from subjects to identify connectivity patterns which have coherent dynamics with a particular contributions across time and found a large-scale connectivity patterns that contain regions from different functional clusters. The

temporal contribution of eigenconnectivities can reveal their importance in explaining dynamic FC over time. Moreover, their results complement traditional resting-state analysis and reveal new insights into brain connectivity dynamics.

Nonetheless, the performance of the new eigenconnectivities approach of Leonardi et al. (2013) [3] has not been evaluated yet. As such, it is unknown whether this method is able to extract the most important dynamic connectivity features underlying fMRI data. Moreover, no clear framework exists that can be used to evaluate this (and similar) method(s). Therefore, the goal of this project is two-fold. First, a framework for testing methods that try to capture dynamic FC will be developed. Second, based on this framework, the eigenconnectivities method of Leonardi et al. (2013) [3] will be evaluated.

The thesis is organized as follows. Chapter 2 contains a detailed explanation of the methodology of Leonardi et al. (2013) [3]. The full design and procedure of the simulation study are discussed in Chapter 3, along with the results of the simulation study. Chapter 4 involves general remarks on the simulation results, a conclusion and some ideas for future work. The R-code and some further results are presented in Appendix.

Chapter 2

Methodology

The purpose of the research of Leonardi et al. (2013) is to develop a data-driven approach to reveal characteristic fluctuations of the whole-brain FC over time [3]. Leonardi et al. (2013) estimated the dynamic FC by calculating the correlations over time between the BOLD activity of anatomically defined brain regions covering the whole brain -as obtained by an atlas or Independent Component Analysis- using sliding time-windows [3, 17]. Finally, they identified FC patterns , representing eigenconnectivities, by means of PCA that reflect meaningful patterns in FC fluctuations over time.

To this end, they averaged the fMRI signal over all voxels to estimate regional mean time series in each brain region [3]. Some time series were discarded due to ventral signal dropout in some subjects, resulting in N brain regions. Before applying PCA, the pairwise Pearson correlations between the N brain regions were calculated using a sliding-window technique yielding a $N \times N$ correlation matrix for each window. The function

$$\rho_{xy}(\Delta t) = corr(x[t, t + \Delta t], y[t, t + \Delta t]) \tag{2.1}$$

gives the sliding-window correlation between the time series x and y, where Δt was the window length in TRs. The repetition time (TR) is the length of time between consecutive points on a repeating series of pulses and echoes [18]. Correlations ρ were Fisher r-to-z transformed (z = atanh(ρ)) to make them approximately normally distributed. Only the upper triangular part

of the correlation matrix was kept due to symmetry when vectorizing the correlation matrices. Then a dynamic FC matrix C was built for each subject, where each column included a vectorized correlation matrix. The result is a $(N^2 - N)/2 \times T_s$ matrix, where T_s is the number of windows for subject s (see Figure 2.1a). Note that in this study the number of windows per subject T_s was kept constant across subjects.

Next, PCA was applied. First, features were centered, that is, each dimension of the data is subtracted from the mean of that dimension; here "dimension" refers to a connectivity feature (i.e., a correlation between two specific brain regions). Thus, the row-wise means of C was subtracted (see Figure 2.1b). Next, the data of all s subjects were concatenated along the temporal dimension:

$$X' = [C_1 - \bar{C}_1, C_2 - \bar{C}_2, ..., C_s - \bar{C}_s] \tag{2.2}$$

where X' is a new matrix that contains the mean of each row is subtracted from the data; i.e. $X' = X - \bar{X}$. X is a matrix including the dynamic FC matrices of all s subjects: $X = [C_1, C_2, ..., C_s]$, and $\bar{X}$ is a matrix containing the rowwise means: $\bar{X} = [\bar{C}_1, \bar{C}_2, ..., \bar{C}_s]$. Secondly, they calculated the covariance matrix of X'. Thirdly, the eigenvectors and eigenvalues of the covariance matrix were found by

$$X'X'^T = U\Lambda U^T \tag{2.3}$$

where U is a matrix with orthonormal eigenvectors in its columns and of size $(N^2 - N)/2 \times (N^2 - N)/2$, Λ is a diagonal matrix with the corresponding eigenvalues on its diagonal. These eigenvectors are called "eigenconnectivities", because they catch connectivity patterns. Lastly, the eigenconnectivities were ordered from large to small. Only the K eigenconnectivities with the K largest eigenvalues were used to approximate the data:

$$X'_{approx} = U_{red}U_{red}^T X' \tag{2.4}$$

where U_{red} contained only the first K eigenconnectivities. The first K eigenconnectivities were

assumed as $\lambda_1, ..., \lambda_K$, resulting in a retained variance of

$$\sum_{k=1}^{K} \lambda_k \bigg/ \sum_{k=1}^{\left(N^2 - N\right)/2} \lambda_k \qquad (2.5)$$

After that, A set of time-dependent weights W_s for each subject was calculated by projecting the centered dynamic FC matrix onto the first K eigenconnectivities:

$$W_s = U_{red}^T (C - \bar{C}) \qquad (2.6)$$

where W_s was a $K \times T_s$ matrix including the weights of K eigenconnectivities at all T_s windows (see Figure 2.1c). Thus, a set of eigenconnectivities can be interpreted as important building blocks of dynamic FC.
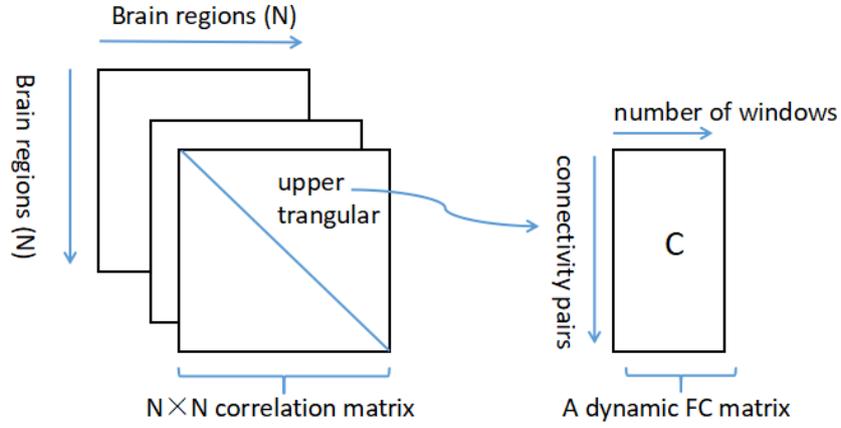
(a) Each column of a dynamic FC matrix C contained the upper triangular part of a $N \times N$ vectorized correlation matrix
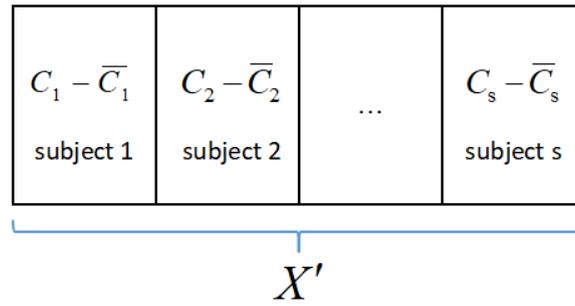
(b) The row-wise means were subtracted for each subject before concatenating all subjects
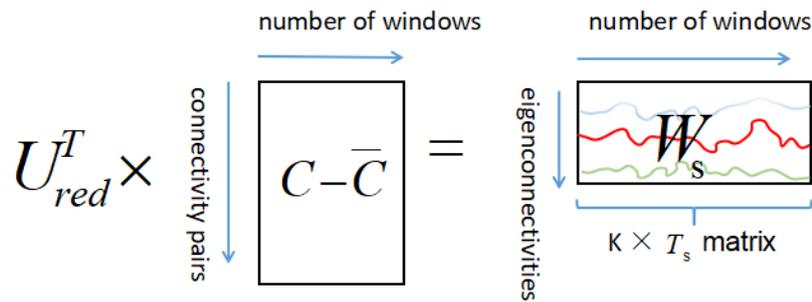
(c) A set of time-dependent weights W_s for each subject was calculated by projecting the centered dynamic FC matrix onto the K eigenconnectivities

Figure 2.1: Processing of Leonardi et al. (2013) method

Chapter 3

Simulation study

3.1 Problem

The performance of the new eigenconnectivities approach of Leonardi et al. (2013) [3] has not been evaluated yet. Hence, it is unknown whether this approach can extract the most important dynamic connectivity features underlying fMRI data. In addition, there is no clear framework that exists that can be used to evaluate this (and similar) approach(es).

Therefore, a simulation study is needed. In statistics, properties of statistical methods must be established so that the methods may be used with confidence. The simulation study is a numerical technique for conducting experiments on the computer. It involves random sampling from probability distributions. It may be a conceptual model of how the elements of the system interact and communicate, or it may be a structural model that shows how the components are in a hierarchy [19]. The process with which we generate dynamic FC patterns to describe the changes in FC is outlined in detail in the design and procedure below.

3.2 Design and procedure

One of the most commonly used strategies for examining dynamics in FC is the sliding window approach, i.e. a repeated application of a (static) FC metric on a gradually changing time period [17]. In the simulation, we investigate whether the FC framework can recover the correlation

structure in the data. A framework with dynamic FC patterns need to be generated to describe the changes in FC, so first, we have to build true profiles with regions which can estimate BOLD activity in the subsection below. The expectation of the simulation study is that eigenconnectivities match true profiles well. In other words, the Leonardi et al. (2013) [3] method should be a good approximation to the true profile under the conditions of interest.

3.2.1 True profiles construction

We assume that, for each time window, BOLD activation is measured for four brain regions: A, B, C and D. The mean activation for each region is 0 and the variance is defined as 1. Therefore, there is a true profile matrix containing six correlations between the four regions: AB, AC, AD, BC, BD and CD. In other words, the dimensions of the true profile matrix are the number of windows $\times$ 6 (correlations) (see Figure 3.1).
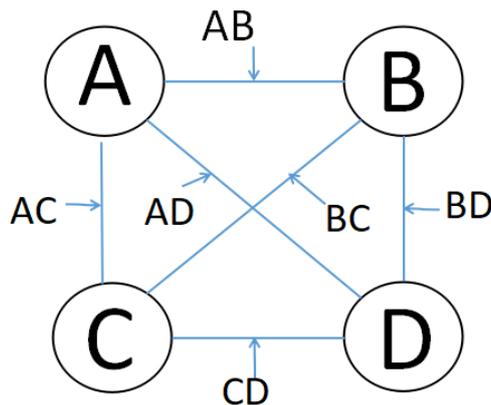
Figure 3.1: Four regions and six correlations of each window

These correlations belong to one type out of three types of link: 0-link, constant link and varying link. The correlations AD, BD and CD are assumed to be 0-link, in other words, the values of AD (correlation between A and D), BD (correlation between B and D) and CD (correlation between C and D) are 0 for each window. AB is a constant link as the value of AB (correlation between A and B) is .50 for each window. BC and AC are supposed to be varying links. In particular, AC (correlation between A and C) has an increasing pattern with values

increasing from .10 to .90 across the windows. BC (correlation between B and C) has a V-pattern, with decreasing value from .80 to .10 in the first half number of windows and increasing values from .10 to .80 for the other half of windows (see Figure 3.2).
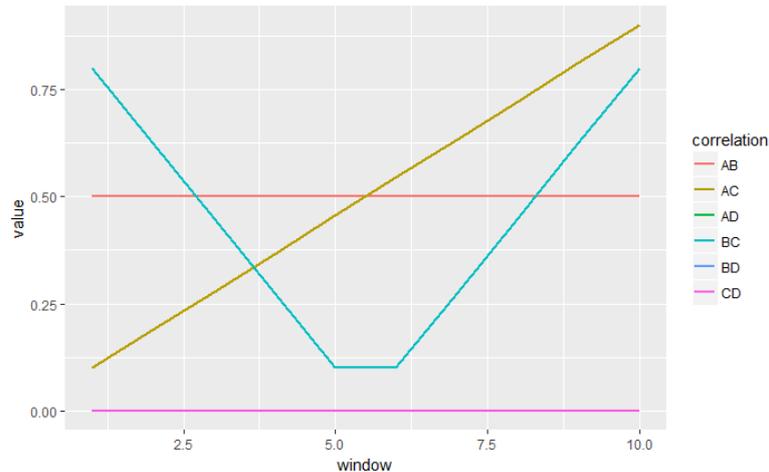
Figure 3.2: Six correlations between four regions across 10 windows. AC and BC are varying links, AB has a constant link and AD, BD and CD have 0-link.

3.2.2 Conditions definition

In addition, several factors may affect the performance of estimation. When applying the simulation, some conditions need to be assumed, which are the number of the windows (window), the length of each window (length), the number of subjects (subject) and the signal-to-noise ratio (SNR). The size of each factor is also varied. They are defined in the different levels as follows:

- the number of windows, at 2 levels: 10, 50;

- the window length, at 2 levels: 30, 100;

- the number of subjects, at 2 levels: 20, 50;

- the Signal to Noise ratio, at 3 levels: 1, 2, 5.

The number of windows is investigated by simulating at two different levels: 10, 50. It has to make sure that whether have more number of windows produces better estimates of FC or

not. According to previous study, the number of windows depends on window size that the use of small size windows requires a larger number of windows to maintain performance [20].

Another one of the most important parameters of sliding window approach is the choice of the employed window size: ideally, the window should be large enough to permit robust estimation of FC and yet be small enough to detect potential transitions [21]. Correlations estimated from short windows have more variance due to the small number of samples used but are more sensitive to changes in FC since long windows reduce temporal variability and may miss anticorrelations [17]. Given the temporal resolution of conventional fMRI (2-3 s), possible window sizes are limited. A higher temporal resolution increases the possible number of samples per window, which in turn could provide a more robust estimation of FC. In the simulation study, the window length was simulated for two different levels: 30, 100.

The number of subjects is investigated for two different levels: 20, 50. The reason is that we would like to know that how many number of subjects (and as a consequence data) can yield estimates of FC well.

Moreover, white noise consists of Gaussian (uncorrelated) error that is added under three single-trial SNR conditions: 1, 2 and 5. SNR is the mean signal divided by the standard deviation of the noise. The white noise added to the signal is normally distributed with zero mean, and it contains $window \times length$ numbers of observations. Hence, the standard deviation of the noise is found by $1/SNR$, which has three levels (1, 1/2, 1/5).

Thus, there are in total 2 (number of windows) $\times$ 2 (window length) $\times$ 2 (number of subjects) $\times$ 3 (SNR) = 24 conditions in the simulation study. For each condition, 50 replicate data sets are created, resulting in the generation of $24 \times 50 = 1200$ data sets.

3.2.3   Data generation

The simulation is performed by using R function mvrnorm from the R-package MASS [22]. The function simulates from the specified multivariate normal distribution to produce one or more samples. Arguments for n (the number of samples), mu (a vector giving the means of the variables) and $Sigma$ (a positive-definite symmetric matrix specifying the covariance matrix of the variables) are required when using the function. In the simulation design, we first use the

function mvrnorm to simulate a true profile matrix (number of windows × 6 correlations). The factor window length is the number of samples. The means of the variables had defined as 0. A positive-definite symmetric matrix specifying the covariance matrix of the variables has to be calculated. For example, the covariance of AB is determined by:

$$cov(AB) = cor(AB) \times \sqrt{cov(AA) \times cov(BB)} \qquad (3.1)$$

The covariances of AA, BB, CC and DD are assumed as 1. As such, the covariance of AB, AC, AD, BC, BD and CD can be computed separately (for each window). Next, we merge results vertically (across windows) to get a data matrix for each subject. Finally, a data set for all subjects is generated by combining the data of each subject.

3.2.4 Principal components

After that, the Leonardi et al. (2013) [3] method which was explained in Chapter 2 is performed to analysis the generated data and to get the first K eigenconnectivities. K equals the number of principal components to retain when performing PCA. We want the proportion of variation explained by the first K principal components to be large. Usually, it is acceptable to take the first K components that together capture at least 85% of the total sum of the eigenvalues. In this simulation study, the first $K = 5$ eigenconnectivities can explain above 90% of the cumulative variation with all 24 conditions, it may even get 99% of variation under certain conditions.

3.2.5 Experiments

The true profiles, which has six correlations (see Figure 3.2), are compared to the retrieved eigenconnectivities to check which true profile matches with which eigenconnectivity. To this end, the correlation between the true profile and the eigenconnectivities is computed. Because AC and BC are time-varying links, the correlations between the first K eigenconnectivities and AC and BC are computed. Next, the maximum of absolute value of the correlations is found, and, as such, AC and BC are matched to a single eigenconnectivity for each subject. The other links are different than AC and BC in that these links have true profiles that are constant over time and thus no correlation between these true profile and the eigenconnectivities can be computed.

As a solution, first the eigenconnectivity matching AB is sought and next the eigenconnectivities matching with AD/BD/CD. To do this, the two eigenconnectivities with the lowest standard deviation are selected; next, the mean of these two eigenconnectivities is computed to find out the one matching AB (with largest mean) and the other one (with lowest mean) matching AD/BD/CD. The next step is to look at the standard deviation of the eigenconnectivity for AB and AD/BD/CD separately. This standard deviation is zero when AB and AD/BD/CD are recovered perfectly.

After this matching step, analysis of variance (ANOVA) is applied to determine how the estimated correlation (for AC and BC) and the estimated standard deviation (for AB and AD/BD/CD) varies as a function of the manipulated factors and their interactions. ANOVA is a statistical technique that suits comparing (testing) three or more means (groups) for statistical significance [23]. In this project, four-way ANOVA is used with the estimated correlation (for AC and BC) or the estimated standard deviation (for AB and AD/BD/CD) as the dependent variable and the factors number of windows, window length, number of subjects, SNR and their interactions as independent variables.

3.2.6   Software

The whole analysis of the project was done with the open source software environment R, using the integrated development environment RStudio. Different software packages were used. For coding the simulation, the R-package MASS [22] was used. Plots were made with the R-package ggplot2 [24], grid [25], gridExtra [26] and dplyr [27]. An additional product in this thesis is the R-code of the simulation study which can be used for reproducibility (see Appendix B).

3.3   Results

In this section, we present the results from the simulation study. First, we illustrate the average correlation (for AC and BC) and standard deviation (for the other links) across all manipulated data sets (Figure 3.3 and Figure 3.4). Then, the ANOVA results (Table 3.1 and Figure 3.5) for the correlation (for AC and BC) and the standard deviation (for the other links) are shown to see that whether there are factors influencing this result. Because the true profile for BD and

15

CD is the same as the true profile for AD and similar results have been found for these three profiles, only results for the AD profile are presented. The full ANOVA tables for AC, BC, AB and AD are shown in Appendix A.

Figure 3.3 shows the boxplots of the mean correlation as a function of SNR, number of windows, window length and number of subjects for varying links (AC and BC) in different number of subjects. The subject factor does not have an effect on the correlation between the varying links, because boxplots are the same for different number of subjects. Each correlation value for the increasing-pattern (AC) is close to each correlation value for the V-pattern (BC) and this for all 24 conditions from the simulation design. The conditions with median or high SNR, large window length and large number of windows lead to the highest correlations. Moreover, boxplots being comparatively short suggest low variability in the conditions with large number of window, large window length and high SNR. That is, more windows, longer windows and less noise can reduce variability in the estimation.

Figure 3.4 displays the boxplots of the standard deviation as a function of SNR, number of windows, window length and number of subjects for the constant link (AB) and the 0-link (AD) for different number of subjects. The subject factor does not have an effect on the standard deviation for the constant link and the 0-link, because boxplots are the same for different number of subjects. Only the condition of window length has a strong effect on the standard deviation. Larger window lengths lead to a smaller standard deviation and lower variability. The standard deviation for the constant link has a little bit larger values, while the standard deviation for the 0-link has quite small values for all situations.

Table 3.1 shows the Eta Square value for all manipulated factors and their interactions for the links AC, BC, AB and AD. To determine the most important main and interaction effects, the eta squared is used as effect size. The eta squared is the proportion of the total variability in the dependent variable that is accounted for by the variation in the independent variable. It is the ratio of the sum of squares for each group level to the total sum of squares. It can be interpreted as a percentage of variance accounted for by a variable [28]. It is useful beyond significance tests (p-values) because it estimates the magnitude of effects, independent from sample size (p-values alone do not indicate the size of an effect). The general rules of thumb are given by Miles and Shevlin (2001) [29] for eta squared, which uses the total sum of squares in the denominator. The

16
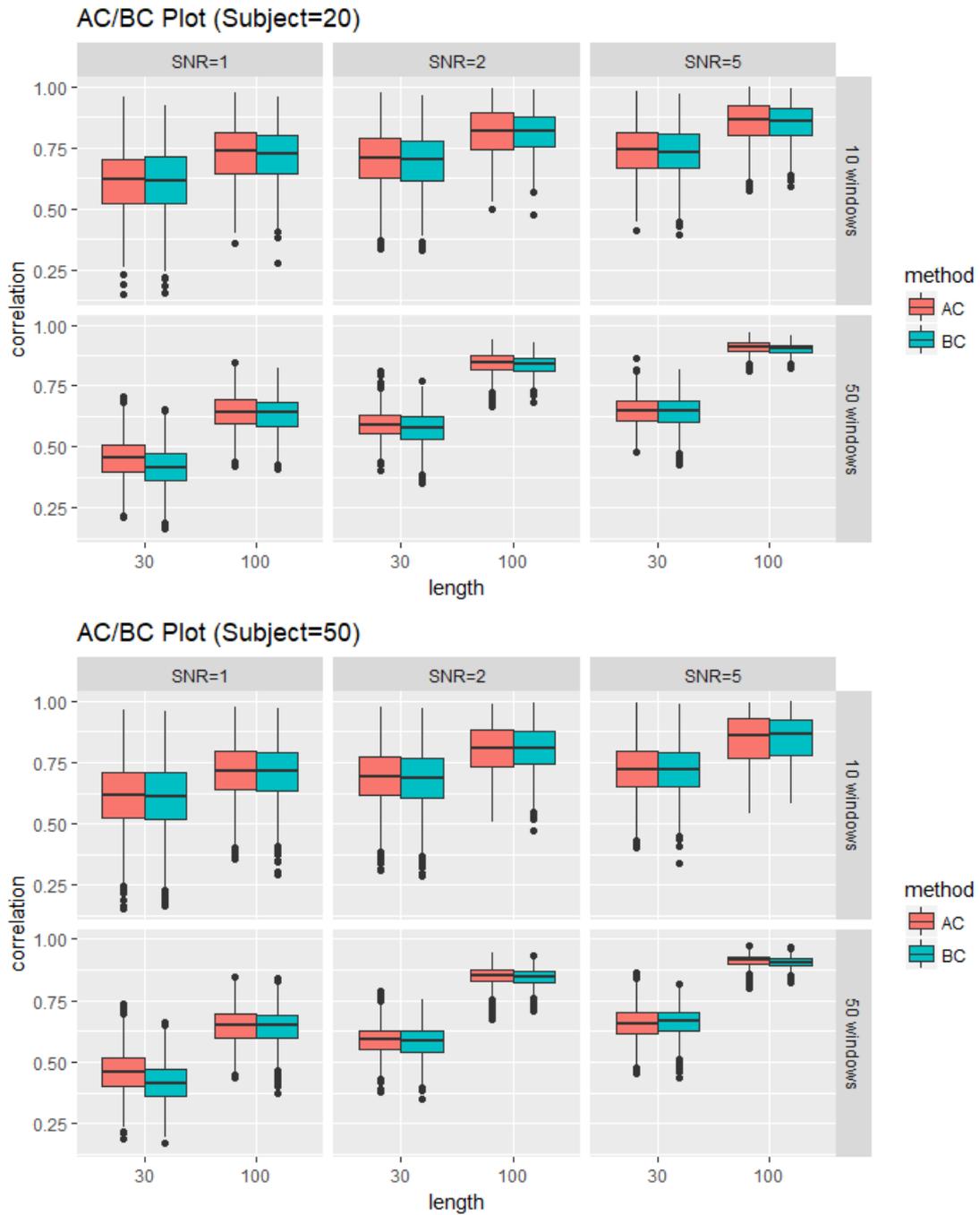
Figure 3.3: Boxplots of the mean correlation as a function of SNR (columns), number of windows (rows), window length (x-axis) and number of subjects (upper panel: 20 subjects; lower panel: 50 subjects) for AC and BC links.
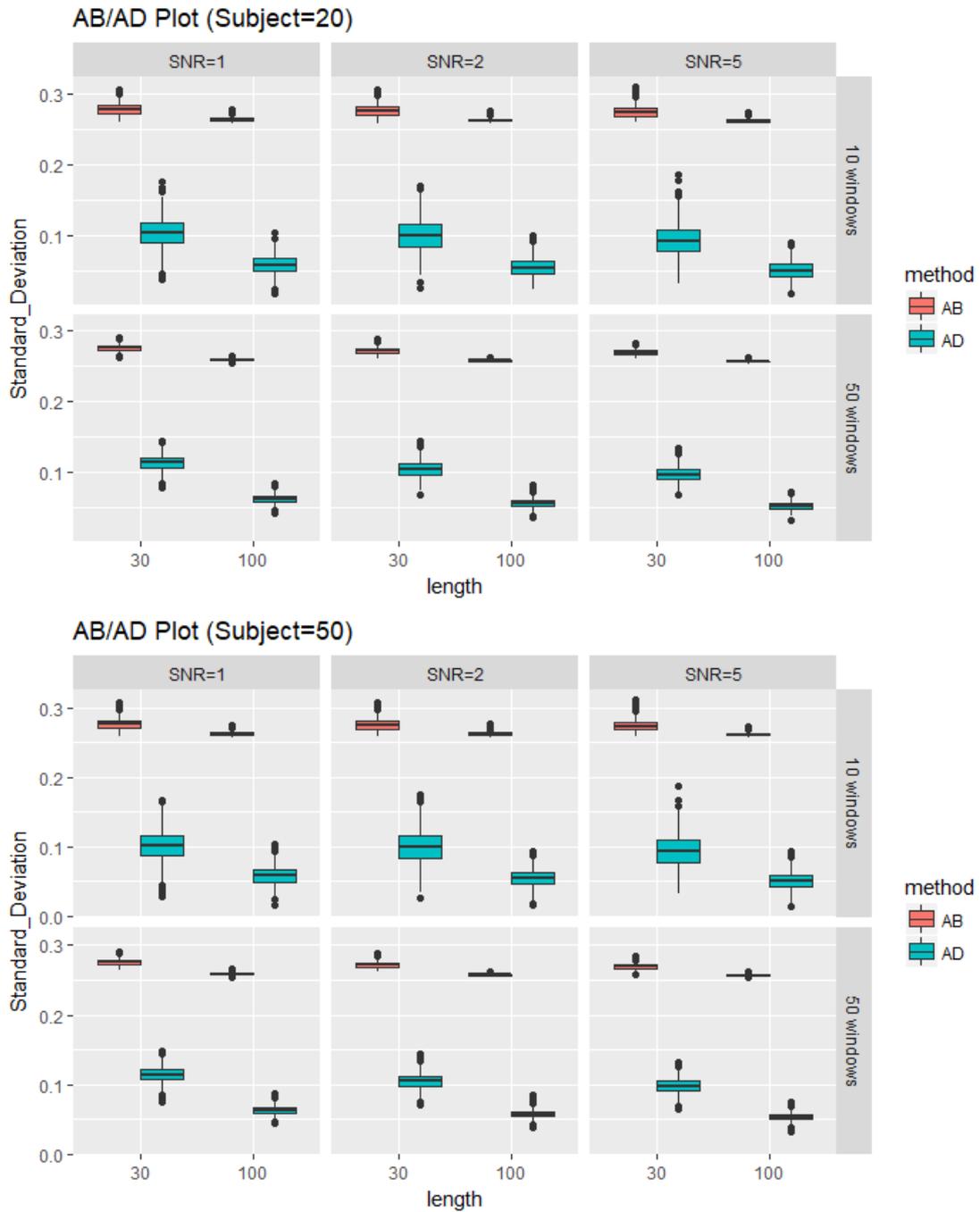
Figure 3.4: Boxplots of the standard deviation as a function of SNR (columns), number of windows (rows), window length (x-axis) and number of subjects (upper panel: 20 subjects; lower panel: 50 subjects) for AB and AD links.

Table 3.1: Eta Square table for links AC, BC, AB and AD

	AC	BC	AB	AD
SNR	.183 ***	.193 ***	.022 ***	.023 ***
window	.025 ***	.030 ***	.046 ***	.009 ***
length	.345 ***	.361 ***	.625 ***	.696 ***
subject	.000 ***	.000 ***	.000	.000
SNR:window	.015 ***	.021 ***	.001 ***	.001 ***
SNR:length	.001 ***	.000 *	.004 ***	.001 ***
window:length	.036 ***	.031 ***	.002 ***	.001 ***
SNR:subject	.000	.000	.000	.000
window:subject	.000 ***	.000 ***	.000	.000
length:subject	.000 **	.000 ***	.000	.000
SNR:window:length	.000 ***	.000 ***	.001 ***	.000 ***
SNR:window:subject	.000 ***	.000 **	.000	.000
SNR:length:subject	.000 ***	.000	.000	.000
window:length:subject	.000 *	.000	.000	.000
SNR:window:length:subject	.000 *	.000 **	.000	.000

Significant codes: *** p<.001; ** p<.01; * p<.05

values for .01, .06 and .14 indicate a small, medium and large effect size, respectively. In this case, to determine which main and interaction effects of the manipulated factors are important, a value of eta squared larger than .01 is chosen.

It appears from Table 3.1 that the factor window length has the largest effect size among all factors and this for every link (i.e., an Eta Square value above .30 for all links). When the window length increases, the correlation between links and eigenconnectivities will increase. As the window length decreases, the standard deviation between links and eigenconnectivities will decrease. The factor SNR represents a large effect for varying links (AC and BC) -values of Eta Square are .183 and .193 for increasing-pattern (AC) and V-pattern (BC), while it is a small effect for the constant link (AB) and the 0-link (AD) with eta square values of .022 and .023. The factor number of windows has a small effect size for both varying links and the constant link with eta square values of .025, .030 and .046, respectively, but it does not affect the 0-link. Otherwise, the factor number of subjects does not seem to be an important effect for any of the links. The correlation (or standard deviation) between links and eigenconnectivities will not change if the number of subjects increases (or decrease). The interaction between SNR and the number of windows has a small effect size for both varying links with eta square values of .025 and .030 respectively, but it does not affect the constant link and the 0-link. The interaction

between the number of windows and window length has also as small effect for the two varying links with eta square values of .036 and .031; it has no effect on the constant link and the 0-link. The remaining second-, third- and fourth order interactions among factors are not important effects for any of the links.

In Figure 3.5, plots of important interaction effects of varying links (AC and BC) are presented. Interaction plots between SNR and the number of windows are shown in the left panel of Figure 3.5. They are almost the same for the increasing-pattern (AC) and the V-pattern (BC). The correlation increases with increased SNR. In the high SNR situation, small number of windows and large number of windows have nearly the same correlation. In contrast, when SNR is low or medium, the small number of windows always has a higher correlation than the large number of windows, with this effect being the strongest when SNR is low. The right panel of Figure 3.5 shows plots of the important interaction between window length and the number of windows for varying links (AC and BC). They are almost the same for the increasing-pattern (AC) and the V-pattern (BC). The correlation increases with increased length. In the large window length situation, small number of windows and large number of windows have almost the same correlation. However, the small number of windows has a higher correlation than the large number of windows in the small window length situation.
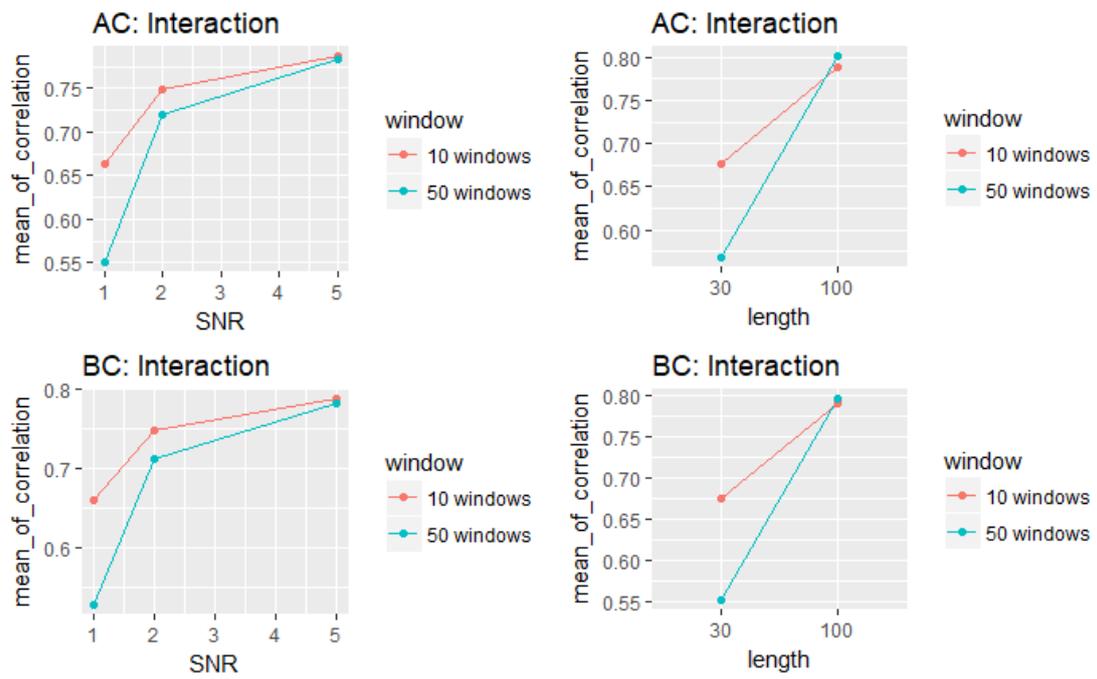
Figure 3.5: The mean correlation as a function of the interaction between SNR (x-axis) and the number of windows (colours), and the interaction between window length (x-axis) and the number of windows (colours) for AC and BC links

Chapter 4

Discussion

Leonardi et al. (2013) [3] proposed a novel approach to estimate building blocks of dynamic brain connectivity. In this thesis, a framework was built to test the performance of such methods for capturing dynamic FC. The new PCA-based eigenconnectivity approach of Leonardi et al. (2013) [3] was evaluated by means of a simulation study following the proposed framework.

Results of simulations have shown that the new method can recover correlations between eigenconnectivities and the framework data we made. More windows, longer windows and less noise increase the estimated correlation. Higher correlation means that there is a strong linear relationship between eigenconnectivities and varying links of true profile. 0-links are closely distributed around the mean values of eigenconnectivities. However, eigenconnectivities and constant link are spread out widely. This indicates that discrimination is not very effective, and the recognition ability of the eigenconnectivities is a bit poor. When using the new approach to analysis dynamic FC patterns, some useful recognized information of fMRI data may be lost. The generalization ability of the new method is not very strong. Specifically, there should be enough windows and windows should be long enough. Moreover, noise should avoided as much as possible in order to permit a robust estimation of FC when using the new PCA approach.

To conclude, FC analysis with PCA can be a helpful tool to study the time-varying connectivities between dynamic brain regions during rest. The advantage is that the new approach evaluated the contributions of connectivity patterns to dynamic FC at any time point. In addi-

tion, it provides an efficient linear representation and effective dimension reduction.

The approach proposed by Leonardi et al. (2013) [3] has its own limitation in that PCA is a linear decomposition and as such is only sensitive to linear relations between connectivity pairs. Moreover, the eigenvalues were close in magnitude, because it might require to specify regions-of-interest a-priori or to rely on a GLM analysis that is stringently corrected for multiple comparisons [30]. An option would be to use Kernel PCA, a PCA expansion, to identify FC patterns based on nonlinear relations [31]. The fMRI data is first converted into a new space by the kernel function and then processed by PCA. Future work may consider using nonlinear dimension reduction methods such as Kernel PCA to estimate the FC model.

The simulation study of the project also has some limitations. First, more extensive simulations could be executed, ensuring robust inference on the performance of the algorithm. Further work can assume more regions, define more time-varying correlation patterns between brain regions, add more factors, and assign more different sizes for each factor to make the simulation more credible. Second, we defined noise by using single-trial SNR from a normal distribution. White noise can be alternatively added, for instance, according to the method described in Weeda et al. (2009) [32] in which the amplitude of the signals was sampled from a multivariate normal distribution with correlation matrix between regions which are modeled by multivariate Gaussian [30]. Third, the simulation study was not tested by real data in this thesis. Its availability and effectiveness are yet to be verified in the future.

Finally, it must be also mentioned that other simulation studies should be explored to evaluate the dynamics of eigenconnectivity time-dependent weights. More complex estimators of dynamic FC could easily be used with the proposed data-driven approach.

Bibliography

[1] UC San Diego Center for Functional MRI. "What is fMRI?" *Center for Functional MRI.* `http://fmri.ucsd.edu/Research/whatisfmri.html` (accessed March 21, 2018).

[2] Lindquist, Martin A. "The Statistical Analysis of FMRI Data." *Statistical Science* 23, no. 4 (2008): 439-64.

[3] Leonardi, Richiardi, Gschwind, Simioni, Annoni, Schluep, Vuilleumier, and Van De Ville. "Principal Components of Functional Connectivity: A New Approach to Study Dynamic Brain Connectivity during Rest." *NeuroImage* 83 (2013): 937-50.

[4] Raichle, Marcus E., Ann Mary Macleod, Abraham Z. Snyder, William J. Powers, Debra A. Gusnard, and Gordon L. Shulman. "A Default Mode of Brain Function." *Proceedings of the National Academy of Sciences of the United States of America* 98, no. 2 (2001): 676-82.

[5] Zeng, Ling-Li, Shen, Hui, Liu, Li, Wang, Lubin, Li, Baojuan, Fang, Peng, Zhou, Zongtan, Li, Yaming, and Hu, Dewen. "Identifying Major Depression Using Whole-brain Functional Connectivity: A Multivariate Pattern Analysis." *Brain* 135, no. 5 (2012): 1498-507.

[6] Larson-Prior, Linda J, John M Zempel, Tracy S Nolan, Fred W Prior, Abraham Z Snyder, and Marcus E Raichle. "Cortical Network Functional Connectivity in the Descent to Sleep." *Proceedings of the National Academy of Sciences of the United States of America* 106, no. 11 (2009): 4489-94.

[7] J. L. Vincent, G. H. Patel, M. D. Fox, A. Z. Snyder, J. T. Baker, D. C. Van Essen, J. M. Zempel, L. H. Snyder, M. Corbetta, and M. E. Raichle. "Intrinsic Functional Architecture in the Anaesthetized Monkey Brain." *Nature* 447, no. 7140 (2007): 83-86.

[8] Cribben, Haraldsdottir, Atlas, Wager, and Lindquist. "Dynamic Connectivity Regression: Determining State-related Changes in Brain Connectivity." *NeuroImage* 61, no. 4 (2012): 907-20.

[9] Biswal, Bharat, F. Zerrin Yetkin, Victor M. Haughton, and James S. Hyde. "Functional Connectivity in the Motor Cortex of Resting Human Brain Using Echo-planar Mri." *Magnetic Resonance in Medicine* 34, no. 4 (1995): 537-41.

[10] Bassett, Danielle S, Nicholas F Wymbs, Mason A Porter, Peter J Mucha, Jean M Carlson, and Scott T Grafton. "Dynamic Reconfiguration of Human Brain Networks during Learning." *Proceedings of the National Academy of Sciences of the United States of America* 108, no. 18 (2011): 7641-6.

[11] Britz, Van De Ville, and Michel. "BOLD Correlates of EEG Topography Reveal Rapid Resting-state Network Dynamics." *NeuroImage* 52, no. 4 (2010): 1162-170.

[12] Allen, Elena A., Eswar Damaraju, Sergey M. Plis, Erik B. Erhardt, Tom Eichele, and Vince D. Calhoun. "Tracking Whole-Brain Connectivity Dynamics in the Resting State." *Cerebral Cortex* 24, no. 3 (2014): 663-76.

[13] Zhang, Fu, Chan, Hung, Motta, Di, and Biswal. "Adaptive window selection in estimating dynamic functional connectivity of resting-state fMRI." *ICICS 2013 - Conference Guide of the 9th International Conference on Information, Communications and Signal Processing.*

[14] Ozgen, Mehmet Tankut. "Extension of the Capon's Spectral Estimator to Timefrequency Analysis and to the Analysis of Polynomial-phase Signals." *Signal Processing* 83, no. 3 (2003): 575-92.

[15] Zhang, Z., G. Chan, and S. Ho. "On Bandwidth Selection in Local Polynomial Regression Analysis and Its Application to Multi-resolution Analysis of Non-uniform Data." *Journal of Signal Processing Systems* 52, no. 3 (2008): 263-80.

[16] Hutchison, Womelsdorf, Allen, Bandettini, Calhoun, Corbetta, Della Penna, Duyn, Glover, Gonzalez-Castillo, Handwerker, Keilholz, Kiviniemi, Leopold, De Pasquale, Sporns, Walter,

and Chang. "Dynamic Functional Connectivity: Promise, Issues, and Interpretations." *NeuroImage* 80 (2013): 360-78.

[17] Chang, and Glover. "Timefrequency Dynamics of Resting-state Brain Connectivity Measured with FMRI." *NeuroImage* 50, no. 1 (2010): 81-98.

[18] Axel, Leon. "Glossary of MR Terms American College of Radiology." *Magnetic Resonance Imaging* 5, no. 3 (1987): 163-64.

[19] Hall, Eric S. "What are simulation studies?" *ResearchGate* https://www.researchgate.net/post/What_are_simulation_studies (accessed March 10, 2018).

[20] Huifang Elizabeth Wang, Christian G Bnar, Pascale P Quilichini, Karl J Friston, Viktor Ejirsa, and Christophe Ebernard. "A Systematic Framework for Functional Connectivity Measures." *Frontiers in Neuroscience* 8 (2014): Frontiers in Neuroscience, 01 December 2014, Vol.8.

[21] Sakoglu, Unal, Godfrey Pearlson, D. Kiehl, Kent Wang, A. Michael, and Y. Calhoun. "A Method for Evaluating Dynamic Functional Network Connectivity and Task-modulation: Application to Schizophrenia." *Resonance Materials in Physics, Biology and Medicine* 23, no. 5 (2010): 351-66.

[22] Ripley, Brian D. *Stochastic Simulation*. Wiley Series in Probability and Mathematical Statistics 810651580. New York ; Chichester: Wiley, 1987.

[23] Fox, John. *Applied Regression Analysis and Generalized Linear Models*. 2nd [rev.] ed. Los Angeles, CA [etc.]: Sage Publications, 2008.

[24] Wickham, Hadley. *Ggplot2 Elegant Graphics for Data Analysis*. Use R! 325763100. New York [etc.]: Springer, 2009.

[25] Murrell, Paul. *R Graphics*. 2nd[ext. Upd.] ed. Chapman & Hall/CRC Computer Science and Data Analysis Series ; 7. Boca Raton, FL: CRC, 2011.

[26] Auguie, Baptiste, and Antonov, Anton. *gridExtra: Miscellaneous Functions for "Grid" Graphics*. `https://cran.r-project.org/web/packages/gridExtra/index.html`, R package version 2.2.1, 2016.

[27] Wickham, H., Francois, R., Henry, L., Mller, K. and RStudio. *dplyr: A Grammar of Data Manipulation*. `https://cran.r-project.org/web/packages/dplyr/index.html`, R package version 0.7.4, 2017.

[28] Levine, Timothy R., and Hullett, Craig R. "Eta Squared, Partial Eta Squared, and Misreporting of Effect Size in Communication Research." *Human Communication Research* 28, no. 4 (2002): 612-25.

[29] Miles, Jeremy, and Mark Shevlin. *Applying regression and correlation : A guide for students and researchers*. London [etc.]: SAGE, 2001.

[30] Weeda, Waldorp, Grasman, Van Gaal, and Huizenga. "Functional Connectivity Analysis of FMRI Data Using Parameterized Regions-of-interest." *NeuroImage* 54, no. 1 (2011): 410-16.

[31] Schoolkopf, Bernhard, Smola, Alexander, and Mller, Klaus-Robert. "Nonlinear Component Analysis as a Kernel Eigenvalue Problem." *Neural Computation* 10, no. 5 (1998): 1299-319.

[32] Weeda, Wouter D., Lourens J. Waldorp, Ingrid Christoffels, and Hilde M. Huizenga. "Activated Region Fitting: A Robust Highpower Method for FMRI Analysis Using Parameterized Regions of Activation." *Human Brain Mapping* 30, no. 8 (2009): 2595-605.

Appendix A

Summary of ANOVA results

Table A.1: ANOVA table for AC

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
SNR	1	182.0	182.0	19623.612	< 2e-16	***
window	1	25.1	25.1	2702.423	< 2e-16	***
length	1	342.1	342.1	36876.752	< 2e-16	***
subject	1	.4	.4	48.051	4.21e-12	***
SNR:window	1	14.9	14.9	1607.956	< 2e-16	***
SNR:length	1	1.2	1.2	124.705	< 2e-16	***
window:length	1	36.0	36.0	3879.142	< 2e-16	***
SNR:subject	1	.0	.0	2.361	.124389	
window:subject	1	.4	.4	45.786	1.34e-11	***
length:subject	1	.1	.1	6.780	.009224	**
SNR:window:length	1	.2	.2	22.962	1.66e-06	***
SNR:window:subject	1	.1	.1	11.680	.000633	***
SNR:length:subject	1	.1	.1	12.844	.000339	***
window:length:subject	1	.0	.0	4.754	.029241	*
SNR:window:length:subject	1	.1	.1	6.544	.010526	*
Residuals	41984	389.4				

Significant codes: *** p<.001; ** p<.01; * p<.05

Table A.2: ANOVA table for BC

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
SNR	1	203.0	203.0	22363.928	< 2e-16	***
window	1	31.6	31.6	3479.385	< 2e-16	***
length	1	378.7	378.7	41728.846	< 2e-16	***
subject	1	.1	.1	11.398	0.000736	***
SNR:window	1	22.0	22.0	2422.144	< 2e-16	***
SNR:length	1	.0	.0	4.992	.025463	*
window:length	1	32.3	32.3	3556.954	< 2e-16	***
SNR:subject	1	.0	.0	.424	.515185	
window:subject	1	.2	.2	25.279	4.98e-07	***
length:subject	1	.1	.1	11.920	.000556	***
SNR:window:length	1	.3	.3	37.725	8.22e-10	***
SNR:window:subject	1	.1	.1	8.103	.004421	**
SNR:length:subject	1	.0	.0	.000	.988017	
window:length:subject	1	.0	.0	.272	.601667	
SNR:window:length:subject	1	.1	.1	7.545	.006021	**
Residuals	41984	381.0				

Significant codes: *** p<.001; ** p<.01; * p<.05

Table A.3: ANOVA table for AB

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
SNR	1	.0697	.0697	3033.849	< 2e-16	***
window	1	.1483	.1483	6459.532	< 2e-16	***
length	1	2.0183	2.0183	87891.402	< 2e-16	***
subject	1	.0000	.0000	0.205	.650	
SNR:window	1	.0047	.0047	205.093	< 2e-16	***
SNR:length	1	.0127	.0127	550.940	< 2e-16	***
window:length	1	.0079	.0079	344.712	< 2e-16	***
SNR:subject	1	.0000	.0000	.005	.942	
window:subject	1	.0000	.0000	.390	.532	
length:subject	1	.0000	.0000	.004	.951	
SNR:window:length	1	.0022	.0022	96.732	< 2e-16	***
SNR:window:subject	1	.0000	.0000	.073	.787	
SNR:length:subject	1	.0000	.0000	.019	.891	
window:length:subject	1	.0000	.0000	.398	.528	
SNR:window:length:subject	1	.0000	.0000	.040	.841	
Residuals	41984	.9641				

Significant codes: *** p<.001; ** p<.01; * p<.05

Table A.4: ANOVA table for AD

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
SNR	1	.713	.713	3.526e+03	< 2e-16	***
window	1	.286	.286	1.414e+03	< 2e-16	***
length	1	21.950	21.950	1.085e+05	< 2e-16	***
subject	1	.000	.000	1.200e-02	.912	
SNR:window	1	.035	.035	1.725e+02	< 2e-16	***
SNR:length	1	.023	.023	1.153e+02	< 2e-16	***
window:length	1	.036	.036	1.766e+02	< 2e-16	***
SNR:subject	1	.000	.000	1.958e+00	0.162	
window:subject	1	.000	.000	3.740e-01	.541	
length:subject	1	.000	.000	2.700e-02	.870	
SNR:window:length	1	.009	.009	4.599e+01	1.2e-11	***
SNR:window:subject	1	.000	.000	9.180e-01	.338	
SNR:length:subject	1	.000	.000	1.462e+00	.227	
window:length:subject	1	.000	.000	3.600e-01	.549	
SNR:window:length:subject	1	.000	.000	1.600e-01	.689	
Residuals	41984	8.494				

Significant codes: *** p<.001; ** p<.01; * p<.05

Appendix B

R code of simulation study

```r
library(MASS)
#performe simulation
#corr.mat:truen, k largest eigenvalues/first k eigenconnectivities
estimate.vec <- function(nSubj, nWindows, WindowLength, nRegion, SDNoise,
corr.mat, k, vec, vec2, startseed) {
  set.seed(startseed)
  L <- list()
   for(K in 1:nSubj ){
     simulate <- numeric(0)
      for(i in 1:nrow(corr.mat)){
        mean <- rep(0, nRegion)
        var <- 1
        #calculate covariance(Sigma), cov(AB) = cor(AB)*sqrt(cov(AA)*cov(BB));
        #assume cov(AA) = cov(BB) = cov(CC) = cov(DD) = 1
        a <-corr.mat[i,]
        b <- matrix(1, nRegion , nRegion)
        b[lower.tri(b, diag=FALSE)] <- a
        b[upper.tri(b)] <- t(b)[upper.tri(b)] #b is a covariance matrix (Sigma)
```

```r
    simulate <- rbind(simulate, mvrnorm(n=WindowLength, mu=mean, Sigma=b))
  }
   L[[K]] <- simulate


 #ADD NOISE TO DATA
  for(i in 1:ncol(L[[K]])) {
    L[[K]][,i] <- L[[K]][,i] + rnorm(n = nWindows*WindowLength, mean = 0,
     sd = SDNoise)
    }
}


# data processing before applying PCA
dat <- list()
b <- list()
mydata <- list()
meanRemoved <- NULL
for(j in 1:length(L)){
  dat <- L[[j]]
  # compute 4 by 4 correlation matrix for each window
   for(i in 1:nWindows){
     w <- dat[((i-1)*WindowLength+1) : (i*WindowLength), ]
     c <- cor(w)
     mydata[[i]] <- c[upper.tri(c)]
   }
  cs <- sapply(mydata, cbind)
  cs.bar <- rowMeans(cs)
  b[[j]] <- cs-cs.bar
  meanRemoved <- cbind(meanRemoved, b[[j]])
}
```

```r
#preforme PCA
#keep k largest eigenvalues.
 covdat <- as.matrix(meanRemoved) %*% t(as.matrix(meanRemoved))
 eigendat <- eigen(covdat)
 eigenValue <- eigendat$values
 eigenVector <- eigendat$vectors
 order_value <- order(eigenValue, decreasing = T)
 order_value <- order_value[1:k]
 redEigVects <- eigenVector[, order_value]
 WMat <- list()
 for(j in 1:nSubj){
   WMat[[j]] <- t(redEigVects) %*% as.matrix(b[[j]])
 }

 #code to match eigenconnectivities with AC or BC
 #vec <- as.matrix(cbind(AC,BC))
 maxcor = m = vector(mode="list", dim(vec)[2])
 for(col in 1:dim(vec)[2])
 {
 maxcor[[col]] = numeric()
 m[[col]] = numeric()
 for(j in 1:nSubj){
   o <- WMat[[j]]
   abscor = numeric()
   for(i in 1:k){ #first k eigenconnectivities
     corr <- cor(vec[, col], o[i,])
     abscor[i] <- abs(corr)
   }
```

```r
        maxcor[[col]][j] <- max(abscor)
        m[[col]][j] <- which.max(abscor)[1]
    }
}


    #code to match eigenconnectivities with AB or AD/BD/CD
    SDmAB= SDmAD = vector( mode="list", dim(vec2)[2] )
        for( col in 1:dim(vec2)[2] ){
            SDmAB[[col]] = numeric()
            SDmAD[[col]] = numeric()
            for(j in 1:nSubj){
                o <- WMat[[j]]
                SD <- apply(o, 1, sd)
                ordersd <- order(SD, decreasing = F)[1:2]
                mean1 <- rowMeans(o[ordersd,], na.rm=F)
                ordermean <- order(mean1, decreasing = F)
                eigenmAB <- o[ordersd,][ordermean[2],]
                eigenmAD <- o[ordersd,][ordermean[1],]
                SDmAB[[col]][j] <- sd(c(vec2[,col], eigenmAB))
                SDmAD[[col]][j] <- sd(c(vec2[,col], eigenmAD))
            }
        }
    return(list(maxcor, m, SDmAB, SDmAD))
}
```