FUNCTIONAL CONNECTIVITY ANALYSIS OF FMRI TIME-SERIES DATA

by

Dongli Zhou

B.S. in Statistics, Beijing Normal University, China, 2005

M.A. in Statistics, University of Pittsburgh, 2007

Submitted to the Graduate Faculty of

the Arts & Science in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH

ARTS & SCIENCE

This dissertation was presented

by

Dongli Zhou

It was defended on

October 28th 2010

and approved by

Yu Cheng, PhD, Assistant Professor, Department of Statistics, School of Arts and Science, University of Pittsburgh

Wesley Thompson, PhD, Professor, Department of Psychiatry, University of California, San Diego

Leon Gleser, PhD, Professor, Department of Statistics,
School of Arts and Science, University of Pittsburgh

Robert Krafty, PhD, Assistant Professor, Department of Statistics,
School of Arts and Science, University of Pittsburgh

Greg Siegle, PhD, Associate Professor, Department of Psychiatry,
School of Medicine, University of Pittsburgh

Dissertation Director:

Yu Cheng, PhD, Assistant Professor, Department of Statistics, School of Arts and Science, University of Pittsburgh

Yu Cheng, PhD

Wesley Thompson, PhD

FUNCTIONAL CONNECTIVITY ANALYSIS OF FMRI TIME-SERIES DATA

Dongli Zhou, PhD

University of Pittsburgh, 2010

The term "functional connectivity" is used to denote correlations in activation among spatially-distinct brain regions, either in a resting state or when processing external stimuli. Functional connectivity has been extensively evaluated with several functional neuroimaging methods, particularly PET and fMRI. Yet these relationships have been quantified using very different measures and the extent to which they index the same constructs is unclear. We have implemented a variety of these functional connectivity measures in a new freely-available MATLAB toolbox. These measures are categorized into two groups: whole time-series and trial-based approaches. We evaluate these measures via simulations with different patterns of functional connectivity and provide recommendations for their use. We also apply these measures to a previously published fMRI dataset in which activity in dorsal anterior cingulate cortex (dACC) and dorsolateral prefrontal cortex (DLPFC) was evaluated in 32 healthy subjects during a digit sorting task. Though all implemented measures demonstrate functional connectivity between dACC and DLPFC activity during event-related tasks, different participants appeared to display qualitatively different relationships.

We also propose a new methodology for exploring functional connectivity in slow event-related designs, where stimuli are presented at a sufficient separation to examine the dynamic responses in brain regions. Our methodology simultaneously determines the level of smoothing to obtain the underlying noise-free BOLD response and the functional connectivity among several regions. Smoothing is accomplished through an empirical basis via functional principal components analysis. The coefficients of the basis are assumed to be correlated across regions, and the nature and strength of functional connectivity is derived from this correla-

tion matrix. The model is implemented within a Bayesian framework by specifying priors on the parameters and using a Markov Chain Monte Carlo (MCMC) Gibbs sampling algorithm. We demonstrate this new approach on a sample of clinically depressed subjects and healthy controls in examining relationships among three brain regions implicated in depression and emotion during emotional information processing. The results show that depressed subjects display decreased coupling between left amygdala and DLPFC compared to healthy subjects and this may potentially be due to inefficient functioning in mediating connectivity from the rostral portion Brodmanns area24 (BA24).

Keywords Functional magnetic resonance imaging; Event-related design; Functional connectivity; Functional data analysis; B-spline; Mixed-effects model; Principal components; Reduced rank model; Bayesian method; Markov chain monte carlo.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

First of all, I would like to express my gratitude to my advisor Dr. Wesley Thompson and co-advisor Dr. Yu Cheng, for their guidance, encouragement and continuous support throughout the years of my PhD program. I would also like to thank Dr. Leon Gleser, Dr. Robert Krafty and Dr. Greg Siegle for their questions, comments, suggestions and time spent discussing my work. Special thanks to Dr. Wesley Thompson and Dr. Greg Siegle for providing me an intriguing topic as the starting point of my research and their advice helped me get through many difficulties of my research work.

I have enjoyed my time in Pittsburgh, and must thank for all the professors who helped me to build a solid foundation for my current and future career through their classes. I also want to thank my classmates, friends and colleagues for the good time we have had, and the support they could provide. My family, especially my husband Xing Yuan, my parents, and my grandparents, deserves special thanks for their love and unfailing support, which has enabled me to bring this work to fruition. I will be forever grateful.

1.0 MATLAB TOOLBOX FOR FUNCTIONAL CONNECTIVITY

1.1 INTRODUCTION

The brain forms a distributed network, whereby specialized regions communicate with each other to process information [15, 72]. The attempt to identify and quantify such inter-regional relationships has been termed "connectivity" analysis [24]. Functional connectivity (FC), in particular, is defined as the statistical association or dependency among two or more anatomically distinct time-series [26, 37, 63]. Measures of FC are agnostic regarding causality or direction of connections.

FC analyses were first performed on human brain functional data using positron emission tomography (PET) [14, 38, 27], and have since expanded into functional Magnetic Resonance Imaging (fMRI), Magnetoencephalography (MEG), Electroencephalography (EEG), and peripheral physiological measures [24, 70, 63, 18, 11, 41, 54, 32]. FC has also been assessed with a variety of different experimental designs, including block designs [36, 45, 3] and event-related designs [59, 22, 68, 2]. Block designs alternate periods of stimulus types, with each period presenting a given stimulus type multiple times, whereas event-related designs present stimuli individually separated by an inter-stimulus interval (ISI), termed a "trial". More recently, resting state studies, particularly involving the "default mode" network [7, 50, 35, 23], have also become popular for determining connectivity. Methods for computing FC from resting state data usually assume that the time-series are stationary (i.e., probabilistically unchanging across time), and utilize information from the entire time-series of fMRI scans ("whole time-series" approaches). Conversely, methods for event-related designs do not require stationarity, and FC is often computed based on associations obtained by examining data divided into individual trials ("trial-based" approaches). Block designs may be consid-

ered locally stationary and hence are intermediate between resting state and event-related designs.

The MATLAB toolbox described in this paper includes measures of FC from both whole time-series and trial-based approaches, including zero-order and cross-correlation [7, 4, 68], cross-coherence [70, 63, 53], mutual information [41, 54, 63], peak correlation [59], and functional canonical correlation [75, 68]. We also implement optional temporal smoothing steps in the toolbox. Many of these measures have not been previously available in a user-friendly package aimed at neuroscientists and this toolbox provides a quick and easy means to compute and compare results from different FC measures.

The implemented techniques specifically attempt to characterize the relationships between time-series extracted from two or more regions above and beyond zero-order correlational relationships, potentially controlling for one or more other time-series. The toolbox does not include other types of measures of FC. Following Cattel (1952),[10] data can be clustered along dimensions of occasions (time), variables (for fMRI, space or region), and person. We have concentrated on characterizing relationships between occasions across regions; e.g., lagged cross-correlation analysis characterizes relationships of a time-series at one lag to another time-series at other lags. FC has been alternately defined in the literature based on other ways of clustering fMRI data. For example, techniques such as principal components analysis (PCA) or independent components analysis (ICA) generally attempt to cluster voxels or regions (i.e., variables). ICA, in particular, extracts latent time-series which characterize the behavior of sets of voxels (e.g., Formisano et al. 2004 [21]). The degenerate case of spatial PCA with just two regions is the zero-order correlation between the regions. Thus, these techniques are more appropriate when large numbers of voxels or regions are examined. Other techniques such as the examination of psychophysiological interactions [25] do not control for third-variable time-series, but rather, examine interactions with them. Such alternate techniques generally focus on zero-order relationships between time-series, but could be generalized to account for the types of relationships examined in this paper.

The rest of this chapter is organized as follows. A detailed summary of statistical methodologies commonly used for FC is introduced in Section 1.2 along with a description of the

Matlab toolbox. In Section 1.3, we evaluate the performances of different FC measures via simulations of different patterns of relationships between regional time-series. In this section, we also analyze data from a small application to an empirical neuroimaging dataset. In Section 1.4, we present a short discussion and suggest areas of further development.

1.2 MATERIALS AND METHODS

1.2.1 Overview of How to Use the Toolbox

The Functional Connectivity Toolbox is developed in MATLAB (Mathworks, Inc.) as an open source package. It is designed to use existing routines in the MATLAB distribution with an additional freely-available toolbox for functional data analysis (Ramsay 2005; http://www.psych.mcgill.ca/misc/fda/index.html). Each FC measure listed above is coded into a function in MATLAB. The inputs of the functions are equal length time-series data from brain regions' responses within a single subject. For slow event-related designs, the user must also indicate the number of scans per trial. For example, if $Y_1(t)$ and $Y_2(t)$ are one subject's fMRI responses from two brain regions for an slow event-related design with T scans per trial, then the peak correlation between Y_1 and Y_2 can be obtained by entering $corrpeak(Y_1, Y_2, T)$, where *corrpeak* is the function's name, as defined in MATLAB. For fast event-related designs, stimulus time-series should be indicated instead of scans per trial. Calling features for each of the functions in the toolbox, along with its arguments are described in the Appendix A.

1.2.2 Smoothing

We also include an optimized smoothing step as a noise-reduction procedure because not all noise can be removed or cleaned [47]. Temporal smoothing is particularly important in connectivity analyses because reliable detection of FC between brain responses can require high signal-to-noise ratio (SNR) [5]. FC estimates are therefore strongly dependent on the level of temporal smoothing; too much smoothing yields over-estimates of relation-

ships between time-series whereas too little smoothing underestimates these relationships. We have implemented an optional Bayesian temporal smoothing technique using a Markov Chain Monte Carlo (MCMC) Gibbs sampling algorithm with roughness penalty parameters treated as components of variance and estimated from the data. This technique utilizes a cubic B-spline basis expansion with equally spaced knots, and includes the ability to handle white or lag-1 autocorrelated noise. An optional pre-whitening step is also available with lag-1 autocorrelated noise. The level of autocorrelation has to be predetermined; the default value is set to 0.7. We chose a cubic B-spline basis because it produces smooth yet flexible fits and for efficient computation of the roughness penalty parameter. A simple default choice of the number of knots for the basis is $min(\frac{1}{4} \times \text{length of data}, 35)$ [62].

1.2.3 Whole Time-Series Approaches

Whole time-series approaches aim to examine the relationships contained within the entire time-series of fMRI scans, based on the assumption that the time-series is stationary. Stationarity of the time-series guarantees that the relationships among them are probabilistically consistent over time [66].

1.2.3.1 Cross-Correlation Zero-order correlation measures the simultaneous linear coupling relationship between two time-series. When the time courses are highly positively correlated, this implies that the two regions are on average more or less active at the same times. Conversely, a high negative correlation implies that when one region is more active the other is less active. Zero-order correlation has been used often to measure inter-regional relationships in fMRI, e.g., in Biswal et al. (1995)[7].

Lagged cross-correlation can also be used to evaluate inter-regional relationships [68]. Lagged cross-correlations capture the lagged or delayed linear relationships between regions. Cross-correlation between brain regions A and B at positive lags indicate a relationship between activity of region A and subsequent activity of region B, or vice versa. One study that used lagged cross-correlations [68] found that depressed people had attenuated correlation of dorsolateral prefrontal cortex activity with amygdala activity 3 to 6 sec later.

Cross-correlation of any two individual time-series (i, j), at lag h, $\rho_{ij}(h)$, is given by

$$\rho_{ij}(h) = \frac{cov_{ij}(t, t+h)}{\sqrt{var_i(t)var_j(t+h)}}, \tag{1.1}$$

where $\rho_{ij}(h) = \rho_{ji}(-h)$, and is restricted to the $[-1, 1]$ interval. $h = 0$ corresponds to zero-order correlation.

Correlations are often subtended by low-frequency (less than 0.1 Hz) components of the data, as has been shown in several studies [7, 50, 16]. Biswal et al. (1995)[7], for example, reported that low-frequency (below 0.08 Hz) correlations existed between the bilateral primary motor cortices (M1) and the supplementary motor area (SMA) during resting state scans. We include the ability to apply a low-frequency filter as an option parameter to the cross-correlation routine. By default this filter is set to 0.1 Hz.

1.2.3.2 Cross-Coherence While correlation is defined in the time domain, coherence measures are in the frequency domain. Coherence has been repeatedly shown to be a useful statistic for investigating FC across brain regions [48, 70, 63]. Coherence measures implicitly account for lags in the effects of one region on another. If a time-series in one region is broadly similar to that in another, but with a time delay, then the ordinary zero-order correlation between the two will be moderate or low; the coherence, by contrast, will be high within the bandwidth of the curve. The squared coefficient of coherence can be interpreted as the proportion of the power in one of the two time series (at a selected frequency), which can be explained by its linear regression to the other time course.

The concept of coherence of time-series was introduced by Wiener (1949)[74] and extensively developed and described by Rosenberg et al. (1989)[61] for its applicability to functional imaging data. Spectral coherence for determination of FC was applied to motor experiments by Sun et al. (2004)[70].

Coherence $Coh_{ij}(\lambda)$ between any two individual time-series (i, j) at frequency λ is defined as

$$Coh_{ij}(\lambda) = |R_{ij}(\lambda)|^2 = \frac{|f_{ij}(\lambda)|^2}{f_{ii}(\lambda)f_{jj}(\lambda)}, \tag{1.2}$$

where $R_{ij}(\lambda)$ is the complex valued coherency of Y_i and Y_j; $f_{ij}(\lambda)$ is the cross-spectral density between Y_i and Y_j; and $f_{ii}(\lambda)$ and $f_{jj}(\lambda)$ are the spectral densities of Y_i and Y_j. Coherence

is a positive function, it is symmetric in i and j(e.g., $Coh_{ij}(\lambda) = Coh_{ji}(-\lambda)$), and bounded by 0 and 1.

1.2.3.3 Mutual Information Theoretically, correlation and coherence measure the linear dependence between two time-series, whereas mutual information is a statistical measure of both linear and nonlinear dependence [64]. Mutual information quantifies the shared information between two time-series. For example, if the two time-series are independent, there is no shared information and hence the mutual information is zero. At the other extreme, if one time-series is a deterministic one-to-one function of the other, then they share the same information: in this case their mutual information is infinite.

Jeong et al. (2001)[41] used an entropy-based measure of mutual information to investigate FC among time-series from different cortical areas in both Alzheimer's disease patients and healthy controls. Salvador et al. (2005) employed mutual information based on coherence and showed that FC lay mainly in low frequencies. Chen et al. (2008)[11] developed a conjoined time-frequency analytical-based method of mutual information to explore brain neural connectivity by MEG during a self-paced finger lifting task. In this toolbox, we implement mutual information based on coherence in the frequency domain [63], which is defined as

$$\delta_{ij} = -\frac{1}{2\pi}\int_{\lambda 1}^{\lambda 2} \log(1 - Coh_{ij}(\lambda))\, d\lambda, \tag{1.3}$$

where $[\lambda 1, \lambda 2]$ specify the frequency band within which to integrate the infomation and $-\pi \leq \lambda 1 < \lambda 2 \leq \pi$. This formula assumes time-series are Gaussian.

This integral is unbounded, ranging from 0 to infinity. A simple transformation can be applied, however, to obtain a normalized mutual information [42, 34, 63], with scores in the interval $[0, 1]$. This is implemented in the toolbox as

$$\phi_{ij} = [1 - \exp(-2\delta_{ij})]^{\frac{1}{2}}. \tag{1.4}$$

1.2.3.4 Partial Cross-Correlation/Coherence/Mutual Information With analyses including more than two brain regions, one question is whether an observed dependence between any two regional time-series is attributable to a direct connection between the two

brain regions or to an indirect relationship involving other regions. This question may be addressed by measuring the association between the two regions (i, j) after accounting for the relationship of each to other reference time-series $1, \ldots, P \backslash i, j$ (time-series from region 1 to P except time-series from region i and j). This is called conditional dependence. A discussion of how to apply bivariate cross-correlation/coherence/mutual information analysis to multivariate time-series was provided by Salvador et al. (2005)[63]. Our toolbox includes similar routines for computing partial cross-correlation/coherence/mutual information. These measures yield the partial relationship of each pair of time-series accounting for the remaining time-series in a matrix.

1.2.4 Trial-Based Approaches

Trial-based approaches evaluate trial-to-trial relationships and are usually applied to event-related experimental designs. Time-series of brain regions in trial-based approaches does not have to meet the assumption of stationarity.

1.2.4.1 Peak Correlation Peak correlation captures the coupling relationship of peaks in activitation in pairs of brain regions associated with discrete events (trials). We implement this by first creating functional versions of trial-related time-series by temporally smoothing the trial time courses: each trial yields one curve per region. Separate peak estimates are computed from the functional responses for each individual trial and for each region and consequently used as the data in a correlation analysis. A high positive value of peak correlation between two regions indicates that two regions are more or less active than average during the same trials. Similarly, a high negative value implies that when one region has a higher-than-average peak, the other region has a lower-than-average peak.

Similarly, instead of trial peaks, other trial trajectory characteristics can also be estimated for constructing trail-by-trail relationship between brain regions. Rissman, et al. (2004) [59] derived a new method which is capable of characterizing stage-specific functional interactions. They constructed a standard general linear model (GLM) in which every stage (cue, delay, and probe) of every trial is modeled with a separate covariate, so that trial-

to-trial separate parameter estimates relating to cue, delay, and probe are computed. The parameter estimates are sorted according to the stage from which they were derived and then use those as the dependent data in a correlation analysis to obtain FC for different stages of the task respectively.

1.2.4.2 Functional Canonical Correlation Functional canonical correlation seeks to investigate which modes of variation between pairs of observed random curves are most associated with one another. Functional versions of trial-related time-series are created by smoothing the trials and the canonical correlation between these functional responses is computed. The functional canonical correlation [56] between any pair of individual time-series (i, j) is given by

$$ccorsq_{ij}(\xi, \eta) = \frac{cov(\int \xi(t)Y_{im}(t)dt\,, \int \eta(t)Y_{jm}(t)dt\,)^2}{(var\int \xi(t)Y_{im}(t)dt\, + \lambda\|D^2(\boldsymbol{\xi})\|^2)(var\int \eta(t)Y_{jm}(t)dt\, + \lambda\|D^2(\boldsymbol{\eta})\|^2)}. \quad (1.5)$$

Here, $\xi(t)$ is the weight function for Y_i and Y_{im} refers to the mth trial of region i; similarly, $\eta(t)$ is the weight function for Y_j and Y_{jm} refers to mth trial of region j. λ is a smoothing parameter, chosen via cross-validation, that describes the smoothness constraint on the weight functions. $\|D^2(\boldsymbol{\xi})\|^2$ and $\|D^2(\boldsymbol{\eta})\|^2$ represent the roughness of the weight functions, where $D^2(\cdot)$ is the second derivative operator.

Qualitative relationships between the two regions can be explored by examining the weight functions. For example, weight functions may indicate that sustained activity on one region is related to peak activity in another region. Siegle et al. (2007)[68] used functional canonical correlation analysis to measure FC between amygdala and DLPFC responses to negative words.

1.2.4.3 Use of Whole Time-Series Approaches with Nonstationary Designs The extent to which whole time-series approaches are applicable to nonstationary (e.g., trial-based) designs is unclear. The whole time-series approaches generally assume stationarity, and thus in the case of a recurring effect of a stimulus on two regions, may overestimate true connectivity. This over-estimation might be attenuated by examining connectivity at

frequency bands outside the trial range (e.g., using a low-pass filter). Alternately, if effects of the stimulus are assumed to be constant across trials, residual relationships between regions might be inferred to reflect connectivity. In this case, stimulus effects can be explicitly accounted for before computing connectivity, e.g., via residual regressions or partial correlation analysis, entering a stimulus-related response as a covariate. With a fast event-related or jittered design with catch trials, the covariate waveform could be constructed by deriving an impulse response function (e.g., via deconvolution) which could be convolved with the design. In the case of a fast event-related design in which catch trials are not presented, it would be possible to covary a series of canonical responses convolved with the design from the waveforms in both regions. With slow event-related designs, if responses within a region are assumed to have a canonical shape, the covariate waveform could be constructed by convolving a canonical response with a delta function at the stimulus frequency. But systematic deviations from the canonical waveform in either structure might then create a spurious pattern of connectivity. Rather, using the mean responses in each candidate waveform as covariates minimizes these effects by assuming only that trial-to-trial deviations from the mean response in both regions reflect the effects of connectivity. We have adopted this last approach in the simulations described below.

1.3 RESULTS

1.3.1 Simulation 1

In Simulation 1, we performed two Monte Carlo studies to illustrate how the implemented FC measures detect different patterns of association between two regions. The first study simulated stationary (resting sate) fMRI time-series data for use with whole time-series FC measures. Since any stationary time-series can be represented as the random superposition of sines and cosines oscillating at various frequencies [66], each regional time-series was generated by a linear combination of sine waveforms of different frequencies and phases; the weights for each component were generated from a normal distribution. The regional

time-series were then convolved with an hemodynamic response function (double gamma) to produce a temporally smoothed BOLD time-series. Connectivity between regional time-series was achieved by linking the weights of sine waveforms at certain frequencies and phases. Simulated data was generated as single run for each of fifty subjects, consisting of 140 scans (TR=2s). Gaussian white noise was added to each run. Functional SNR (the ratio between the intensity of a signal associated with changes in brain function and the variability in the data due to all sources of noise [39, 66]) was set to approximately 3 here and in the second study. Simulation and FC results for the first study are given in Figure 1.1.

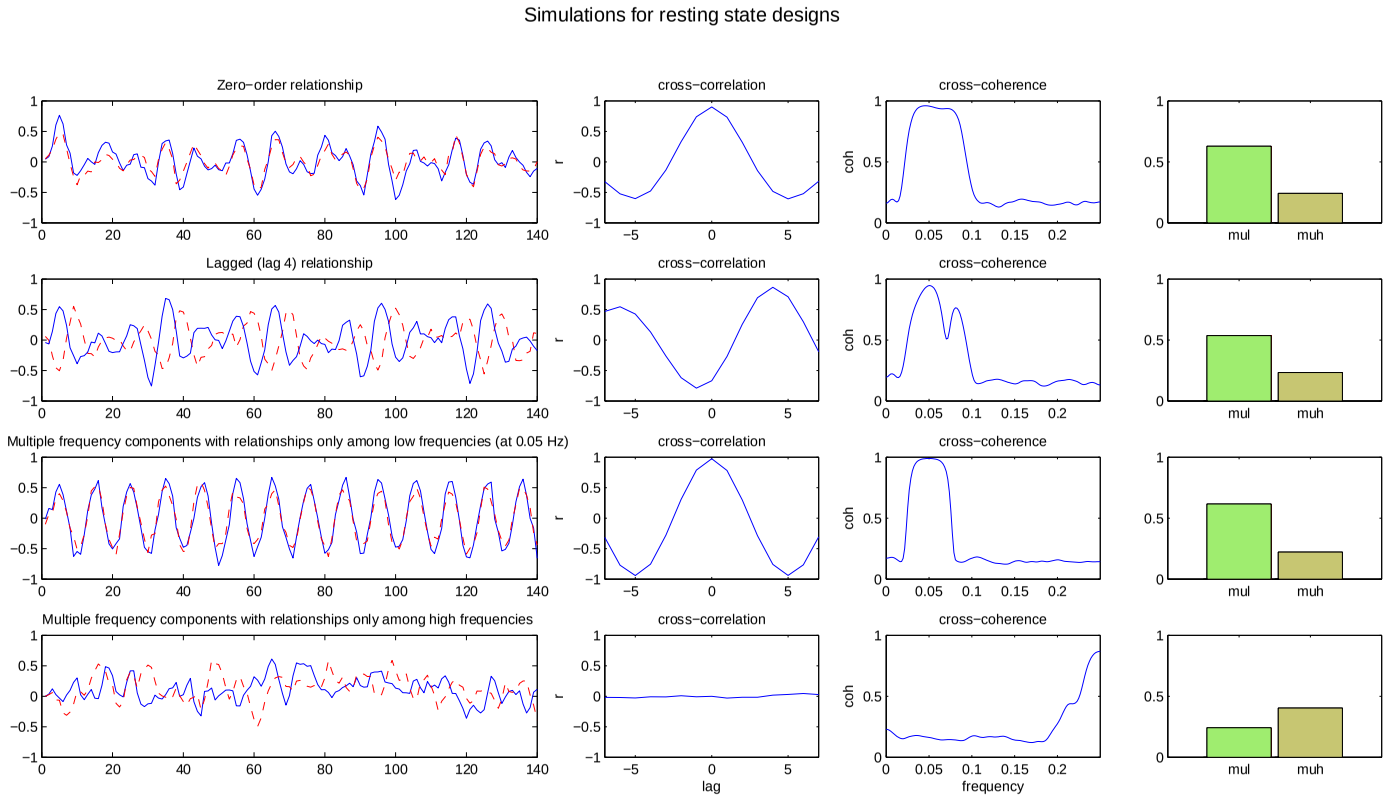
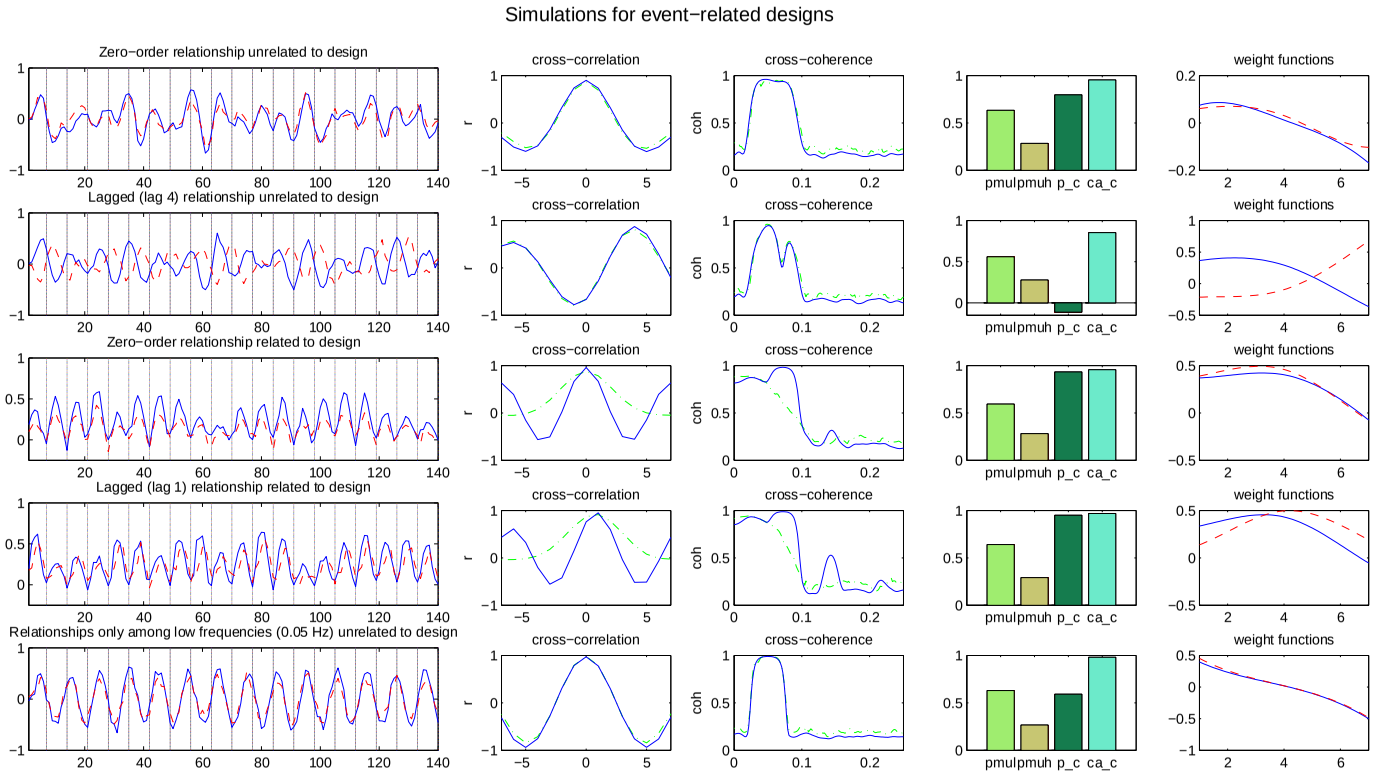


Figure 1.1: Simulations for resting state designs

The plots from top to bottom on the far left show one subject's whole-waveform-based simulated time-series (140 scans) of two regions (blue lines and red lines) for 4 different patterns of association. Results of cross-correlation and cross-coherence shown here are the overall mean across 50 subjects. mu_l, mu_h represent mean mutual information over low frequency band and high frequency band respectively.

The second study simulated nonstationary (event-related) fMRI brain responses. The BOLD activation curve for each trial was generated by a linear combination of B-spline basis functions. Weights for the B-splines were generated from multivariate normal distributions with predefined mean and variance. An association between two regions was created by correlating the weights of B-spline basis functions. Simulations were generated with 20 trials and 7 scans per trial (TR=$2s$) for each of fifty subjects with additive white Gaussian noise. Simulation and FC results for the second study are given in Figure 1.2.

In the second study, we also considered partial correlation/coherence/mutual information by including reference waveforms to control for co-activation of regions from application of the stimulus. Correlation and coherence measures can be dominated by the stimulus-locked response in event-related designs. When a stimulus is presented, the stimulus-locked neural response may cause an increase in the BOLD signal in both regions simultaneously (co-activation). This is not necessarily due to an intrinsic task-induced functional coupling, but may be due to the response in both regions to the externally-applied stimuli. For example, in the simulations of event-related design data shown in Figure 1.2, coherence is particularly high at the trial frequency even when there is no inherent connectivity. By including stimulus reference waveforms, partialling methods estimate any remaining relationship between two time-series after taking co-activation into account. For these simululations we chose two stimulus reference waveforms generated by repeating the mean trial-averaged responses across all trials for each region. Partial results are shown in Figure 1.2 as dashed red lines.



continued on next page

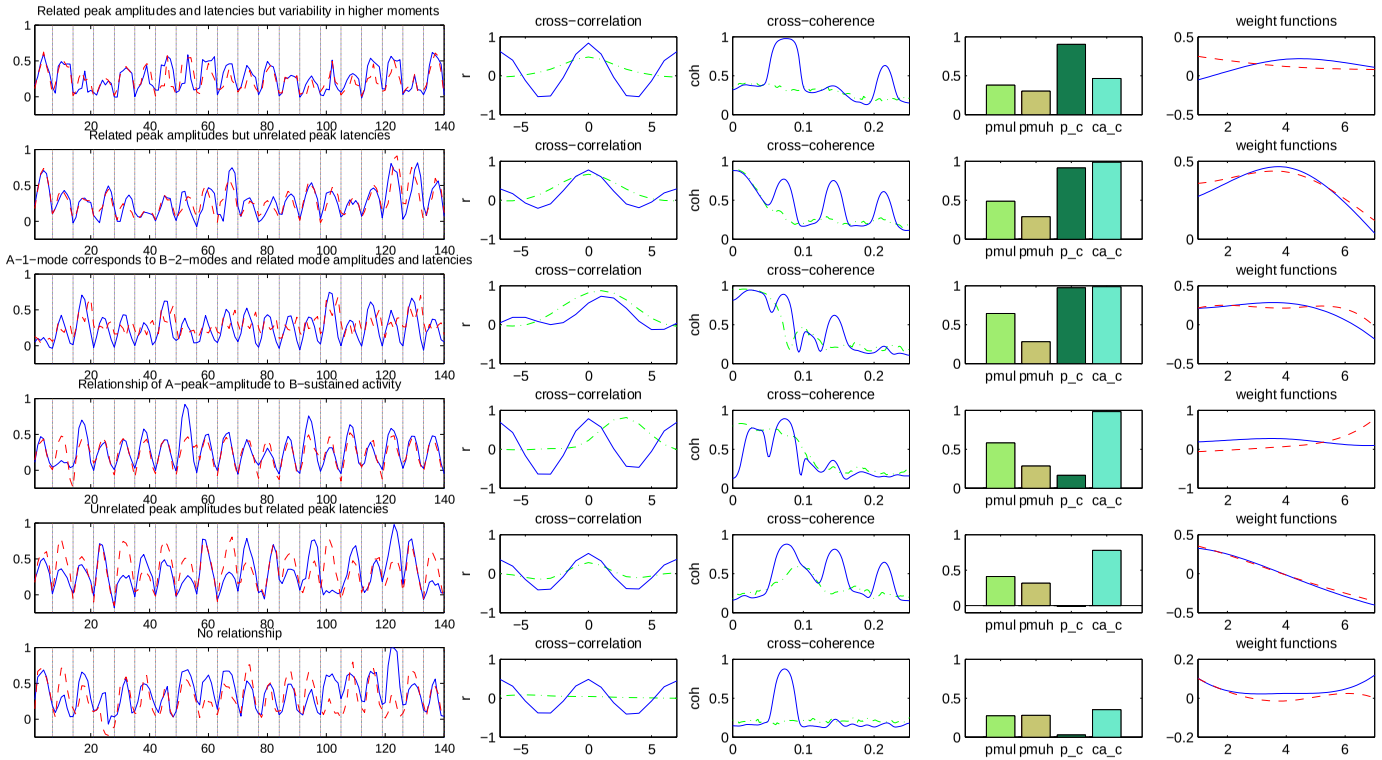


Figure 1.2: Simulations for event-related designs

The plots from top to bottom on the far left show one subject's trial-based simulated time-series(20×7 scans) of two regions (blue line and red line) for 11 different patterns of association. Results of FC measures shown here are overall mean across 50 subjects. mu_l, mu_h, p_c and ca_c represents mutual information over low frequency band, mutual information over high frequency band, peak correlation, and functional canonical correlation respectively. Weight function of each region(blue line and red line) for canonical correlation are the overall mean of weight functions across subjects. Cross-correlation, coherence and mutual information are taken after centering the time-series data from the overall median of each run(20 trials).

Tables 1.1 and 1.2 summarize the computed FC measures for different patterns of association between two regions, based on the simulation results. Statements within brackets are partial results accounting for stimulus waveforms. For simulations of non-design related connectivity, including the mean stimulus waveforms as covariates does not change the value of correlation and coherence (e.g., zero-order (lag 1) relationship unrelated to the design; relationships only among low frequencies (0.05 Hz) unrelated to design). For all types of relationships related to the design, coherence at the trial frequency is attenuated after including stimulus waveforms. In contrast, cross-correlation is still high at lag 0 (or 1) for relationships of design-induced peak amplitude between two regions (e.g., zero-order (lag 1) relationship related to the design; related peak amplitudes but unrelated latencies; and A-1-mode and B-2-modes) because the two regions become more or less active than the average response during the same trials, a relationship remains after accounting for stimulus waveforms reflecting true connectivity. Cross-correlation attenuates a bit for data with related peak amplitudes and latencies but variability in higher moments, because for some trials region 1 may activate before region 2 and for other trials the reverse may happen. The relationship of A-peak amplitude to B-sustained activity is phase-lagged after accounting for the stimulus waveform, since when region 1 has a higher than average peak, region 2 has higher than average sustained later activity. Partial cross-correlation and partial cross-coherence are very low for data generated with unrelated peak amplitudes but related peak latencies and essentially zero for data generated with stimulus co-activation but no FC relationship. Note, however, that canonical correlation weighting functions for data generated with unrelated peak amplitudes but related peak latencies do indicate the nature of the FC relationship. The two weight functions for this relationship (displayed in Figure 1.2) are identical and put most weight in the beginning and end of the trials, indicating that the primary mode of covariation lies in shifting the timing of the peaks.

Table 1.1: Interpreting data based on simulation results (whole waveform)

Zero-order corr.	Lagged corr.	Cross coh.	Mutual info.(L)	Mutual info.(H)	Interpretation
High	Medium or low	High at low freq.	Medium	Low	Zero-order relationship
Medium or low	High	High at low freq.	Medium	Low	Lagged (lag 4) relationship
Could be high	Could be high	High at low freq.	Medium	Low	Multiple frequency components with relationships only among low frequencies
Low	Low or high	High at high freq.	Low	Medium	Multiple frequency components with relationships only among high frequencies

Table 1.2: Interpreting data based on simulation results (trial-based waveform)

Zero-order corr.	Lagged corr.	Cross coh.	Mutual info. (L)	Mutual info. (H)	Peak corr.	Canonical corr.	Interpretation
High (high)	Medium or low (medium or low)	High at low freq. (high at low freq.)	Medium (medium)	Low (low)	Medium	High	Zero-order relationship unrelated to design
Medium or low (medium or low)	High (high)	High at low freq. (high at low freq.)	Medium (medium)	Low (low)	Low	High	Lagged relationship unrelated to design
High (high)	Medium or low (medium or low)	High at trial freq. (high at low freq.)	Medium (medium)	Low (low)	High	High	Zero-order relationship related to design
Medium or low (medium or low)	High (high)	High at trial freq. (high at low freq.)	Medium (medium)	Low (low)	High	High	Lagged relationship related to design
Could be high (could be high)	Could be high (could be high)	High at 0.05 Hz (high at 0.05 Hz)	Medium (medium)	Low (low)	Medium or low	High	Multiple frequency components with relationships only among low frequencies (0.05Hz) unrelated to design
High	Medium	High at	Medium	Medium	High	Medium	Related peak

Continued on next page

Table 1.2 – continued from previous page

Zero-order corr.	Lagged corr.	Cross coh.	Mutual info. (L)	Mutual info. (H)	Peak corr.	Canonical corr.	Interpretation
or medium (medium)	or low (low)	trial freq. (low)	(low)	or low (low)			amplitudes and latencies but variability in higher moments (e.g.,skew, kurtosis)
Medium (medium)	Low (low)	Medium at trial freq. (low)	Medium (medium)	Medium or low (low)	High	High	Related peak amplitudes but unrelated peak latencies
Medium (high or medium)	Medium or low (high at certain lag)	High at trial freq. (high at low freq.)	Medium (medium)	Low (low)	High	High	A-1-mode corresponds to B-2-mode and related mode amplitudes and latencies
Medium (medium or low)	Low (high at certain lag)	High at trial freq. (high at low freq.)	Medium (medium)	Low (low)	Low	High	Relationship of A-peak -amplitude to B-sustained activity
Medium (low)	Low (low)	High at trial freq. (low)	Medium (low)	Low (low)	Low	High	Unrelated peak amplitudes but related peak latencies
Low (low)	Low (low)	High at trial freq. (low)	Low (low)	Low (low)	Low	Low	No connectivity

1.3.2 Tree-Building

Tables 1.1 and 1.2 show that the combination of all FC measures did not exhibit the same pattern for any two simulated connectivity relationships between the two regional time-series. Thus, it appears that quantifying more than one FC measure may be useful in understanding the precise nature of connectivity between two fMRI time-series. To explore this idea further, we determined a set of if-then logical (split) conditions permitted accurate predictions of

association type between regions (from the 11 patterns listed in Table 1.2) from a set of FC measure predictors (partial correlation/coherence/mutual information are implemented instead of correlation/coherence/mutual information). We conducted a classification analysis via a tree-building algorithm [8] using the values of the FC measures to move through the tree (until reaching a terminal node) to predict the category (1-11) shown for that node. This classification tree is shown in Figure 1.3. The tree was built based on 80% of the simulation data and the pruning parameter was chosen by 10-fold cross-validation [73]. The proportion of correctly classified simulations for the other 20% was 92.7%. The inclusion of multiple measures in the tree suggested that multiple FC measures are useful in determining the true nature of a functional relationship between two regions.

1.3.3 Simulation 2

In Simulation 2, we performed two Monte Carlo studies to illustrate the effect on computed FC relationships when different types of noise are added to related BOLD signals. BOLD signals were generated as in Simulation 1. In study 1, BOLD signals from two regions were stationary with a zero-order relationship. In study 2, BOLD signals were event-related with zero-order relationship. Lag-1 autocorrelated noise with low SNR (functional SNR around 1.5) and lag-1 autocorrelated noise with high SNR (functional SNR around 2.5) were added to the coupled BOLD responses. In each of the studies, FC measures were computed after smoothing with and without explicitly modeling autocorrelated noise. FC results from the two studies are shown in Figure 1.4 and 1.5 respectively. These results show that as SNR decreases, FC measures tend to become attenuated. However, even under a high level of noise inter-regional relationsips were detectable for some FC methods; e.g., functional canonical correlation is still high (around 0.8) when applied to smoothed fMRI responses with low SNR (see second row of Figure 1.5).

In addition, the results illustrate that smoothing fMRI responses with a method that explicitly models autocorrelated noise can be important for detecting FC when noise is in fact autocorrelated, especially when the SNR is low. For example, in Figure 1.5, partial cross-coherence between the two regions is quite small (around 0.2) at low frequencies when

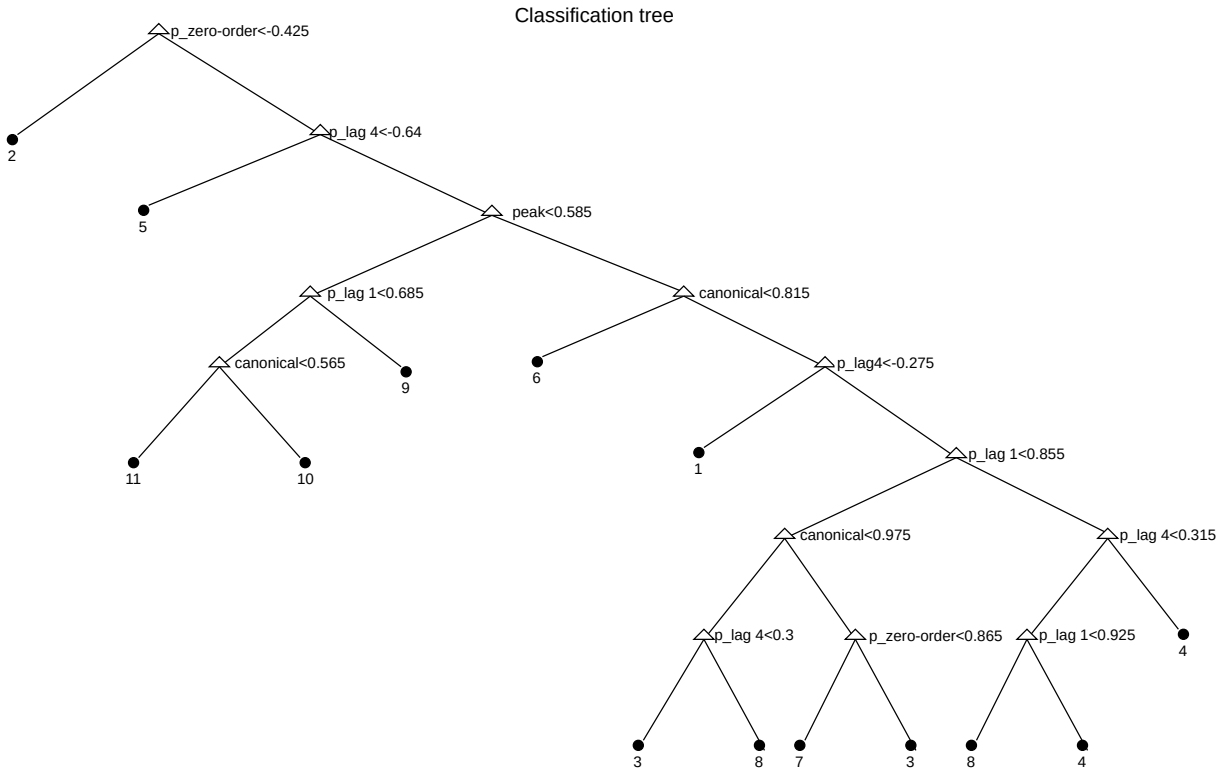


Figure 1.3: Multiple trials for one subject

Classification tree of 11 pattern of associations. Here, 1-11 represent: 1.Zero-order relationship unrelated to design; 2.Lagged (lag 4) relationship unrelated to design; 3.Zero-order relationship related to design; 4.Lagged (lag 1) relationship related to design; 5.Relationships among low frequencies unrelated to design; 6.Related peak amplitudes and latencies but variability in higher moments; 7.Related peak amplitudes but unrelated peak latencies; 8.A-1-mode B-2-modes; 9.Relationship of A-peak-amplitude to B-sustained activity; 10.Unrelated peak amplitudes but related peak latencies; and 11.No connectivity. Covariates put in this classification tree were partial zero-order correlation('p_zero-order'), partial lag-1 correlation('p_lag 1'), partial lag-4 correlation('p_lag 4'), partial mutual information over low frequency band('pmu_l'), partial mutual information over high frequency band('pmu_h'), peak correlation('peak'), and functional canonical correlation('canonical').

the SNR is low and smoothed without modeling autocorrelated noise, but increases to around 0.7 when the data are smoothed with a method that models autocorrelated noise. Smoothing data after pre-whitening (not shown here) resulted in similar improved estimates of FC in the presence of autocorrelated noise. In contrast, we found little difference in the FC measures for the stationary design (Figure 1.4).

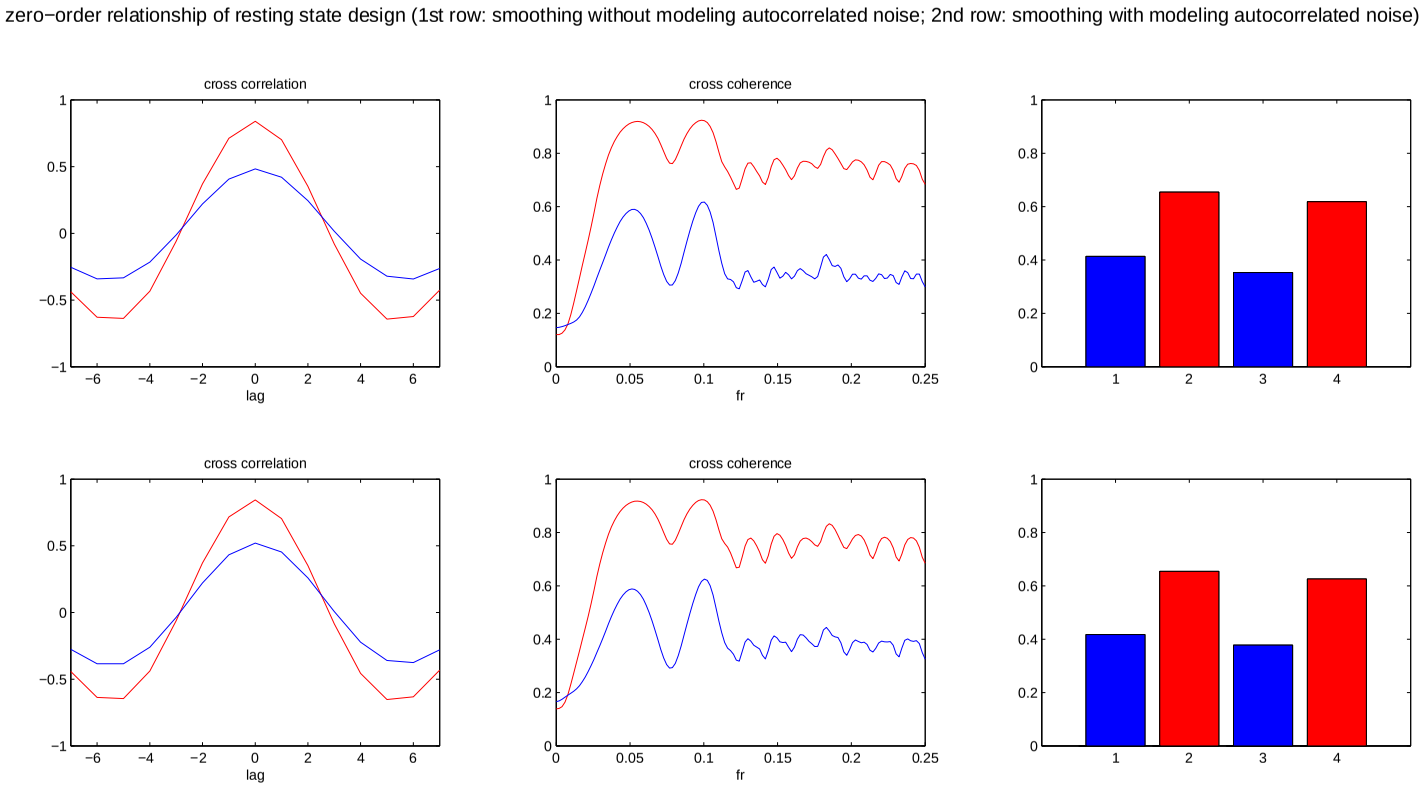


Figure 1.4: Mean results of FC on Simulations with zero-order relationship (stationary design)

Mean results of different FC measures applied to simulations with zero-order relationship (stationary design) between regions. FC measures were computed after smoothing with and without explicitly modeling autocorrelated noise. Results of data with high SNR are shown in red and results of data with low SNR are shown in blue. 1st row: smoothing without modeling autocorrelated noise; 2nd row: smoothing with modeling autocorrelated noise.

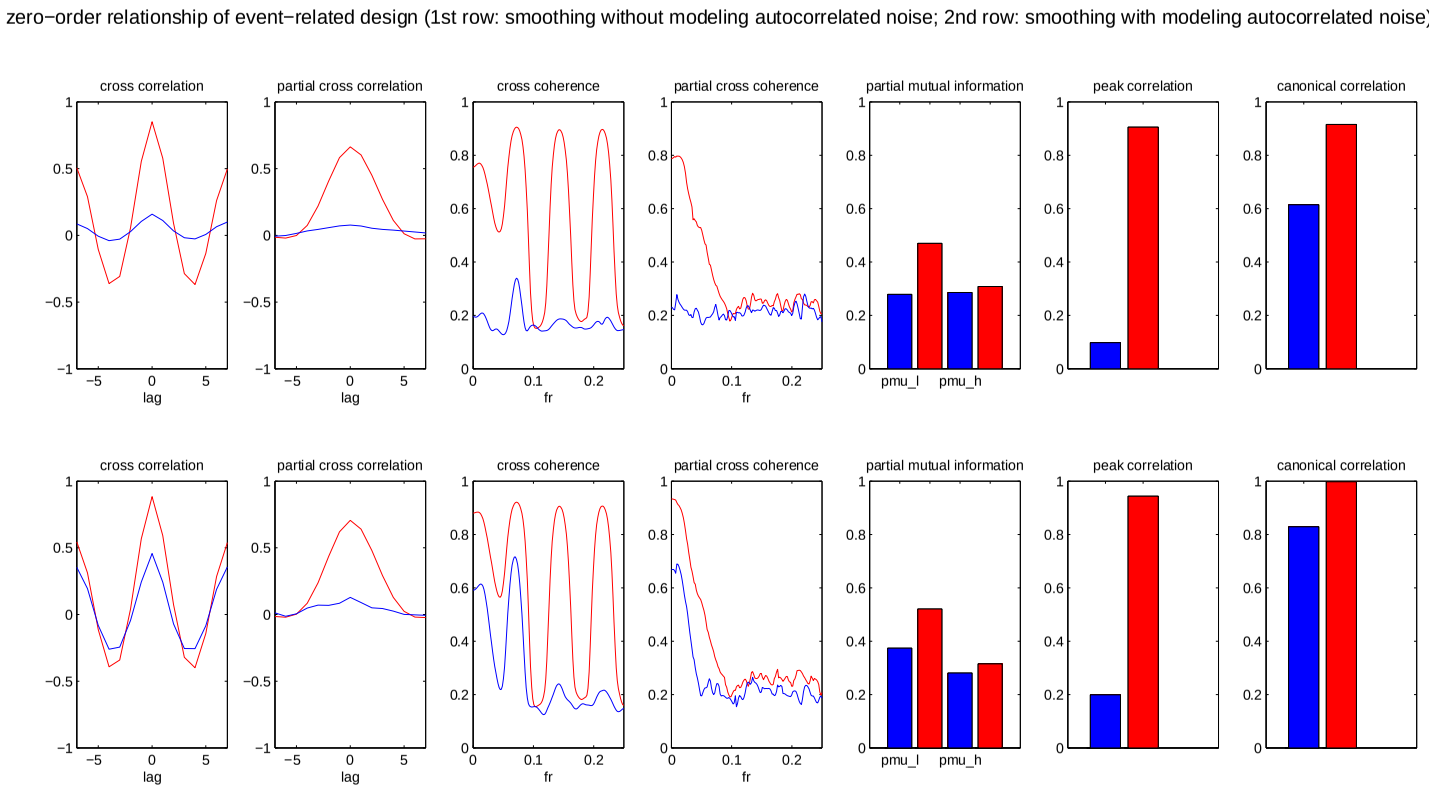


Figure 1.5: Mean results of FC on Simulations with zero-order relationship (event-related design)

Mean results of different FC measures applied to simulations with zero-order relationship (event-related design) between regions. FC measures were computed after smoothing with and without explicitly modeling autocorrelated noise. Results of data with high SNR are shown in red and results of data with low SNR are shown in blue. 1st row: smoothing without modeling autocorrelated noise; 2nd row: smoothing with modeling autocorrelated noise.

1.3.4 Application to An Empirical Neuroimaging Dataset

This analysis involves data from 32 healthy adult subjects, acquired as part of a larger study. The goal was to determine functional relationships between the dorsal anterior cingulate cortex (dACC) and dorsolateral prefrontal cortex (DLPFC) during executive control. Theoretical models [15] and initial neuroimaging analyses [51, 2] suggest that bidirectional relationships should be apparent - that is, the DLPFC is involved in executive control necessary to prime the dACC, and dACC activity should spur future DLPFC activity. But these relationships have not been tested using multiple measures of connectivity, so the true nature of relationships has not been determined. In addition, no study of relationships between these regions has considered the idea that different healthy subjects may be characterized by qualitatively different functional relationships among these regions. In 36 slow event-related trials (one subject with 33), participants viewed a fixation mask (1 sec), followed by a set of three, four or five digits (2 sec), followed by another fixation mask (5 sec). Then, a 'target' digit from the previously presented set appeared (10 sec). Participants were told to push a button indicating whether the target was the middle digit of the previously presented set or not. The fMRI data were gathered every 1.5 sec. The full experimental design as well as preprocessing of these data are described by Siegle et al. (2007). Briefly, data were subjected to motion correction, detrending within blocks, and temporal smoothing, cross registered to an image within the dataset, and subjected to spatial smoothing, 6mm FWHM Gaussian kernel. The reference brain was then transformed into Talairach space using AFNI (Cox, 1996)[17] to extract anatomical masks. Regions involved in this analysis were dACC restricted in BA32 and DLPFC restricted in BA9 (Figure 1.6), with significant scan $\times$ condition interactions ($p < 0.0001$) using a repeated measures Analysis of Variance (ANOVA) in which subject was a random factor and scan and condition (number of presented digits) were fixed factors, subject to an empirically derived contiguity threshold of 105 voxels. Significant regions were restricted to those in BA32 and BA9 using Talairach masks and are shown in Figure 1.6. The averaged fMRI signal from each ROI(region of interest) were acquired and normalized by subtracting and then divided by the median regional activation across the whole time-series within stimulus types and subjects. Figure 1.7 shows one subject's BOLD

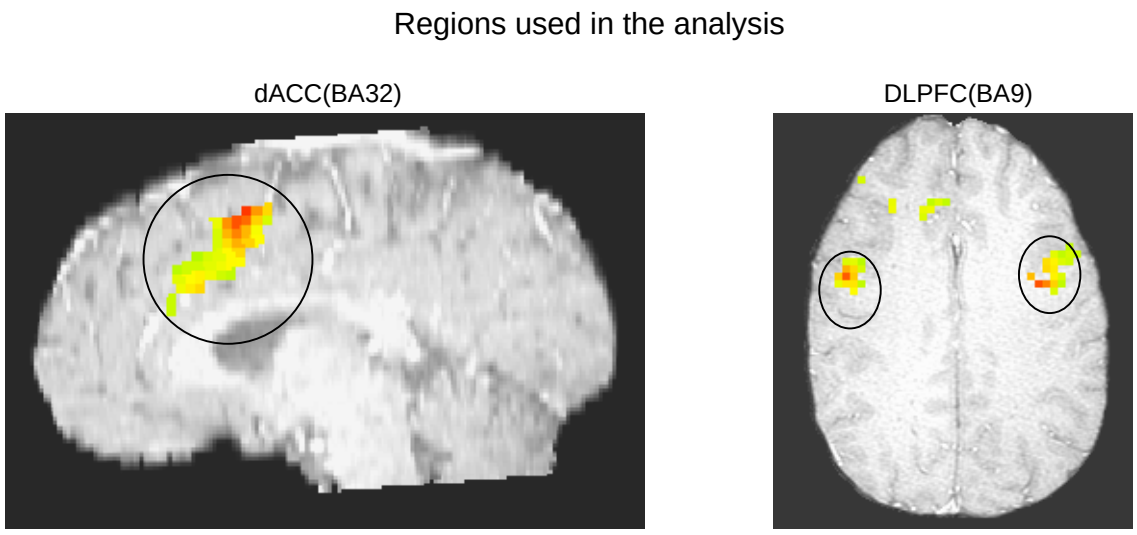


Figure 1.6: Regions used in the analysis

Identified regions of interest used in the analysis: dACC restricted in BA32 (left) and DLPFC restricted in BA9 (right).

responses of the ROIs.

We applied all of the FC measures (including partial correlation methods, controlling for the mean effect from the external stimulus in each region) to dACC(BA32) and DLPFC(BA9) within subjects. Cross-correlation and partial cross-correlation both reached their maximum at lag 0 and the associations were not attenuated when controlling the effects from the external stimulus. Taking out the external stimulus waveform reduced cross-coherence between these two regions at the trial frequency. A strong relationship was observed for both peak correlation and canonical correlation.

Based on the results of different FC measures on this real dataset, we classified the 32 subjects according to the classification tree shown in Figure 1.3. There were 6 detected relationships: 3. Zero-order relationship unrelated to design; 6. Related peak amplitudes and latencies but variability in higher moments; 8. A-1-mode B-2-modes; 9. Relationship of A-peak-amplitude to B-sustained activity; 10. Unrelated peak amplitudes but related peak latencies; and 11. No relationship. Most subjects (27 of 32) were classified to relationship 6, 8 and 10 (Figure 1.8). Average results of the computed FC within three main groups are shown in Figure 1.9 (partial results are shown in dashed red lines). Subjects classified

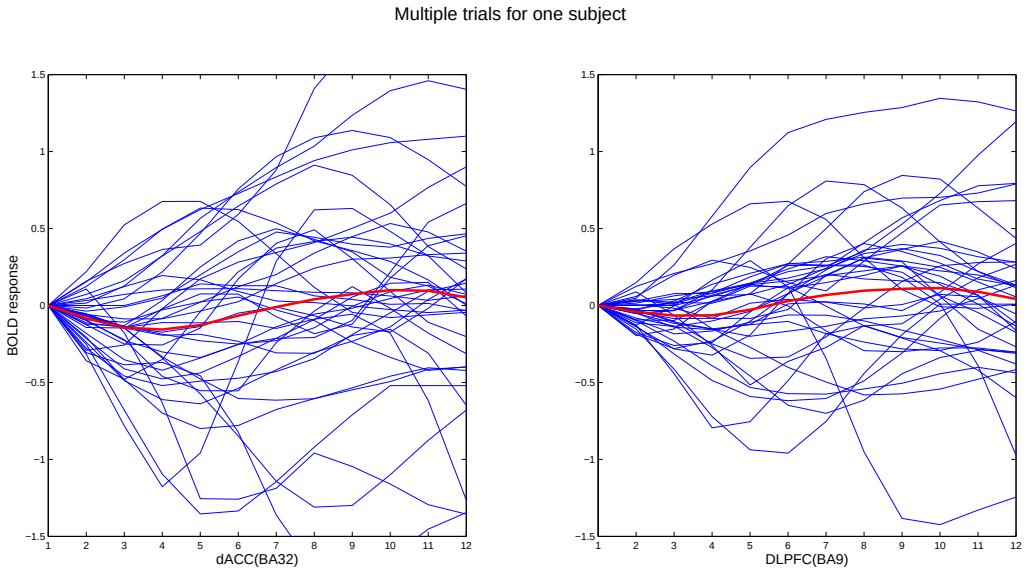


Figure 1.7: Multiple trials for one subject

Thirty-six experimental trials for one subject. Blue lines are trials in dACC(BA32) (left panel) and DLPFC(BA9) (right panel). Bold red lines are subject's mean responses for each ROI.

to relationships 6 and 8 showed higher cross-correlation and peak correlation than subjects classified to relationship 10 because they had related peak amplitude activity between the two regions. Subjects classified to relationship 8 showed higher cross-coherence over low frequency band than the other groups.

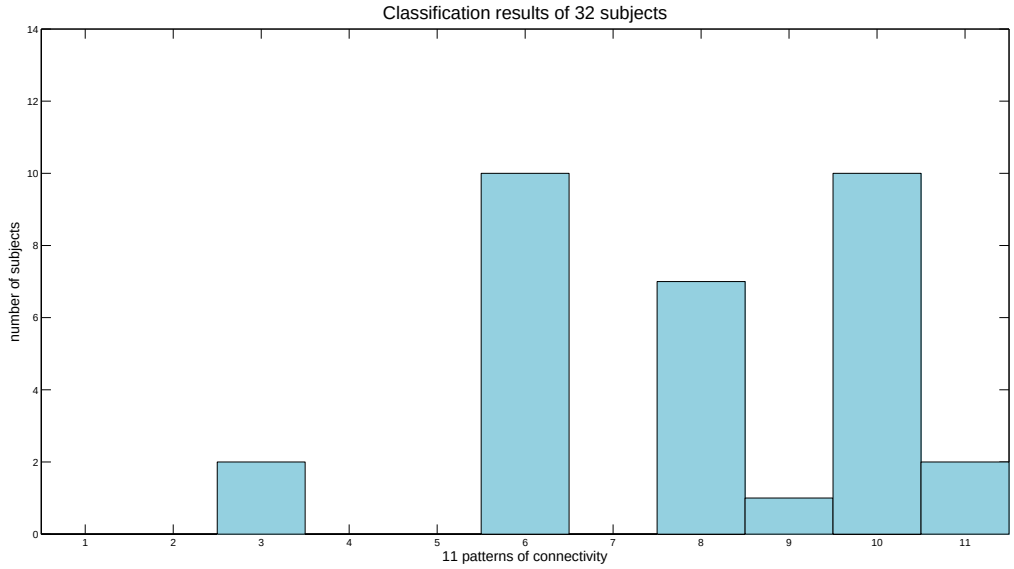


Figure 1.8: Classification results of 32 subjects

Classification results for 32 subjects. X axis represents the 11 types of association corresponding to Figure 1.3. Y axis represents the number of subjects classified to each relationship.

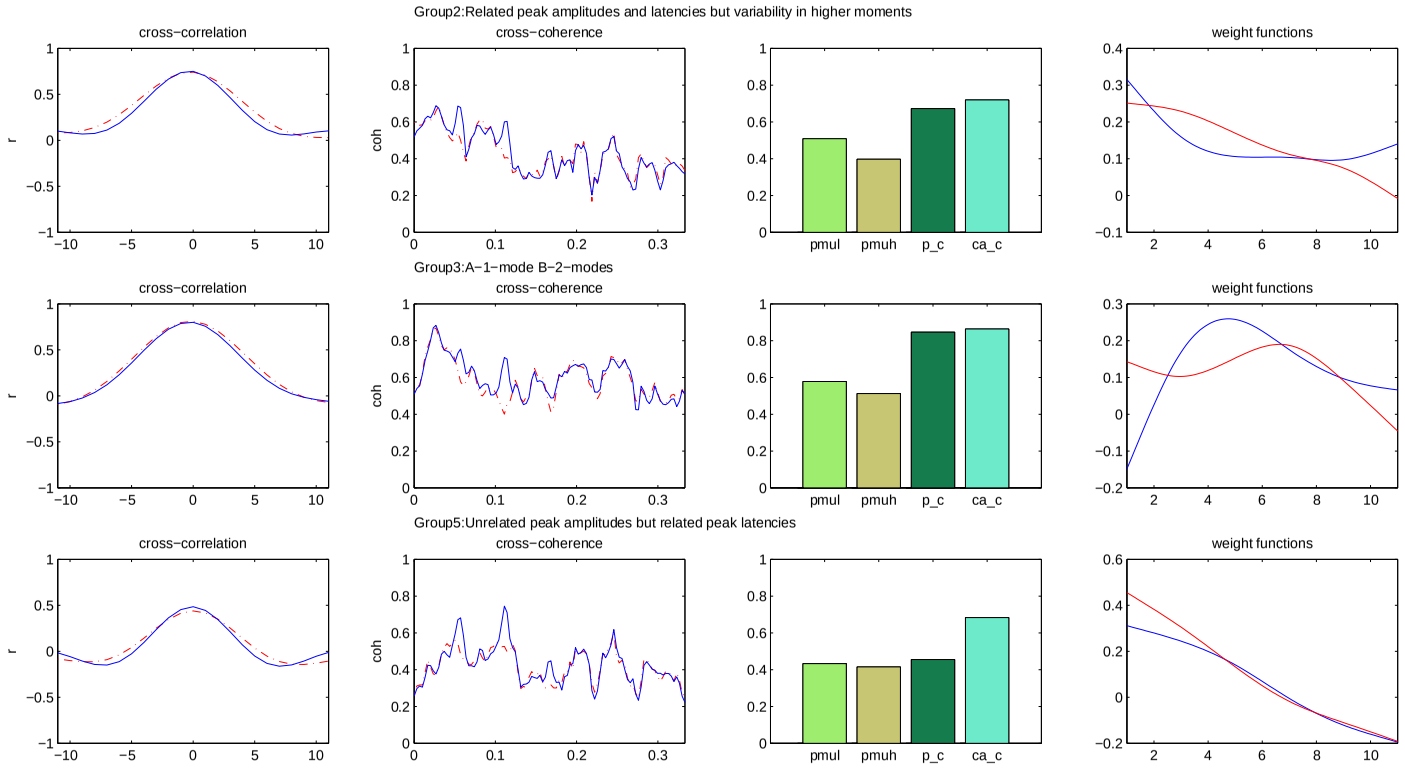


Figure 1.9: Mean results of FC on Empirical Dataset

Mean results of different FC measures between dACC(BA32) and DLPFC(BA9) for three main classified groups. Red dashed lines shown are partial correlation or coherence. 'pmul', 'pmuh', 'p_c', 'ca_c' represent partial mutual information over low frequency band, partial mutual information over high frequency band, peak correlation, and functional canonical correlation respectively.

1.4 DISCUSSION

We have developed a MATLAB toolbox for performing FC analyses which includes many of the most commonly-used approaches researchers have utilized to date for the identification of condition-dependent functional interactions between fMRI time-series obtained from two or more brain regions [7, 59, 63, 68]. The approaches are either bivariate or multivariate methods defined in time or frequency domains that emphasize distinct features of relationships among the time-series. An optional pre-smoothing step is also implemented which allows empirically-derived temporal smoothing of the data before performing FC analyses. The FC toolbox enables ease of comparison and greater flexibility for choosing among FC measures, and may potentially lead to a greater understanding of the precise nature of FC relationships manifested in a given dataset. The simulation results illustrated that using multiple FC measures could effectively detect and classify regional associations and provide more information about the type of FC than any single measure.

We applied these methods to an fMRI study to determine FC between dACC(BA32) and DLPFC(BA9) during a digit sorting task. We found strong relationships between these two ROIs. Relationships between the regions were 1) heterogeneous across subjects, 2) related to task, and 3) more complex than would have been detected using simple zero-order statistics such as correlation. Following the classification tree (Figure 1.3) 27 of 32 subjects were classified to relationship 6, 8 and 10 (Figure 1.8). This indicated that the most common FC relationship in the sample involved a higher peak response in dACC(BA32) in response to a higher peak response in DLPFC(BA9). But some subjects displayed a prolongation of response in dACC(BA32) in response to a higher peak response in DLPFC(BA9). These relationships were not trivial, and certainly were more complex than would be revealed by zero-order correlation alone. At the most basic level, we can conclude that in this study, the dACC(BA32) and DLPFC(BA9) were strongly related among nearly all subjects - this point would not have been possible without using multiple connectivity measures. Future research is necessary to suggest whether the different observed patterns of relationships have psychological and biological importance, e.g., whether subjects with different patterns of connectivity differed in other important ways such as their performance.

The toolbox is flexible, taking brain activity data as inputs but also able to accept peripheral physiological measures (i.e., blood pressure, heart rate, etc.) into the FC function, with the requirement that all time-series inputs should be on the same resolution . An interpolation function is available in this toolbox that would allow time courses with different resolutions to be applied that are thereafter altered to be on the same resolution as the fMRI time courses. Furthermore, this toolbox could be implemented for determining whole brain network structure in which, instead of doing FC analysis between ROIs, researchers could do FC analysis between each pair of voxels over the whole brain.

We plan future work in several areas. First, an important question not addressed in the smoothing step is the estimation of the autocorrelation in the noise. We intend to implement an improved smoothing step which estimates the autocorrelation as well as smoothing in a future version of the toolbox. Second, the simulation studies implemented in this paper only considered direct relationships between two regions; multivariate relationships involving three or more regions are, of course, important. We intend to perform further simulations involving more than two regions to examine the behavior of these algorithms under indirect regional associations. Finally, our simulation studies only included 11 distinct patterns of inter-regional connectivity; since there may be many more types of connectivity relationships in actual data, results obtained from our classification tree may be misleading. However, we have used this limited set of patterns to demonstrate some possible associations, and to show that for understanding plausible relationships it may be useful to compute multiple FC measures. Moreover, the website for this toolbox is open to the public so that users can provide information on this issue. Specifically, we have created an area in which new empirical or simulated datasets can be uploaded with known relationships. We plan to update the classification tree regularly based on these data.

The Matlab Functional Connectivity toolbox is freely-availabe at http://groups.google.com /group/fc-toolbox.

2.0 A BAYESIAN APPROACH ON SMOOTHING AND MAPPING FUNCTIONAL CONNECTIVITY FOR EVENT-RELATED FMRI DESIGNS

2.1 INTRODUCTION

Functional magnetic resonance imaging (FMRI) has provided a useful technique to study activation of the human brain in tasks of cognition, emotion, and behavior. During an fMRI experiment, the subject performs a sequence of tasks while magnetic resonance images of his or her brain are acquired at regular intervals, which yields a sequence of three-dimensional images of the subject's brain over time. Thus, fMRI data is collected in four dimensions, three spatial and one temporal, and the time dimension measures the blood oxygen level-dependent (BOLD) response to reflect neural activity. FMRI experiments are often employed to study activation and functional connectivity (FC), which seeks to characterize the dynamics of BOLD responses and determine relationships among specialized brain regions in processing information under various experimental conditions, or subgroups of subjects. The aim of this chapter is to present an exploratory model to analyze fMRI data from multiple brain regions in slow event-related fMRI experimental designs. This model simultaneously estimates the BOLD responses and FC, while making minimal assumptions about the shape of BOLD responses and the nature of the inter-regional relationships.

Slow event-related designs are well-suited for exploring temporal dynamics of activity in regions involved in complex cognition or emotion tasks, which may take several seconds to process [71]. In a slow event-related experimental paradigm, stimuli or trials are presented individually, separated by inter-stimulus interval, which is usually on the order of 10 to 20 seconds. This duration is usually long enough to allow the neural activation following the stimulus returning to its baseline [33], hence allowing researchers to learn about the BOLD

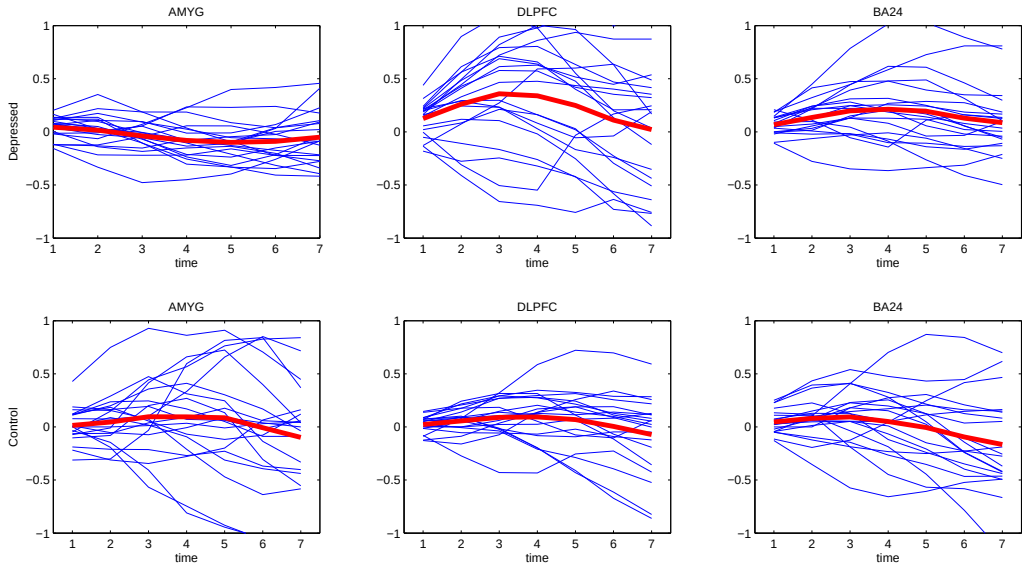


Figure 2.1: Two subjects' trial activation trajectories

Subjects' stimulus-locked trial activation trajectories in three brain regions in response to 20 negative word stimuli. Regional activation trajectories were determined by averaging activation trajectories over voxels within the region. Blue lines indicate individual stimulus-locked trajectories, red lines indicate subjects' averaged activation trajectories.

response and the post-stimulus inter-regional relationships. Experimental stimuli can vary trial by trial and are usually presented in a random order, which permits the examination of the inter-regional relationships for different stimulus types.

There are two special aspects of the data resulting from this type of experiment. First, the responses are inherently functional in nature, sampled discretely at a finite number of time points. Second, the data resulting from this experiment are nonlinear and nonstationary, vary from trial to trial, person to person and group to group. A motivating experiment of slow event-related design published in 2007 by Siegle, et al [68] is shown in Figure 2.1; data in this figure are stimulus-locked responses trajectories for three regions of interest (ROIs) of a mentally-healthy control subject and a depressed subject with major depressive disorder. The subjects were presented with a list of 60 personally-relevent negative/neutral/positive words and the presented trials are from the negative stimuli.

In traditional analyses, curve averaging within subjects or trial types is used for studying stimulus-locked activation in fMRI study. This technique provides a simple method for

learning the nonstationarities in stimulus-locked BOLD responses. However, it ignores the within-subject variability. Another commonly used approach for studying stimulus-locked activation is to use parametric models with a variety of covariates of interest, for example, categorical variables describing the type of multiple tasks, group membership, and so forth. Usually, the design matrix will also include terms representing the predicted shape of BOLD response curve, specifically a convolution of the stimulus time courses with a model for the hemodynamic response function (HRF), typically a simple gamma [46], Possion [28], or Gaussin model [55]. But fMRI responses are frequently observed that they do not match the HRF model [68, 47]. In addition, study designs are becoming increasingly complex and the signals under investigation are closer to the limits of detection. Thus, nonparametric approaches are proposed to model fMRI responses [19, 13, 71], which require no or very limited assumptions to be made about the data and are therefore applicable in a wider variety of situations.

In this chapter, a nonparametric approach is going to be introduced. As a point of general statistical methodology, this new approach is concerned with multivariate functional data analysis, with aims of smoothing and learning the covariance structure of multiple curves, where the curves are observed longitudinally over equally spaced time points. The statistical framework of functional data analysis (FDA) is a term introduced by Ramsay and Silverman [56], where the basic unit of information is the entire function, such as a curve or image. There has been an increasing interest in the nonparametric analysis of data that are in the form of functions. Rice and Silverman(1991) [57] used smoothing splines to model the mean function and modeled the covariance functions in terms of eigenfunctions; Brumback and Rice (1998) [9] introduced a penalized smoothing spline mixed model which used the mixed-effects model as a basis for extending the smoothing spline model for individual curves (Kimmeldrof and Wahba 1970 [44]; Silverman 1985 [69]) to a sample of curves; Shi, et al. (1996) [65], Rice and Wu (2001) [58], and James, et al. (2001) [40] proposed B-splines to model the individual curves with random coefficients through mixed-effects model; Rice and Wu (2001) and James et al. (2001) [40] suggested using principal component (PC) functions to characterize the dominant modes of variation of a sample of random trajectories around an overall mean trend function. Our approach is most closely related to that of James,

et al. (2001) [40] and Rice and Wu (2001)[58]. Unlike the standard nonparametric FDA literature, the major novelty in our approach is that it focuses on a set of functions which are not necessarily independent. Smoothing is accomplished through a few PC functions via the one-dimensional reduced rank mixed-effect model proposed by James, et al. (2001) [40], and then modeling the association of curves by jointly modeling the PC scores. The model is employed in a Bayesian formulation and we use Markov chain Monte Carlo (MCMC) methods to sample from the posterior distribution of the model parameters. Our model also allows for straightforward computation of pointwise Bayesian posterior credible regions for both the mean curve, the individual curves, and the PC curves.

The outline for the remainder of this chapter is as follows. In Section 2.2, we describe the proposed model, while Section 2.3 introduces a new FC measure. Section 2.4 outlines the sampling scheme, model selection criteria as well as the posterior inference for the new approach. In Section 2.5 we provide the results of a simulation study to evaluate the method, and an application to analyze data from a psychiatric neuroscience experiment. Finally some discussion related to the work is given.

2.2 THE MODEL AND PRIOR SPECIFICATION

2.2.1 The Model

2.2.1.1 The Mixed-Effects Model Mixed-effects models have been widely used in the analysis of curve data; Shi, et al. (1996) [65] and Rice and Wu (2001)[58] suggested using a mixed-effects approach to solve the functional principal components problem. in their model, a set of smooth basis functions $\phi_k(t)$ $(k = 1, \ldots, K)$, such as B-splines, are used to represent the curves, where the spline coefficients are assumed to be random to capture the individual-specific effects. Let $Y_i(t)$ be the measurement at time t for the ith individual or curve and write

$$Y_i(t) = \mu(t) + h_i(t) + \epsilon_i(t), \quad 0 \leqslant t \leqslant T, \quad i = 1, \ldots, N \tag{2.1}$$

where $\mu(t)$ is the overall mean curve for the population, $h_i(t)$ represents the random departure from the mean for subject i and $\epsilon_i(t)$ is the random measurement of error with mean zero and variance σ^2. Let $\boldsymbol{\phi}(t) = (\phi_1(t), \phi_2(t), \ldots, \phi_K(t))^T$ be the vector of K spline basis functions evaluated at time t. Denote $\boldsymbol{\beta}$ be the unknown but fixed vector of spline coefficients, and let $\boldsymbol{\gamma}_i$ be a random vector of spline coefficients for each curve with population covariance matrix Γ. Then, $\mu(t)$ and $h_i(t)$ are modelled by a linear combination of basis functions, and the resulting mixed-effects model has the form

$$Y_i(t) = \boldsymbol{\phi}(t)^T\boldsymbol{\beta} + \boldsymbol{\phi}(t)^T\boldsymbol{\gamma_i} + \epsilon_i(t), \quad i = 1, \ldots, N. \tag{2.2}$$

The principal patterns of variation about the mean curve are referred to as functional principal component curves. Rice and Wu (2000) [58] suggested modeling the patterns of variation of the basis coefficients, $\boldsymbol{\gamma_i}$, and then transform back to the original space. Since Γ is the covariance matrix of the $\boldsymbol{\gamma_i}$, this is achieved by multiplying the eigenvectors of Γ by $\mathbf{b}(t)$.

2.2.1.2 Single-Curve Reduced Rank Mixed-Effects Model If the dimension of the spline basis is K then in fitting the covariance matrix Γ, $K(K+1)/2$ different parameters must be estimated, and the large number of parameters may make the model's fit deteriorate. James, et al. (2000) [40] pointed out this problem by developing a reduced rank model in which the individual departure from the mean curve is modeled by a small number of PC curves. This reduced rank model can be interpreted as a submodel of the mixed effects model. Let f_q be the qth PC function and let $\mathbf{f}(t) = (f_1(t), f_2(t), \ldots, f_Q(t))^T$ be the vector of PC functions evaluated at time t. The reduced rank model is defined as

$$Y_i(t) = \mu(t) + \mathbf{f}(t)^T\boldsymbol{\alpha_i} + \epsilon_i(t) \quad i = 1, \ldots, N \tag{2.3}$$

subject to the orthogonality constraint $\int f_j f_l = \delta_{jl}$, with δ_{jl} being the Kronecker δ. These orthogonal PC functions characterize the major modes of variation in the individual curves.The components of the random vector $\boldsymbol{\alpha_i}$ for the ith individual give the relative weight of the PC functions. The $\boldsymbol{\alpha_i}$'s and ϵ_i's are assumed to be uncorrelated with each other and the ϵ_i's are temporally uncorrelated with each other as well. The α_i's are taken to a common covariance matrix and the ϵ_i's are assumed to have a common covariance σ^2.

Similar to the mixed-effects model 2.2, in order to fit this model when the data are measured at only a finite number of time points, James, et al. (2000) [40] chose to represent $\mu(t)$ and $\mathbf{f}(t)$ using a basis of spline functions. Let $\boldsymbol{\phi}(t) = (\phi_1(t), \phi_2(t), \ldots, \phi_K(t))^T$ be the vector of K spline basis functions evaluated at time t. Let $\boldsymbol{\theta_\mu}$ and Θ be, respectively, a K-dimensional vector and a K by Q matrix of spline coefficients. Then,

$$\begin{aligned} \mu(t) &= \boldsymbol{\phi}(t)^T \boldsymbol{\theta_\mu} \\ \mathbf{f}(t)^T &= \boldsymbol{\phi}(t)^T \Theta \end{aligned}$$

The reduced rank model then has the form

$$\begin{aligned} Y_i(t) &= \boldsymbol{\phi}(t)^T \boldsymbol{\theta_\mu} + \boldsymbol{\phi}^T \Theta \boldsymbol{\alpha_i} + \epsilon_i(t) \quad i = 1, \ldots, N \\ & \boldsymbol{\alpha_i} \sim (0, \Sigma_\alpha), \quad \epsilon_i(t) \sim (0, \sigma^2) \end{aligned} \tag{2.4}$$

subject to

$$\Theta^T \Theta = I, \quad \int \boldsymbol{\phi}(t)^T \boldsymbol{\phi}(t) dt = 1, \quad \int \int \boldsymbol{\phi}(t)^T \boldsymbol{\phi}(s) dt ds = 0. \tag{2.5}$$

The equations in 2.5 impose orthogonality constraints on the PC curves. Σ_α is restricted to be diagonal, otherwise, α_i's, Θ and Σ will be confounded [40]. According to Zhou, et al. 2008 [76], the orthogonality constraints imposed on $b(t)$ could be achieved approximately by choosing $\boldsymbol{\phi}(t)$ such that $(L/l)B^T B = I$. Here, $B = (\boldsymbol{\phi}(t_1), \ldots, \boldsymbol{\phi}(t_l))^T$ is the matrix of basis functions evaluated at a fine grid of l time-points $t_1, \ldots, t_l$ and L is the length of the interval in which these grid points are taken.

As we have mentioned at the beginning of this section, the reduced rank model is a submodel of the mixed effects model defined in Equation 2.2. Reparameterizing $\boldsymbol{\gamma_i}$ in 2.2 as

$$\begin{bmatrix} \Theta & \Theta^* \end{bmatrix} \begin{pmatrix} \boldsymbol{\alpha_i} \\ \boldsymbol{\alpha_i}^* \end{pmatrix} \tag{2.6}$$

where Θ and $\boldsymbol{\alpha_i}$ are defined as above, Θ^* is a K by $K - Q$ matrix which is orthogonal to Θ, and $\boldsymbol{\alpha_i}^*$ is a random vector of length $K - Q$ with a diagonal covariance matrix. As a result the mixed effects model in 2.2 can be written as

$$Y_i(t) = \boldsymbol{\phi}(t)^T \boldsymbol{\theta_\mu} + \boldsymbol{\phi}(t)^T \Theta \boldsymbol{\alpha_i} + \boldsymbol{\phi}(t)^T \Theta^* \boldsymbol{\alpha_i}^* + \epsilon_i(t) \quad i = 1, \ldots, N \tag{2.7}$$

In the reduced rank model the $\boldsymbol{\alpha}_i^*$'s are set to zero and no attempt is made to estimate the additional parameters Θ^*. To fit Q principle component curves Rice and Wu (2000)[58] suggested to calculate the first Q eigenvectors of Γ; in other words, even though Θ^* is estimated it is never used. By employing the reduced rank model, the principal component curves are estimated directly rather than estimating an entire covariance matrix and computing the first Q eigenvectors.

2.2.1.3 Multiple-Curve Reduced Rank Mixed-Effects Model For data consisting of multiple curves, an important problem of interest is modeling the association among them. Zhou, et al. 2008 [76] discussed jointly modeling paired curves and the idea we use is similar to theirs, but emphasizing the multivariate case. We first model each curve via the single reduced rank PC model as discussed in Section 2.2.1.2, and the association among the curves is modeled through the association among the PC scores corresponding to the underlying variables. To be more specific, we assume that the scores of each PC function are realizations of a stationary process.

Suppose $\mathbf{Y}_i(t)$ is a $P \times 1$ vector of response variables on subject i at time t, $i = 1, \ldots, N$, and consider the model

$$\begin{aligned} \mathbf{Y}_i(t) &= \boldsymbol{\mu}(t) + \boldsymbol{h}_i(t) + \boldsymbol{\epsilon}_i(t), \quad 0 \leqslant t \leqslant T, \quad i = 1, \ldots, N, \\ &= \boldsymbol{\mu}(t) + \mathbf{f}(t)^T \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_i(t) \end{aligned} \tag{2.8}$$

In Equation 2.8, $\boldsymbol{\mu}(t) = (\mu_1(t), \ldots, \mu_P(t))'$, $\mathbf{f}(t)^T = \text{diag}\left(\mathbf{f}_1(t)^T, \ldots, \mathbf{f}_P(t)^T\right)$ where $\mathbf{f}_p(t) = \left(f_{p1}(t), \ldots, f_{pQ_p}(t)\right)^T$, $p = 1, \ldots, P$, is the $Q_p \times 1$ vector of PC functions corresponding to the underling variable evaluated at time t, $\boldsymbol{\epsilon}_i(t) = (\epsilon_{i1}(t), \ldots, \epsilon_{iP}(t))'$ is the vector of measurement errors at time t. The measurement errors ϵ_i's are assumed to be temporally uncorrelated with mean zero and covariance matrix $\Sigma_\omega = \text{diag}(\sigma^2_{\epsilon,1}, \ldots, \sigma^2_{\epsilon,P})$. The random vector $\boldsymbol{\alpha}_i^T = \left(\boldsymbol{\alpha}_{i1}^T, \ldots, \boldsymbol{\alpha}_{iP}^T\right)$ and $\boldsymbol{\alpha}_{ip}$, $p = 1, \ldots, P$, gives the relative weights on the PC scores for the pth variable and ith individual. It is also assumed that the $\boldsymbol{\alpha}_i$'s and $\boldsymbol{\epsilon}_i$'s are mutually independent. The PC functions for each single curve are subject to the orthogonality constraints $\int f_{pq} f_{pl} = \delta_{p_{ql}}$.

As discussed in Section 2.2.1.2, for the purpose of identifiability, the PC scores α_{ipq}, $q = 1, \ldots, Q_p$, for the pth curve should be independent with each other for $p = 1, \ldots, P$. The strength of association among the curves is modeled on the smoothed response level via correlations between the scores of PC functions. To be specific, $\boldsymbol{\alpha}_i \sim N(0, \Sigma)$ and Σ is restricted to the form as

$$\begin{pmatrix} \Sigma_1 & C_{12} & \cdots & C_{1P} \\ C_{12}^T & \Sigma_2 & \cdots & C_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1P}^T & C_{2P}^T & \cdots & \Sigma_P \end{pmatrix},$$

where $\Sigma_1, \ldots, \Sigma_P$ are diagonal matrix. Σ can also be broken down in terms of its corresponding standard deviations and correlation matrix. Specifically, we write

$$\Sigma = \text{diag}(D) R \text{diag}(D) \tag{2.9}$$

where $D = (D_1', \ldots, D_P')'$ is the $Q \times 1$ ($Q = Q_1 + \ldots + Q_P$) vector of standard deviations and R is restricted to the form as

$$\begin{pmatrix} I_{Q_1} & R_{12} & \cdots & R_{1P} \\ R_{12}^T & I_{Q_2} & \cdots & R_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ R_{1P}^T & R_{2P}^T & \cdots & I_{Q_P} \end{pmatrix},$$

with I being the identity matrix.

Let $\boldsymbol{\phi}(t) = (\phi_1(t), \phi_2(t), \ldots, \phi_K(t))^T$ be the K-dimensional vector of orthogonal spline basis functions evaluated at time t and $\Phi^b(t)$ is a P by $P \times K$ matrix defined as

$$\Phi^b(t) = \begin{bmatrix} \boldsymbol{\phi}^T(t) & \ldots & \mathbf{0}^T \\ \vdots & \ddots & \vdots \\ \mathbf{0}^T & \ldots & \boldsymbol{\phi}^T(t) \end{bmatrix}. \tag{2.10}$$

Then the model for the observed data can be written as

$$\mathbf{Y}_i(t) = \boldsymbol{\Phi}^b(t)\boldsymbol{\theta}_{\boldsymbol{\mu}} + \boldsymbol{\Phi}^b(t)\Theta\boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_i(t). \tag{2.11}$$

In Equation 2.11, $\boldsymbol{\theta_\mu}$ and Θ are, respectively, a $P \times K$-dimensional vector and a $P \times K$ by $P \times Q$ matrix of spline coefficients where $\boldsymbol{\theta_\mu} = (\boldsymbol{\theta_{\mu_1}}', \ldots, \boldsymbol{\theta_{\mu_P}}')'$ and

$$\Theta = \begin{bmatrix} \Theta_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \Theta_P \end{bmatrix} \tag{2.12}$$

subject to $\Theta_p^T \Theta_p = I, p = 1, \ldots, P$. Θ_p is a $K \times Q_p$ matrix.

2.2.1.4 Adaptation of the Multiple-Curve Reduced Rank Model to Slow Event-Related Design

As discussed in Section 2.1, in a slow event-related design, the fMRI data is observed for multiple subjects at multiple brain regions, with multiple trials per subject and multiple scans per trial. Suppose there are N subjects, with subject i completing M_i trials. Each trial consists of a short stimulus followed by an intertrial period consisting of S fMRI scans. The fMRI activations for each trial are obtained from P pre-specified brain regions. For trial j nested within subject i, denote the P-dimensional observed fMRI response at the tth time (scan) following stimulus presentation by $\mathbf{Y}_{ij}(t)$. The fMRI response $\mathbf{Y}_{ij}(t)$ can be decomposed into two parts

$$\begin{gathered} \mathbf{Y}_{ij}(t) = \mathbf{B}_{ij}(t) + \boldsymbol{\epsilon}_{ij}(t), \\ i = 1, \ldots, N, \quad j = 1, \ldots, M_i, \quad t = 1, \ldots, S, \end{gathered} \tag{2.13}$$

where $\mathbf{B}_{ij}(t)$ represents the underlying BOLD response and $\boldsymbol{\epsilon}_{ij}(t)$ represents the noise which is assumed to be temporally uncorrelated with mean zeros and variance covariance matrix $\Sigma_\omega = \text{diag}(\sigma^2_{\epsilon,1}, \ldots, \sigma^2_{\epsilon,P})$. $\mathbf{B}_{ij}(t)$ can be written as

$$\mathbf{B}_{ij}(t) = \boldsymbol{\mu}_i(t) + \mathbf{h}_{ij}(t), \tag{2.14}$$

$\boldsymbol{\mu}_i(t)$ is the overall mean function for subject i evaluated at time t and $\mathbf{h}_{ij}(t)$ represents the smoothed deviation of the jth trial at time t. Following model 2.11, each dimension of $\mathbf{h}_{ij}(t)$ is summarized by a set of PC functions which characterize the major modes of variation

in the trial waveforms and we assume the sources of variation are constant across subjects. Then, model 2.14 becomes

$$\mathbf{B}_{ij}(t) = \mathbf{\Phi}^b(t)\boldsymbol{\theta}_{\mu_i} + \mathbf{\Phi}^b(t)\Theta\boldsymbol{\alpha}_{ij}. \tag{2.15}$$

$\mathbf{\Phi}^b(t)$, $\boldsymbol{\theta}_{\boldsymbol{\mu_i}}$ and Θ are defined the same as in Model 2.11. The random vector $\boldsymbol{\alpha}_{ij}$ gives the relative weights on the PC scores for the jth trial of ith individual and it is assumed that $\boldsymbol{\alpha}_{ij} \sim N(0, \Sigma_i)$. It is also assumed that $\boldsymbol{\alpha}_{ij}$'s and $\boldsymbol{\epsilon}_{ij}$'s are mutually independent. Writing $\Sigma_i = \text{diag}(\mathbf{D_i})R_i\text{diag}(\mathbf{D_i})$, $i = 1, \ldots, N$, where R_i is the correlation matrix and $\mathbf{D_i}$ is the vector of standard deviations in subject i. We allow subject-dependent vectors of standard deviations, but a common correlation matrix across subjects. Thus, $R_i = R$ for all i, and $\mathbf{D_i}$'s are unrestricted except being component-wise positive. This assumption of the correlation structure keeps the number of unknown parameters to a much smaller size which is more manageable. Clinically, we assume that the communication structure among brain regions is similar across subjects which is reasonable if all the subjects are in the same categorical group, for example, if they are all depressed patients. Here, the inter-regional association is assumed within trials.

2.2.2 Prior Specification

The parameters of the proposed model are denoted by $\{\boldsymbol{\theta}_{\boldsymbol{\mu_i}}, \Theta, \mathbf{D_i}, R, \Sigma_\omega\}$, where index $1 \leqslant i \leqslant N$ ranges over subjects. Prior specifications for some of the parameters follow the common practice for Bayesian mixed models [29].

Specifically, the prior distributions for spline coefficients $\boldsymbol{\theta}_{\boldsymbol{\mu_i}}$, $i = 1, \ldots, N$, are independent and follow multivariate normal distributions

$$\boldsymbol{\theta}_{\boldsymbol{\mu_i}} \sim N_{P\times K}(\mathbf{0}, c_\theta I_{P\times K}), \tag{2.16}$$

where $I_{P\times K}$ is a identity matrix and the multiplier c_θ is a prespecified constant; large values of c_θ correspond to a vague prior for $\boldsymbol{\theta}_{\boldsymbol{\mu_i}}$.

The spline coefficients $\Theta_p = \left(\boldsymbol{\theta}_{p,1}, \ldots, \boldsymbol{\theta}_{p,Q_p}\right), p = 1, \ldots, P$, is a $K \times Q_p$ matrix and the prior for each column vector $\boldsymbol{\theta}_{p,q}, q = 1, \ldots, Q_p$, is given by $N_K(\mathbf{0}, c_\Theta I_K)$ and set c_Θ to large values to give a diffuse yet proper prior.

A uniform prior distribution on $\log(\sigma_{\omega,p})$,$p = 1, \ldots, P$, is assumed which is equivalent to

$$p(\sigma_{\omega,p}) \propto \sigma_{\omega,p}^{-1} \quad or \quad p(\sigma_{\omega,p}^2) \propto \sigma_{\omega,p}^{-2}, \tag{2.17}$$

and this can be taken as a limit of proper conditionally-conjugate inverse-gamma priors [30].

Since $\mathbf{D_i}$, $i = 1, \ldots, N$, is simply a $Q \times 1$ vector with component-wise non-negativity as the only constraint, the priors is placed on the logarithm of $\mathbf{D_i}$ [6]. The $\log(\mathbf{D_i})$'s, $i = 1, \ldots, N$, are shrinked to the same structure by following a common multivariate normal distribution where

$$\log(\mathbf{D_i}) \sim N_Q(\boldsymbol{\xi}, \Lambda). \tag{2.18}$$

The matrix Λ is chosen to be diagonal, that is, we are choosing independent log normal distributions for each of the standard deviations. The hyperpriors for $\boldsymbol{\xi}$ and Λ are given as

$$\begin{aligned} &\boldsymbol{\xi} \sim N_Q(\mathbf{0}, c_{\boldsymbol{\xi}} I_Q), \\ &\log(\sqrt{\Lambda_q}) \propto 1, \quad q = 1, \ldots, Q \end{aligned} \tag{2.19}$$

with a fixed large value of $c_{\boldsymbol{\xi}}$. Λ_q is the qth diagonal element of Λ.

Let $\mathcal{G}$ be a gaussian bi-directed graph and an absent edge of $\mathcal{G}$ corresponding to marginal independence. We accommodate the common correlation model [49] to correlation matrix subject to $R \in \mathscr{R}^Q$, a positive definite matrix, and $R \in M(\mathcal{G})$, where $M(\mathcal{G})$ is the cone of correlation matrices which fulfill the linear restrictions

$$r_{ql} = 0 \text{ if } (q, l) \notin \mathcal{G}. \tag{2.20}$$

We consider correlation matrix R restricted to the form as

$$\begin{pmatrix} I_{Q_1} & R_{12} & \cdots & R_{1P} \\ R_{12}^T & I_{Q_2} & \cdots & R_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ R_{1P}^T & R_{2P}^T & \cdots & I_{Q_P} \end{pmatrix}.$$

Then, a zero value in R corresponds to an absent edge in $\mathcal{G}$. In the common correlation model, all nonzero correlations r_{ql} are assumed to follow a common normal distribution where

$$r_{ql} \sim N(\mu, \sigma^2),\ 1 \leqslant q < l \leqslant Q \text{ and } (q,l) \in \mathcal{G}. \tag{2.21}$$

Then,

$$f(R|\mu,\sigma^2) = C(\mu,\sigma^2) \prod_{\substack{q<l \\ (q,l)\in\mathcal{G}}} \exp\left\{-\frac{(r_{ql}-\mu)^2}{2\sigma^2}\right\}\ I\left\{R \in \mathscr{R}^Q, R \in M(\mathcal{G})\right\} \tag{2.22}$$

where

$$C^{-1}(\mu,\sigma^2) = \int\limits_{\substack{R\in\mathscr{R}^Q \\ R\in M(\mathcal{G})}} \prod_{\substack{q<l \\ (q,l)\in\mathcal{G}}} \exp\left\{\frac{-(r_{ql}-\mu)^2}{2\sigma^2}\right\}\ dr_{ql}. \tag{2.23}$$

and where $I\{\centerdot\}$ represents an indicator function which introduces dependence among r_{ql}s. We assume the hyperpriors for μ and σ as

$$\begin{aligned} &\mu \sim N(0, c_\mu), \\ &\log(\sigma) \propto 1, \end{aligned} \tag{2.24}$$

where c_μ is a specified constant. The full conditional densities for μ and σ^2 are

$$p(\mu \mid R,\sigma^2) \ \propto\ C(\mu,\sigma^2) \prod_{\substack{q<l \\ (q,l)\in\mathcal{G}}} \exp\left\{-\frac{(r_{ql}-\mu)^2}{2\sigma^2}\right\} \exp\left(\frac{-\mu^2}{2c_\mu}\right), \tag{2.25}$$

$$p(\sigma^2 \mid R,\mu) \ \propto\ C(\mu,\sigma^2) \prod_{\substack{q<l \\ (q,l)\in\mathcal{G}}} \exp\left\{-\frac{(r_{ql}-\mu)^2}{2\sigma^2}\right\} \frac{1}{\sigma^2}. \tag{2.26}$$

2.3 A NEW MEASURE OF FUNCTIONAL CONNECTIVITY

One of the aims of our proposed model is to determine FC among brain regions. A direct way to learn FC is from the entries of the correlation matrix R. A positive value of $R_{p_1p_2}(q,l)$ indicates that the qth PC score for ROI p_1 and the lth PC score for ROI p_2 are positively correlated. This could also help us better understand the type of FC in terms of learning the shape of the corresponding PC functions. However, it is desirable to come up a summarized measure of FC for the purpose of ease of comparison. We proposed a new measure of FC based on the correlation matrix R by calculating the mutual information between the vectors of PC scores from different ROIs. Then the inter-regional relationship between region p_1 and region p_2 is measured via the mutual information between $\boldsymbol{\alpha_{ijp_1}}$ and $\boldsymbol{\alpha_{ijp_2}}$, $1 \leqslant p_1, p_2 \leqslant P$, which is defined as

$$MI\left(\boldsymbol{\alpha_{ijp_1}}, \boldsymbol{\alpha_{ijp_2}}\right) = H\left(\boldsymbol{\alpha_{ijp_1}}\right) + H\left(\boldsymbol{\alpha_{ijp_2}}\right) - H\left(\boldsymbol{\alpha_{ijp_1}}, \boldsymbol{\alpha_{ijp_2}}\right), \tag{2.27}$$

and $H(\mathbf{X}) = -E\left(\log(f_{\mathbf{X}}(\mathbf{x}))\right)$ is the entropy of $\mathbf{X}$. Here, $f_{\mathbf{X}}(\mathbf{x})$ is the probability density function of $\mathbf{X}$. If k dimensional random variable $\mathbf{X}$ follows multivariate normal distribution with variance covariance matrix Σ, then

$$H(\mathbf{X}) = \frac{k}{2} + \frac{k}{2}\log(2\pi) + \frac{1}{2}|\Sigma|[1]. \tag{2.28}$$

Based on the multivariate normal assumption of $\boldsymbol{\alpha_{ij}}$, any subvector of $\boldsymbol{\alpha_{ij}}$ is of the same structure with the correlation matrix being a submatrix of R. Then, the mutual information between $\boldsymbol{\alpha_{ijp_1}}$ and $\boldsymbol{\alpha_{ijp_2}}$ could be simplified as

$$MI\left(\boldsymbol{\alpha_{ijp_1}}, \boldsymbol{\alpha_{ijp_2}}\right) = -\frac{1}{2}\log|R_{\{p_1,p_2\}}|, \tag{2.29}$$

where

$$R_{\{p_1,p_2\}} = \begin{vmatrix} I_{Q_{p_1}} & R_{p_1p_2} \\ R_{p_1p_2}^T & I_{Q_{p_2}} \end{vmatrix}. \tag{2.30}$$

As shown in Model 2.29, FC between region p_1 and region p_2 is subject-independent, which is consistent with our model assumption.

Another good property of this new connectivity measure is that it is easy to derive the conditional FC between any two regions or response variables given the other variables in the model. Learning conditional FC among brain regions has high clinical importance and an application of it in research on depression is going to be presented in Section 2.5.2. We define the conditional FC between ROI p_1 and ROI p_2 by the partial mutual information between $\boldsymbol{\alpha_{ijp_1}}$ and $\boldsymbol{\alpha_{ijp_2}}$, $1 \leqslant p_1, p_2 \leqslant P$, which is given by

$$\begin{aligned}
& MI\left(\boldsymbol{\alpha_{ijp_1}}, \boldsymbol{\alpha_{ijp_2}} | \boldsymbol{\alpha_{ij\{1,\ldots,P\setminus p_1,p_2\}}}\right) \\
= \; & H\left(\boldsymbol{\alpha_{ijp_1}}, \boldsymbol{\alpha_{ij\{1,\ldots,P\setminus p_1,p_2\}}}\right) + H\left(\boldsymbol{\alpha_{ijp_2}}, \boldsymbol{\alpha_{ij\{1,\ldots,P\setminus p_1,p_2\}}}\right) \\
& -H\left(\boldsymbol{\alpha_{ij\{1,\ldots,P\setminus p_1,p_2\}}}\right) - H\left(\boldsymbol{\alpha_{ijp_1}}, \boldsymbol{\alpha_{ijp_2}}, \boldsymbol{\alpha_{ij\{1,\ldots,P\setminus p_1,p_2\}}}\right) \\
= \; & H\left(\boldsymbol{\alpha_{ij\{1,\ldots,P\setminus p_2\}}}\right) + H\left(\boldsymbol{\alpha_{ij\{1,\ldots,P\setminus p_1\}}}\right) \\
& -H\left(\boldsymbol{\alpha_{ij\{1,\ldots,P\setminus p_1,p_2\}}}\right) - H\left(\boldsymbol{\alpha_{ij}}\right) \\
= \; & \frac{1}{2}\log|R_{\{1\ldots,P\setminus p_2\}}| + \frac{1}{2}\log|R_{\{1\ldots,P\setminus p_1\}}| - \frac{1}{2}\log|R_{\{1\ldots,P\setminus p_1,p_2\}}| \\
& -\frac{1}{2}\log|R|, \qquad (2.31)
\end{aligned}$$

where $R_{\{1\ldots,P\setminus p_2\}}$, $R_{\{1\ldots,P\setminus p_1\}}$ and $R_{\{1\ldots,P\setminus p_1,p_2\}}$ are defined in the same way as $R_{\{p_1,p_2\}}$ in 2.30. Details of how Model 2.29 and 2.31 are derived are given in Appendix B.

FC obtained from this way ranges from 0 to infinity. A simple transformation can be applied, however, to obtain a normalized mutual information [42, 34, 63], with scores in the interval $[0, 1]$.

2.4 MODEL FITTING, SELECTION AND BAYESIAN INFERENCE

2.4.1 The Sampling Scheme

We implement MCMC methods for model estimation using Gibbs sampling. In this section, the MCMC Gibbs sampling scheme for our proposed model is outlined in a few stages. More details on the sampling scheme and the derivation of the conditional posterior distributions using the priors described in Section 2.2.2 on the model parameters are given in the Appendix

C. Suppose that $\Phi^0 = \{\boldsymbol{\theta_\mu}^0, \boldsymbol{\alpha}^0, \Theta^0, \mathbf{D}^0, R^0, \Sigma_\omega^0\}$ are the current draws for the parameters in the proposed model. The sampling scheme in the following iteration is as follows:

Step 1: Sample $\{\boldsymbol{\theta}_\mu^{\text{new}},\ \boldsymbol{\alpha}^{\text{new}}\}$ from $p(\boldsymbol{\theta}_\mu^{\text{new}},\ \boldsymbol{\alpha}^{\text{new}} \mid \Theta^0,\ \mathbf{D}^0,\ R^0,\ \Sigma_\omega^0,\ \mathbf{Y})$. The parameters $\{\boldsymbol{\theta}_{\mu_i}^{\text{new}}\}_{i=1}^{N}$ and $\{\boldsymbol{\alpha}_{ij}^{\text{new}}\}_{i=1}^{N}{}_{j=1}^{M_i}$ are sampled together to improve mixing and obtain a more efficient algorithm [12].

Step 1(a): Sample $\{\boldsymbol{\theta}_{\mu_i}^{\text{new}}\}_{i=1}^{N}$ from $p(\boldsymbol{\theta}_{\mu_i} \mid \Theta^0,\ \mathbf{D_i}^0,\ R^0,\ \Sigma_\omega^0,\ \mathbf{Y}_i) = \text{MVN}(\boldsymbol{\mu}_{\boldsymbol{\theta}_{\mu_i}|\cdot},\ \Sigma_{\boldsymbol{\theta}_{\mu_i}|\cdot})$ and $\boldsymbol{\mu}_{\boldsymbol{\theta}_{\mu_i}|\cdot}$ and $\Sigma_{\boldsymbol{\theta}_{\mu_i}|\cdot}$ are given by C.2 in Appendix C.

Step 1(b): Sample $\{\boldsymbol{\alpha}_{ij}^{\text{new}}\}_{i=1}^{N}{}_{j=1}^{M_i}$ from $p(\boldsymbol{\alpha}_{ij} \mid \boldsymbol{\theta}_{\mu_i}^{\text{new}},\ \Theta^0,\ \mathbf{D_i}^0,\ R^0,\ \Sigma_\omega^0,\ \mathbf{Y}_i) = \text{MVN}(\boldsymbol{\mu}_{\boldsymbol{\alpha}_{ij}|\cdot},\ \Sigma_{\boldsymbol{\alpha}_{ij}|\cdot})$ and $\boldsymbol{\mu}_{\boldsymbol{\alpha}_{ij}|\cdot}$ and $\Sigma_{\boldsymbol{\alpha}_{ij}|\cdot}$ are given by C.3 in Appendix C.

Step 2: Sample $\{\mathbf{D}_{iq}^{\text{new}}\}_{i=1}^{N}$ from $p(\mathbf{D}_{iq} \mid \boldsymbol{\alpha}_i^{\text{new}},\ \mathbf{D}_{i\{-q\}}\ R^0)$. Since the conditional posterior distribution is not standard, we use the griddy Gibbs sampler strategy [60, 6] to make the draws of D_i because it is easy to program. We draw each of the components of D_i one at a time and the only constraint of each D_{iq}, $q = 1, \ldots, Q$ is being positive. Details are discussed in Appendix C.

Step 3: Sample hyperparameters $\boldsymbol{\xi}_q^{\text{new}}$ from $p(\boldsymbol{\xi}_q \mid \mathbf{D}_{iq}^{\text{new}},\ \Lambda_q^0) = \text{N}(\mu_{\boldsymbol{\xi}_q|\cdot}, \sigma^2_{\boldsymbol{\xi}_q|\cdot})$, $q = 1, \ldots, Q$, where $\mu_{\boldsymbol{\xi}_q|\cdot}$ and $\sigma^2_{\boldsymbol{\xi}_q|\cdot}$ are given by C.7 in Appendix C. Sample hyperparameters Λ_q^{new} from $p(\Lambda_q \mid \mathbf{D}_{iq}^{\text{new}},\ \boldsymbol{\xi}_q^{\text{new}}) = \text{IG}(c_{\Lambda_q|\cdot},\ d_{\Lambda_q|\cdot})$, $q = 1, \ldots, Q$, where $c_{\Lambda_q|\cdot}$ and $d_{\Lambda_q|\cdot}$ are given by C.8 in Appendix C

Step 4: Sample $r_{q,l}^{\text{new}}$ from $p(r_{ql} \mid \boldsymbol{\alpha}^{\text{new}},\ \mathbf{D}^{\text{new}},\ R_{\{-ql\}})$. Each r_{ql}, $1 \leqslant q < l \leqslant Q$ and $(q, l) \in \mathcal{G}$, is drawn one at a time using the Metropolis-Hastings step. The positive definiteness of R constrains r_{ql} to an interval (l_{ql}, u_{ql}), and the proposal density could be the uniform density on this interval. Details are given in Appendix C.

Step 5: Sample hyperparameter μ^{new} from $p(\mu \mid R^{\text{new}},\ (\sigma^2)^0)$ and sample $(\sigma^2)^{\text{new}}$ from $p(\sigma^2 \mid R^{\text{new}},\ \mu^{\text{new}})$. Since the densities of μ and σ^2 are not conjugate and with an additional factor of the normalizing constant $C(\mu, \sigma^2)$, we will use Metropolis-Hastings step to update them, as detailed in Appendix C.

Step 6: Sample $\{(\sigma^2_{\omega,p})^{\text{new}}\}_{p=1}^{P}$ from $p(\sigma^2_{\omega,p} \mid \boldsymbol{\theta}_{\mu_p}^{\text{new}},\ \boldsymbol{\alpha}_p^{\text{new}},\ \Theta^0,\ \mathbf{Y}_p\ ) = \text{IG}(c_{\sigma^2_{\omega,p}|\cdot},\ d_{\sigma^2_{\omega,p}|\cdot})$ with $c_{\sigma^2_{\omega,p}|\cdot}$ and $d_{\sigma^2_{\omega,p}|\cdot}$ are given by C.14 in Appendix C.

Step 7: Sample $\{\Theta_p^{\text{new}}\}_{p=1}^{P}$ from $p(\text{vec}(\Theta_p)\ |\boldsymbol{\theta}_{\mu_p}^{\text{new}},\ \boldsymbol{\alpha}_p^{\text{new}},\ \sigma_{\omega,p}^{\text{new}},\ \mathbf{Y}_p) = \text{MVN}(\boldsymbol{\mu}_{\text{vec}(\Theta_p)}, \Sigma_{\text{vec}(\Theta_p)})$, where $\boldsymbol{\mu}_{\text{vec}(\Theta_p)}$ and $\Sigma_{\text{vec}(\Theta_p)}$ are given by C.15 in Appendix C. The matrix

Θ_p obtained in this step does not have to be orthogonal. We orthogonalize it in the way that firstly compute the pooled variance matrix

$$\hat{\Sigma}_{p,pool} = \frac{\sum M_i D_{ip}^{\text{new}} (D_{ip}^{\text{new}})'}{\sum M_i} \tag{2.32}$$

and then compute $\Theta_p^{\text{new}} \hat{\Sigma}_{p,pool} \Theta_p^{\text{new}}$ and reset Θ_p^{new} equal to the first Q_p eigenvectors.

2.4.2 Model Selection

It is particular important to identify the number of important PCs in functional PC analysis. As discussed by James, et al. (2000) [40] and Zhou, et al. 2008 [76], choosing to fit too many PCs can degrade the fit of all the PCs because they are not independent with each other. We follow a similar approach as discussed in [40, 76] to choose the number of PC functions which is to calculate the proportion of variability explained by each PC. Firstly, we apply the single-curve reduced rank mixed-effects model to each variable to select the number of PC for each variable. The model starts with one PC and then adding one more PC one at a time. If the proportion of the variability explained by the PCs already in the model does not change much after adding the new PC and the proportion of the variability explained by the newly added PC is much smaller than those already in the model, the process stops. The variability explained by the each PC is calculated by pooling the variance of the PC score across subjects. Then, we fit the multiple-curve reduced rank mixed-effects model using the chosen numbers of significant PCs from fitting the single-curve models. This procedure is tested in the simulated datasets and will be discussed in details in Section 2.5.1.

2.4.3 Posterior Inferences

In the MCMC procedure, samples from the joint posterior density of model parameters are produced which are then summarized for the purposes of inference. After a burn-in period, suppose L iterations of samples are produced, where $\{\boldsymbol{\theta_\mu}^l, \boldsymbol{\alpha}^l, \Theta^l, \mathbf{D}^l, R^l, \Sigma_\omega^l\}$, $1 \leqslant l \leqslant L$. The mean curve for subject i at a given time point t is obtained by averaging over the draws:

$$\hat{\boldsymbol{\mu}}_i(t) = \frac{1}{L} \sum_{l=1}^{L} \boldsymbol{\Phi}^T(t) \boldsymbol{\theta}_{\mu_i}^{[l]}. \tag{2.33}$$

Then $\hat{\boldsymbol{\mu}}_i(\centerdot)$ could be obtained by varying t on a fine grid on the interval $[0, T]$. The estimated functional response $\mathbf{f}_{ij}(t)$ for the jth trial of subject i at time t could be given by

$$\hat{\mathbf{f}}_{ij}(t) = \frac{1}{L}\sum_{l=1}^{L}(\boldsymbol{\Phi}^T(t)\boldsymbol{\theta}_{\mu_i}^{[l]} + \boldsymbol{\Phi}^T(t)\Theta^{[l]}\boldsymbol{\alpha}_{ij}^{[l]}). \tag{2.34}$$

The pointwise posterior credible intervals of the mean and individual functions are easy to obtain. For example, the credible interval for $\hat{\boldsymbol{\mu}}_i(\centerdot)$ at time t with approximate probability content $(1-\alpha)$ is the $\alpha/2$ and $1-\alpha/2$ quantiles of the L draws of $\boldsymbol{\Phi}^T(t)\boldsymbol{\theta}_{\mu_i}^{[l]}$.

The estimator of correlation matrix R with respect to the squared error loss function is the posterior mean estimator which is estimated from the MCMC samples in the standard manner. The estimated functional connectivity for region p_1 and p_2 could be obtained in a similar fashion where

$$\hat{MI}(\boldsymbol{\alpha}_{ijp_1},\ \boldsymbol{\alpha}_{ijp_2}) = \frac{1}{L}\sum_{l=1}^{L}(-\frac{1}{2}ln|R_{\{p_1,p_2\}}^{[l]}|), \tag{2.35}$$

and

$$\begin{aligned}\hat{MI} \quad & (\boldsymbol{\alpha}_{ijp_1},\ \boldsymbol{\alpha}_{ijp_2} \mid \boldsymbol{\alpha}_{ij\{1,\ldots,P\setminus p_1,p_2\}}) \\ = \quad & \frac{1}{L}\sum_{l=1}^{L}(\frac{1}{2}\log|R_{\{1\ldots,P\setminus p_2\}}^{[l]}| + \frac{1}{2}\log|R_{\{1\ldots,P\setminus p_1\}}^{[l]}| \\ & -\frac{1}{2}\log|R_{\{1\ldots,P\setminus p_1,p_2\}}^{[l]}| - \frac{1}{2}\log|R^{[l]}|).\end{aligned} \tag{2.36}$$

The posterior credible intervals of these connectivity measures could be obtained by determining the quantiles from the posterior draws as well.

2.5 SIMULATION AND REAL DATA ANALYSIS

2.5.1 Simulation Study

Here we illustrate the performance of the proposed MCMC sampling algorithm and model selection procedure in fitting the multivariate reduced rank model and the utility of the proposed model in estimating the connectivity coefficients through a small simulation study. In the study, the dataset was generated from the model

$$\begin{aligned} \mathbf{Y}_{ij}(t) &= \mathbf{B}_{ij}(t) + \boldsymbol{\epsilon}_{ij}(t) \\ &= \boldsymbol{\mu}_i(t) + \mathbf{f}(t)^T \boldsymbol{\alpha_{ij}} + \boldsymbol{\epsilon}_{ij}(t), \\ &\quad \boldsymbol{\epsilon}_{ij}(t) \sim \mathrm{N}\left(\mathbf{0}, \Sigma_\epsilon\right), \\ &\quad \boldsymbol{\alpha_{ij}} \sim \mathrm{N}\left(\mathbf{0}, \Sigma_i\right) \text{ with } \Sigma_i = D_i R D_i. \end{aligned} \tag{2.37}$$

Here $\mathbf{Y}_{ij}(t)$ represents fMRI responses from $P = 4$ ROIs and $\mathbf{f}(t)$ is defined as in 2.8 where $\mathbf{f}(t)^T = \mathrm{diag}\left(\mathbf{f}_1(t)^T, \ldots, \mathbf{f}_4(t)^T\right)$ and $\mathbf{f}_p(t) = \left(f_{p1}(t), \ldots, f_{pQ_p}(t)\right)^T$, $p = 1, \ldots, 4$, is the $Q_p \times 1$ vector of PC functions corresponding to the pth ROI evaluated at time t. We assume $Q_1 = 1$, $Q_2 = Q_3 = 2$ and $Q_4 = 1$. The PC functions are normalized such that $\int f_{pq}^2(t)dt = 1$, $p = 1, \ldots, 4$ and $q = 1, \ldots, Q_p$. The PC curves for each ROI are orthogonal to each other. The simulated data was generated mimicking a typical event-related fMRI design. In the simulation, we have $N = 50$ subjects ($i = 1, \ldots, 50$), $M_i = 20$ trials per subject ($j = 1, \ldots, 20$), $S = 7$ scans per trial ($t = 1, \ldots, 7$). The time points t are uniformly distributed over the unit interval.

We chose to fit the mean functions and the PC functions using orthogonal cubic B-splines with one internal knot at 0.5 quantile of the given time interval. Then, five basis functions was used and this was deemed more than sufficient to provide a good fit for a curve consisting of 7 points. Firstly, the single-curve model was applied to each ROIs response to choose the number of PC functions. A sequence of models with difference numbers of PC functions were considered, and the corresponding variances of PC scores for these models are given in Table 2.1. Based on these, we successfully picked 1 PC function for ROI 1 and ROI 4 and 2 PC functions for ROI 2 and ROI 3. Then, the MCMC fitting algorithm was applied

Table 2.1: Posterior estimated variance of PC scores for models with difference number of PCs in the simulation study.

Number of PC	1	2		3		
PC	1	1	2	1	2	3
$\Sigma_{ROI1,pool}$	0.136	0.136	0.003	0.136	0.004	0.002
$\Sigma_{ROI2,pool}$	0.100	0.103	0.016	0.103	0.016	0.002
$\Sigma_{ROI3,pool}$	0.084	0.086	0.015	0.086	0.015	0.002
$\Sigma_{ROI4,pool}$	0.124	0.124	0.004	0.125	0.004	0.003

to fit the multiple-curve model which was run for 10,000 iterations with a burn-in period of 5,000. The conditional posterior distributions on the parameters were derived using the prior distributions specied in Section 2.2.2. Convergence was monitored by initializing the chains at multiple random starting values and observing that the posterior distributions of parameters had converged to the same space [29]. Figure 2.2 presents the posterior estimate of the trial-based BOLD response (right-hand panel) along with the simulated underline BOLD response (left-hand panel) and simulated fMRI response (middle panel) from a randomly chosen subject. As observed from Figure 2.2, the recovered BOLD responses are close to the underlying BOLD responses. Figure 2.3 compares the true PC curves in generating the data (upper panel) and the fitted PC curves (lower panel) with 95% credible intervals and the estimated PC curves are very close to the true curves with fairly narrow credible intervals over time. We also compared the true and estimated correlation parameters in R which is presented in Table 2.2 and our fitting algorithm could successfully estimate the correlation matrix with the assigned common normal prior distribution.

To assess the utility of our proposed model in estimating FC, we compared FC obtained from our method and FC obtained from functional canonical correlation described at Section 1.2.4.2. Functional canonical correlation seeks to investigate which modes of variation between pairs of observed random curves are mostly associated with one another and the qualitative relationships between the two responses can be explored by examining the canonical variate weight functions. Since functional canonical correlation examines paired

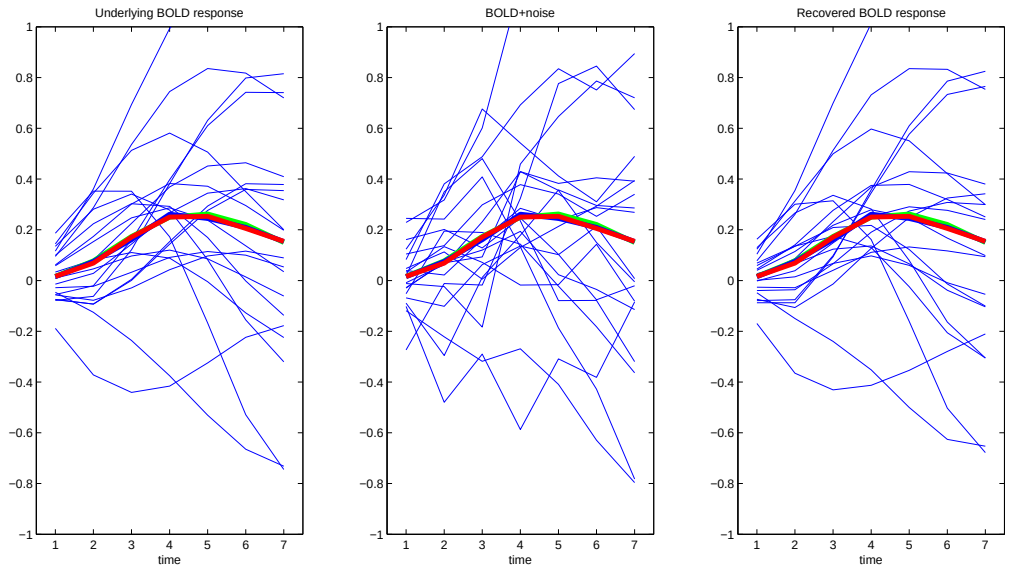


Figure 2.2: True and fitted PC functions in simulation data

From left to right: Underlying BOLD responses; Simulated fMRI response (underlying BOLD response plus error); Recovered BOLD responses; . Blue lines are trials. Heavy green, red and blue line are subject means for underlying BOLD responses, and simulated fMRI response, and recovered BOLD responses, respectively.

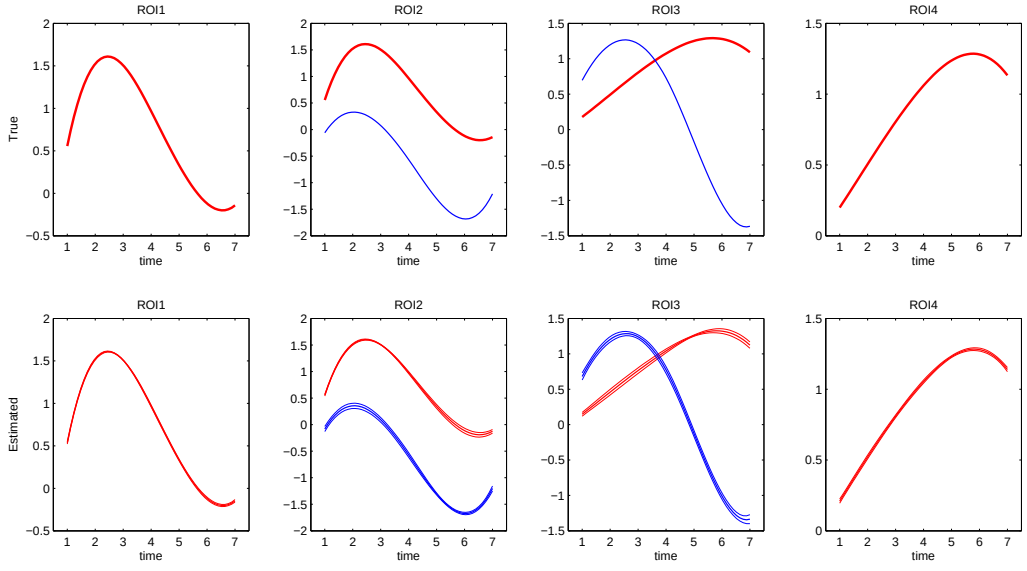


Figure 2.3: True and fitted PC functions in simulation data

Comparing true PC curves in generating the data (upper panel) and the fitted principal component curves (lower panel) with 95% credible intervals. The first PC function is plotted in red and the second one is plotted in blue

Table 2.2: Comparing true and estimated correlation parameters in R.

True	1.00	0.62	0.51	0.47	-0.27	0.16
Mean		0.64	0.50	0.49	-0.30	0.23
True		1.00	0	0.53	-0.05	0.32
Mean				0.54	-0.04	0.34
True		0	1.00	-0.05	-0.43	0.05
Mean				-0.05	-0.47	0.09
True				1.00	0	0.50
Mean						0.53
True				0	1.00	-0.25
Mean						-0.30
True						1.00

relationship, we only applied it to ROI 1 and ROI 2 in the simulated dataset. By using the Matlab Functional Connectivity Toolbox introduced in Chapter 1, we obtained the first canonical correlation between ROI 1 and ROI 2 was 0.47 and the weighted functions are presented in Figure 2.4. It should be noted that we calculated functional canonical correlation over all people at once, rather than person-by-person. The canonical correlation weighting function for ROI 1 puts most weight at early scans around scan 2 to 3 and the canonical correlation weighting function for ROI 2 puts most weight weight in the beginning and end of the trial but with opposite signs, indicating that early activities in ROI 1 is mostly related to early and later activities in ROI 2. By applying our proposed model in the simulation dataset, firstly, we found one PC function for ROI 1 and two PC functions for ROI 2 that characterized the major modes of variation in the trial-based activations. The PC functions are displayed in Figure 2.3. The PC function in ROI 1 accounts for the variability of activation throughout the whole trial but peaks around scan 2 to 3. The first PC function in RIO 2 accounts the variability of activation over the whole trial with the magnitude peaks around scan 2 to 3 as well and the second PC function in ROI 2 contributes to the variability over later scans. Secondly, from the estimated correlation matrix R, we found the score corresponding to the first PC function of ROI 1 is positively correlated with both the scores corresponding to the first (mean=0.64) and second (mean=0.50) PC functions of ROI 2. By observing all these, we could conclude that ROI 1 and ROI 2 have excitatory relationships especially over scan 2 to 3 and decreased activity in ROI 1, primarily at early scans, tends

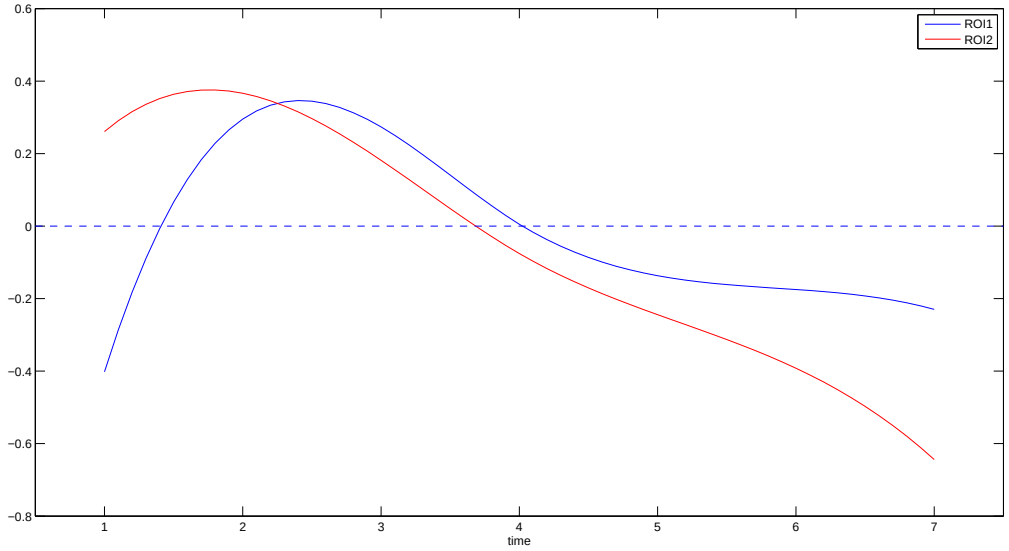


Figure 2.4: Estimated canonical weight functions

Estimated canonical variate weight functions for ROI 1 (in red) and ROI 2 (in blue) in the simulation data.

to lead more sustained activity in ROI 2. The summarized connectivity between ROI 1 and ROI 2 obtained from the new FC measure is 0.54. Comparing to the conclusion obtained by applying functional canonical correlation to ROI 1 and ROI 2, this new measure tends to provide a more accurate picture of the type of connectivity between the selected regions. In addition, we can further apply this new measure to calculate the conditional FC between ROI 1 and ROI 2 given activations from ROI 3 and ROI 4 being partialled out where other existing event-related FC measures could not be applied directly.

2.5.2 Real Data Analysis

As discussed in 2.1, we applied our model to data from a psychiatric neuroscience experiment designed to test differences in relationships of functioning among brain regions between unipolar major depressed subjects and never-depressed healthy controls [68]. In particular, we were interested in examining relationships among a candidate mechanism of three ROIs during emotional information processing, activity in left amygdala, a brain region linked with recognizing the emotionality of information and generating emotional reactions, and

two cortical regions: dorsolateral prefrontal cortex (DLPFC), a brain region associated with executive control and initiating emotion regulation, and the rostral portion Brodmann's area 24 (BA24), a regional associated with processing self-relevant information and emotion regulation, particularly inhibition of the amygdala [71]. Relationships among these areas have been hypothesized by several researchers. For example, if executive control is necessary for emotion regulation and, specially, if the DLPFC initiates a process of emotion regulation that results in inhibition of limbic regions such as the amygdala [52], sustained emotional reactivity might result from decreased DLPFC function. Indeed, the increased and sustained amygdala activity has been linked to decreased DLPFC activity in healthy [20] and depressed individuals [67]. But it is not hypothesized that there would be a negative correlation between DLPFC and amygdala activity on tasks involving emotional stimuli; if the DLPFC is important for emotion regulation, it would become active following amygdala activity, leading to a positive correlation and the sustained amygdala activity would be explained by decreased strength of such coupling relationship [68]. There are no direct relationships between DLPFC and amygdala. Rather, it might be mediated by connections from the ventromedial regions such as BA24 to the amygdala [31]. The regulatory communication of DLPFC with amygdala is impaired in depression, possibly through decreased functioning in mediation from ventromedial regions such as BA24.

Thirty patients with major depressive disorder and 28 healthy control subjects participated in 60 slow event-related trials and completed tasks designed to provoke limbic reactivity to emotional stimuli in depression. During each trial, participants viewed a fixation cue (1 sec) followed by a positive, negative, or neutral word (200 msec), followed by a mask (row of Xs; 10.8 sec). Participants pushed a button for whether the word was relevant, somewhat relevant, or not relevant to them or their lives, as quickly and accurately as they could. We examined results from 20 trials using negative words. The fMRI data were generated per 1.5 sec. More details on experimental design and subject characteristics can be found in Siegle, et al. (2007) [68].

Briefly, data were pre-processed in several steps of motion correction, detrending within blocks, outliers rescaled, cross registered to an reference brain, and spatially smoothed. The reference brain was then transformed into Talairach space using AFNI (Cox, 1996) [17] to

extract anatomical masks. The left amygdala was identified anatomically in the functional data. DLPFC and BA24 were identified empirically. Specifically, the DLPFC regions were derived from a group × scan random effects voxelwise ANOVA in the time course of response to putting digits in order on a sorting-task. Identified regions were with significant group × scan interactions, as shown in Siegle, et al. (2007), Figure 5 [68]. BA24 were identified which differentiated depressed and healthy individuals in the time course of response to negative words, in the context of a group × valence × scan ANOVA (shown in Siegle, et al. 2007 [68]). Figure 2.1 shows two subjects' 20 negative word trial trajectories from the three ROIs along with subjects' averaged activation trajectories. Within-trial regional activations were normalized to the first scan regional BOLD activation within each trial so the resulting trial trajectories should begin at zeros. We discarded the first scan in each trial and 7 scans' BOLD response per trial were analyzed.

We ran the multiple-curve reduced rank model on the groups (healthy vs depressed) separately. The mean and the PC functions were fitted using five orthogonal cubic B-splines with one internal knot at the 0.5-quantile of the given time interval. The MCMC sampling algorithm described in Section 2.4 was applied for 10,000 iterations with a burn in of 5000 iterations. Convergence was monitored using iteration plots. Following the method described in Section 2.4.2, firstly, we run the single-curve reduced rank model on each ROI separately, and the corresponding variances of PC scores are given in Table 2.3. We chose two PC functions for each ROI so that the model was fitted with $Q_1 = Q_2 = Q_3 = 2$ for each group. But the second PC is less important than the first one given that the variability explained by the first PC is around three times of the variability explained by the second one.

If Figure 2.5, we plot the estimated two PC curves from each ROI for depressed and healthy groups, respectively. The first PC curves correspond to a level shift from the mean curve and the magnitude of shifting increases with time. The second PC curves change sign during the time period and correspond to opposite departures from the mean at the beginning and the end of the trial. The effect on subject's mean curve is displayed in Figure 2.6 by adding and subtracting a multiple of each of the PC curves. We did not know the shape of the PC curves prior to the analysis, but it turns out that the shape of the estimated PC curves are rather similar across regions. This may be caused by the high level

Table 2.3: Posterior estimated variance of PC scores for models with difference number of PCs in the fMRI study.

Group	Number of PC	1	2		3		
	PC	1	1	2	1	2	3
Depressed	$\Sigma_{AMYG,pool}$	0.434	0.441	0.184	0.442	0.188	0.080
	$\Sigma_{DLPFC,pool}$	0.335	0.340	0.125	0.341	0.128	0.059
	$\Sigma_{BA24,pool}$	0.305	0.308	0.111	0.310	0.114	0.051
Control	$\Sigma_{AMYG,pool}$	0.412	0.417	0.143	0.419	0.149	0.079
	$\Sigma_{DLPFC,pool}$	0.348	0.354	0.146	0.355	0.149	0.064
	$\Sigma_{BA24,pool}$	0.323	0.328	0.139	0.328	0.141	0.054

of autocorrelation in the data and the limited observations per trial.

Figure 2.7 presents the data on six individual trials from one subject, along with the estimated trajectories and 95% pointwise posterior credible intervals. The dashed lines indicate subject's mean trajectories and circles indicate actual observed data points for each trial. As observed, the variations in the trail waveforms could be characterized very well by the PC functions and the pointwise credible intervals for the individual functions are fairly narrow over time.

The estimated correlation parameters in R from the two groups are presented in Table 2.4 and 2.5 along with the corresponding 95% posterior credible intervals in brackets. The correlation matrix estimated from the healthy group implies that any two of the three ROIs are positively correlated in such a way that the first PC scores from the corresponding two ROIs are positively correlated with each other and the second PC scores from the corresponding two ROIs are positively correlated with each other, which suggests the excitatory influence between them. The correlation matrix estimated from the depressed group exhibits similar patterns, but with some exceptions. Any two of the three ROIs are positively correlated in the same way that the first PC scores from the corresponding two ROIs are positively correlated with each other as well as the second PC scores. However, the correlation parameters are smaller than those estimated from the healthy group. In addition, the

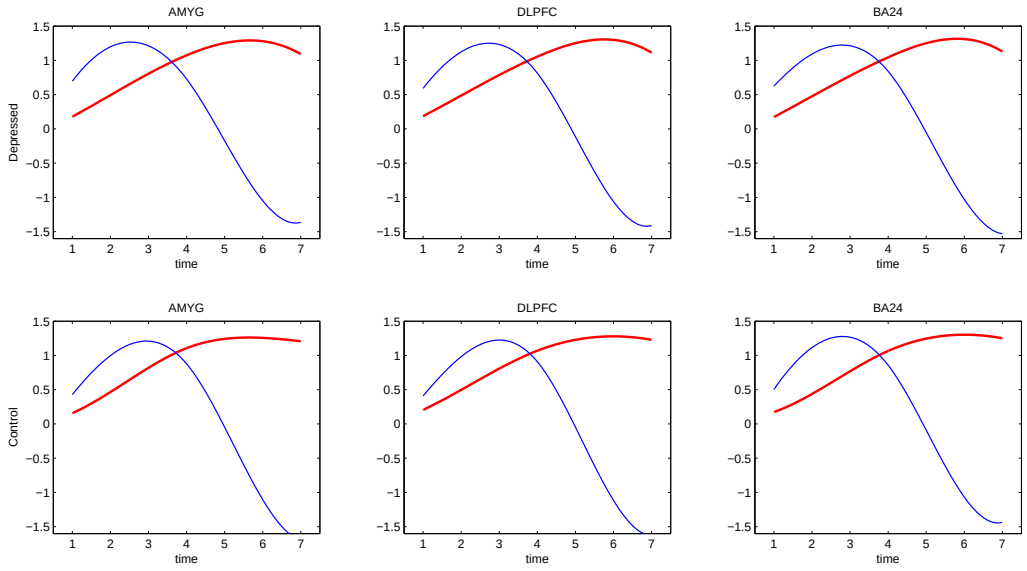


Figure 2.5: Estimated PC functions from three ROIs

Estimated PC functions from amygdala, DLPFC and BA24 for the depressed (upper panel) and the healthy (lower panel) groups. The first PC function is plotted in red and the second one is plotted in blue.

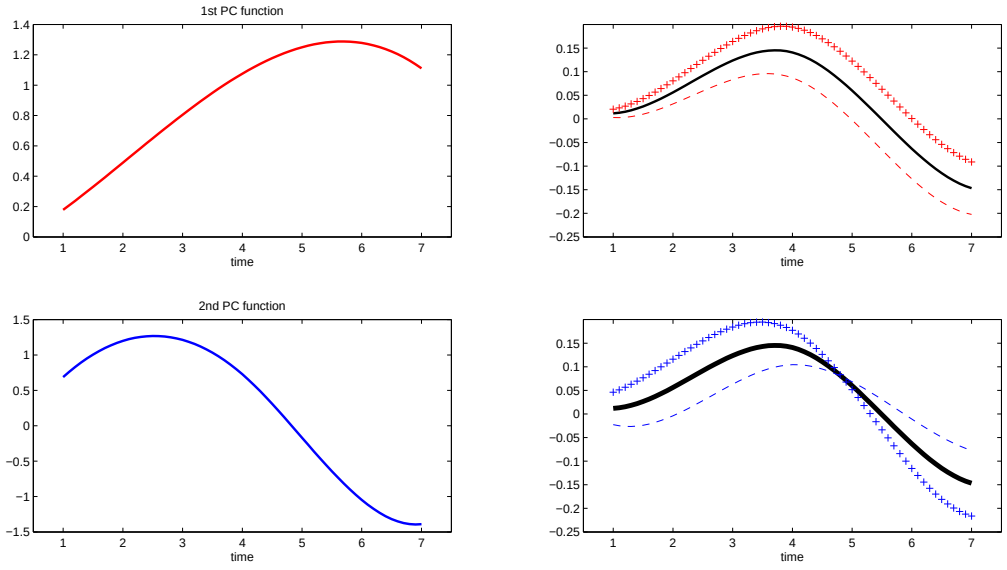


Figure 2.6: Effect of PC functions on the mean

Estimated PC functions (left panel) and the effects on the mean curves of adding (plus signs) and subtracting (minus signs) a multiple of each of the PCs (right panel).

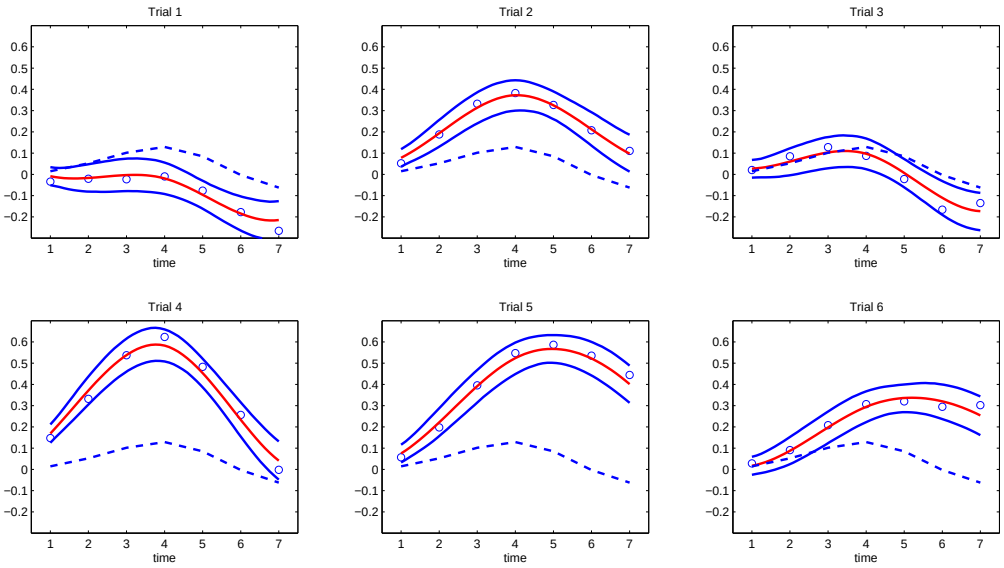


Figure 2.7: Estimated individual trajectories

Estimated individual trajectories with 95% posterior credible intervals for six trials from one subject. Circles indicate actual data points and dashed lines indicate subject's mean trajectory.

first PC score from amygdala and the second PC score from BA24 are negatively correlated (mean=-0.16, PCI=[-0.26 -0.06]), which suggests a positive score on the second PC of BA24 tends to be associated with a negative score on the first PC scores of amygdala. In other words, when BA24 has higher early activity, respectively lower, than the mean response, amygdala tends to have lower activity, respectively higher, than the mean response primally at later scans. Thus, in depressed subjects, early activity in BA24 tends to damp amygdala activity especially at later scans, and vice versa. This is not found significantly from the healthy group.

We applied the new measure of FC introduced in Section 2.3 to amygdala and DLPFC (see Figure 2.8). Estimated FC between amygdala and DLPFC sharply reduces in the depressed group and the difference between the two groups is -0.28 with posterior credible interval being [-0.41, -0.14]. Conditional FC between amygdala and DLPFC given the effects from BA24 being partialled out slightly reduces in the depressed group but sharply reduces in the healthy group, suggesting that BA24 activity mediates the functional relationship between amygdala and DLPFC strongly in the healthy group but weakly in the

Table 2.4: Posterior estimates of correlation parameters in R for the depressed group with corresponding 95% posterior credible intervals in brackets.

		AMYG		DLPFC		BA24	
		PC 1	PC 2	PC1	PC2	PC1	PC2
AMYG	PC1	1.00	0	0.16	-0.07	0.45	-0.16
				(0.07 0.25)	(-0.15 0.03)	(0.37 0.53)	(-0.26 -0.06)
	PC2	0	1.00	-0.03	0.21	0.06	0.45
				(-0.12 0.06)	(0.11 0.30)	(-0.01 0.14)	(0.36 0.53)
DLPFC	PC1	0.16	-0.03	1.00	0	0.53	0.09
						(0.47 0.60)	(0.00 0.17)
	PC2	-0.07	0.21	0	1.00	-0.04	0.66
						(-0.11 0.06)	(0.61 0.72)
BA24	PC1	0.45	0.06	0.53	-0.04	1.00	0
	PC2	-0.16	0.45	0.09	0.66	0	1.00

Table 2.5: Posterior estimates of correlation parameters in R for the healthy group with corresponding 95% posterior credible intervals in brackets.

		AMYG		DLPFC		BA24	
		PC 1	PC 2	PC1	PC2	PC1	PC2
AMYG	PC1	1.00	0	0.39	-0.02	0.50	-0.03
				(0.29 0.48)	(-0.10 0.11)	(0.42 0.57)	(-0.13 0.07)
	PC2	0	1.00	0.05	0.44	0.04	0.57
				(-0.07 0.16)	(0.32 0.54)	(-0.06 0.15)	(0.46 0.66)
DLPFC	PC1	0.39	0.05	1.00	0	0.68	0.02
						(0.61 0.74)	(-0.08 0.09)
	PC2	-0.02	0.44	0	1.00	0.06	0.79
						(-0.01 0.14)	(0.73 0.83)
BA24	PC1	0.50	0.04	0.68	0.06	1.00	0
	PC2	-0.03	0.57	0.02	0.79	0	1.00

depressed group. Thus, the findings further support the hypothesis that depressed subjects display decreased coupling relationship between amygdala and DLPFC than healthy subjects and the impaired communication between them in depression is possibly through decreased functioning in mediation from ventromedial regions such as BA24.

2.6 DISCUSSION

To date, the most common approaches applied to FC analysis for event-related designs are peak correlation [59] and functional canonical correlation [68], which are introduced in Section 1.2.4. Peak correlation tends to capture the coupling relationship of peaks in activitation in pairs of brain regions associated with discrete events, whereas functional canonical correlation seeks to investigate which modes of variation between pairs of observed post-stimulus responses are most associated with one another. While both methods have been successfully applied to FC analysis in various contexts, they also have some disadvantages. Firstly, both of them are correlation approaches which are not appropriate for inferences and simulation. Secondly, they both deal solely with bivariate cases from which the conditional dependence can not be derived directly. Moreover, the determination of BOLD responses and FC are implemented separately when applying both methods to event-related fMRI designs. It is desirable to encompass simultaneous determination of BOLD responses and FC since the level of smoothing could impact estimates of connectivity. We have proposed a Bayesian model tailored for smoothing and exploring FC of multiple brain regions in slow event-related fMRI designs, which has several key advantages over the alternative approaches. First of all, our approach consists of a full probability model and the use of a Bayesian paradigm provides a range of flexible inferences. The MCMC estimation procedure produces samples from the joint posterior distribution of all of the model parameters, which facilitates estimation of and inferences about response functions and FC parameters. In addition, our proposed model allows incorporation of multivariate responses, where the conditional bivariate inter-regional relationship given other regions' responses or covariates of interest being partialled out could be derived easily. Exploring conditional FC has high clinical importance in depression re-

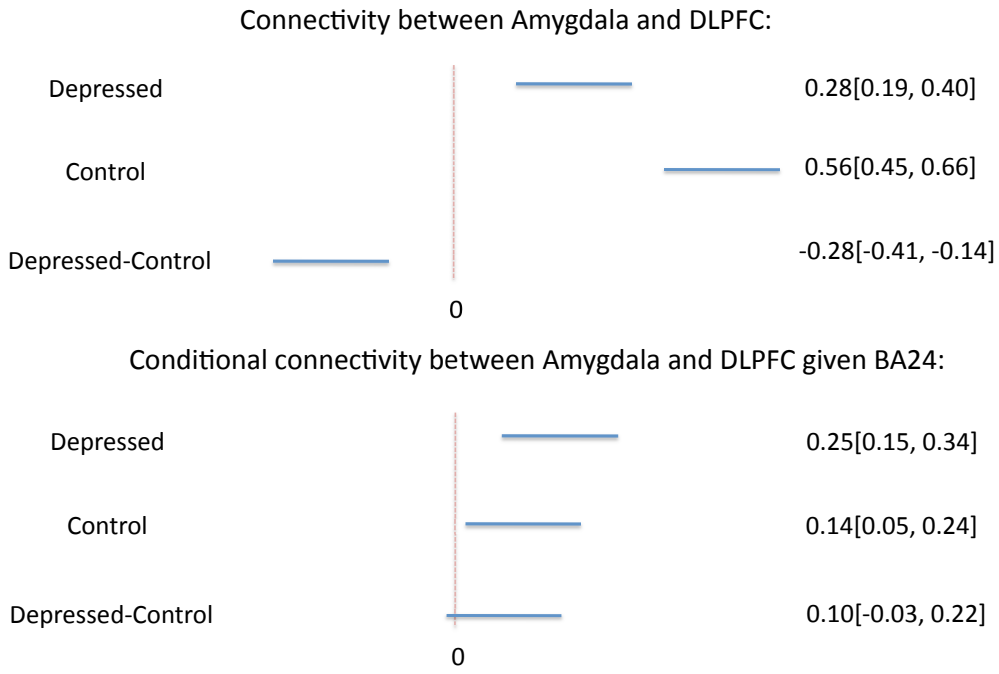


Figure 2.8: Estimated FC between amygdala and DLPFC

Estimated FC (upper panel) and conditional FC (lower panel) with 95% posterior credible intervals between amygdala and DLPFC for depressed and healthy groups.

search. Also, the new approach provides a unified framework to obtain neural activation inferences as well as FC inferences, rather than treating them as distinct analytical objectives. Trial waveforms are modeled while accounting for inter-regional relationships. This model is also constructed to exploit important characteristics of slow event-related fMRI designs. It allows for nonstationarities in stimulus-locked trial waveforms and individual variations in activations. Another incremental utility of our proposed technique above and beyond the existing methods for exploring inter-regional relationships is that it could explore a more concrete picture of the communication structure.

We demonstrated the usefulness of our proposed methodology, including the fitting algorithm and prior distribution specifications, by applying this model in simulations. It has shown that the proposed methodology could successfully estimate the model parameters and capture the underlying BOLD responses. As we mentioned before, our aim in the new approach is two fold. Besides the aim of uncovering BOLD responses, we also seek to identify the inter-regional correlations. We applied both functional canonical correlation and our method to the simulated data with known inter-regional relationships and demonstrated that our new approach provide more information about the type of FC.

The utility of this methodology was demonstrated by application to a real-life psychiatric neuroscience experiment looking at the functional relationships among multiple brain regions. In particular, we examined relationships among amygdala, DLPFC and BA24. The results suggested that depressed subjects exhibits decreased strength of coupling relationship between amygdala and DLPFC as hypothesized (e.g. [68]). Moreover, in depressed subjects, it is likely that the early activity in BA24 tends to damp amygdala activity especially at later scans, suggesting the inhibitory effects of BA24 on amygdala for emotion regulation. We also calculated the conditional FC between amygdala and DLPFC by covaring out the effects from BA24 and found that the conditional FC between amygdala and DLPFC slightly reduces in the depressed group but strongly reduces in the healthy group. This observations suggested that the impaired relationship between amygdala and DLPFC in depressed subjects may potentially due to the decreased functioning in mediation FC from BA24 (e.g. [43]).

One important issue not addressed in this approach is to allow individual subject dif-

ferences in FC. This could enable researchers to obtain subject-level indices which is of considerable interest to determine whether individual variation in FC estimates are predictive of clinical measures of depression, such as Beck Depression Inventory (BDI), a clinical measure of depressive severity. To accommodate this, we could allow subject differences in the correlation matrix R. Then, in the common correlation model, all correlations $r_{ql,i}$ for subject i are assumed to follow a common normal distribution where

$$r_{ql,i} \sim N(\mu_i, \sigma_i^2),\ 1 \leqslant q < l \leqslant Q \text{ and } (q,l) \in \mathcal{G}, \tag{2.38}$$

and

$$\begin{aligned} \mu_i &\sim N(\delta_\mu, \nu_\mu^2), \\ \log(\sigma_i) &\sim N(\delta_\sigma, \nu_\sigma^2). \end{aligned} \tag{2.39}$$

This could be easily implemented from the proposed model.

An important methodological question in implementing this multiple-curve reduced rank model is the choice of number and placement of the knots for the basis spline functions. We have done some sensitivity analysis by choosing different numbers and spacing of knots and found out that our choice of $K = 5$ was more than sufficient to provide a good fit for a curve consisting of 7 points. Also, we could allow different basis functions for subject's mean curve and PC curves instead of assuming that they have the same underling level of smoothness.

Another area for further research is to implement an automatic procedure to determine the number of PC functions and estimate the model parameters simultaneously under the same disciplined framework. Our current approach requires a separate model selection procedure that the single-curve reduced rank model is applied to each variable to select the number of PC functions. This would result in extensive computation as more response variables or time points involved. One way to accomplish this is through reversible-jump MCMC by regarding each PC function as a one-dimensional space and employing birth-death moves in the reversible jump methodology among different spaces. We intend to implement this procedure into our approach in the futher.

APPENDIX A

LIST OF FUNCTIONS OF THE MATLAB FUNCTIONAL CONNECTIVITY TOOLBOX

Cross-correlation :

[xcorr,xcorr_lag]=lagged(y_1,y_2,lag)

Inputs:

y_1 and y_2 (required): Time-series of brain region's response. They should be column-wise.

lag (optional): A positive integer indicates the maximum lag of the cross-correlation. The default value is 10.

Outputs:

xcorr : A vector of estimated cross-correlations at different lags.

xcorr_lag : A vector of integers indicates the lags corresponding to the estimated cross-correlations.

Partial cross-correlation :

[Pxcorr,Pxcorr_lag]=Plagged(y,lag)

Inputs:

y (required): A matrix of brain regions' response. Each region's time-series is column-wise.

lag (optional): A positive integer indicates the maximum lag of the partial cross-correlation. The default value is 10.

Outputs:

Pxcorr : A matrix of estimated partial cross-correlations at different lags. Pxcorr(i,j,:) corresponds to the partial cross-correlations between region i and j.

xcorr_lag : A vector of integers indicates the lags corresponding to the estimated partial cross-correlations.

Cross-coherence :

[Coh,lambda_Coh]=coh(y_1,y_2,l,sr)

Inputs:

y_1 and y_2 (required): Time-series of brain region's response. They should be column-wise.

l (optional): A positive integer specifies the length of the cross-coherence. The default value equals the length of the input time-series.

sr (optional): A positive number specifies the sampling rate (in Hz). The default value is 1.

Outputs:

Coh : A vector of estimated cross-coherence at different frequencies.

lambda_Coh : A vector of positive values indicates the frequencies corresponding to the estimated cross-coherence.

Partial cross-coherence :

[PCoh,lambda_PCoh]=Pcoh(y,l,sr)

Inputs:

y (required): A matrix of brain regions' response. Each region's time-series is column-wise.

l (optional): A positive integer specifies the length of the partial cross-coherence. The default value equals the length of the input time-series.

sr (optional): A positive number specifies the sampling rate (in Hz). The default value is 1.

Outputs:

PCoh : A matrix of estimated partial cross-coherence at different frequencies. PCoh(i,j,:) corresponds to the partial cross-coherence between region i and j.

lambda_PCoh : A vector of positive values indicates the frequencies corresponding to the estimated partial cross-coherence.

Mutual information :

phi=mutualinf(y_1,y_2,sr,lambdamin,lambdamax)

Inputs:

y_1 and y_2 (required): Time-series of brain region's response. They should be column-wise.

sr (optional): A positive number specifies the sampling rate (in Hz). The default value is 1.

lambdamin,lambdamax : [lambdamin, lambdamax] specifies the frequency boundaries (in Hz) within which to integrate the information. The default values are 0 and $\frac{1}{2}$ respectively.

Outputs:

phi : Estimated mutual information which is between 0 and 1.

Partial mutual information :

Pphi=Pmutualinf(y,sr,lambdamin,lambdamax)

Inputs:

y (required): A matrix of brain regions' response. Each region's time-series is column-wise.

sr (optional): A positive number specifies the sampling rate (in Hz). The default value is 1.

lambdamin,lambdamax : [lambdamin, lambdamax] specifies the frequency boundaries (in Hz) within which to integrate the information. The default values are 0 and $\frac{1}{2}$ respectively.

Outputs:

Pphi : A matrix of estimated partial mutual information. Pphi(i,j) corresponds to the partial mutual information between region i and j.

Peak correlation :

pcorr=corrpeak(y_1,y_2,scan,k,norder)

Inputs:

y_1 and y_2 (required): Time-series of brain region's response. They should be column-wise.

scan (required): A positive integer specifies the number of scans per trial.

k (optional): An integer variable specifies the number of B-spline basis functions when smoothing the trial-based curve. The default value is calculated by $min(1/4 \times \text{scan}, 35) + \text{norder}$.

norder (optional): An integer specifies the order of B-spline functions when smoothing the trial-based curve. The default order is 4, and this defines splines that are piecewise cubic.

Outputs:

pcorr : Estimated peak correlation which is between -1 and 1.

Peak correlation (fast event-related design) :

pcorr=corrpeak_fast(y_1,y_2,st,k,norder)

Inputs:

y_1 and y_2 (required): Time-series of brain region's response. They should be column-wise.

st (required): Column-wise vector of stimulus time-series.

k (optional): An integer variable specifies the number of B-spline basis functions when smoothing the trial-based curve. The default value is calculated by $min(1/4 \times \text{scan}, 35) + \text{norder}$.

norder (optional): An integer specifies the order of B-spline functions when smoothing the trial-based curve. The default order is 4, and this defines splines that are piecewise cubic.

Outputs:

pcorr : Estimated peak correlation which is between -1 and 1.

Scan of interest correlation :

scancorr=corrscan(y_1,y_2,scan,scanofinterest,k,norder)

Inputs:

y_1 and y_2 (required): Time-series of brain region's response. They should be column-wise.

scan (required): A positive integer specifies the number of scans per trial.

scanofinterest (require): A positive integer within [0, scan] specifies the scan number at which something interesting is hypothesized to occur.

k (optional): An integer variable specifies the number of B-spline basis functions when smoothing the trial-based curve. The default value is calculated by $min(1/4 \times \text{scan}, 35) + \text{norder}$.

norder (optional): An integer specifies the order of B-spline functions when smoothing the trial-based curve. The default order is 4, and this defines splines that are piecewise cubic.

Outputs:

scancorr : Estimated correlation of activity at the scan of interest which is between -1 and 1.

Scan of interest correlation (fast event-related design) :

scancorr=corrscan_fast($\mathbf{y_1}$,$\mathbf{y_2}$,st,scanofinterest,k,norder)

Inputs:

$\mathbf{y_1}$ and $\mathbf{y_2}$ (required): Time-series of brain region's response. They should be column-wise.

st (required): Column-wise vector of stimulus time-series.

scanofinterest (require): A positive integer within [0, scan] specifies the scan number at which something interesting is hypothesized to occur.

k (optional): An integer variable specifies the number of B-spline basis functions when smoothing the trial-based curve. The default value is calculated by $min(1/4 \times \text{scan}, 35) + \text{norder}$.

norder (optional): An integer specifies the order of B-spline functions when smoothing the trial-based curve. The default order is 4, and this defines splines that are piecewise cubic.

Outputs:

scancorr : Estimated correlation of activity at the scan of interest which is between -1 and 1.

Functional canonical correlation :

[cr,u,v,lambda]=ccorr_cv(y_1,y_2,scan,k,norder,lambda_max)

Inputs:

y_1 and y_2 (required): Time-series of brain region's response. They should be column-wise.

scan (required): A positive integer specifies the number of scans per trial.

k (optional): An integer variable specifies the number of B-spline basis functions when smoothing the trial-based curve. The default value is calculated by $min(1/4 \times \text{scan}, 35) + \text{norder}$.

norder (optional): An integer specifies the order of B-spline functions when smoothing the trial-based curve. The default order is 4, and this defines splines that are piecewise cubic.

lambda_max : [0, lambda_max] specifies the boundaries within which to choose the smoothing parameter of the weight functions.

Outputs:

cr : Estimated functional canonical correlation which is between -1 and 1.

u and v : Estimated weight functions for the two regions respectively.

lambda : The smoothing parameter which is chosen via cross-validation.

Functional canonical correlation (fast event-related design) :

[cr,u,v,lambda]=ccorr_cv_fast(y_1,y_2,st,k,norder,lambda_max)

Inputs:

y_1 and y_2 (required): Time-series of brain region's response. They should be column-wise.

st (required): Column-wise vector of stimulus time-series.

k (optional): An integer variable specifies the number of B-spline basis functions when smoothing the trial-based curve. The default value is calculated by $min(1/4 \times \text{scan}, 35) + \text{norder}$.

norder (optional): An integer specifies the order of B-spline functions when smoothing the trial-based curve. The default order is 4, and this defines splines that are piecewise cubic.

lambda_max : [0, lambda_max] specifies the boundaries within which to choose the smoothing parameter of the weight functions.

Outputs:

cr : Estimated functional canonical correlation which is between -1 and 1.

u and v : Estimated weight functions for the two regions respectively.

lambda : The smoothing parameter which is chosen via cross-validation.

Get all Measures :

s=getallconnectivityinds(y_1,y_2,scan,tr,justeventmeas,scanofinterest)

Inputs:

y_1 and y_2 (required): Time-series of brain region's response. They should be column-wise.

scan (required): A positive integer specifies the number of scans per trial.

tr (optional): A positive number specifies the number of seconds between samples (scans). The default value is 1.5.

justeventmeas (optional): If 1, only trial-based measures should be computed. It 0, only whole time-series measures should be computed. The default value is 0.

scanofinterest (optional): A specific scan number within trials at which something interesting is hypothesized to occur.

Outputs:

s : A cell array with all the estimated FC measures assigned to it.

Low pass filter :

y_lp=lowpass(y,sr,f,order)

Inputs:

y (required): Time-series of brain region's response.

sr (required): A positive number specifies the sampling frequency (in Hz).

f (required): A positive number specifies the cut-off frequency (in Hz) which should be between 0 and half of the sampling frequency.

order (optional): An integer specifies the order of the filter. The default value is 10.

Outputs:

y_lp : The new time-series after being low pass filtered.

Smoothing (resting state design) :

Y_sm=smoothing_whole(Y,P,N,scan,auto,nloop,k,norder)

Inputs:

Y (required): A three-dimensional matrix storing brain regions' response.

Its dimensions are:

1. time ... size = no. of scans per trial
2. variables ... size = no. of regions
3. replications ... size = no. of subjects × no. of trials per subject

P (required): A positive integer indicates the number of regions.

N (required): A positive integer indicates the number of subjects.

scan (required): A positive integer specifies the number of scans per trial.

auto (optional): A number within [-1 1] specifies the autocorrelation of the noise. The default value is 0.7.

k (optional): An integer variable specifies the number of B-spline basis functions when smoothing the data. The default value is calculated by $min(1/4 \times \text{scan}, 35) + \text{norder}$, where n equals the length of the input time-series.

norder (optional): An integer specifies the order of B-spline functions when smoothing the data. The default order is 4, and this defines splines that are piecewise cubic.

Outputs:

Y_sm : A three-dimensional matrix storing smoothed brain regions' response.

Smoothing (event-related design) :

Y_sm=smoothing(Y,P,N,M,scan,auto,nloop,k,norder)

Inputs:

Y (required): A three-dimensional matrix storing brain regions' response.

Its dimensions are:

1. time ... size = no. of scans per trial
2. variables ... size = no. of regions
3. replications ... size = no. of subjects × no. of trials per subject

P (required): A positive integer indicates the number of regions.

N (required): A positive integer indicates the number of subjects.

M (required): A vector of positive integers with size N×1 where the ith position of M indicates the number of trials of subject i.

scan (required): A positive integer specifies the number of scans per trial.

auto (optional): A number within [-1 1] specifies the autocorrelation of the noise. The default value is 0.7.

k (optional): An integer variable specifies the number of B-spline basis functions when smoothing the data. The default value is calculated by $min(1/4 \times \text{scan}, 35) + \text{norder}$, where n equals the length of the input time-series.

norder (optional): An integer specifies the order of B-spline functions when smoothing the data. The default order is 4, and this defines splines that are piecewise cubic.

Outputs:

Y_sm : A three-dimensional matrix storing smoothed brain regions' response.

Pre-whitening :

y_prewhiten=prewhiten(y,auto)

Inputs:

y (required): Input time-series data.

auto (optional): A number within [-1 1] specifies the autocorrelation of the noise. The default value is 0.7.

Outputs:

y_prewhiten : The new time-series after being pre-whitened.

Interpolation: y_scale=gsresample(y,origHz,newHz)

Inputs:

y (required): Input time-series data.

origHz (required): Original sampling rate.

newHz (required): New sampling rate.

Outputs:

y_scale : The new time-series after being rescaled.

APPENDIX B

DERIVATIONS OF THE NEW FUNCTIONAL CONNECTIVITY MEASURE

$$
\begin{aligned}
& MI\left(\boldsymbol{\alpha_{ijp_1}}, \boldsymbol{\alpha_{ijp_2}}\right) \\
= \; & H\left(\boldsymbol{\alpha_{ijp_1}}\right) + H\left(\boldsymbol{\alpha_{ijp_2}}\right) - H\left(\boldsymbol{\alpha_{ijp_1}}, \boldsymbol{\alpha_{ijp_2}}\right) \\
= \; & \frac{Q_{p_1}}{2} + \frac{Q_{p_1}}{2}\log(2\pi) + \log|\mathrm{diag}(\mathbf{D}_{ip_1})| \\
& + \frac{Q_{p_2}}{2} + \frac{Q_{p_2}}{2}\log(2\pi) + \log|\mathrm{diag}(\mathbf{D}_{ip_2})| \\
& - \frac{Q_{p_1}+Q_{p_2}}{2} - \frac{Q_{p_1}+Q_{p_2}}{2}\log(2\pi) \\
& - \frac{1}{2}\log\left|\mathrm{diag}(\mathbf{D}_{i,\{p_1,p_2\}})\right|\left|R_{\{p_1,p_2\}}\right|\left|\mathrm{diag}(\mathbf{D}_{i,\{p_1,p_2\}})\right| \\
= \; & -\frac{1}{2}\log\left|R_{\{p_1,p_2\}}\right| \qquad \text{(B.1)}
\end{aligned}
$$

where

$$
\mathbf{D}_{i,\{p_1,p_2\}} = \left(\mathbf{D}'_{ip_1} \; \mathbf{D}'_{ip_2}\right), \qquad \text{(B.2)}
$$

and

$$
\left|R_{\{p_1,p_2\}}\right| = \begin{vmatrix} I & R_{p_1p_2} \\ R^T_{p_1p_2} & I \end{vmatrix}. \qquad \text{(B.3)}
$$

$$
\begin{aligned}
& MI\left(\boldsymbol{\alpha}_{ijp_1}, \boldsymbol{\alpha}_{ijp_2} | \boldsymbol{\alpha}_{ij\{1,\ldots,P\setminus p_1,p_2\}}\right) \\
= \; & H\left(\boldsymbol{\alpha}_{ij\{1,\ldots,P\setminus p_2\}}\right) + H\left(\boldsymbol{\alpha}_{ij\{1,\ldots,P\setminus p_1\}}\right) \\
& -H\left(\boldsymbol{\alpha}_{ij\{1,\ldots,P\setminus p_1,p_2\}}\right) - H\left(\boldsymbol{\alpha}_{ij}\right) \\
= \; & \frac{1}{2}\log\left|\mathrm{diag}(\mathbf{D}_{i,\{1\ldots,P\setminus p_2\}})\right|\left|R_{\{1\ldots,P\setminus p_2\}}\right|\left|\mathrm{diag}(\mathbf{D}_{i,\{1\ldots,P\setminus p_2\}})\right| \\
& +\frac{1}{2}\log\left|\mathrm{diag}(\mathbf{D}_{i,\{1\ldots,P\setminus p_1\}})\right|\left|R_{\{1\ldots,P\setminus p_1\}}\right|\left|\mathrm{diag}(\mathbf{D}_{i,\{1\ldots,P\setminus p_1\}})\right| \\
& -\frac{1}{2}\log\left|\mathrm{diag}(\mathbf{D}_{i,\{1\ldots,P\setminus p_1,p_2\}})\right|\left|R_{\{1\ldots,P\setminus p_1,p_2\}}\right|\left|\mathrm{diag}(\mathbf{D}_{i,\{1\ldots,P\setminus p_1,p_2\}})\right| \\
& -\frac{1}{2}\log\left|\mathrm{diag}(\mathbf{D}_i)\right|\left|R\right|\left|\mathrm{diag}(\mathbf{D}_i)\right| \\
= \; & \frac{1}{2}\log|R_{\{1\ldots,P\setminus p_2\}}| + \frac{1}{2}\log|R_{\{1\ldots,P\setminus p_1\}}| - \frac{1}{2}\log|R_{\{1\ldots,P\setminus p_1,p_2\}}| \\
& -\frac{1}{2}\log|R|. \qquad \text{(B.4)}
\end{aligned}
$$

APPENDIX C

DETAILS OF MCMC SAMPLER

Step 1: Generating $\{\boldsymbol{\theta}_\mu,\ \boldsymbol{\alpha}\}$ conditional on $\{\Theta,\ \mathbf{D},\ R,\ \Sigma_\omega,\ \mathbf{Y}\}$. This is done in two steps, using the factorization

$$\begin{aligned} p(\boldsymbol{\theta}_\mu,\ \boldsymbol{\alpha} \mid \Theta,\ \mathbf{D},\ R,\ \Sigma_\omega,\ \mathbf{Y}) \ &\propto\ p(\boldsymbol{\theta}_\mu \mid \Theta,\ \mathbf{D},\ R,\ \Sigma_\omega,\ \mathbf{Y}) \\ &\quad \times p(\boldsymbol{\alpha} \mid \boldsymbol{\theta}_\mu,\ \Theta,\ \mathbf{D},\ R,\ \Sigma_\omega,\ \mathbf{Y}). \end{aligned} \tag{C.1}$$

Step 1(a): Generating $\{\boldsymbol{\theta}_{\mu_i}\}_{i=1}^{N}$ conditional on $\{\Theta,\ \mathbf{D}_i,\ R,\ \Sigma_\omega,\ \mathbf{Y}_i\}$. Let $\mathbf{Y}_{ij} = (\mathbf{Y}_{ij}^T(1), \ldots, \mathbf{Y}_{ij}^T(S))^T$ and $\Phi^b = (\Phi^b(1)^T, \ldots, \Phi^b(S)^T)^T$. The conditional posterior distributions of the $\{\boldsymbol{\theta}_{\mu_i}\}_{i=1}^{N}$ are independent MVN $\left(\boldsymbol{\mu}_{\boldsymbol{\theta}_{\mu_i}|\mathbf{.}},\ \Sigma_{\boldsymbol{\theta}_{\mu_i}|\mathbf{.}}\right)$, where

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{\theta}_{\mu_i}|\mathbf{.}} \ &=\ \Sigma_{\boldsymbol{\theta}_{\mu_i}|\mathbf{.}} \left(\sum_{j=1}^{M_i} (\Phi^b)'(\mathrm{diag}(\mathbf{D}_i) R \mathrm{diag}(\mathbf{D}_i))^{-1} \mathbf{Y}_{ij}\right) \quad \text{and} \\ \Sigma_{\boldsymbol{\theta}_{\mu_i}|\mathbf{.}} \ &=\ \left(M_i (\Phi^b)'(\mathrm{diag}(\mathbf{D}_i) R \mathrm{diag}(\mathbf{D}_i))^{-1} \Phi^b + \frac{1}{c_\theta} I\right)^{-1}. \end{aligned} \tag{C.2}$$

Step 1(b): Generating $\{\boldsymbol{\alpha}_{ij}\}_{i=1\ j=1}^{N\ M_i}$ conditional on $\{\boldsymbol{\theta}_{\mu_i},\ \Theta,\ \mathbf{D}_i,\ R,\ \Sigma_\omega,\ \mathbf{Y}_{ij}\}$. The conditional posterior distributions of the $\{\boldsymbol{\alpha}_{ij}\}_{i=1\ j=1}^{N\ M_i}$ are independent MVN $\left(\boldsymbol{\mu}_{\boldsymbol{\alpha}_{ij}|\mathbf{.}},\ \Sigma_{\boldsymbol{\alpha}_{ij}|\mathbf{.}}\right)$, where

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{\alpha}_{ij}|\mathbf{.}} \ &=\ \Sigma_{\boldsymbol{\alpha}_{ij}|\mathbf{.}} \left(\sum_{t=1}^{S} \Theta'(\Phi^b(t))' \Sigma_\omega^{-1} (\mathbf{Y}_{ij}(t) - \Phi^b(t) \boldsymbol{\theta}_{\mu_i}\right) \quad \text{and} \\ \Sigma_{\boldsymbol{\alpha}_{ij}|\mathbf{.}} \ &=\ \left(\sum_{t=1}^{S} \Theta'(\Phi^b(t))' \Sigma_\omega^{-1} \Phi^b(t) \Theta + (\mathrm{diag}(\mathbf{D}_i) R \mathrm{diag}(\mathbf{D}_i))^{-1}\right)^{-1}. \end{aligned} \tag{C.3}$$

Step 2: Generating each of the components of $\{\mathbf{D}_i\}_{i=1}^N$ one at a time. The conditional density of $\{\mathbf{D}_{iq}\}_{i=1}^N$ is

$$\begin{aligned} & p(\mathbf{D}_{iq} \mid \boldsymbol{\alpha}_i,\ \mathbf{D}_{i\{-q\}},\ R) \\ \propto\ & p(\boldsymbol{\alpha}_i \mid \mathbf{D}_i,\ R) p(\mathbf{D}_{iq}) \\ \propto\ & |\mathbf{D}_{iq}|^{-M_i-1} \exp\left\{-\frac{1}{2}\left[\frac{(S_i)_{qq}(R^{-1})_{qq}}{\mathbf{D}_{iq}^2} + 2\frac{1}{\mathbf{D}_{iq}}\sum_{q'\neq q}^{Q}(S_i)_{qq'}(R^{-1})_{qq'}\frac{1}{\mathbf{D}_{iq'}}\right]\right\} \\ & \exp\left\{-\frac{(\log(\mathbf{D}_{iq}) - \xi_q)^2}{2\Lambda_q}\right\}, \end{aligned} \tag{C.4}$$

where

$$S_i = \sum_{j=1}^{M_i} \boldsymbol{\alpha}_{ij}\boldsymbol{\alpha}_{ij}'. \tag{C.5}$$

Since this conditional posterior distribution is not standard and hard to sample from, we implement the griddy Gibbs strategy to generate $\mathbf{D}_{iq}$ according to [60, 6]. The Griddy Gibbs sampler constructs an approximation to the conditional density numerically, by evaluating the posterior density on a grid over the support of values for $\mathbf{D}_{iq}|\boldsymbol{\alpha}_i, \mathbf{D}_{i\{-q\}}, R$ (that is, keeping the conditioning parameters $\boldsymbol{\alpha}_i, \mathbf{D}_{i\{-q\}}, R$ constant). To sample from a general density function p with cumulative distribution function P , we can take a drawing u from a uniform distribution, and apply the inverse CDF P^{-1} to arrive at a drawing $\mathbf{D}_{iq} = P^{-1}(u)$ from the original distribution. We put down the grid $d_1, \ldots, d_g$ uniformly distributed and force the density of either end of the grid is less than 5% of the maximum value, which is

$$\begin{aligned} p(d_1 \mid \boldsymbol{.}) &< \max\left(p(d_1 \mid \boldsymbol{.}),\ \ldots,\ p(d_g \mid \boldsymbol{.})\right) \times 5\% \quad \text{and} \\ p(d_g \mid \boldsymbol{.}) &< \max\left(p(d_1 \mid \boldsymbol{.}),\ \ldots,\ p(d_g \mid \boldsymbol{.})\right) \times 5\%. \end{aligned} \tag{C.6}$$

Step 3: Generating $\boldsymbol{\xi}_q$, $q = 1, \ldots, Q$, conditional on $\{\mathbf{D}_{iq},\ \Lambda_q\}$. The conditional posterior distribution of $\boldsymbol{\xi}_q$ is $\mathrm{N}(\mu_{\boldsymbol{\xi}_q|\boldsymbol{.}}, \sigma^2_{\boldsymbol{\xi}_q|\boldsymbol{.}})$ where

$$\begin{aligned} \mu_{\boldsymbol{\xi}_q|\boldsymbol{.}} &= \sigma^2_{\boldsymbol{\xi}_q|\boldsymbol{.}}\left(\sum_{i=1}^{N}\frac{\mathbf{D}_{iq}}{\Lambda_q}\right), \quad \text{and} \\ \sigma^2_{\boldsymbol{\xi}_q|\boldsymbol{.}} &= \left(\frac{N}{\Lambda_q} + \frac{1}{c_\xi}\right)^{-1}. \end{aligned} \tag{C.7}$$

Generating Λ_q, $q = 1, \ldots, Q$, conditional on $\{\mathbf{D}_{iq},\ \boldsymbol{\xi}_q\}$. The conditional posterior distribution of Λ_q is IG($c_{\Lambda_q|\bullet}$, $d_{\Lambda_q|\bullet}$) where

$$\begin{aligned} c_{\Lambda_q|\bullet} &= \frac{N}{2} + 1 \quad \text{and} \\ d_{\Lambda_q|\bullet} &= \frac{1}{2}\sum_{i=1}^{N} \left(\mathbf{D}_{iq} - \boldsymbol{\xi}_q\right)^2. \end{aligned} \tag{C.8}$$

Step 4: Generating each r_{ql}, $1 \leqslant q < l \leqslant Q$ and $(q, l) \in \mathcal{G}$ one at a time. The conditional density of r_{ql} is

$$\begin{aligned} p(r_{ql} \mid \boldsymbol{\alpha},\ \mathbf{D},\ R_{\{-ql\}}) &\propto p(\boldsymbol{\alpha} \mid \mathbf{D},\ R)p(r_{ql}) \\ &\propto |R|^{-\frac{\sum_{i=1}^{N} M_i}{2}} \exp\left\{-\frac{1}{2}\text{trace}(R^{-1}B)\right\} \exp\left\{-\frac{(r_{ql}-\mu)^2}{2\sigma^2}\right\} \\ &\quad I\left\{R \in \mathscr{R}^Q, R \in M(\mathcal{G})\right\}, \end{aligned} \tag{C.9}$$

with $B = \sum_{i=1}^{N}\sum_{j=1}^{M_i} \text{diag}(\mathbf{D}_i)^{-1}\boldsymbol{\alpha}_{ij}\boldsymbol{\alpha}_{ij}'\text{diag}(\mathbf{D}_i)$. Let $R(r)$ be the matrix obtained from R by changing the (q, l)th correlation to r and let $f(r) = |R(r)|$. According to [49], the set of values of r_{ql} preserving the positiveness of R when $R \in M(\mathcal{G})$ are those in the interval (l_{ql}, u_{ql}) which is determined by the roots of quadratic function $ar^2 + br + c$, where $a = [f(1) + f(-1) - 2f(0)]/2$, $b = [f(1) - f(-1)]/2$, and $c = f(0)$. The new proposal of r_{ql} is generated from Uniform(l_{ql}, u_{ql}). **Step 5:** Generating μ conditional on $\{R,\ \sigma^2\}$ and generating σ^2 conditional on $\{R,\ \mu\}$. The conditional densities of μ and σ^2 are similar to the conjugate densities but with an additional factor of the normalizing constant $C(\mu, \sigma^2)$. The normalizing constant C is proportional to the integral of a product of univariate normal densities restricted to a constrained space and could be evaluated using the strategy of importance sampling [49]: generate $r_{ql}^m \sim \text{N}(\mu, \sigma^2)$ for $q < l$ and $(q, l) \in \mathcal{G}$, $m = 1, \ldots, M$, define $R^m = (r_{ql}^m)$ and use

$$\widehat{C}(\mu, \sigma^2) = \frac{1}{M}\sum_{m=1}^{M} I\left\{R^m \in \mathscr{R}^Q, R \in M(\mathcal{G})\right\}. \tag{C.10}$$

The proposal density for μ is $\mathrm{N}(\mu_{pro},\ \sigma^2_{pro})$ and

$$\begin{aligned}\mu_{pro} &= \sigma^2_{pro}\ (\sum_{\substack{q<l\\(q,l)\in\mathcal{G}}} \frac{r_{ql}}{\sigma^2}), \qquad \text{(C.11)}\\ \sigma^2_{pro} &= (\frac{Q_1(Q_2+\ldots+Q_P)+\cdots+Q_{P-1}Q_P}{\sigma^2}+\frac{1}{c_\mu})^{-1}. \qquad \text{(C.12)}\end{aligned}$$

The proposal density for σ^2 is $\mathrm{IG}(c_{pro},\ d_{pro})$ and

$$\begin{aligned}c_{pro} &= \frac{Q_1(Q_2+\ldots+Q_P)+\cdots+Q_{P-1}Q_P}{2}+1 \quad \text{and}\\ d_{pro} &= \sum_{\substack{q<l\\(q,l)\in\mathcal{G}}} \frac{(r_{ql}-\mu)^2}{2}. \qquad \text{(C.13)}\end{aligned}$$

Step 6: Generating $\sigma^2_{\omega,p}$, $p=1,\ldots,P$, conditional on $\{\boldsymbol{\theta}_{\mu_p},\ \boldsymbol{\alpha}_p,\ \Theta_p,\ \mathbf{Y}_p\}$. The conditional posterior distributions is $\mathrm{IG}(c_{\sigma^2_{\omega,p}|\bullet},\ d_{\sigma^2_{\omega,p}|\bullet})$ where

$$\begin{aligned}c_{\sigma^2_{\omega,p}|\bullet} &= S\frac{\sum_{i=1}^N M_i}{2}+1 \quad \text{and} \quad \text{and}\\ d_{\sigma^2_{\omega,p}|\bullet} &= \sum_{i=1}^N\sum_{j=1}^{M_i}\sum_{t=1}^{S}\left(\mathbf{Y}_{ijp}(t)-\boldsymbol{\phi}^T(t)\boldsymbol{\theta}_{\boldsymbol{\mu}_{ip}}-\boldsymbol{\phi}^T(t)\Theta_p\boldsymbol{\alpha}_{ijp}\right). \qquad \text{(C.14)}\end{aligned}$$

Step 7: Generating $\mathrm{vec}(\Theta_p)$ conditional on $\{\boldsymbol{\theta}_{\mu_p},\ \boldsymbol{\alpha}_p,\ \sigma_{\omega,p},\ \mathbf{Y}_p\}$. The conditional posterior distribution is $\mathrm{MVN}(\boldsymbol{\mu}_{\mathrm{vec}(\Theta_p)},\ \Sigma_{\mathrm{vec}(\Theta_p)})$ where

$$\begin{aligned}\boldsymbol{\mu}_{\mathrm{vec}(\Theta_p)} &= \Sigma_{\Theta_p}\left(\sum_{i=1}^N\sum_{j=1}^{M_i}\sum_{t=1}^{S}\frac{(I_{Q_p}\otimes\boldsymbol{\phi}^T(t))'\boldsymbol{\alpha}_{ijp}\left(\mathbf{Y}_{ijp}(t)-\boldsymbol{\phi}^T(t)\boldsymbol{\theta}_{\boldsymbol{\mu}_{ip}}\right)}{\sigma^2_{\omega,p}}\right) \quad \text{and}\\ \Sigma_{\mathrm{vec}(\Theta_p)} &= \left(\sum_{i=1}^N\sum_{j=1}^{M_i}\sum_{t=1}^{S}\frac{(I_{Q_p}\otimes\boldsymbol{\phi}^T(t))'\boldsymbol{\alpha}_{ijp}\boldsymbol{\alpha}'_{ijp}(I_{Q_p}\otimes\boldsymbol{\phi}^T(t))}{\sigma^2_{\omega,p}}+\frac{1}{c_\Theta}I\right)^{-1}. \qquad \text{(C.15)}\end{aligned}$$

BIBLIOGRAPHY

[1] N. Ahmed, D. Gokhale. Entropy expressions and their estimators for multivariate distributions. *IEEE Transactions on Information Theory*, 35(3):688–692, 1989.

[2] H. J. Aizenstein, M. A. Butters, M. Wu, L. M. Mazurkewicz, V. A. Stenger, P. J. Gianaros, J. T. Becker, C. F. Reynolds, C. S. Carter. Altered functioning of the executive control circuit in late-life depression: episodic and persistent phenomena. *The American Journal of Geriatric Psychiatry: Official Journal of the American Association for Geriatric Psychiatry*, 17(1):30–42, 2009.

[3] A. Anand, Y. Li, Y. Wang, J. Wu, S. Gao, L. Bukhari, V. P. Mathews, A. Kalnin, M. J. Lowe. Activity and connectivity of brain mood regulating circuit in depression: A functional magnetic resonance study. *Biological Psychiatry*, 57(10):1079–1088, 2005.

[4] A. Anand, Y. Li, Y. Wang, J. Wu, S. Gao, L. Bukhari, V. P. Mathews, A. Kalnin, M. J. Lowe. Antidepressant effect on connectivity of the mood-regulating circuit: an FMRI study. *Neuropsychopharmacology*, 30(7):1334–1344, 2005.

[5] L. Astolfi, F. Cincotti, D. Mattia, C. Babiloni, F. Carducci, A. Basilisco, P. Rossini, S. Salinari, L. Ding, Y. Ni, B. He, F. Babiloni. Assessing cortical functional connectivity by linear inverse estimation and directed transfer function: simulations and application to real data. *Clinical Neurophysiology*, 116(4):920–932, 2005.

[6] J. Barnard, R. McCulloch, X. Meng. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10(4):1281–1312, 2000.

[7] B. Biswal, F. Z. Yetkin, V. M. Haughton, J. S. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, 34(4):537–541, 1995.

[8] L. Breiman. *Classification and Regression Trees*. Chapman & Hall, 1993.

[9] B. A. Brumback, J. A. Rice. Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association*, 93(443):961–976, Wrze. 1998. ArticleType: primary_article / Full publication date: Sep., 1998 / Copyright 1998 American Statistical Association.

[10] R. B. Cattell. The three basic factor-analytic research designs-their interrelations and derivatives. *Psychological Bulletin*, 49(5):499–520, 1952.

[11] C. Chen, J. Hsieh, Y. Wu, P. Lee, S. Chen, D. M. Niddam, T. Yeh, Y. Wu. Mutual-information-based approach for neural connectivity during self-paced finger lifting task. *Human Brain Mapping*, 29(3):265–280, 2008.

[12] S. Chib, B. P. Carlin. On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing*, 9:17–26, 1998.

[13] P. Ciuciu, J. B. Poline, G. Marrelec, J. Idier, C. Pallier, H. Benali. Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment. *IEEE Transactions on medical imaging*, 22(10):12351251, 2003.

[14] C. M. Clark, R. Kessler, M. S. Buchsbaum, R. A. Margolin, H. H. Holcomb. Correlational methods for determining regional coupling of cerebral glucose metabolism: a pilot study. *Biological Psychiatry*, 19(5):663–78, 1984.

[15] J. D. Cohen, F. Tong. NEUROSCIENCE: the face of controversy. *Science*, 293(5539):2405–2407, 2001.

[16] D. Cordes, V. M. Haughton, K. Arfanakis, J. D. Carew, P. A. Turski, C. H. Moritz, M. A. Quigley, M. E. Meyerand. Frequencies contributing to functional connectivity in the cerebral cortex in "Resting-state" data. *American Journal of Neuroradiology*, 22(7):1326–1333, 2001.

[17] R. W. Cox. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29(3):162–173, 1996.

[18] O. David, D. Cosmelli, K. J. Friston. Evaluation of different measures of functional connectivity using a neural mass model. *Neuroimage*, 21(2):659–673, 2004.

[19] I. Dimatteo, C. R. Genovese, R. E. Kass. Bayesian curve-fitting with free-knot splines. *Biometrika*, 88(4):1055–1071, 2001.

[20] F. Dolcos, G. McCarthy. Brain systems mediating cognitive interference by emotional distraction. *The Journal of Neuroscience*, 26(7):2072–2079, 2006.

[21] E. Formisano, F. Esposito, F. D. Salle, R. Goebel. Cortex-based independent component analysis of fMRI time series. *Magnetic Resonance Imaging*, 22(10):1493–1504, 2004.

[22] M. D. Fox, A. Z. Snyder, J. M. Zacks, M. E. Raichle. Coherent spontaneous activity accounts for trial-to-trial variability in human evoked brain responses. *Nature Neuroscience*, 9(1):23–25, 2006.

[23] P. Fransson. Spontaneous low-frequency BOLD signal fluctuations: An fMRI investigation of the resting-state default mode of brain function hypothesis. *Human Brain Mapping*, 26(1):15–29, 2005.

[24] K. J. Friston. Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 2(1-2):56–78, 1994.

[25] K. J. Friston, C. Buechel, G. R. Fink, J. Morris, E. Rolls, R. J. Dolan. Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, 6(3):218–229, 1997.

[26] K. J. Friston, C. D. Frith, P. Fletcher, P. F. Liddle, R. S. J. Frackowiak. Functional topography: multidimensional scaling and functional connectivity in the brain. *Cerebral Cortex*, 6(2):156–164, 1996.

[27] K. J. Friston, C. D. Frith, P. F. Liddle, R. S. J. Frackowiak. Functional connectivity: the principal-component analysis of large(PET) data sets. *Journal of cerebral blood flow and metabolism*, 13(1):5–14, 1993.

[28] K. J. Friston, P. Jezzard, R. Turner. Analysis of functional MRI time-series. *Human Brain Mapping*, 1(2):153–171, 1994.

[29] A. Gelman. *Bayesian Data Analysis*. CRC Press, 2004.

[30] A. Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1:1–19, 2006.

[31] H. T. Ghashghaei, H. Barbas. Pathways for emotion: interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. *Neuroscience*, 115(4):1261–1279, 2002.

[32] P. J. Gianaros, S. W. Derbyshire, J. C. May, G. J. Siegle, M. A. Gamalo, J. R. Jennings. Anterior cingulate activity correlates with blood pressure during stress. *Psychophysiology*, 42(6):627–635, 2005.

[33] G. H. Glover. Deconvolution of impulse response in Event-Related BOLD fMRI1. *Neuroimage*, 9(4):416–429, 1999.

[34] C. Granger, J. L. Lin. Using the mutual information coeffcient to identify lags in nonlinear models. *Journal of Time Series Analysis*, 15(4):371–384, 1994.

[35] M. D. Greicius, K. Supekar, V. Menon, R. F. Dougherty. Resting-State functional connectivity reflects structural connectivity in the default mode network. *Cerebral Cortex*, strona bhn059, 2008.

[36] M. Hampson, B. S. Peterson, P. Skudlarski, J. C. Gatenby, J. C. Gore. Detection of functional connectivity using temporal correlations in MR images. *Human Brain Mapping*, 15(4):247–262, 2002.

[37] B. Horwitz. The elusive concept of brain connectivity. *NeuroImage*, 19(2):466–470, 2003.

[38] B. Horwitz, R. Duara, S. I. Rapoport. Intercorrelations of glucose metabolic rates between brain regions: application to healthy males in a state of reduced sensory input. *Journal of Cerebral Blood Flow and Metabolism*, 4(4):484–499, 1984.

[39] S. A. Huettel, A. W. Song, G. McCarthy. *Functional Magnetic Resonance Imaging*. Sinauer Associates, 2004.

[40] G. M. James, T. J. Hastie, C. A. Sugar. Principal component models for sparse functional data. *Biometrika*, 87(3):587–602, 2000.

[41] J. Jeong, J. C. Gore, B. S. Peterson. Mutual information analysis of the EEG in patients with alzheimer's disease. *Clinical Neurophysiology*, 112(5):827–835, 2001.

[42] H. Joe. Relative entropy measures of multivariate dependence. *Journal of the American Statistical Association*, 84(405):157–164, 1989.

[43] H. Johansen-Berg, D. A. Gutman, T. E. J. Behrens, P. M. Matthews, M. F. S. Rushworth, E. Katz, A. M. Lozano, H. S. Mayberg. Anatomical connectivity of the subgenual cingulate region targeted with deep brain stimulation for Treatment-Resistant depression. *Cerebral Cortex*, 18(6):1374–1383, 2008.

[44] G. S. Kimeldorf, G. Wahba. A correspondence between bayesian estimation on stochastic processes and smoothing by splines. *The Annals of Mathematical Statistics*, 41(2):495–502, 1970.

[45] H. Koshino, P. A. Carpenter, N. J. Minshew, V. L. Cherkassky, T. A. Keller, M. A. Just. Functional connectivity in an fMRI working memory task in high-functioning autism. *NeuroImage*, 24(3):810–821, 2005.

[46] N. Lange, S. L. Zeger. Non-linear fourier time series analysis for human brain mapping by functional magnetic resonance imaging. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46(1):1–29, 1997.

[47] N. A. Lazar. *The Statistical Analysis of Functional MRI Data*. Springer, wydanie 1, Lip. 2008.

[48] D. A. Leopold, Y. Murayama, N. K. Logothetis. Very slow activity fluctuations in monkey visual cortex: Implications for functional brain imaging. *Cerebral Cortex*, 13(4):422–433, 2003.

[49] J. C. Liechty, M. W. Liechty, P. Muller. Bayesian correlation estimation. *Biometrika*, 91(1):1–14, 2004.

[50] M. J. Lowe, B. J. Mock, J. A. Sorenson. Functional connectivity in single and multislice echoplanar imaging using Resting-State fluctuations. *Neuroimage*, 7(2):119–132, 1998.

[51] A. W. MacDonald, J. D. Cohen, V. A. Stenger, C. S. Carter. Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, 288(5472):1835–1838, 2000.

[52] H. S. Mayberg, M. Liotti, S. K. Brannan, S. McGinnis, R. K. Mahurin, P. A. Jerabek, J. A. Silva, J. L. Tekell, C. C. Martin, J. L. Lancaster, P. T. Fox. Reciprocal Limbic-Cortical function and negative mood: Converging PET findings in depression and normal sadness. *The American Journal of Psychiatry*, 156(5):675–682, 1999.

[53] M. Murias, S. J. Webb, J. Greenson, G. Dawson. Resting state cortical connectivity reflected in EEG coherence in individuals with autism. *Biological Psychiatry*, 62(3):270–273, 2007.

[54] S. H. Na, S. H. Jin, S. Y. Kim, B. J. Ham. EEG in schizophrenic patients: mutual information analysis. *Clinical Neurophysiology*, 113(12):1954–1960, 2002.

[55] J. C. Rajapakse, F. Kruggel, J. M. Maisog, D. Y. V. Cramon. Modeling hemodynamic response for analysis of functional MRI time-series. *Human Brain Mapping*, 6(4):283–300, 1998.

[56] J. Ramsay, B. W. Silverman. *Functional Data Analysis*. Springer, Czerw. 1997.

[57] J. A. Rice, B. W. Silverman. Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(1):233–243, 1991.

[58] J. A. Rice, C. O. Wu. Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, 57(1):253–259, 2001.

[59] J. Rissman, A. Gazzaley, M. D'Esposito. Measuring functional connectivity during distinct stages of a cognitive task. *NeuroImage*, 23(2):752–763, 2004.

[60] C. Ritter, M. Tanner. Facilitating the gibbs sampler: The gibbs stopper and the Griddy-Gibbs sampler. *Journal of the American Statistical Association*, 87(419):861–868, 1992.

[61] J. R. Rosenberg, A. M. Amjad, P. Breeze, D. R. Brillinger, D. M. Halliday. The fourier approach to the identification of functional coupling between neuronal spike trains. *Progress in Biophysics and Molecular Biology*, 53(1):1–31, 1989.

[62] D. Ruppert, M. P. Wand, R. J. Carroll. *Semiparametric regression*. Cambridge Univ Pr, 2003.

[63] R. Salvador, J. Suckling, C. Schwarzbauer, E. Bullmore. Undirected graphs of frequency-dependent functional connectivity in whole brain networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457):937–946, 2005.

[64] C. E. Shannon, W. Weaver. A mathematical theory of communications. *Bell System Technical Journal*, 27(2):632–656, 1948.

[65] M. Shi, R. E. Weiss, J. M. G. Taylor. An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 45(2):151–163, 1996.

[66] R. Shumway, D. Stoffer. *Time Series Analysis and Its Applications: With R Examples*. Springer, 2006.

[67] G. J. Siegle, S. R. Steinhauer, M. E. Thase, V. A. Stenger, C. S. Carter. Can't shake that feeling: event-related fMRI assessment of sustained amygdala activity in response to emotional information in depressed individuals. *Biological Psychiatry*, 51(9):693–707, 2002.

[68] G. J. Siegle, W. Thompson, C. S. Carter, S. R. Steinhauer, M. E. Thase. Increased amygdala and decreased dorsolateral prefrontal BOLD responses in unipolar depression: Related and independent features. *Biological Psychiatry*, 61(2):198–209, 2007.

[69] B. W. Silverman. Some aspects of the spline smoothing approach to Non-Parametric regression curve fitting. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(1):1–52, 1985.

[70] F. T. Sun, L. M. Miller, M. D'Esposito. Measuring interregional functional connectivity using coherence and partial coherence analyses of fMRI data. *Neuroimage*, 21(2):647–658, 2004.

[71] W. K. Thompson, G. Siegle. A stimulus-locked vector autoregressive model for slow event-related fMRI designs. *NeuroImage*, 46(3):739–748, 2009.

[72] A. W. Toga, J. C. Mazziotta. *Brain Mapping: The Methods*. Academic Press, 2002.

[73] W. Venables, B. D. Ripley. *Modern Applied Statistics with S*. Springer, Wrze. 2003.

[74] N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications*. MIT Press, 1949.

[75] K. J. Worsley, A. Charil, J. Lerch, A. C. Evans. Connectivity of anatomical and functional MRI data. *Neural Networks, 2005. IJCNN '05. Proceedings. 2005 IEEE International Joint Conference on*, wolumen 3, strony 1534–1541, 2005.

[76] L. Zhou, J. Z. Huang, R. J. Carroll. Joint modelling of paired sparse functional data using principal components. *Biometrika*, 95(3):601–619, 2008.