Background Connectivity– Understanding the brain's functional organization

by

Mikayla J Holmes
BSc, University of Victoria, 2021

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Mathematics and Statistics

Background Connectivity–
Understanding the brain's functional organization

by

Mikayla Jayne Holmes
BSc, University of Victoria, 2021

Supervisory Committee

Dr. Michelle F Miranda, Supervisor
(Department of Mathematics and Statistics)

Dr. Mary Lesperance, Additional Member
(Department of Mathematics and Statistics)

ABSTRACT

Task-state fMRI (tfMRI) and rest-state fMRI (rfMRI) surface data from the Human Connectome Project (HCP) was examined with the goal of better understanding the nature of background activation signatures and how they compare to the functional connectivity of a brain at rest. In this paper we use a hybrid—decomposition and seed-based—approach to calculate functional connectivity of both rfMRI data and the estimated residual data from a Bayesian spatiotemporal model. This model accounts for local and global spatial correlations within the brain by applying two levels of data decomposition methods. Moreover, long-memory temporal correlations are taken into account by using the Haar discrete wavelet transform. Modifications applied to the original spatiotemporal model that facilitate the use of surface and volumetric (whole-brain) data – in the CIFTI file format – are what make this analysis novel. Motor task data from the HCP is modelled, followed by an analysis of the residuals, which provide details regarding the brain's background functional connectivity. These residual connectivity patterns are assessed using a manual procedure and through studying the induced covariance matrix of the model's error term. When we compare these activation signatures to those found for the same subject at rest we found that regions within the subcortex displayed strong connections in both states. Regions associated with the default mode network also displayed statistically significant connectivity while the subject was at rest. In contrast, the pre-central ventral and mid-cingulate regions had strong functional patterns in the background activation signatures that were not present in the rest-state data. This modelling technique combined with a hybrid approach to assessing functional activation signatures provides valuable insights into the role background connections play in the brain. Moreover, it is easily adaptable which allows for this research to be extended across a variety of tasks and at a multi-subject level.

Contents

Supervisory Committee i

Abstract ii

Table of Contents iii

List of Tables v

List of Figures vi

Acknowledgements viii

Dedication ix

Chapter 1 Introduction 1

1.1 Problem Overview 1

Chapter 2 Data 5

2.1 Resting State Data 6

2.2 Task State Data 8

Chapter 3 Methods 11

3.1 Resting State Analysis 14

3.2 Motor Task Analysis 17

3.2.1 General Model Construct 20

3.2.2 Spatiotemporal Transformations 21

3.2.3 Model Assumptions 25

3.3 Analyzing Background Connectivity 27

Chapter 4 Results 29

Chapter 5 Conclusion 39

5.1 Limitations 39
5.2 Discussion and Future Work 40

Appendix A Acronyms **42**

References **44**

List of Tables

Table 3.1 Discrete wavelet specifications for temporal transformation 24

Table 4.1 Region and network pairs from the resting state connectivity matrix that have $t_{RV} > \tau_{.9997}$ 34
Table 4.2 Pairs of regions (w/ hemisphere) and networks corresponding to the top 25 RV coefficients from the resting state connectivity matrix. 35
Table 4.3 Pairs of regions (w/ hemisphere) and networks corresponding to the top 25 t_{RV} values from the motor task background connectivity matrix. 37

Table A.1 Parcel abbreviations and corresponding full names. 42
Table A.2 Network abbreviations and corresponding full names. 43

List of Figures

Figure 1.1 Examples of different magnetic resonance imaging scan types as displayed on p.4 of Buimer et al. (2020). In order from left to right and top to bottom: **A**) T1-weighted structural MRI scan, subcortical and cortical segmentation and regions of interest of the right cortex. **B**) Resting-state functional MRI scan, top view of cortical region with two regions of interest displaying their BOLD contrast values over time. **C**) Task-state fMRI scan along with a dynamic visualization of how the BOLD signal changes while the subject is completing a task. 2

Figure 2.1 Six still images (with four angles per image) taken from an interactive widget displaying CIFTI surface data from one subject while completing the HCP motor task. By default, the surface data is plotted onto an "averaged inflated surface[s]"(Pham et al. (2022), pg.3) 7

Figure 2.2 Slices of the CIFTI volumetric data for one subject completing the HCP motor task. This data provides us with the subcortical and cerebeller regions superimposed onto a 2mm smoothed anatomical mapping. The default is chosen to be the mapping from Montreal Neurological Institute and Hospital (MNI), but can be modified after importing other mappings. (Pham et al., 2022). 8

Figure 2.3 15 of the 19 Subcortical regions specified by the CIFTI file meta data, as displayed in figure 1 of Gao et al. (2021). The brain stem is the one subcortical region which is not divided into left and right counterparts, moreover, the cerebellum and diencephalon regions are not highlighted here and account for the remaining 4 regions when divided into their left and right hemisphere parts. 9

Figure 2.4 Motor task details. 10

Figure 3.1 Anatomical mapping of the brain's left and right hemispheres into 114 regions. 15
Figure 3.2 Design matrix $\mathbf{X}$ of the motor task. Here, h denotes the hrf and $\mathbb{1}$ is the indicator value for stimulus i. Going left to right across the horizontal axis, these stimuli acronyms correspond to left foot, left hand, right foot, right hand and tongue, respectively. 22
Figure 3.3 Visualization of the primary components of our final model assumptions prior to estimating our coefficients with the MCMC algorithm 26

Figure 4.1 Z-scores calculated from the Geweke convergence diagnostic test, corresponding to the 8000 iterations of MCMC for the left foot (LF), left hand (LH), right foot (RF), right hand (RH) and tongue (T) stimuli. 30
Figure 4.2 Top left: resting state RV connectivity matrix ordered by region. Top right: Resting state RV connectivity matrix clustered. Bottom left: motor task residual RV connectivity matrix ordered by region. Bottom right: motor task residual RV connectivity matrix clustered. 31
Figure 4.3 Visualization of the top three pairs of regions (and networks) corresponding to the three largest t_{RV} values from the rest state analysis. 33
Figure 4.4 Visualization of the top three pairs of regions (and networks) contained entirely within the left and right hemispheres from the background connectivity analysis. Since the precentral ventral cortex is present in each of these pairs, it is displayed in the top left corner with the arrows indicating the three other regions it has high connectivity with. 36
Figure 4.5 Background activation signatures found using the composite hybrid spatial basis induced residual covariance matrix. 38

ACKNOWLEDGEMENTS

I would first like to express my utmost gratitude to my supervisor, Michelle Miranda, for her compassion, understanding and wisdom throughout my degree. I also want to recognize all of the professors whom I have had the pleasure of being inspired by throughout my years in the mathematics and statistics department at UVIC. This endeavour would not have been possible without the frequent words of encouragement I received in so many ways from each and every one of you. I also want to acknowledge my deepest appreciation for my family. They picked me up, reminded me of my dreams and were never more than a phone call away when I needed support. Last, but not least, I would like to acknowledge my two fur children, Brewster and Misty. Your sitting on my keyboard and papers, chasing the mouse on the screen and knocking erasers onto the floor were an endless reminder to take breaks and laugh throughout these years.

Data were provided [in part] by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

DEDICATION

I would like to dedicate this work to my parents. You have both been an inspiration to me for as long as I can remember. Thank you, from the bottom of my heart, for your endless encouragement and for always believing in me.

Chapter 1
Introduction

1.1 Problem Overview

The desire to gain a better understanding of the human brain and all its intricacies is one that has motivated scientists for decades. Throughout the years, technology advancements have rapidly increased the rate of discoveries and our overall understanding of the brain. A fundamental technological advancement is that of the magnetic resonance imaging (MRI) scanner. The MRI scanner was first used 50 years ago and is now one of the most common brain imaging devices due to it being non-invasive. That is, an MRI scan does not require any medical instruments to enter the body. The original MRI—also known as—structural MRI scanner looks at the anatomical makeup of the brain and how its tissues and organs are organized. The structural MRI images provide this information by measuring fluid characteristics and the rate at which protons in water molecules return to equilibrium after a radio frequency has been sent through the scanner. Another type of MRI—and the one of focus for this research— is called functional MRI (fMRI) and was first used in 1990. It is a more advanced brain imaging technique as it measures the blood-oxygen level dependent (BOLD) value across the brain while a subject is at rest or completing a task. The reason that the oxygen level is the measure of interest is because of a known association between brain region activation and blood oxygen levels. Similar to muscle contractions, activated regions in the brain require more energy to fire and transmit signals effectively. This energy requirement is satisfied by increasing the cerebral metabolic rate of oxygen (CMRO2) (Glover, 2011). Functional MRI is central in providing information about how each region of the brain is functioning due to this correlation between cognitive up-regulation and blood flow. In the figure below, we show the differences between structural and functional MRI outputs. Structural MRI outputs depend on how long the magnetization is allowed to recover before the magnetic resonance value is collected. In contrast, figure 1.1 displays an example BOLD time series and the way it fluctuates across time regardless of whether a subject is directly presented with a task stimulus or is simply at rest.

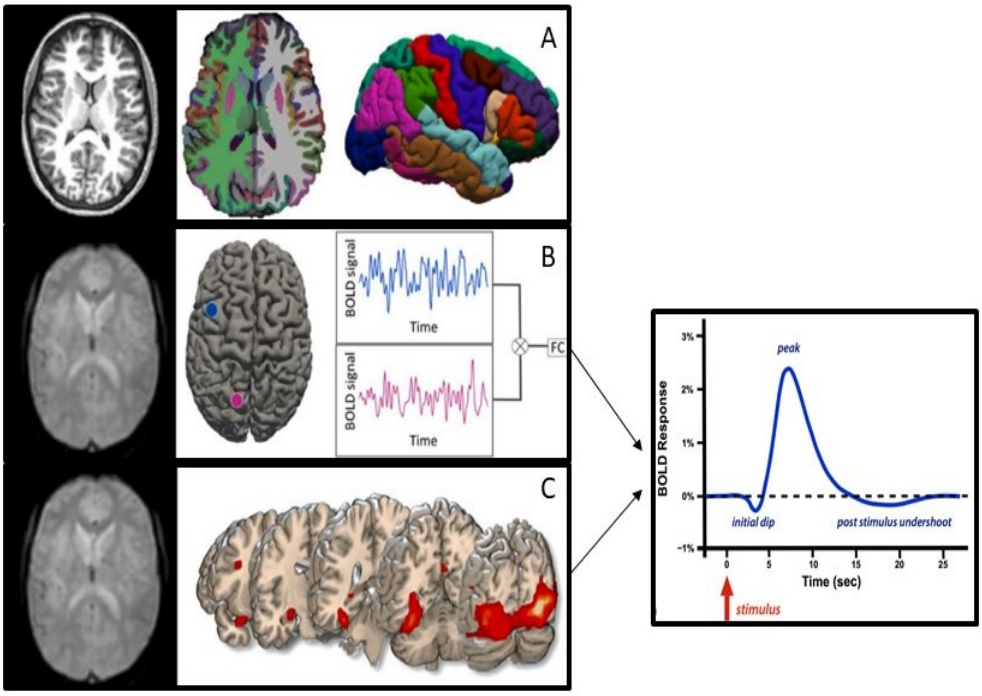

Figure 1.1: Examples of different magnetic resonance imaging scan types as displayed on p.4 of Buimer et al. (2020). In order from left to right and top to bottom: **A**) T1-weighted structural MRI scan, subcortical and cortical segmentation and regions of interest of the right cortex. **B**) Resting-state functional MRI scan, top view of cortical region with two regions of interest displaying their BOLD contrast values over time. **C**) Task-state fMRI scan along with a dynamic visualization of how the BOLD signal changes while the subject is completing a task.

Akin to structural MRI, studies using functional MRI became more popular due to the technique being non-invasive. Consequently fMRI studies became less expensive and from this, came an increase in research done on functional connectivity. More formally, *functional connectivity* is the study of activation signatures in the brain when a subject is at rest or completing a task (Fox and Raichle, 2007). These activation signatures allow for regions of the brain to be identified, analyzed and correlated in terms of their respective functions (Miranda and Morris, 2021). When a subject is at rest, the activation signatures found are referred to as *resting state connectivity*. At rest, there are a variety of functions still ongoing in the brain, therefore, resting state connectivity consists of spurious BOLD signals which we assume to be functionally correlated. Additionally, when a subject is completing a task we can look at either the functional connections that arise due to task-dependent stimulus or the connections which can be identified as background connections. These background connections—

called *background connectivity*—are found once the task-specific fluctuations have been accounted for vis a vis a modelling process.

At the onset of functional connectivity research, very simple modelling strategies were used for both rest-state and task-state studies. The original task models, built by Karl Friston in the 1990's , treated every *voxel* in the brain independently (Friston et al., 1993). The *voxel* can be thought of as a three dimensional pixel - or cube - that has an associated BOLD value. Until recently most fMRI data was analyzed using this volumetric data structure. These datasets are very large and computationally inefficient to manipulate when accounting for time as an additional dimension. In these original models each voxel was treated as independent and the BOLD value time series was used as the predictor variable. This procedure was then applied to every voxel in the brain – motivating the term "generalized" linear model. The final estimation was then created by combining N (independent) column vectors, each containing the estimated coefficients corresponding to the times series at locations 1,...,N. With volumetric data, the number of locations (N) was often over 900 thousand which demonstrates the computational demands with this data. The generalized linear model was used for both rest-state and task-state data. In the former, the BOLD values were modeled as a time series of a subject at rest and in the latter, they were modeled as a function of external stimuli in a task. The assumption of independence was necessary to simplify the modelling processes, however, this is a massive limitation when modelling a biological structure of such high complexity. Assuming independence results in a model that does not adequately account for spatial dependencies within the brain. Additionally, these linear models do not incorporate temporal correlations that are innate in BOLD time series. These issues are the foundation for research into more advanced statistical modelling techniques as well as improved fMRI data structures. These advancements are discussed in more detail in the methods section.

The modelling process of task state fMRI data can by done in a number of ways and the use of a Bayesian framework has been increasingly popular in recent years. Bayesian modelling allows for prior assumptions to be made which take into account empirical knowledge about the brain's structure and attributes. Some of these prior assumptions take into account characteristics of the brain such as: long memory processing and blood flow response time. These biological attributes have been shown to follow specific patterns which can be represented by mathematical functions which are then incorporated into Bayesian priors to improve the accuracy of the models built. Additionally, it is of value to improve the accuracy of task state models because this directly impacts how reliable and representative the background connectivity is of the underlying activation patterns. Improving statistical models directly benefits those who intend to use fMRI data in predicting disease progression, drug response, and the impact of psychological therapies. There are also many other benefits for neuroscience specific fields. As is the case with any data driven fields, the progress and benefits

gained are impacted greatly by the quality of the data. With this in mind, we next discuss the Human Connectome Project, which is the source of data used in the research presented here.

The Human Connectome Project (HCP) is the first of its kind when it comes to fMRI data and is paving the way for meaningful advancements in our understanding of the brain's functional and structural neural connections. The original five year project began in 2010 and was awarded over 1.3 million dollars by the National Institute of Health (NIH). This funding supported the 1200 Subjects Release data from Washington University, University of Minnesota and Oxford University (denoted as the Wu-Minn Ox consortium). This specific release involved over 1200 healthy subjects who were scanned by state-of-the-art brain imaging technologies. One primary benefit of this project has been and is the accessibility of the data. All data is widely available along with detailed reference documents that outline all pre-processing steps taken to produce the datasets. This in and of itself has provided scientists with a magnitude of clean data that has never before been available.

These advancements have motivated a deeper look into the relationship between the brain's background connectivity and resting state connectivity. The remaining sections will outline all details of this research problem and have the following structure. Chapter 2 covers all specifics on the datasets used for the analyses. Chapter 3 outlines the different methods applied to resting and task data from the HCP and chapters 4 and 5 present the results and discussion as well as suggestions for future work.

Chapter 2
Data

The primary goal of this research is to gather reputable evidence pertinent to the nature of background connectivity in the human brain. Exploring this goal involves analyzing both resting state and task state fMRI data for a single subject from HCP's Wu-Minn 1200 Subject Release. Within this research we focus on the data provided for Subject 100307. Subject 100307 is a female, aged 26-30 and is the example subject for all rest and task datasets when only a single subjects data is downloaded from the HCP database. Behavioural, altertness and other toolbox assessment data is also available for this subject with familial and genetic information available upon request. The full dataset includes task state fMRI (tfMRI), resting state fMRI (rfMRI), diffusion MRI (dMRI), structural imaging, as well as behavioural and demographic data on 1206 adults who are all considered to be healthy (David et al., 2013). For the purpose of studying background connectivity, we use rfMRI and tfMRI data. Within the 1200 Subjects Release, there is also a variety of preprocessed datasets available depending on the needs and goals of the study. Complete details on the processing pipelines applied to all HCP data is found in Glasser et al. (2013). Different file structures are also available on the human connectome database, however, the HCP reference manual encourages the use of the more modern Connectivity Informatics Technology Initiative (CIFTI) file structure[1].

CIFTI files were first used in 2012 on Freesurfer: an open-source software package built for analyzing brain imaging data. The major benefit of using CIFTI files is that they reduce computation time due to the number of data points being just over 90 thousand (as opposed to almost 1 million for NIFTI files). Additionally, this surface-based structure mitigates the spatial correlation problems that can arise with volumetric data (Pham et al., 2022). This improvement allows for results between different subjects to be more comparable as well as the anatomical properties to be more adequately accounted for. Pham et al. (2022) iterate a ten-fold improvement when these surface based datasets are also combined with Bayesian modelling capac-

[1] `https://www.humanconnectome.org/storage/app/media/documentation/s1200/HCP_S1200_Release_Reference_Manual.pdf`

ities. The HCP reference manual also outlines that for all releases after 2018, only surface hemisphere data and subcortical volumetric data will be available. This decision was made to encourage the use of CIFTI files on all future analyses due to the many strengths they have over volumetric datasets. Moreover, CIFTI files contain cortical, subcortical and cerebellar gray matter data (displayed in figures 2.1 and 2.2), therefore, we are able to assess connectivity in regions that otherwise are difficult to account for with volumetric data. Figure 2.3, from p.2620 of Gao et al. (2021), displays the eight subcortical regions that CIFTI files have label data for within their nested list structure. This is novel in that most anatomical mappings used to analyze functional connectivity only take into account the left and right hemispheres. These advantages motivated performing all analyses using CIFTI files for rest state and task state data. All datasets used in this research are publicly available at `https://db.humanconnectome.org/data/projects/HCP_1200`. Also, extensive information on all task and rest state fMRI data can be found in the HCP 1200 Subjects Data Release Reference Manual. Although we used the same file structure and subject to obtain resting state and task state data, the more specific details differ slightly and thus are presented in sections 2.1 and 2.2.

2.1 Resting State Data

Within the HCP project, resting state data was collected during two sessions. The subject was in the fMRI scanner for a total of 30 minutes split into two 15 minute runs during each session. Within these runs, an image was taken every 720ms. The specific resting state dataset used from this data collection process was the FIX-Denoised (Extended) version. This extended version contained both volumetric and surface-based data for all four (two for left-right and two for right-left phase encoding) runs and was pre-processed using the independent component analysis FMRIB's ICA-based X-noisifier (ICA-FIX) pipeline. This pipeline originated with HCP and effectively cleans the data of structural noise by identifying independent components as "good" or "bad", where the "bad" components are eventually removed from the data. This is done in what's called a "non-aggressive" manner, which means that all components (good and bad) are initially included in the time-series regression so as to not hastly remove too much information from the dataset (Smith et al., 2013b). Moreover, the extended dataset contains files which outline the different signals and noise within rfMRI data, which are found by looking at the variability between different processing stages. Further details on the HCP pre-processing pipeline techniques can be found in Glasser et al. (2013). As mentioned above, only the CIFTI file structure within this dataset was used for the current analysis, and more specifically the left-to-right (LR) phase encoding data from one run. The CIFTI file structure can be thought of as a nested list with the first element of the list containing three matrices, for the left cortex,

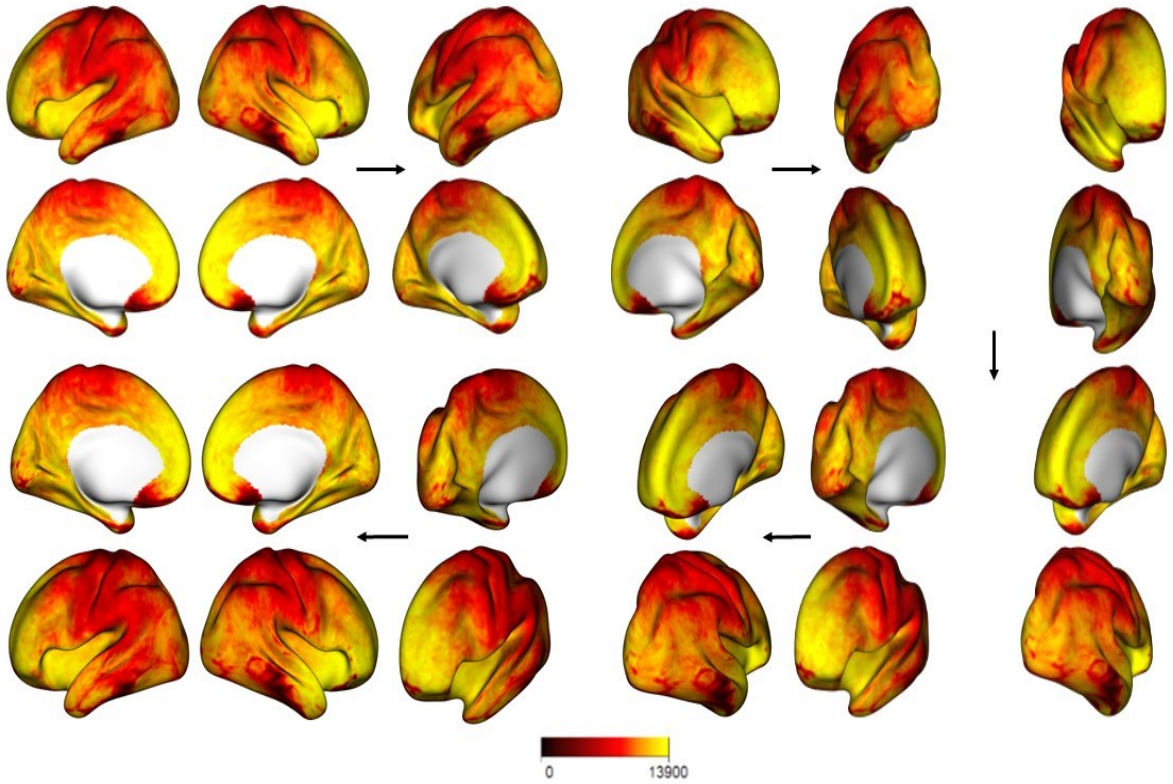

Figure 2.1: Six still images (with four angles per image) taken from an interactive widget displaying CIFTI surface data from one subject while completing the HCP motor task. By default, the surface data is plotted onto an "averaged inflated surface[s]"(Pham et al. (2022), pg.3)

right cortex and subcortical regions of the brain. Although the subcortical region is still available as volumetric (rather than surface-based) data, the CIFTI file converts these 3-dimensional regions into 2-dimensional data via a transformation matrix with homogeneous coordinates and units (Pham et al., 2022). This conversion allows for all regions of the brain to be assessed, thus improving the evidence gained from the analysis. The three matrices are of size T x N_h for the left and right hemispheres and of size T x N_c for the subcortical region. For all three matrices $T = 1200$ while $N_h = 32,492$ and $N_c = 31,870$. T is the number of time points at which scans were taken and N_* are the number of locations that BOLD contrast values are collected at each time step. It is important to note that certain locations have no data available and are called the medial wall vertices. There are $2,796$ and $2,776$ points within the left and right hemispheres respectively, that contain no data. These medial wall locations are removed and accounted for via a function within the CiftiTools R packages which was built to specifically manage the CIFTI file metadata in a clean and concise manner

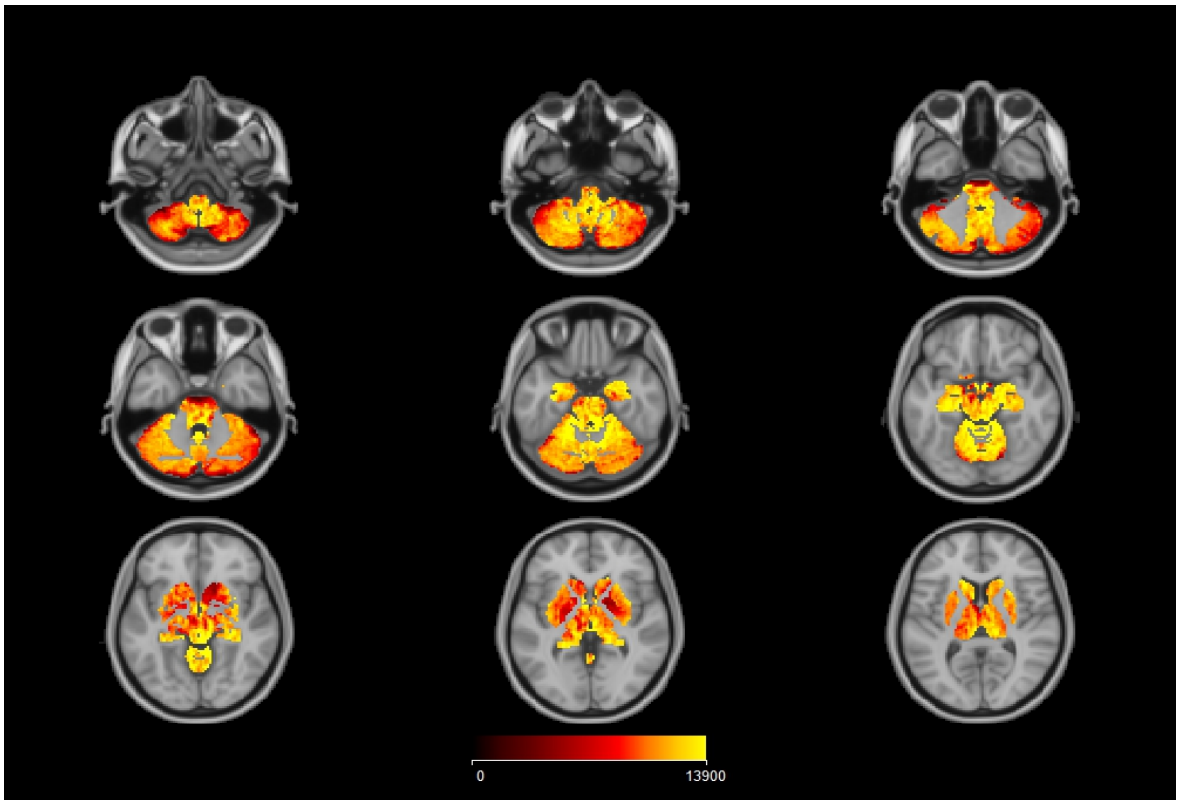

Figure 2.2: Slices of the CIFTI volumetric data for one subject completing the HCP motor task. This data provides us with the subcortical and cerebeller regions superimposed onto a 2mm smoothed anatomical mapping. The default is chosen to be the mapping from Montreal Neurological Institute and Hospital (MNI), but can be modified after importing other mappings. (Pham et al., 2022).

(Pham et al., 2022). Lastly, the BOLD matrices for each hemisphere and subcortical region are combined to create a T x N_ζ matrix of size $1,200$ x $91,146$. All statistical analyses and methods (detailed in section 3.1) are carried out on this dataset to gain valuable insights on resting state connectivity for Subject 100307.

2.2 Task State Data

To analyze background connectivity, the data from HCP's motor task was used for Subject 100307. During the motor task the subject was instructed to move their left and right hand, left and right foot and tongue for 12 seconds after being given a three second visual cue. 284 images (T) were collected during each of the two runs. Within these runs, the subject was asked to do each movement two times as well as being given a 15 second rest period at three points throughout the run. Similar to the rfMRI data, scans were taken every 720ms (David et al., 2013; Glasser et al., 2013). A visualization of the motor task structure for a single run is displayed in figure 2.4 below.

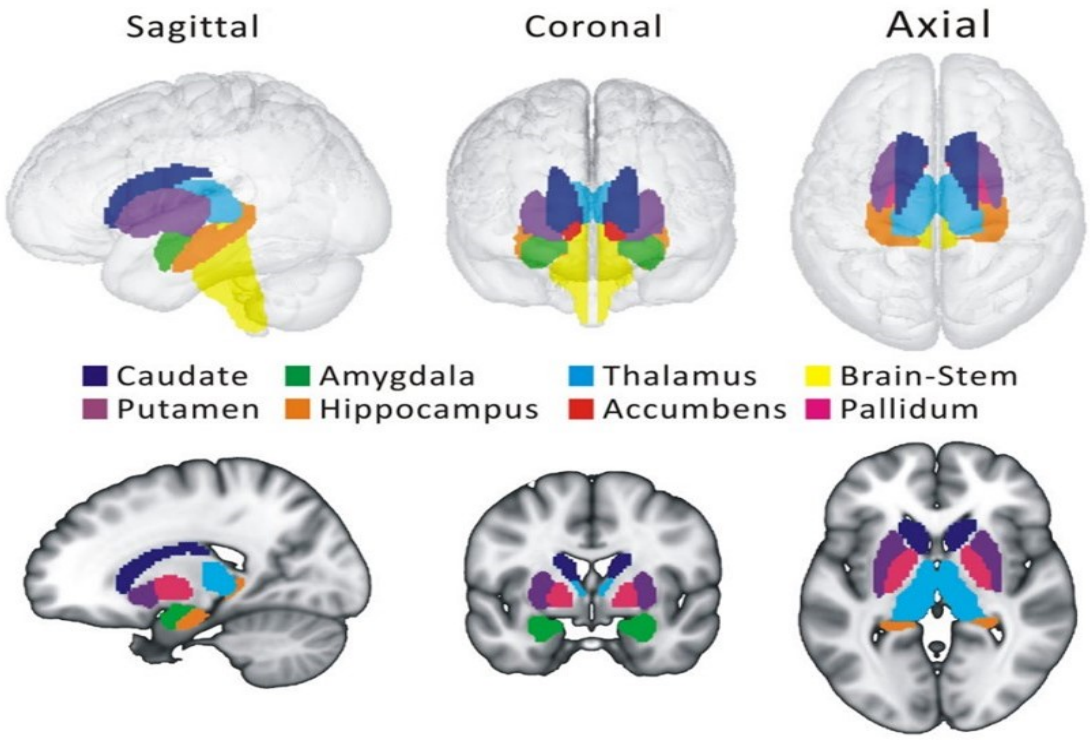

Figure 2.3: 15 of the 19 Subcortical regions specified by the CIFTI file meta data, as displayed in figure 1 of Gao et al. (2021). The brain stem is the one subcortical region which is not divided into left and right counterparts, moreover, the cerebellum and diencephalon regions are not highlighted here and account for the remaining 4 regions when divided into their left and right hemisphere parts.

The left-to-right phase encoding on minimally pre-processed tfMRI motor data was used for this analysis. The minimal smoothing parameter of 2mm was chosen to avoid losing important information prior to implementing all model and dimension reduction transformations. This smoothing was done on all tfMRI datasets during level 2 of the pre-processing pipeline. Prior to the 2mm minimal FWHM smoothing on the volumetric data, motion and distortion corrections were applied as well as "registration to standard space and [the] generation of grayordinate time series for each individual task run" (p.101 of HCP Reference Manual). Similar to the rfMRI data, the CIFTI surface data was used which contained images from the left cortex, right cortex and subcortical regions. For the motor task, these matrices (contained in the CIFTI nested list) were of dimension T x N_h for the left and right hemispheres and of size T x N_c for the subcortical region. For all three matrices $T = 284$ while $N_h = 32,492$ and $N_c = 31,870$. These matrices have the same number of medial wall vertices and are treated identically to that discussed above for the rfMRI data, thus giving a combined matrix of size T x N_ζ (284 x 91,146). This cleaned and combined matrix is the

response matrix ($\mathbf{Y}$) of BOLD values for the task model. The design matrix structure, model assumptions and transformations applied to $\mathbf{Y}$ for the modelling procedure will be discussed in detail in section 3.2. Additional information regarding the tfMRI parameters used on all HCP task datasets are available in the HCP Reference Manual.

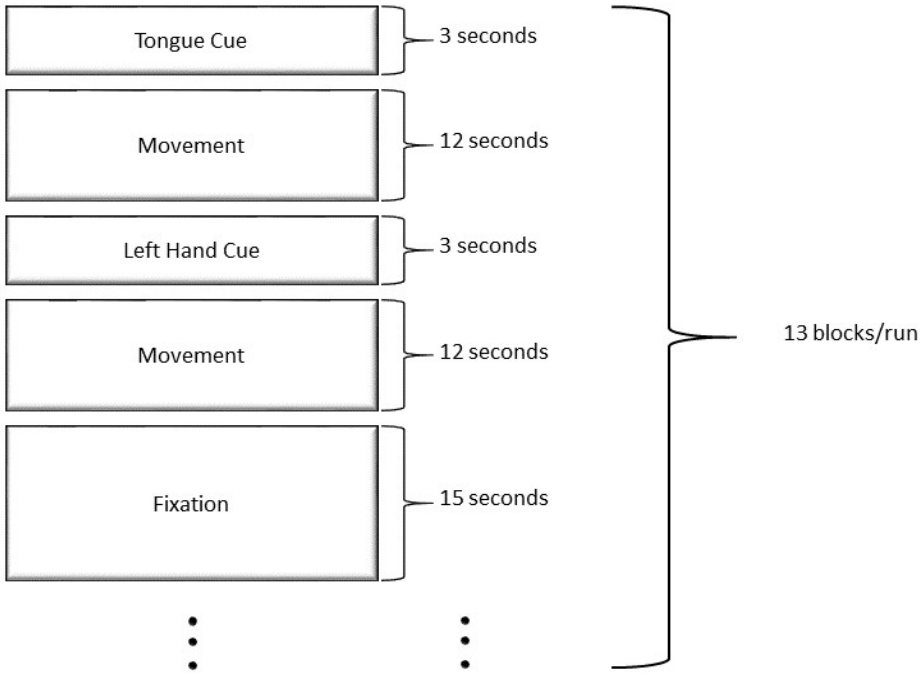

Figure 2.4: Motor task details.

Chapter 3
Methods

A comparative approach was taken to investigate the different patterns and nature of background connectivity versus resting state connectivity in the brain. This was done using data from Subject 100307 completing the motor task and while at rest. Since we were able to use data from the same subject in rest and task states, we do not need to emphasize the caveat of individual variability that is known to exist across different brains (Ting et al., 2020; Hart et al., 2017). The methodology used here can be broken down into three phases. The first phase involved analysing the functional connectivity of Subject 100307 during rest, following the hybrid techniques discussed by Shahhosseini and Miranda (2022) and Ting et al. (2020). In the next phase, we modeled the motor task BOLD response values for Subject 100307. We chose the composite hybrid spatiotemporal model, presented by Miranda and Morris (2021), due to its proven efficacy of accounting for location and time correlations within task data. Based on this model, the third phase consisted of estimating background functional connectivity in two ways. In the first approach, we manually calculated the residuals and then applied the same method used to estimate resting state functional connectivity. In the second approach, we used the estimated covariance matrix from the models residuals to estimate background functional connectivity.

These connectivity matrices were then empirically studied for important patterns, similarities and differences. These characteristics provide us with information on the nature of the neural connections at work in the background of the brain. Moreover, they can inform us as to the roles these activation patterns have in preparing or assisting the brain while it is in a task state. Furthermore, we can assess how this regulation compares to the resting state trends of the same brain.

All processing, modelling and analyzing was done using the statistical software R. This analysis was performed on a 1.19 GHz Intel Core i5 HP Envy with 8GB of RAM. The resting state analysis took 40 minutes to run and the task state analysis took 3 hours. During the processing stage of the task model MATLAB was also used to acquire the discrete wavelet transformation (DWT) matrix—discussed in detail below. All code and files used can be found in the **Supplementary Material**. Before presenting the

specific methods applied to rest-state and motor-task data for Subject 100307, both benefits and drawbacks of techniques used to measure functional connectivity will be addressed. Lastly, the details of the task model as well as all transformations applied to both the rfMRI and tfMRI data will be presented in sections 3.1 and 3.2.

Functional Connectivity

Functional connectivity helps us understand which locations in the brain are interacting and how important these interactions are. A recent paper written by Shahhosseini and Miranda (2022) presents many of the most common methods used to calculate functional connectivity. Some of these techniques are: seed-based, decomposition and hybrid methods. Seed-based methods are the most simplistic which use anatomical mappings to outline regions of interest (ROI) and then apply various measures of pairwise correlation between these regions or between a single region (chosen by a seed) and all others (Shahhosseini and Miranda, 2022). Here, we can choose between many correlation quantities depending on what type of linear dependence we are wanting to calculate. A disadvantage to seed-based methods is that they are computationally inefficient if we are wanting to assess the pairwise connectivity between all regions of the brain, especially when applied to the high dimensional volumetric data. Decomposition methods have an advantage of reducing the data dimension prior to assigning correlation values. The most common techniques used here are principal component analysis (PCA) and independent component analysis (ICA). Shahhosseini and Miranda (2022) emphasize a benefit of PCA that reduces the amount of information lost when projecting the data into a lower dimension. This projection, however, puts a limitation on the ease of interpretation in the projected space as the new locations no longer represent specific brain locations. A major advantage of PCA and ICA is that the transformations use orthogonal matrices and therefore the locations in the projected data space can be assumed to be independent. Specifically within the case of task-state data, this assumption facilitates the use of simple regression methods on very complex structures. The final method I will discuss is that of a hybrid approach and is the one used in the following analysis. Ting et al. (2020) use this approach and demonstrate how it reaps the benefits of both decomposition and seed-based approaches by combining the two techniques.

Ting et al. (2020) applied a factor analysis multi scale approach to HCP resting state data. Generally speaking this method divides the brain into regions and then measures the connectivity between those regions using what's called the RV coefficient. A primary feature of this study was the method used to effectively account for spatial correlation and high dimensionality. The dimension reduction method used was PCA combined with an adaptive projection threshold. That is, each region was transformed into a lower dimension using a different number of principal components. Incorpo-

rating flexibility into the standard PCA projection method enables us to account for the individual variation and complexity within each ROI. This reduces the subjectivity that inherently arises when a certain threshold is fixed for all regions, regardless of known differences. The RV coefficient, used by Ting et al. (2020) to measure functional connectivity after dimension reduction, is an extension of the Pearson correlation measure. Some of its characteristics and deviation from the standard correlation measure are discussed below.

The RV coefficient for two matrices X and Y:

$$RV = \frac{tr(XX^TYY^T)}{\sqrt{tr((XX^T)^2)tr((YY^T)^2)}} \tag{3.1}$$

can be thought of as a generalization of the standard squared correlation coefficient. The numerator takes into account the variance between different regions—in this study—after they have been projected into a lower dimension (Shahhosseini and Miranda, 2022). The interaction or inner product between these matrices are block diagonal and they account for the cross-covariance between regions i and j, thus providing an overall understanding of the functional connectivity. Although this measure of likeness between two matrices is different than the standard correlation equation, some commonalities remain. First, the numerator of the RV coefficient is exactly the sum of square covariance (Josse et al., 2008). Additionally, Josse et al. (2008) emphasize that an RV value of 1 can be thought of in terms of the dissimilarity between two matrices. That is, if the dissimilarity of matrices X and Y is 0, then the RV coefficient will be 1.

A potential pitfall of the approach taken by Ting et al. (2020) is the assumption of asymptotic normality of the RV coefficient test statistic t_{RV}. Josse et al. (2008) test four alternative methods for identifying RV correlation coefficients of statistical significance. These methods are all based on what they call the 'point cloud shape' (Josse et al., 2008). The point cloud can be thought of as individual points plotted in 3D space as a way to study the distribution or configuration. From here, we can more easily assess the relationships between different complex sets of variables. Although the RV coefficient can be used to analyze this relationship, when the point cloud distributions of two regions are skewed Josse et al. (2008) noted diminished efficacy with respect to the assumption of asymptotic normality. Specifically, this limitation arises when this test statistic is used to determine where the correlation between two matrices is statistically significant. Moreover, all approximations tested by Josse et al. (2008) can result in negative densities for the point cloud. This introduces a component of error as we then need to take the absolute value prior to comparing with a critical value. These constraints and weaknesses of the RV coefficient were mitigated here in a number of ways and will be discussed further in 5. Next, we consider the particular approach taken to assess functional connectivity on the Subject 100307 rest state data.

3.1 Resting State Analysis

In general, resting state functional connectivity consists of spontaneous signals emerging throughout the brain which are then considered to be functionally associated. The measure of association is being made by assessing the corresponding BOLD values across local or global spatial regions (Woodward and Cascio, 2015). When analysing resting state data there is no need to remove the temporal correlations from the data before calculating connectivity patterns. We do, however, want to take into account the similarities in connectivity patterns across the brain (Ting et al., 2020). With that in mind, we analyzed the activation signatures of Subject 100307 at rest by following the hybrid connectivity method highlighted in Shahhosseini and Miranda (2022) and discussed by Ting et al. (2020). Ting et al. (2020) applied this method to rfMRI data from HCP as well. Before calculating a measure of correlation the original resting state matrix $\mathbf{R}$ (T x N_ζ) was broken into 133 regions of interest based on an anatomical mapping defined by Schaefer et al. (2018). This anatomical mapping divided the left and right cortices into 114 regions (see figure 3.1[1]). Only mapping hemispheres and not the subcortical region is common practice and therefore the majority of functional connectivity studies do not include connections within or between the subcortical region of the brain. This is one of the major advantages of using the CIFTI surface data. CIFTI files—as mentioned above—include both cortical and subcortical data, however, they are easily combined to get a more holistic dataset. Additionally, the CIFTI file has metadata that can be used to separate the subcortical region into 19 distinct regions (shown in figure 2.3). Thus, combining the anatomical mapping of Schaefer et al. (2018) and the subcortical regions, we were able to assess the functional connectivity between 133 regions of interest.

Dividing the data into regions allows for the brain's functional nature and biological characteristics to be incorporated into the analysis. This mitigates errors that can arise when the spatial correlation innate to our brain is not taken into account. The process of parcellating the brain into 133 ROI's occurred in an iterative fashion by taking the columns of $\mathbf{R}$ that correspond to each region—one region at a time—and applying singular value decomposition (SVD). The number of locations in each region ranged from 48 (region 24) to 9,144 (region 123), where the 19 subcortical regions (115-133 contained the largest number of locations).

The hybrid connectivity method, as mentioned above, combines decomposition and seed-based methods in a way that mitigates the weaknesses of either method on their own (Shahhosseini and Miranda, 2022). Here, the decomposition technique, SVD, is applied to each region individually followed by finding the measure of correlation between each of the decomposed regions. Consider the resting state matrix $\mathbf{R}$ and

[1]The white center is the subcortical region which is not included in Yeo's anatomical mapping

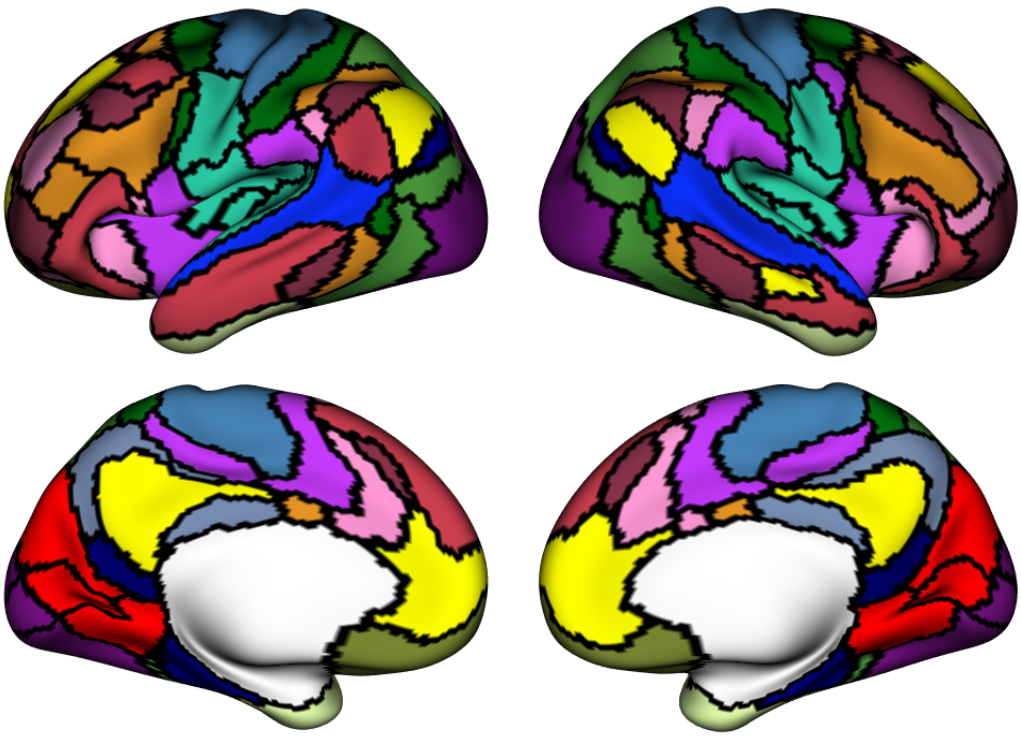

Figure 3.1: Anatomical mapping of the brain's left and right hemispheres into 114 regions.

define $\mathbf{R}^{(k)}$ as the sub matrix containing the columns from $\mathbf{R}$ that correspond to the locations for region k. Each sub matrix $\mathbf{R}^{(\mathbf{k})}$ is projected into a lower dimension. Following the procedures done by Ting et al. (2020) and Miranda and Morris (2021) we projected each region into the data space spanned by the number of components which explain at least 50% of the regions variation (denoted by p_k for $k = 1, \ldots, 133$). That is, if we let $\{\lambda_1, \ldots, \lambda_{S_k}\}$ be the set of ordered singular values for region k, then

$$p_k = \min \left\{ L | \frac{\sum_{l=1}^{L} \lambda_l^{2(k)}}{\sum_{s=1}^{S_k} \lambda_s^{2(k)}} \geq .5 \right\}$$

Ting et al. (2020) used PCA as their decomposition method and assessed the squared error associated with the RV coefficient for adaptive threshold values of 1%, 25%, 50% and 75% of the variance explained. We chose a threshold of 50% as it displayed minimal squared error while still maintaining a reasonable computation time - which increases when we retain many components. SVD (like PCA) is a beneficial decomposition technique because the singular value matrix is orthogonal, therefore, we can assume independence between and within the regions after they are transformed. This advantage will be discussed more in the task data section as it allows us to relax

certain modelling assumptions.

The next step in acquiring our functional connectivity matrix was to use a measure of correlation to describe the strength of connection between each of the 133 ROIs. Here we chose to use the RV coefficient as done by Ting et al. (2020) and discussed in Josse et al. (2008). After all locations (N_ζ) have gone through SVD we end up with 133 matrices of size p_k x p_k, where p_k is defined above. Then we let the RV coefficient between region i and j be given by the following:

$$RV_{ij} = \frac{tr(C_{f_i f_j} C_{f_j f_i})}{\sqrt{tr(C_{f_i f_i} C_{f_i f_i}) tr(C_{f_j f_j} C_{f_j f_j})}} \tag{3.2}$$

Here, $C_{f_i f_j} = (\Sigma_{f_i f_i})^{-\frac{1}{2}} \Sigma_{f_i f_j} (\Sigma_{f_j f_j})^{-\frac{1}{2}}$ and $\Sigma_{f_i f_i}$ are block diagonal matrices of dimension p_i x p_i. Each $\Sigma_{f_i f_i}$ is the lower dimension projection space for the region i. Depending on the dimension of each region—following the adaptive approach from Ting et al. (2020)—these projections were done using right or left singular vector matrices found when we applied SVD to our original data matrix $\mathbf{R}^T\mathbf{R}$. The result of this RV calculation was a 133 x 133 symmetric matrix which displayed a measure of connectivity between each pair of regions for Subject 100307 at rest. Although the RV coefficient provides valuable information about the the functional nature of our brain at rest, Ting et al. (2020) point out that high RV values may not necessarily imply statistical significance. Josse et al. (2008) also emphasize that in order to assess the importance of these connections, we can use asymptotic properties of the standardized RV correlation coefficient. With the test statistic denoted t_{RV}, we can test the following hypothesis for each pair of regions:

$$H_o : RV_{ij} = 0 \qquad vs. \qquad H_1 : RV_{ij} > 0$$

t_{RV} is defined by:

$$t_{RV_{ij}} = \frac{R\hat{V}_{ij} - E(R\hat{V}_{ij})}{\sqrt{Var(R\hat{V}_{ij})}} \tag{3.3}$$

The details of how to find the expected value and variance of the estimated RV correlation coefficient are in Ting et al. (2020). t_{RV} is asymptotically standard normal, therefore we use $N(0, 1)$ to find threshold values. If we let τ be the threshold value, then regions i, j such that $t_{RV_{ij}} > \tau$ are deemed statistically significant. The outcomes that result from these hypothesis tests while considering two threshold values are discussed further in our results (4) and conclusion (5) sections.

3.2 Motor Task Analysis

Background connectivity activation signatures of Subject 100307 were analyzed after the task-dependent BOLD values were effectively modelled and removed from the response matrix. The general steps taken to analyse the motor task data are as follows. We first applied three levels of transformations to the original dataset. This processed response matrix was decorrelated both spatially and temporally which allowed us to relax assumptions necessary for the modelling process. Specifically, we could assume spatial independence between the columns of our transformed response matrix. This assumption streamlines both modelling components as well as the procedure used to sample from the joint posterior distribution. Next, we modelled the transformed data and used the Monte Carlo Markov Chain (MCMC) algorithm to acquire our Bayesian regression coefficient estimates. After estimating the regression coefficients, the design matrix and its associated coefficients were projected back to the original data space so that the residual matrix could be calculated. The residual matrix consisted of the persisting BOLD values present in the brain because we had already taken into account the magnitude and trends within BOLD values due to task-dependent stimulus. The final step we took was to derive the functional connectivity matrix from the residual structure in the same way as the resting state data. Maintaining consistency in processing these datasets allowed us to effectively compare the resulting activation patterns. The structure of the motor task data is summarized in section 2.2 and all other details on the modelling assumptions and processes are outlined below. Specifically, section 3.2.1 presents the basic model structure. Section 3.2.2 covers the spatial and temporal decorrelation techniques used. Lastly, section 3.2.3 provides the model assumptions and details the full conditional distributions sampled by MCMC. This section is concluded by discussing the ways background connectivity and task-data models have changed and advanced and how these shifts motivated the methods chosen for the research presented here.

Background Connectivity and Task State Models

Background connectivity refers to the on-going activation patterns in the brain which are not due to any conscious task-dependent stimulus. Some think of background connectivity as the connections that prepare the brain or help it maintain task states, but which are not directly caused by the task someone is taking part in (Elkhetali et al., 2019). When assessing background connectivity of task-state data it is crucial that the model adequately account for spatial and temporal correlation within the BOLD response values or otherwise the estimated model residuals—used to measure background connectivity—will not accurately represent the remaining BOLD value signatures. Here, we outline the progression of how background connectivity has been

studied which naturally leads to a discussion on the developments made in modelling fMRI task data.

A recent paper by Elkhetali et al. (2019) analyzed the effects on background connectivity activation signatures when subjects were given task commands by either visual or auditory stimulus. How the modalities of stimulus impacted background connectivity was thought to help in understanding the role of background connectivity in initiating or maintaining task states. An important result gleaned by Elkhetali et al. (2019) was that task commands given via visual stimulus gave rise to significantly higher levels of BOLD contrast values within the background connectivity. This result is of importance because subjects within the HCP were given visual cues for the motor task. Additionally, Elkhetali et al. (2019) noted the magnitude of impact that external factors can have on background connectivity. Some of these factors include behavioural state and type of task being accomplished. They refer to Arieli et al. (1996); Cardoso et al. (2012); He (2013) and Sylvester et al. (2009) when discussing a potential premise regarding the role background connectivity plays. That is, the brain needs to enter into a form of holding pattern prior to being able to execute a task, therefore, it is supposed that a function of background connectivity is to modify and control these task-dependent connection patterns. Another paper by Fox and Raichle (2007) outlines the variation found in background connectivity BOLD values are not displaying a random nature, but rather, display a necessary baseline that is consistent in subjects at rest. They also note the impact discrepancies between subjects have on background connectivity patterns, thus, providing motivation for comparing activation signatures for the same subject completing a task and at rest, as was done here.

The study of background connectivity is commonly addressed using a two-stage procedure; first, estimate model coefficients and task-specific activation patterns and second, calculate residuals and estimate their connectivity. Although this is a valid procedure and one used below, it is negatively impacted if the original model does not adequately describe the task-specific stimuli. That is, the importance of a model which sufficiently accounts for the biological properties of the brain is of utmost importance when the goal is to assess background connectivity functionality. Therefore, the succeeding discussion will cover critical advancements made on Bayesian fMRI models.

One way to assess and evaluate the evolution of Bayesian models on fMRI data and subsequently task models, is how well they are able to deal with biological characteristics of the brain. Although there are many such characteristics, the three challenging ones focused on in the current study are: spatial and temporal correlation amongst brain regions as well as high data dimensionality. Spatial correlation refers to the inherent similarities in function present in local and global neighborhoods within the brain. Temporal correlation is inherent in all BOLD value times series as there is a natural repetition process occurring with respect to the blood flow across regions of the brain. Lastly, the computational power required to effectively implement analyses on

fMRI data is quite large due to the high dimension of the data. As an example, the file size of one subject's rest and task data from the HCP project is over eight gigabytes, which highlights the importance of using robust dimension reduction techniques.

One of the first improvements made to the simple generalized linear models discussed above, was to assume auto correlations present in the models' error terms. This adaptation allowed the model to more accurately account for noise within fMRI time series data caused by physiological movements (Worsley and Friston, 1995). Further enhancements were made by Penny et al. (2005) which involved assuming a prior distribution on the models' regression coefficients. These priors accounted for spatial homogeneity within local regions but were not adequate for handling temporal correlation within the data. One of the simplest ways in which temporal correlation was first integrated into these models was by assuming an "autoregressive (AR) structure of order q with a prior distribution on the AR coefficients" (Miranda and Morris (2021), pg.2), (Lee et al., 2014). An alternative approach was to incorporate the temporal structure by assuming that the residual time series resembled independent noise. Although this is a step in the right direction, these priors do not adequately account for long-memory processes within the brain (Miranda and Morris, 2021).

Two important pieces of literature that display the benefits of applying Bayesian modelling to task fMRI data are that of Zhang et al. (2014) and Zhang et al. (2016). The 2014 framework used a model that concurrently located the activated regions responding to task stimulus while also detecting the groups of regions with similar activation signatures. This bimodal goal was achieved by using a dirichlet process prior for the model. This choice of prior induced a clustering process between time series that showed comparable variance and temporal attributes. Akin to the process applied in this analysis, Zhang et al. (2014) built their design matrix by taking the convolution between the hemodynamic response function (hrf) and a task state indicator matrix. Additionally, they incorporated temporal dependencies into their model using a discrete wavelet transform (DWT). In the 2016 study, this unified nonparametric Bayesian model was expanded to a multi subject framework. Here, the variational Bayes algorithm was used instead of Monte Carlo Markov Chain (MCMC). This optimization technique produced adequate estimations of means but was poor at estimating correlation structure and posterior variance. For that reason, MCMC algorithm was chosen as the sampling method for estimating posterior means in this analysis.

The final iteration of models to mention here are those known as *spatiotemporal* models. These models assume a prior on the regression coefficients to account for the spatial correlation as well as a prior—some of which were mentioned above—to account for temporal correlation. The most common spatial priors are Gaussian Markov random fields, used by Quirós et al. (2010), and Ising priors, used by Penny et al. (2005). Ting et al. (2020) also motivated the work done by Miranda and Morris

(2021) which used a closely related bimodal modelling goal. Miranda and Morris (2021) present a *composite-hybrid spatiotemporal* Bayesian model that is capable of modelling the spatial and temporal brain characteristics while simultaneously inducing a residual structure which exhibits the background connectivity activation patterns. Two levels of PCA are applied to account for spatial similarities among regions followed by a DWT. The model priors also take into account long memory processes and other anatomical characteristics. This model provides the basis on which the current analysis is built, thus, we will outline more details of this model and how it sufficiently accounts for characteristics of the brain in section 3.2.3.

3.2.1 General Model Construct

As previously mentioned, the motor task data was modelled using the Bayesian composite hybrid spatiotemporal regression model. The model was originally built and conceptualized for volumetric (NIFTI) data and the working memory task from the HCP (Miranda and Morris, 2021). Thus, some adjustments were made in the model building process to account for a different task and the use of surface (CIFTI) data.

As outlined in section 2.2, the task-state dataset is comprised of BOLD time series measured at 91,146 locations within the brain of Subject 100307. These values are modelled by accounting for the way that blood-oxygen levels behave when Subject 100307 is presented with a stimulus. Biologists have determined that the *hemodynamic response functions (hrf)* represents the way oxygen levels peak and decay after a stimulus has occurred. Our model's design matrix was devised by taking a convolution between the hrf and an indicator matrix West et al. (2019), Wald and Polimeni (2015). This convolution models the delay between when the subject is presented with a stimulus—denoted by 1's in an indicator function—and how the brain's oxygenated blood levels respond to this stimulus over time. The indicator matrix, in general, identifies all time points where the subject was given a visual stimulus. The HCP's explanatory variable (EV) files for the motor task, provided time stamps which correspond to instructions being presented (visually) to move the left hand, right hand, left foot, right foot or tongue ($p = 5$ different visual cues). Additionally, the EV files provide the length of time that the stimulus was displayed which was needed to build the indicator matrix. Taking the convolution of this binary matrix with the hrf gives us a design matrix which accurately describes the brain's blood-oxygen levels and their expected trends during the motor task. With this in mind, each time series at the N_ζ locations, for stimulus $i = 1, \ldots, p$ can be written as:

$$y_t(\zeta) = \sum_{i=1}^{p} b_i(\zeta)(h * \mathbb{1}_i)(t) + e(t) \tag{3.4}$$

Where $h(t)$ is the hrf, approximated by[2]:

$$h(t) = \frac{t^5 e^{-t}}{\Gamma(6)} - \frac{t^{15} e^{-t}}{6\Gamma(16)}$$

This model can also be displayed using matrix notation as shown here:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \tag{3.5}$$

Here $\mathbf{Y}$ is the matrix of BOLD contrast values of size T x N_ζ where T is 284 and N_ζ is 91,146 (same number of locations as resting state). The standard regression interpretation remains for $\mathbf{X}$ and $\mathbf{B}$ with $\mathbf{X}$ being the design matrix and $\mathbf{B}$ denoting the coefficients we will estimate using MCMC. The dimensions of $\mathbf{X}$ and $\mathbf{B}$ are T x $2p$ and $2p$ x N_ζ, respectively. A visualization of the design matrix is shown in figure 3.2. Here, we can see the motor task features by looking at the odd columns. These correspond to the convolution described above while the even columns are found by taking the same convolution, except with the first derivative of the hrf. Including the derivative of hrf in the design matrix facilitates the incorporation of additional stimulus variation in the model (Miranda and Morris, 2021).

3.2.2 Spatiotemporal Transformations

Following the methods presented by Miranda and Morris (2021), this original dataset $\mathbf{Y}$ underwent two levels of spatial transformation and one temporal transformation. The motivation behind applying these transformations is to adequately account for the complex structure of the data. When we account for these complexities the model residuals will more accurately represent the background BOLD activation signatures that arise when a subject is completing a task.

Local and Global Spatial Correlations

The spatial correlation present in our observed BOLD values—given by $\mathbf{Y}$—are considered in two parts; first, the local spatial correlation that occurs within each of the regions of interest and second, we refer to global correlation as being the functional patterns that are similar between these regions. Singular value decomposition was used to project $\mathbf{Y}$ into first a local and then global data space, spanned by principal components, which allowed the relaxing of model assumptions of independence.

As we outlined in section 3.1, SVD is useful for many reasons. Most importantly, it reduces our high data dimension. Additionally, it is a decomposition method that uses

[2]The shape and scale parameter choices come from Miranda and Morris (2021)

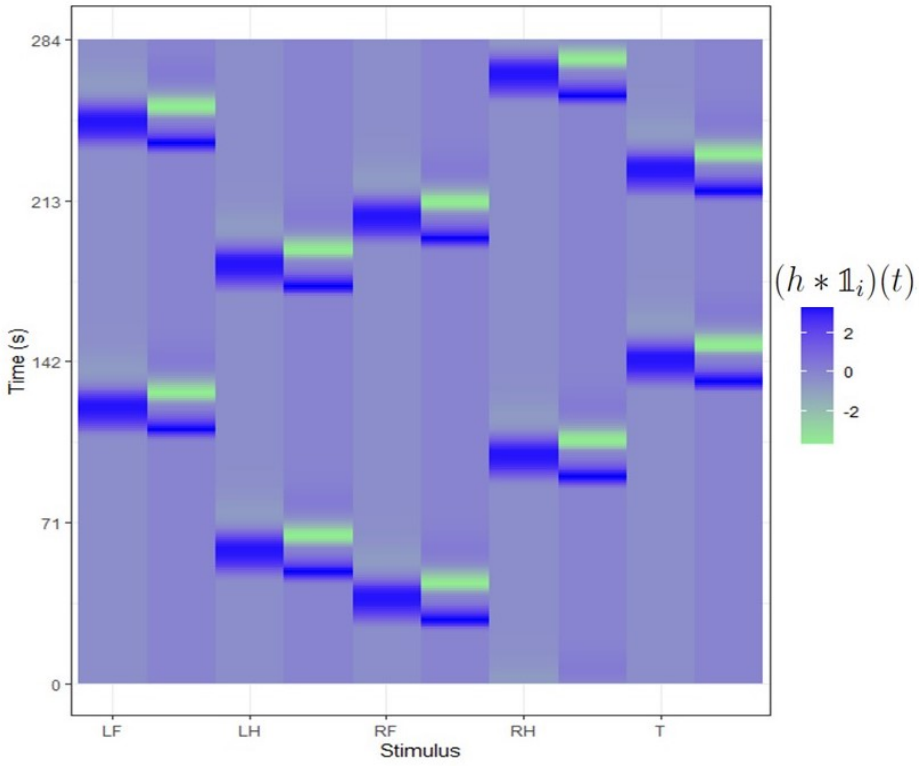

Figure 3.2: Design matrix $\mathbf{X}$ of the motor task. Here, h denotes the hrf and $\mathbb{1}$ is the indicator value for stimulus i. Going left to right across the horizontal axis, these stimuli acronyms correspond to left foot, left hand, right foot, right hand and tongue, respectively.

orthogonal transformation matrices, thus enabling the assumption of independence in the projection space. We can also easily modify the method to account for more or less variation explained by the transformed region. This is advantageous because when we project into the global space—considering the correlation between all regions—it is necessary to account for more variance within the data than when we project into the local space. It is important to note that the original data matrix $\mathbf{Y}$ was centralized with respect to its column means prior to applying any decomposition procedures. Analogous to the resting-state process, the combination of cortical and subcortical mappings were used to divide the N_ζ locations into 114 cortical ROIs and 19 subcortical ROIs (refer back to figures 2.3 and3.1).

The first projection took into account local correlations or functional similarities within each of the 133 ROIs. Let $\mathbf{Y}^{(k)}$ be the sub-matrices for each region $k = 1, \ldots, 133$ of size T x n_k,where n_k is the number of locations from $\mathbf{Y}$ that are in region k. We applied SVD to each of these sub-matrices i5and projected them into a lower dimension using a flexible threshold approach—similar to Miranda and Morris (2021). We projected the data within each region into the space spanned by the principal components

that explain at least 50% of the region's variance. That is, let $\hat{\mathbf{Y}} = \{\mathbf{Y}^{(1)} \dots \mathbf{Y}^{(k)}\}$ be the matrix where all N_ζ columns are sorted according to their respective region. Also, let ϕ^k be the matrix of orthonormal eigenvectors of $\mathbf{Y}^{(k)^T}\mathbf{Y}^{(k)}$. Now, let $\{\lambda_1, \dots, \lambda_{S_k}\}$ be the set of ordered singular values for region k, then

$$p_k = \min\left\{ L | \frac{\sum_{l=1}^{L} \lambda_l^{2(k)}}{\sum_{s=1}^{S_k} \lambda_s^{2(k)}} \geq .5 \right\}$$

Here, p_k varies for each region which provides us with a better understanding of the complexity of the region (Ting et al., 2020). Once this process has been applied to all 133 regions of interest, we define $\mathbf{Y}_L^{(k)} = \mathbf{Y}^{(k)}\phi^k$ which has size T x Σp_k as the matrix of local features (Miranda and Morris, 2021). Thus, in matrix notation, $\mathbf{Y}_L = \left[\mathbf{Y}_L^{(1)}, \dots, \mathbf{Y}_L^{(k)}\right] = \hat{\mathbf{Y}}\phi$ where ϕ is given as:

$$\phi = \begin{pmatrix} \phi^1 & \cdots & 0 \\ \vdots & \ddots & \\ 0 & & \phi^k \end{pmatrix}$$

Next, we consider the global correlations. Again, following the process presented by Miranda and Morris (2021) and Ting et al. (2020), it is beneficial to do a secondary dimension reduction in order to account for the spatial correlations between the 133 anatomically defined regions. Within this second layer, we projected $\mathbf{Y}_L$ based on the number of components that explain at least 95% of the variation in the data. Prior to applying the global transformation, the local feature space was size T x $\Sigma p_k = 1,836$. When we apply SVD to $\mathbf{Y}_L$, however, we reduced the space of local features down to size T x S where $S = 167$. Similar to above, let ψ be the matrix of orthonormal eigenvectors to get our final data matrix in the global feature space: $\mathbf{Y}_G = \mathbf{Y}_L\psi = \hat{\mathbf{Y}}\phi\psi$.

After $\mathbf{Y}$ has been centralized and spatial correlations have been accounted for, we can assume that the columns within $\mathbf{Y}_G$ are independent. The rows, however, are still temporally correlated, due to the innate BOLD value attributes and how they respond to the visual stimulus presented to Subject 100307 throughout the motor task assessment. Therefore, we must decorrelate the times series before estimating our model coefficients (Miranda and Morris, 2021; Zhang et al., 2014, 2016).

Temporal Correlations

The *discrete wavelet transform* (DWT) was chosen as a means to account for the temporal correlation in $\mathbf{Y}_G$ due to its emphasis in Fan (2003); Fadili and Bullmore (2002). The DWT decomposes the original time series into a certain number of sets

which are also time-series[3]. Each of these sets is made up of coefficients that depict the way the original signal evolves over time for a specific frequency band (Hosseinzadeh, 2020). The wavelet type impacts the technical details by which the denoising process is performed. The wavelet characteristics chosen are in table 3.1. Although there are a number of potential wavelet types that could have been used, the Haar wavelet transform was chosen due to its symmetric structure. That is, if we define the Haar wavelet coefficient matrix be $\mathbf{W}$, then we get $\mathbf{W}^{-1} = \mathbf{W}^T$. This attribute is necessary for our specific task model because the inverse DWT is needed to project our estimated coefficients back into the original data space prior to calculating the residual matrix we use to make conclusions on background connectivity.

Wavelet Attribute	**Value**	**Variable Defined**
Wavelet Type	'db1'	L = 2
Number of levels of wavelet transform	8	m
Number of wavelet levels	9	J = (nlevel + 1)
Total number of wavelet coefficients	288	K
Vector of length J with number of compressed wavelet coefficients at each level	[2,2,3,5,9,18,36,71]	K_J

Table 3.1: Discrete wavelet specifications for temporal transformation

Equation 3.6 represents the model after both the response matrix and predictor variables have been transformed by applying the matrix of wavelet coefficients $\mathbf{W}$. Following the notation of Miranda and Morris (2021), we have:

$$\mathbf{Y}^* = \mathbf{X}^*\mathbf{B}^* + \mathbf{E}^* \tag{3.6}$$

In more detail, $\mathbf{Y}^* = \mathbf{W}(\mathbf{Y_G})$ is the new response matrix after having been transformed by two levels of PCA (denoted $\mathbf{Y_G}$) and applying the discrete wavelet coefficient matrix $\mathbf{W}$ of size T^W x T. Similarly, $\mathbf{X}^* = \mathbf{WX}$, $\mathbf{B}^* = \mathbf{B}\phi\psi$ and $\mathbf{E}^* = \mathbf{WE}\phi\psi$, giving the final model dimensions of size T^W x S, T^W x P, P x S and T^W x S for $\mathbf{Y}^*$, $\mathbf{X}^*$, $\mathbf{B}^*$ and $\mathbf{E}^*$, respectively. Relating to the notation of Miranda and Morris (2021), T^W is exactly K in table 3.1.

A desired outcome of the spatial and temporal transformations is that long-memory aspects of the data have been accounted for, which is emphasized as an important part of the modelling process by Zhang et al. (2014) and Ting et al. (2020). Zhang

[3]This value can be user-specified or if J is defined as in table 3.1, then the n level value is found using: p.145 of Percival and Walden (2000), $J = log_2 \frac{T-1}{L-1} + 1$ where L is the length of chosen wavelet filter.

et al. (2014) outline how failing to incorporate this behaviour within the data—via wavelet transforms and informed priors on the error terms—increases the standard error of model estimates. Finally, through assuming independence between columns in $\mathbf{Y}^*$ and having decorrelated each time series, we can now apply standard Bayesian regression techniques. Here, we are using Bayesian regression to model equation 3.6 and subsequently acquire coefficient estimates using MCMC sampling. The final section in our motor task analysis chapter presents specifics on the model assumptions, priors, and hyper parameters as well as the derivation of our Bayesian joint posterior distribution.

3.2.3 Model Assumptions

A visual overview of the model assumptions pertaining to 3.6, is shown in figure 3.3). In order to estimate the $\mathbf{B}^* = \{B_1^*, \ldots, B_S^*, s = 1, \ldots, S\}$ coefficients, using a standard hierarchical Bayesian procedure, we need to sample from the joint posterior distribution for each location s:

$$p(B_s^*, \tau_s^2, \Sigma_s^W, y_s^*)$$

As is often the case, the joint posterior distribution is hard to obtain in closed form, therefore we decompose it into the marginal distributions and subsequently sample from the full conditionals to obtain our model estimates. Therefore, we must derive the full conditional distribution for B_s^*.

Recall, $s = 1, \ldots, 167$ are the independent locations in the global feature space. In order to take full advantage of the Bayesian regression framework, we use empirical insights on the brain's memory characteristics to build an informative prior on B_s^* and $\mathbf{E}^*$. Here, we let $B_s^* \sim N(\mathbf{0}, \tau_s^2)$ where τ_s^2 accounts for the regularization that is required from a functional regression standpoint (Morris, 2015). Moreover, $\tau_s^2 = k_\nu((\mathbf{X}^*)^T(\Sigma_s^W)^{-1}(\mathbf{X}^*))^{-1}$ where k_ν is a constant, chosen to be 100 which specifies the amount of shrinkage applied (Morris, 2015). Similarly, we let $\mathbf{E}^* \sim N(\mathbf{0}, \Sigma^W)$ where $\Sigma^W = diag(\Sigma_s^W); s = 1, \ldots, S$. As the levels show in figure 4.2, $\Sigma_s^W = diag(\psi_s(2^{\alpha_s})^{-m})$. Here, we estimated α_s using method of moments and then fixed our hyper-parameters based on the value of convergence. This follows from the supplementary material in Miranda and Morris (2021). Moreover, we let $\Psi_s \sim InvGamma(a_o, b_o)$ be the prior distribution for Ψ_s within the hierarchical model, where a_o and b_o are fixed constants.

The parameters Ψ_s and α_s are called the *innovation variance* and *long-memory parameter*, which are incorporated into the model due to the task nature of our analysis. These biological and observed brain features were included in the work done by Zhang et al. (2014), Zhang et al. (2016) and Miranda and Morris (2021). The final assumption of note is the likelihood for the transformed dataset. This follows the standard likelihood distribution where we have $s = 1, \ldots, S$ independent columns: $y_s^* \sim N(\mathbf{X}^* B_s^*, \Sigma_s^W)$. Using these assumptions we can find the full conditional distribution as follows:

$$p(B_s^*|\tau_s^2, \Sigma_s^W, y_s^*) \propto p(B_s^*|\tau_s^2)p(y_s^*|B_s^*, \Sigma_s^W)$$
$$\propto N(\mathbf{0}, \tau_s^2)N(\mathbf{X}^* B_s^*, \Sigma_s^W)$$

Our full conditional distribution is then given by:

$$p(B_s^*|\tau_s^2, \Sigma_s^W, y_s^*) \sim N(V_{B_s^*}\mathbf{X}^{*^T}(\Sigma_s^W)^{-1}y_s^*, V_{B_s^*}) \tag{3.7}$$

where $V_{B_s^*}$ is the covariance matrix with formula:

$$V_{B_s^*} = \mathbf{X}^{*^T}(\Sigma_s^W)^{-1}\mathbf{X}^*(I_p + k_\nu^{-1}I_p)^{-1}$$

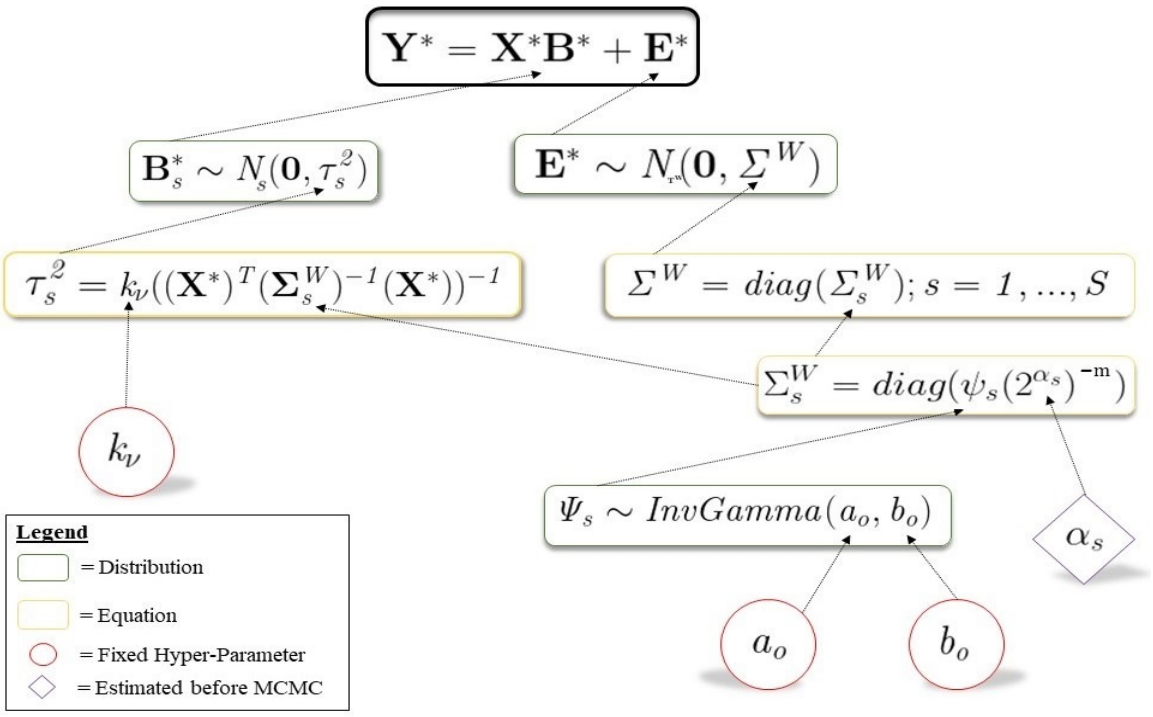

Figure 3.3: Visualization of the primary components of our final model assumptions prior to estimating our coefficients with the MCMC algorithm

For each location s, we did 8,000 iterations using the MCMC algorithm. The hyper-parameter α_s was initialized using a method of moments procedure, as done in Miranda and Morris (2021) and explained in detail in Miranda and Morris (2021) supplementary material. Ψ_s and B_s^* were also initialized and then updated at each of the 8,000 iterations in order to obtain estimated coefficients from the full conditional distribution given by 3.7. In order to acquire stable coefficient values, we used a burn-in

of 2,000 and thinning of 10 before taking the average across the remaining values of B_s^*, for each location s. The final estimated coefficients matrix is defined as $\hat{\mathbf{B}}^*$ and is size P x S, with each column containing the estimated coefficients for the stimuli—displayed by the design matrix–at location s. Another outcome of the MCMC procedure was the acquisition of the estimated covariance matrix that was simultaneously updated at each iteration. Section 3.3 explains the procedure used to obtain manually calculated residuals, the induced covariance matrix, and how these were used to conduct our final analysis on the background connectivity for Subject 100307.

3.3 Analyzing Background Connectivity

The final step prior to reaching conclusions about resting state and background connectivity patterns, was to analyze the background activation signatures. A novel component of the spatiotemporal model in Miranda and Morris (2021) is that it provides the means to estimate background connectivity which are induced by the final basis model assumptions (3.6). We analyzed the covariance matrix of the error terms—Σ^W from $\mathbf{E}^*$—to assess background activation patterns. First, however, background connectivity was evaluated using the manual approach (i.e. calculating model residuals and assessing their connectivity patterns) as this allowed the processing of the residual matrix using the same method for Subject 100307 at rest. Maintaining consistency in the methodology used, to assess functional connectivity, was imperative for our specific research question since we were studying the similarities and differences between these two brain states. The activation signatures that resulted from the induced covariance matrix give further support for both the model's efficacy and our residual calculations of background connectivity and thus will be discussed briefly below.

As mentioned above, the MCMC algorithm was used to sample from the full conditional distribution which provided us with our model's estimated coefficients—given by $\mathbf{B}^*$. These coefficients lie in the transformed data space that model 3.6 is in. Using the residual matrix calculated directly from these estimated coefficients would not facilitate an accurate comparison to the resting states' connectivity patterns, therefore we projected the estimated coefficient matrix $\hat{\mathbf{B}}^*$ back into the original data space (where model 3.5 lies). After this, we computed the estimated residual matrix manually and assessed its connectivity patterns.

Namely, let $\hat{\mathbf{B}}$ be the matrix of estimated coefficients in the original data space, then $\hat{\mathbf{B}} = \hat{\mathbf{B}}^* \Psi^T \Phi^T$. From here, we calculated the estimated residuals in the original data space in the standard way:

$$\hat{\mathbf{E}} = \mathbf{Y_c} - \mathbf{X}\hat{\mathbf{B}}$$

$\hat{\mathbf{E}}$ has dimension T x N_ζ and $\mathbf{Y_c}$ is $\mathbf{Y}$ after centralizing by subtracting the column

means. The residuals were calculated using the centralized dataset because the column means were subtracted from the original dataset before applying the two levels of SVD and thus was the appropriate version from which to acquire our residual matrix.

The final step to acquire a background connectivity matrix was to apply the same dimension reduction procedure to $\hat{\mathbf{E}}$ as was explained in section 3.1 with respect to $\mathbf{R}$. The dataset was iteratively broken into the 133 ROIs and then projected into a lower dimension space using SVD. The dimension was chosen according to the number of principal components that explained at least 50% of the variance (to remain consistent with rest-state procedure). Once projected into these lower dimensions, the RV coefficient was calculated using equation 3.2. Our final symmetric matrix of size 133 x 133 exhibits the background connectivity present in Subject 100307 while completing the HCP motor task. The connectivity matrix acquired using our two-stage method is the primary focus of our results and discussion sections, for reasons mentioned above. For completeness, however, we will briefly outline the processing done to acquire the induced background correlation matrix from the composite hybrid spatiotemporal model.

Here, we are looking at the covariance matrix of $\mathbf{E}^*$. The matrix Σ^W also lies in the transformed data space, therefore was transformed into what is called the "ROI space" (Miranda and Morris (2021), pg.7). Since the two levels of spatial transformations applied to our original model were associated with the correlations between and within our regions of interest, the ROI space can be thought of as the local or intra-region space (Miranda and Morris, 2021). That is, Σ^W is transformed using the inverse discrete wavelet transform and the transpose of our global feature matrix. Specifically, let $\theta_s = diag(mode(\mathbf{W}^T\Sigma_s^W\mathbf{W}))$ and $\Sigma_\theta = diag(\theta_1, \ldots, \theta_S)$. Then, we get $\Sigma_{ROI} = \Psi\Sigma_\theta\Psi^T$ as our matrix of background connectivity patterns in the ROI space. Here, Σ_{ROI} has dimension Σp_k x Σp_k. A slightly modified RV coefficient is calculated for Σ_{ROI} to get a connectivity matrix of size 133 x 133. Further details on the calculation steps for this process are found in the supplementary material of Miranda and Morris (2021). This matrix provided additional information and assisted in assessing the findings obtained using the two-stage procedure. The resulting background connectivity patterns from this induced process will be discussed briefly in section 4 below. More importantly, the detailed results of the residual connectivity analysis obtained using the same procedures as the resting-state date, including hypothesis testing and comparisons are examined in sections 4 and 5.

Chapter 4
Results

Throughout the previous chapters, we have discussed background information, data structure and statistical methods used to analyze fMRI data from the HCP Subject 100307. These techniques, in summary, yielded two functional connectivity matrices which depict interactions and variability between specified ROIs while the subject was in two different brain states. The entries of these symmetric matrices—with rows and columns corresponding to the 133 ROIs—are the basis of our results and discussion section. To reiterate, the goal of this research was to improve our understanding of the patterns and characteristics of the brain's functional connections while at rest or during a task. Specifically, we wanted to better understand the active regions in the brain during rest state and how these compare with task state background connections. Below, we provide the detailed connectivity results as well as how they advance our understanding of these complex components of the human brain. We will also discuss the limitations of these results as well as future work in sections 5.1 and 5.2.

MCMC Convergence

As discussed in section 3.2.3, the MCMC sampling algorithm was used estimate the coefficients of the transformed Bayesian regression model. Due to assumptions of independence between each location of $\mathbf{Y}^*$, the MCMC algorithm was performed, iteratively, on each of the 167 locations. For each location, the estimated coefficient $\mathbf{B}_s^*$ was of size 1 x 10, corresponding to the $2 * p = 2 * 5 = 10$ columns of the design matrix (with even columns corresponding to the first derivative of the hrf BOLD values). In order to assess convergence for each location and coefficient corresponding to the five stimuli, the Geweke convergence diagnostic was performed. Figure 4.1 displays the Geweke z-scores for the coefficients corresponding to the odd columns of the design matrix. This shows that the MCMC samples converge sufficiently for all five of the stimuli coefficients.

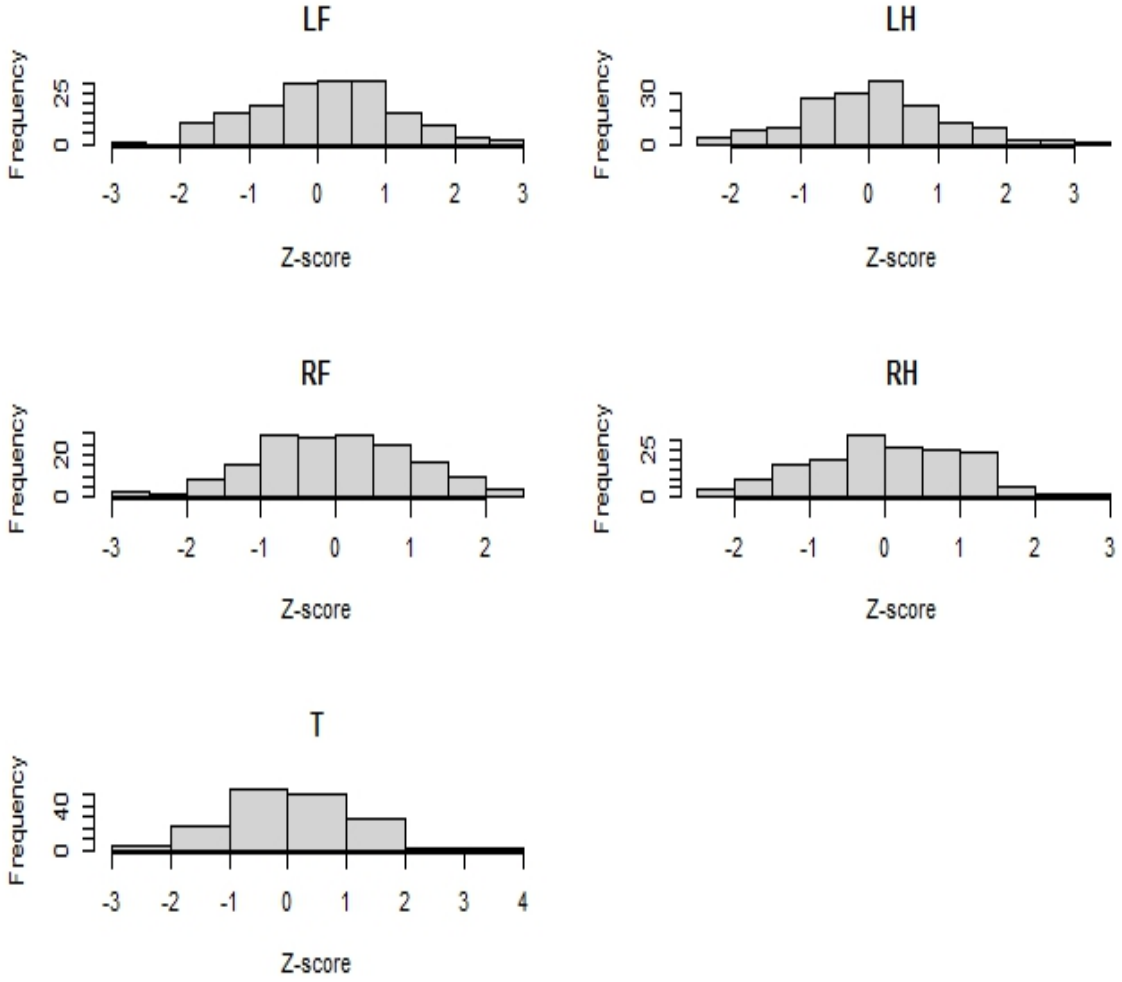

Figure 4.1: Z-scores calculated from the Geweke convergence diagnostic test, corresponding to the 8000 iterations of MCMC for the left foot (LF), left hand (LH), right foot (RF), right hand (RH) and tongue (T) stimuli.

Connectivity Results

In figure 4.2 we display the ordered and unordered connectivity matrices found using equation 3.2 on both rest and motor task residual data (from manual calculation). Unique patterns emerged when comparing each individual region (ordered) versus the general activation signatures found within the brain (unordered) at rest or during a task, thus explaining why both are presented. We can visually compare the connectivity patterns between regions by looking at the ordered matrices (left top, left bottom). A relative peak in intensity among the subcortical regions is evident (ROIs 115-133) on both resting and residual connectivity matrices. There are also distinct activation signatures that are detected in the background connectivity matrices which do not

appear in the resting state connectivity matrices. These similarities and differences are easy to detect visually. Below we present precise assessments of which connections were statistically significant, determined using standardized RV coefficients (3.3), for all pairwise correlations within the rest and background activation signatures.

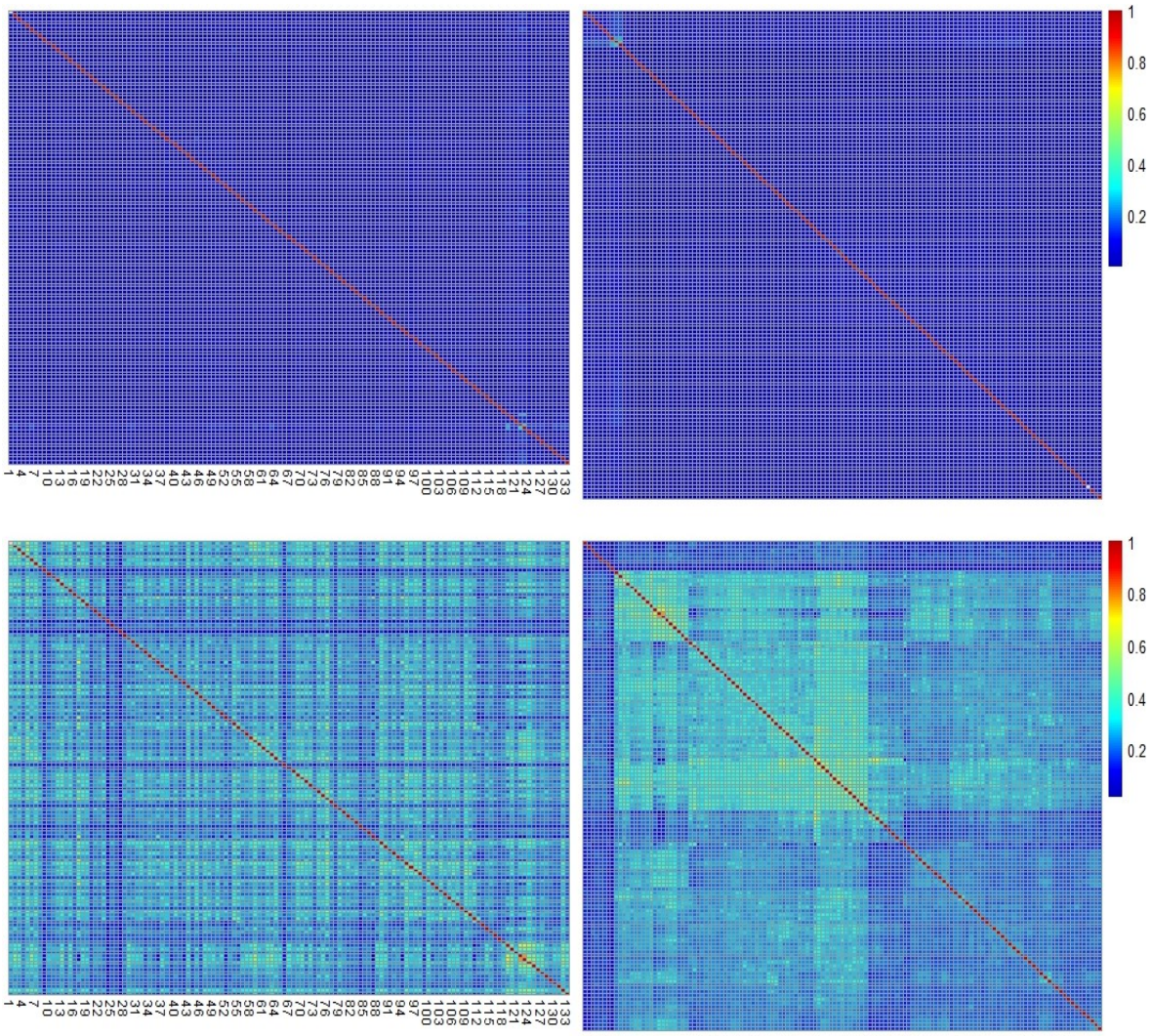

Figure 4.2: Top left: resting state RV connectivity matrix ordered by region. Top right: Resting state RV connectivity matrix clustered. Bottom left: motor task residual RV connectivity matrix ordered by region. Bottom right: motor task residual RV connectivity matrix clustered.

For the ***resting state data***, two threshold values were studied and used to assess which regions were statistically significant. We used $\tau_{.9997} = 3$ — as done by Ting et al. (2020))— which defines the 99.97th percentile from the standard normal distribution. Using this value of τ, we found all pairs of regions i, j such that $t_{RV_{ij}} > \tau_{.9997}$. This leads to 25 regions being identified as statistically significant—displayed in table 4.1. Alternatively, using the Bonferonni correction, we found a more precise threshold of

$\tau_b = 4.3889$. This comes from using $N(0,1)$ to find the critical value given a probability $p = 1 - \frac{\alpha}{(D(D-1))/2}$, where $\alpha = 0.05$ and $D = 133$ is the number of regions compared. Using the Bonferroni correction reduces the number of statistically significant pairs of regions down from 25 to only 2 (displayed as the first two rows in Table 4.1).

An important outcome to note is that regions satisfying the threshold condition for statistical significance with $\tau_{.9997}$ differ from regions which correspond to the 25 largest RV values. This discrepancy is due to a limitation of the RV coefficient (investigated in 5.1) and did not have the same influence with the background connectivity results. Furthermore, valuable insights can be gained by looking at both measures. Thus, table 4.2, displays the top 25 RV values along with their respective regions and networks. To limit the scope of the results discussion we will first explore the three largest t_{RV} values from table 4.1 followed by the three pairs associated with the largest RV values.

The top three pairs of regions from table 4.1 are precisely the ones with $t_{RV_{ij}} > \tau_b$. They include the inferior parietal lobule (IPL) region—for both hemispheres—and the parietal occipital lobe (ParOcc) in the left hemisphere as shown in figure 4.3. It is known that default mode network (DMN) contains the angular gyrus, dorsal medial prefrontal cortex and precuneus (or posterior cingulate cortex) Buckner (2013). Moreover, the angular gyrus is contained within the IPL and the precuneus is bordered by the ParOcc lobe, thus giving a possible indication behind why these regions showed strong functional connectivity within the rest state data (Tanaka and Kirino, 2019). Smith et al. (2013a) saw similar activation patterns when assessing group average resting state connectivity on HCP data. High connectivity was shown between the retrosplenial cortex, parieto-occipital suculus and regions within the DMN. The retrosplenial cortex has deep connections to both cortical and subcortical regions (Mitchell et al., 2018), thus demonstrating an advantage of including subcortical regions in this analysis.

By looking at table 4.2 containing the top 25 RV coefficients, we notice a similar pattern to that mentioned above, as the largest coefficient values are also in subcortical regions. Specifically, the cerebellum and brain stem are the only two subcortical regions present in the top three pairs. The cerebellum and brain stem are foundational regions within the brain and are known to play a role in many of the brain's oldest functions (Klein et al., 2016). This demonstrates the benefit of including whole brain data and supports the finding in Smith et al. (2013a). Finally, we note that although only one of the top three pairs from table 4.2 show up in table 4.1, all three t_{RV} values are very close to the $\tau_{.9997}$ threshold.

The ***background connectivity*** matrices, derived from analyzing residual activation signatures of the motor task, are displayed in the bottom left and bottom right panels of figure 4.2. Using the previously defined Bonferonni correction threshold ($\tau_b = 4.3889$) on the background connectivity t_{RV} values, gave over 2000 pairs of connections being deemed statistically significant. To compare the results to resting state

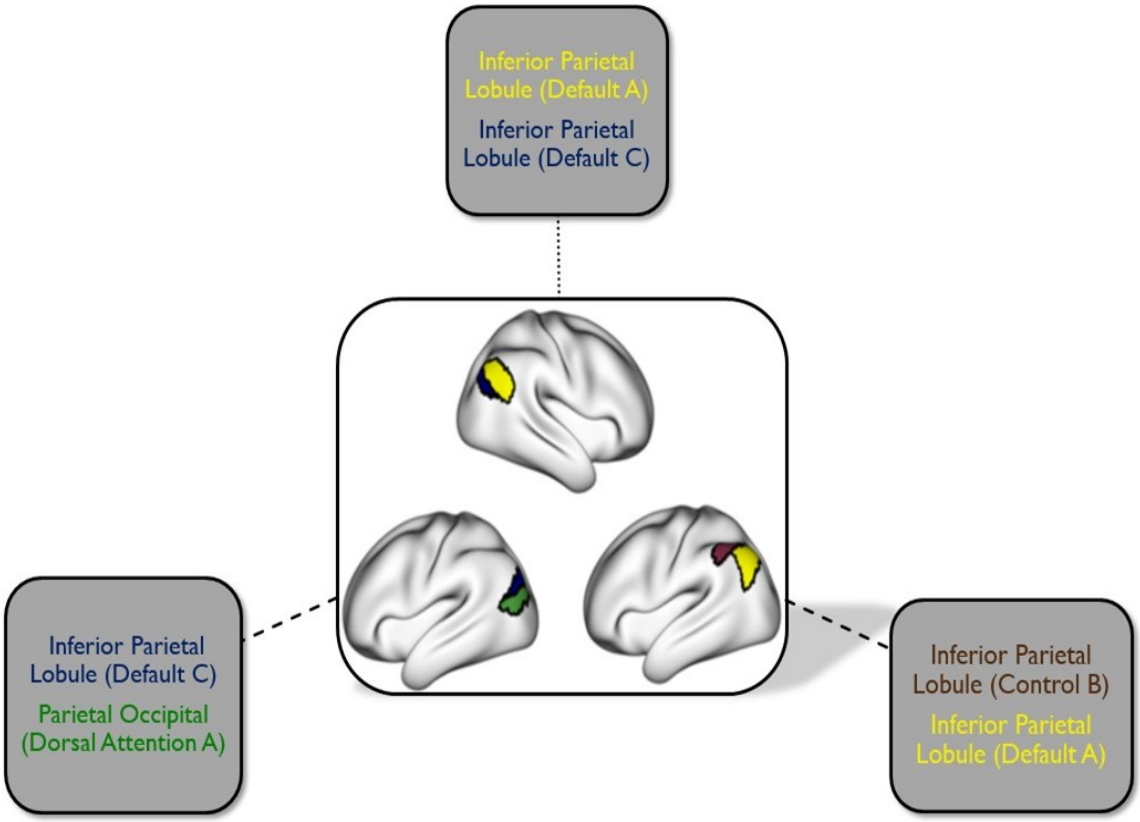

Figure 4.3: Visualization of the top three pairs of regions (and networks) corresponding to the three largest t_{RV} values from the rest state analysis.

values, we displayed the region and network pairs associated with only the 25 largest t_{RV} values in table 4.3. As will be addressed further in the limitations section (5.1), the residual connectivity was less impacted by the instability of the RV coefficients. Consequently, the same pairs of regions and networks show up when looking at the largest 25 RV or t_{RV} values, with the only difference being some variation of ordering.

The six pairs of regions with the strongest connectivity will be discussed in detail. Of the six connections with the largest observed test statistic (t_{RV}), three sets of pairs are contained entirely in the subcortical area and three sets of pairs within the cortical area of the brain (displayed in figure 4.4). Similar to the resting state connectivity, we see that the cerebellum and brain stem are functionally connected and activated with a high level of correlation (RV coefficients of 0.7813, 0.6331 and 0.6002). A key function of the cerebellum is the role it plays in muscle control and movement, thus demonstrating a plausible reason why the strength of connection shows an increase of almost four times the magnitude of those found while the subject was at rest.

It is important to note that of the three pairs of statistically significant regions not contained in the subcortical network (rows 2, 3, and 5 of table 4.3), the left hemisphere's pre-central ventral (PrCv) region appears in all of them. The precentral ventral is the

Table 4.1: Region and network pairs from the resting state connectivity matrix that have $t_{RV} > \tau_{.9997}$

tRV	**RV**	**Correlated Regions**	**Correlated Networks**
5.0908	0.1515	IPL–L, IPL–L	Default A, Control B
4.5355	0.1354	IPL–L, ParOcc–L	Default C, DorsAttnA
4.3761	0.1300	IPL–R, IPL–R	Default C, Default A
4.1440	0.1244	IPL–R, IPL–L	Control B, Control B
4.0803	0.1296	PFCv–R, Ins–R	Default B, Sal/VentAttnB
3.9152	0.1175	IPL–R, IPL–L	Control B, Default A
3.8858	0.1147	IPL–R, IPL–L	Default C, Default C
3.8360	0.1174	IPL–R, IPS–R	Control B, Control A
3.8289	0.1148	IPL–R, IPL–R	Default C, DorsAttnA
3.8057	0.1213	ExStrSup–L, StriCal–L	Visual B, Visual B
3.7308	0.2749	Cerebellum–R, Cerebellum–L	Subcortical, Subcortical
3.6226	0.1080	IPL–L, IPL–L	Default C, Default A
3.4404	0.1044	pCunPCC–L, Cingp–L	Default A, Control C
3.4403	0.1064	IPL–L, IPS–L	Control B, Control A
3.4281	0.1022	IPL–R, IPL–L	Default C, Default A
3.3979	0.1034	IPL–R, IPL–R	Default A, Control B
3.3715	0.1027	IPL–L, PFCl–L	Default A, Control B
3.3698	0.1066	PFCd–R, FEF–R	Control A, DorsAttnB
3.3403	0.1015	IPL–R, IPL–L	Default A, Default A
3.3100	0.0988	IPL–R, IPL–R	Control B, Sal/VentAttnB
3.1926	0.1063	PFCv–R, OFC–R	Default B, Limbic B
3.1524	0.0970	IPL–L, IPL–L	Default B, Control B
3.0225	0.1115	ExStr–R, ExStr–L	Visual A, Visual A
3.0130	0.1453	Cerebellum–L, ExStr–R	Subcortical, Visual A
3.0092	0.0894	IPL–R,TempOcc–R	Sal/VentAttnB, DorsAttnB

frontal part of the precentral gyrus which, itself, contains the motor cortex. The motor cortex is known primarily for its role in executing and planning voluntary movements of the body (Yip and Lui, 2022).

There are also strong correlations between the mid-cingulate region and the PrCv. The mid-cingulate cortex is contained in the anterior cingulate cortex, which is a subset of the frontal area within the cingulate cortex in the brain. The dorsal portion of the anterior cingulate cortex is known for playing a role in top-down and bottom-up processing when the brain needs to assign tasks or orders to other regions (Jumah and Dossani, 2022). These capabilities give us important evidence on how the brain's background connectivity patterns may assist in maintaining the task state. Kahali et al. (2021) saw similar connections arise in their analysis—executed to assess how to effectively parcellate the brain into regions based on their correlation. Specifically, they

Table 4.2: Pairs of regions (w/ hemisphere) and networks corresponding to the top 25 RV coefficients from the resting state connectivity matrix.

tRV	**RV**	**Correlated Regions**	**Correlated Networks**
3.7308	0.2749	Cerebellum–R, Cerebellum–L	Subcortical, Subcortical
2.9708	0.2406	Cerebellum–R, Brain Stem	Subcortical, Subcortical
2.8499	0.2372	Cerebellum–L, Brain Stem	Subcortical, Subcortical
2.5017	0.1571	Cerebellum–R, SomMotA–R	Subcortical, SomMotA
2.4596	0.1558	Cerebellum–R, SomMotA–L	Subcortical, SomMotA
2.4212	0.1546	Cerebellum–L, SomMotA–L	Subcortical, SomMotA
2.3854	0.1539	Cerebellum–L, SomMotA–R	Subcortical, SomMotA
5.0908	0.1515	IPL–L, IPL–R	Default A, Control B
1.6752	0.1507	Thalamus–L, Cerebellum–R	Subcortical, Subcortical
1.4831	0.1499	Thalamus–R, Cerebellum–R	Subcortical, Subcortical
1.4944	0.1499	Thalamus–R, Cerebellum–L	Subcortical, Subcortical
1.6229	0.1491	Thalamus–L, Cerebellum–L	Subcortical, Subcortical
2.8951	0.1474	Cerebellum–R, ExStr–L	Subcortical, Visual A
2.8907	0.1471	Cerebellum–L, ExStr–L	Subcortical, Visual A
1.4089	0.1453	Putamen–L, Cerebellum–L	Subcortical, Subcortical
1.3964	0.1453	Putamen–L, Cerebellum–R	Subcortical, Subcortical
3.0130	0.1453	Cerebellum–L, ExStr–R	Subcortical, Visual A
2.9077	0.1427	Cerebellum–R, ExStr–R	Subcortical, Visual A
1.4177	0.1398	Hippocampus–R, Cerebellum–R	Subcortical, Subcortical
1.4150	0.1394	Hippocampus–R, Cerebellum–L	Subcortical, Subcortical
1.7465	0.1391	Cerebellum– R, OFC–L	Subcortical, Limbic B
1.7181	0.1381	Cerebellum–L, OFC–L	Subcortical, Limbic B
1.7181	0.1371	Brain Stem, SomMotA–R	Subcortical, SomMotA
1.5455	0.1367	Diencephalon–L, Cerebellum–R	Subcortical, Subcortical
4.5355	0.1354	IPL–L, ParOcc–L	Default C, DorsAttnA

looked at how hand, foot and tongue movements were related. The authors reported that regions of high functional connectivity were split across the somatomotor network. Additionally, Kahali et al. (2021) found connections present between the somatomotor network and the mid-cingulate region which aligns with both Cingm-R and Cingm-L displaying high activation values within the background connectivity matrix.

Finally, we will briefly present the background activation signatures that were found when studying the induced covariance matrix from our spatiotemporal model. As mentioned in section 3.3, the induced background connectivity matrix comes from the model assumption that $\mathbf{E}^* \sim N_{TW}(\mathbf{0}, \Sigma^W)$. After transforming this matrix into the ROI space and applying a modified version of the RV coefficient, we obtained the background connectivity matrix, displayed in figure 4.5. It is evident that similar patterns are emerging as those displayed in the bottom left panel of figure 4.2. Al-

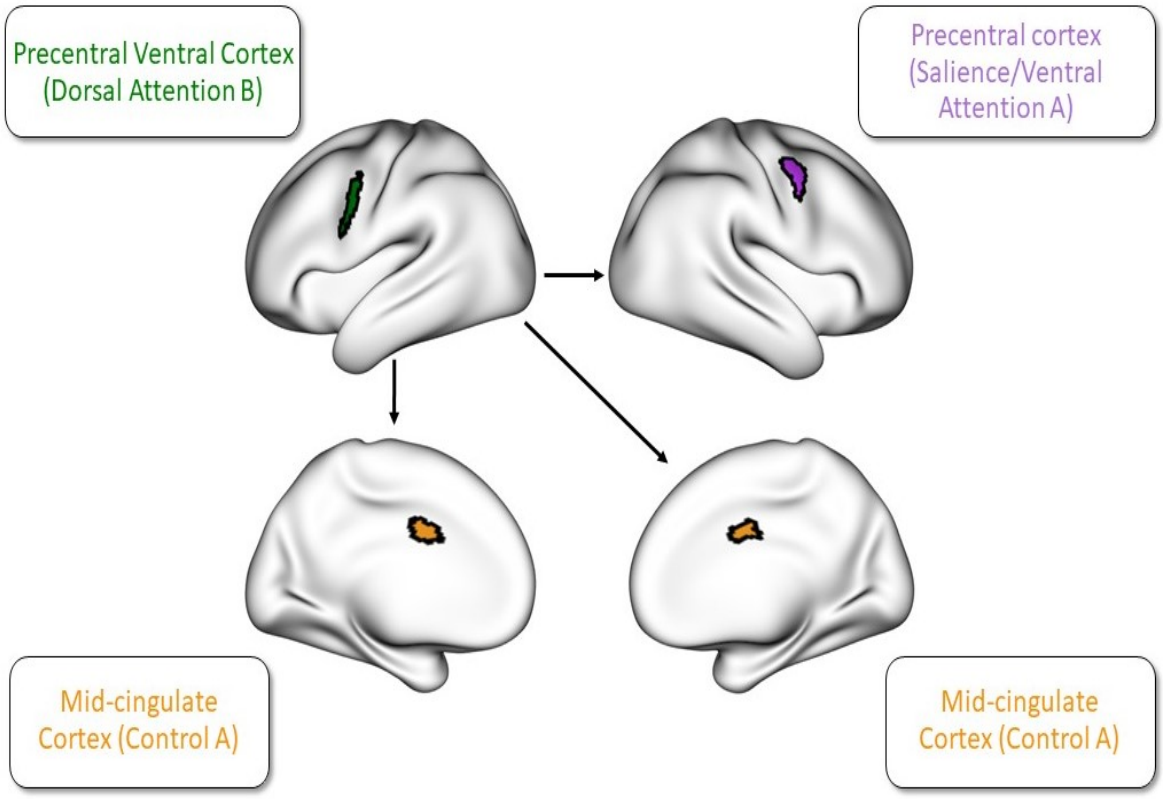

Figure 4.4: Visualization of the top three pairs of regions (and networks) contained entirely within the left and right hemispheres from the background connectivity analysis. Since the precentral ventral cortex is present in each of these pairs, it is displayed in the top left corner with the arrows indicating the three other regions it has high connectivity with.

though there appears to be a relative decrease in strength across all connections, the most prominent areas of connectivity with both the subcortical and cortical regions are analogous to those seen above. The connections with the largest RV value are, consistently, within the subcortical regions 119 and 121. Moreover, we see increased activation in columns 1-7 and 55-65, echoing that seen in the bottom left and right panels of figure 4.2. The similarities in these patterns provide support for the results obtained using the two-stage procedure. Specifically, it emphasizes that the spatial and temporal transformations were effective at accounting for the biological characteristics and complexities in the brain. Consistency in activation patterns affirms the analysis procedure taken and the resulting background connectivity patterns found for the motor task data. In the remaining section we will discuss the primary limitations of this study and the ways in which the dataset structure mitigated some of these weaknesses as well as areas for future work.

Table 4.3: Pairs of regions (w/ hemisphere) and networks corresponding to the top 25 t_{RV} values from the motor task background connectivity matrix.

tRV	**RV**	**Correlated Regions**	**Correlated Networks**
12.9887	0.7813	Cerebellum–R, Cerebellum–L	Subcortical, Subcortical
10.7403	0.6445	PrC–R, PrCv–L	SalVentAttnA, DorsAttnB
10.7210	0.6434	Cingm–R, PrCv–L	ContA, DorsAttnB
10.4849	0.6331	Cerebellum–R, Brain Stem	Subcortical, Subcortical
10.2814	0.6172	Cingm–L, PrCv–L	ContA, DorsAttnB
9.9243	0.6002	Cerebellum–L, Brain Stem	Subcortical, Subcortical
9.3923	0.5657	ExStrSup–L, StriCal–L	VisPeri, VisPeri
9.3594	0.5635	PFCd–R, PrC–R	ContA, SalVentAttnA
9.0055	0.5454	Diencephalon-L, Brain Stem	Subcortical, Subcortical
8.8763	0.5337	Cingp–R, PrCv–L	ContC, DorsAttnB
8.7784	0.5296	ExStrSup–L, Striate–L	VisPeri, VisCent
8.7641	0.5280	PrC–R, PrCv–R	SalVentAttnA, DorsAttnB
8.6756	0.5216	Cinga–R, PrCv–L	SalVentAttnB, DorsAttnB
8.5997	0.5186	OFC–R, ExStrSup–L	LimbicB, VisPeri
8.5929	0.5179	pCun–R, ExStrSup–L	ContC, VisPeri
8.5923	0.5176	PrC–R, Cent–L	SalVentAttnA, SomMotB
8.5726	0.5221	Cerebellum–R, ExStr–R	Subcortical, VisCent
8.5457	0.5142	PFCd–R, PrC–R	DefaultB, SalVentAttnA
8.4945	0.5115	PrC–R, ExStrSup–L	SalVentAttnA, VisPeri
8.4706	0.5145	Diencephalon–L, Cerebellum–R	Subcortical, Subcortical
8.4650	0.5103	PrC–R, Cent–R	SalVentAttnA, SomMotB
8.4241	0.5163	ExStr–R, ExStr–L	VisCent, VisCent
8.3767	0.5061	Cent–R, Cent–L	SomMotB, SomMotB
8.3603	0.5048	ExStrSup–R, ExStrSup–L	VisPeri, VisPeri
8.3390	0.5088	Cerebellum–L, ExStr–R	Subcortical, VisCent

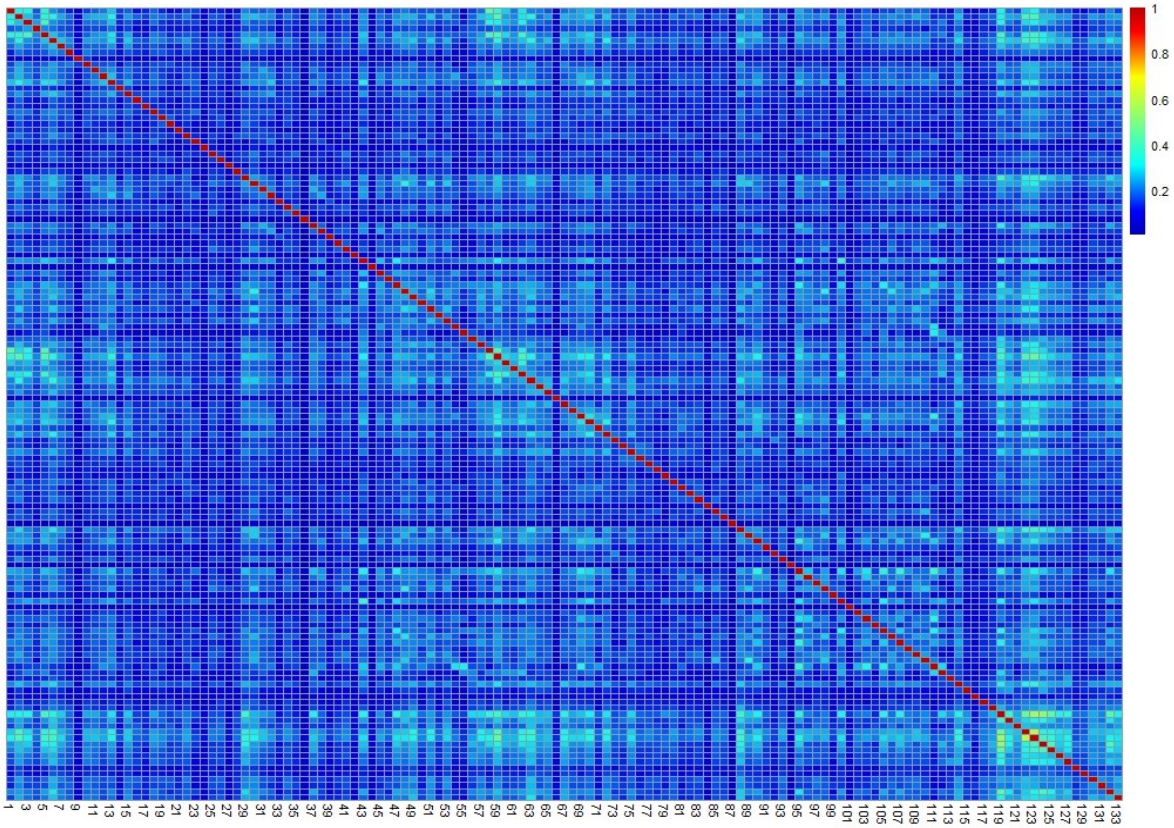

Figure 4.5: Background activation signatures found using the composite hybrid spatial basis induced residual covariance matrix.

Chapter 5
Conclusion

5.1 Limitations

As with all statistical procedures, assumptions and asymptotic properties often have caveats that impact the results obtained, therefore they must be mentioned. First, we look at a limitation of the t_{RV} asymptotic distribution. The t_{RV} values are said to be asymptotically standard normal, however, the null distribution depends on the size of the time-series (T) (Ting et al., 2020). This dependency diminishes the benefit of directly comparing the exact t_{RV} values between our rest and residual connectivity matrices given that T is 1200 and 284 for these two datasets, respectively. Moreover, the RV values have reduced stability when the $\Sigma_{f_i f_i}$ matrices have a high dimension relative to T (Ting et al., 2020).

Generalized versions of the RV coefficient as well as other null distributions which can diminish these tendencies are discussed in Josse and Holmes (2016). These constraints and their impact on the results were reduced due to certain properties of the datasets and projection procedures used here. First, we looked at the dimension of all $\Sigma_{f_i f_i}$ matrices for both rest and residual datasets (refer to equation 3.2 for details on $\Sigma_{f_i f_i}$). Three regions within the resting state data were associated to projection matrices of high dimension and were thus analyzed further. Regions 119, 122, and 123 had projection matrices of size 203, 271 and 269, respectively. Recall that these dimensions are defined by the number of principal components which describe at least 50% of the variation in region k and $\Sigma_{f_i f_i}$ is of size p_k x p_k for $k = 1, \ldots, 133$. All other regions within the resting state transformations had projection dimensions less than 100 with a median dimension of 22. The larger dimensions, however, are reasonable when we consider their biological properties and the number of locations contained in these three regions. That is, 119, 122 and 123 correspond to the subcortical regions brain stem, cerebellum-L and cerebellum-R, which physically account for more neurons than all other regions combined (Klein et al., 2016). Moreover, the relative complexity of these three transformed matrices do not align exactly with the limitation mentioned above due to the fact that these three values are relatively small compared to our

$T = 1200$. We can see, however, that these attributes of the subcortical regions may be, in part, the cause of our resting state data showing slightly different results when we compare the RV and t_{RV} tables. Lastly, we assessed the connectivity patterns for both measures of connection (t_{RV} and RV) which further diminish this undesirable characteristic of the standardized RV coefficient.

When examining the background connectivity processes for the aforementioned properties we found that the $\Sigma_{f_i f_i}$ dimensions did not show any notable inconsistencies. This is consistent with the little variability seen when comparing the background connectivity RV and t_{RV} values. Furthermore, we reduce the impact these inconsistencies may have on the results by interpreting and displaying all outputs in a variety of ways. This allows both the t_{RV} and RV values to illustrate different patterns within the data. A final constraint of note is that the procedures applied above were done so on only one subject. Applying this same procedure to multiple subjects, using the same HCP data for each subject would greatly improve the strength of evidence gained from the results presented above.

5.2 Discussion and Future Work

We have demonstrated the benefits of comparing background connectivity activation signatures with resting state connectivity using the newest nueroimaging file structure combined with a composite-hybrid Bayesian modelling framework. The use of the CIFTI file provides benefits that were two-fold. First, it proved the flexibility of the composite hybrid model, built by Miranda and Morris (2021). This model was originally built to be used on the NIFTI, volumetric data as well as for the working memory task, however, this analysis demonstrated the ease of applying modifications to incorporate the motor task and CIFTI data which neither diminished the effectiveness nor increased computational time. Additionally, the CIFTI data allowed for the whole brain to be part of the analysis process, rather than just the hemispheres. The subcortical region contains critical brain structures which are now substantially easier to incorporate into connectivity analyses, thus providing further evidence and understanding of the entire brain and its connection patterns during rest and task states.

There is a lack of general consensus regarding the nature of our brain's background connectivity and the role it plays in maintaining or inducing task states within the brain—if it does so at all. Here we saw a variety of activation patterns arise when comparing the background connectivity to the resting state connectivity. Certain subcortical regions—specifically the cerebellum and brain stem—were highly active and functionally connected when Subject 100307 was both at rest and completing the motor task. These were the primary similarities between brain states. Alternatively, we saw pairs of regions which are part of the default mode network that displayed sta-

tistically significant activation signatures during rest but were not notably significant within the background connectivity patterns. Another important difference was that the precentral ventral region and mid-cingulate regions showed strong functional connections within the background activation signatures. Most significantly, these regions contain areas which are part of the motor cortex while also being responsible for assigning tasks to other areas of the brain. These factors provide important insights into on-going processes within the brain which are not due to task-dependent stimulus.

The findings discussed here give reason to suggest that background connectivity within the brain may be playing a more important role than previously believed. This conclusion, however, cannot be made without doing further analyses on multiple subjects as well as across a variety of tasks. Using the HCP CIFTI data from the 1200 subjects release makes it straightforward to extend these analyses to a broader context, thus providing insight on both of these areas.

Supplementary Material

All R and MATLAB code used for this analysis can be found here: `https://github.com/Mjholmes13/Background-Connectivity-Code`

Appendix A
Acronyms

Table A.1: Parcel abbreviations and corresponding full names.

Abbreviation	Full Parcel Name	Abbreviation	Full Parcel Name
AntTemp	anterior temporal	PFCd	dorsal prefrontal cortex
Aud	auditory	PFCl	lateral prefrontal cortex
Cent	central	PFCld	lateral dorsal prefrontal cortex
Cinga	cingulate anterior	PFClv	lateral ventral prefrontal cortex
Cingm	mid-cingulate	PFCm	medial prefrontal cortex
Cingp	cingulate posterior	PFCmp	medial posterior prefrontal cortex
ExStr	extrastriate cortex	PFCv	ventral prefrontal cortex
ExStrInf	extra-striate inferior	PHC	parahippocampal cortex
ExStrSup	extra-striate superior	PostC	post central
FEF	frontal eye fields	PrC	precentral
FPole	frontal pole	PrCd	precentral dorsal
FrMed	frontal medial	PrCv	precentral ventral
FrOper	frontal operculum	RSC	retrosplenial cortex
IFG	inferior frontal gyrus	Rsp	retrosplenial
Ins	insula	S2	S2
IPL	inferior parietal lobule	SPL	superior parietal lobule
IPS	intraparietal sulcus	ST	superior temporal
OFC	orbital frontal cortex	Striate	striate cortex
ParMed	parietal medial	StriCal	striate calcarine
ParOcc	parietal occipital	Temp	temporal
ParOper	parietal operculum	TempOcc	temporal occipital
pCun	precuneus	TempPar	temporal parietal
pCunPCC	precuneus posterior cingulate cortex	TempPole	temporal pole

Table A.2: Network abbreviations and corresponding full names.

Network Abbreviation	**Network Full Name**
VisCent	Visual A
VisPeri	Visual B
SomMotA	Somatomotor A
SomMotB	Somatomotor B
DorsAttnA	Dorsal Attention A
DorsAttnB	Dorsal Attention B
SalVentAttnA	Salience/Ventral Attention A
SalVentAttnB	Salience/Ventral Attention B
LimbicB	Limbic B
LimbicA	Limbic A
ContA	Control A
ContB	Control B
ContC	Control C
DefaultA	Default A
DefaultB	Default B
DefaultC	Default C
TempPar	Temporal Parietal

References

Arieli, A., Sterkin, A., Grinvald, A., and Aertsen, A. (1996). Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science*, 273(5283):1868–1871.

Buckner, R. L. (2013). The brain's default network: origins and implications for the study of psychosis. *Dialogues Clin. Neurosci.*, 15(3):351–358.

Buimer, E., Pas, P., Brouwer, R., Froeling, M., Hoogduin, H., Leemans, A., Luijten, P., Nierop, B., Raemaekers, M., Schnack, H., Teeuw, J., Vink, M., Visser, F., Pol, H., and Mandl, R. (2020). The youth cohort study: Mri protocol and test-retest reliability in adults. *Developmental Cognitive Neuroscience*, 45:100816.

Cardoso, M. M. B., Sirotin, Y. B., Lima, B., Glushenkova, E., and Das, A. (2012). The neuroimaging signal is a linear sum of neurally distinct stimulus- and task-related components. *Nat. Neurosci.*, 15(9):1298–1306.

David, C., Van Essen, S. M., Smith, D. M., Barch, T. E. J., and Behrens, E. (2013). The WU-Minn human connectome project: An overview. *NeuroImage*, 80:62–79.

Elkhetali, A. S., Fleming, L. L., Vaden, R. J., Nenert, R., Mendle, J. E., and Visscher, K. M. (2019). Background connectivity between frontal and sensory cortex depends on task state, independent of stimulus modality. *Neuroimage*, 184:790–800.

Fadili, M. J. and Bullmore, E. T. (2002). Wavelet-generalized least squares: a new BLU estimator of linear regression models with 1/f errors. *Neuroimage*, 15(1):217–232.

Fan, Y. (2003). On the approximate decorrelation property of the discrete wavelet transform for fractionally differenced processes. *IEEE Transactions on Information Theory*, 49(2):516–521.

Fox, M. D. and Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat. Rev. Neurosci.*, 8(9):700–711.

Friston, K. J., Frith, C. D., Liddle, P. F., and Frackowiak, R. S. (1993). Functional connectivity: the principal-component analysis of large (PET) data sets. *J. Cereb. Blood Flow Metab.*, 13(1):5–14.

Gao, Z., Liu, X., Zhang, D., Liu, M., and Hao, N. (2021). Subcortical structures and visual divergent thinking: a resting-state functional MRI analysis. *Brain Struct. Funct.*, 226(8):2617–2627.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., and Jenkinson, M. (2013). The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, 80:105–124. Mapping the Connectome.

Glover, G. H. (2011). Overview of functional magnetic resonance imaging. *Neurosurg. Clin. N. Am.*, 22(2):133–9, vii.

Hart, B., Cribben, I., Fiecas, M., and for the Alzheimer's Disease Neuroimaging Initiative (2017). A longitudinal model for functional connectivity networks using resting-state fmri. *bioRxiv*.

He, B. J. (2013). Spontaneous and task-evoked brain activity negatively interact. *Journal of Neuroscience*, 33(11):4672–4682.

Hosseinzadeh, M. (2020). 4 - robust control applications in biomedical engineering: Control of depth of hypnosis. In Azar, A. T., editor, *Control Applications for Biomedical Engineering Systems*, pages 89–125. Academic Press.

Josse, J. and Holmes, S. (2016). Measuring multivariate association and beyond. *Statistics Surveys*, 10(none):132 – 167.

Josse, J., Pagès, J., and Husson, F. (2008). Testing the significance of the rv coefficient. *Comput. Stat. Data Anal.*, 53(1):82–91.

Jumah, F. R. and Dossani, R. H. (2022). *Neuroanatomy, Cingulate Cortex*. StatPearls Publishing.

Kahali, S., Raichle, M. E., and Yablonskiy, D. A. (2021). The role of the human brain neuron-glia-synapse composition in forming resting-state functional connectivity networks. *Brain Sci.*, 11(12):1565.

Klein, A. P., Ulmer, J. L., Quinet, S. A., Mathews, V., and Mark, L. P. (2016). Nonmotor functions of the cerebellum: An introduction. *AJNR Am. J. Neuroradiol.*, 37(6):1005–1009.

Lee, K.-J., Jones, G. L., Caffo, B. S., and Bassett, S. S. (2014). Spatial bayesian variable selection models on functional magnetic resonance imaging time-series data. *Bayesian Anal.*, 9(3):699–732.

Miranda, M. and Morris, J. (2021). Novel bayesian method for simultaneous detection of activation signatures and background connectivity for task fmri data.

Mitchell, A. S., Czajkowski, R., Zhang, N., Jeffery, K., and Nelson, A. J. D. (2018). Retrosplenial cortex and its role in spatial cognition. *Brain Neurosci. Adv.*, 2:2398212818757098.

Morris, J. S. (2015). Functional regression. *Annual Review of Statistics and Its Application*, 2(1):321–359.

Penny, W. D., Trujillo-Barreto, N. J., and Friston, K. J. (2005). Bayesian fMRI time series analysis with spatial priors. *Neuroimage*, 24(2):350–362.

Percival, D. and Walden, A. (2000). Wavelet methods for time series. *Wavelet Methods for Time Series Analysis*.

Pham, D. D., Muschelli, J., and Mejia, A. F. (2022). ciftitools: A package for reading, writing, visualizing, and manipulating cifti files in r. *NeuroImage*, 250:118877.

Quirós, A., Diez, R. M., and Gamerman, D. (2010). Bayesian spatiotemporal model of fMRI data. *Neuroimage*, 49(1):442–456.

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., and Yeo, B. T. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex*, 28(9):3095–3114.

Shahhosseini, Y. and Miranda, M. F. (2022). Functional connectivity methods and their applications in fMRI data. *Entropy (Basel)*, 24(3):390.

Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D. A., Griffanti, L., Harms, M. P., Kelly, M., Laumann, T., Miller, K. L., Moeller, S., Petersen, S., Power, J., Salimi-Khorshidi, G., Snyder, A. Z., Vu, A. T., Woolrich, M. W., Xu, J., Yacoub, E., Uğurbil, K., Van Essen, D. C., Glasser, M. F., and WU-Minn HCP Consortium (2013a). Resting-state fMRI in the human connectome project. *Neuroimage*, 80:144–168.

Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D. A., Griffanti, L., Harms, M. P., Kelly, M., Laumann, T., Miller, K. L., Moeller, S., Petersen, S., Power, J., Salimi-Khorshidi, G.,

Snyder, A. Z., Vu, A. T., Woolrich, M. W., Xu, J., Yacoub, E., Uğurbil, K., Van Essen, D. C., and Glasser, M. F. (2013b). Resting-state fmri in the human connectome project. *NeuroImage*, 80:144–168. Mapping the Connectome.

Sylvester, C. M., Shulman, G. L., Jack, A. I., and Corbetta, M. (2009). Anticipatory and stimulus-evoked blood oxygenation level-dependent modulations related to spatial attention reflect a common additive signal. *J. Neurosci.*, 29(34):10671–10682.

Tanaka, S. and Kirino, E. (2019). Increased functional connectivity of the angular gyrus during imagined music performance. *Frontiers in Human Neuroscience*, 13.

Ting, C.-M., Ombao, H., Salleh, S.-H., and Latif, A. Z. A. (2020). Multi-scale factor analysis of high-dimensional functional connectivity in brain networks. *IEEE Transactions on Network Science and Engineering*, 7(1):449–465.

Wald, L. and Polimeni, J. (2015). High-speed, high-resolution acquisitions. In Toga, A. W., editor, *Brain Mapping*, pages 103–116. Academic Press, Waltham.

West, K. L., Zuppichini, M. D., Turner, M. P., Sivakolundu, D. K., Zhao, Y., Abdelkarim, D., Spence, J. S., and Rypma, B. (2019). Bold hemodynamic response function changes significantly with healthy aging. *NeuroImage*, 188:198–207.

Woodward, N. D. and Cascio, C. J. (2015). Resting-state functional connectivity in psychiatric disorders. *JAMA Psychiatry*, 72(8):743.

Worsley, K. J. and Friston, K. J. (1995). Analysis of fMRI time-series revisited–again. *Neuroimage*, 2(3):173–181.

Yip, D. W. and Lui, F. (2022). Physiology, motor cortical. In *StatPearls*. StatPearls Publishing.

Zhang, L., Guindani, M., Versace, F., Engelmann, J. M., and Vannucci, M. (2016). A spatiotemporal nonparametric Bayesian model of multi-subject fMRI data. *The Annals of Applied Statistics*, 10(2):638 – 666.

Zhang, L., Guindani, M., Versace, F., and Vannucci, M. (2014). A spatio-temporal nonparametric bayesian variable selection model of fMRI data for clustering correlated time courses. *Neuroimage*, 95:162–175.