

Optimising Network Modelling Methods for fMRI

Usama Pervaiz

Supervised by: Stephen Smith, Mark Woolrich

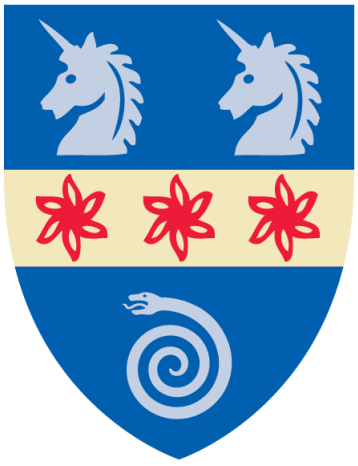
St Hilda's College

University of Oxford

A thesis presented for the degree of

Doctor of Philosophy in Clinical Neuroscience

Trinity 2021



I dedicate this to my loving parents and my partner ...

Copyright © 2021 by Usama Pevaiz
All Rights Reserved

All truths are easy to understand once they are discovered; the point is to discover them.

— Galileo Galilei

ACKNOWLEDGEMENTS

Firstly, I would like to express utmost gratitude to my supervisors, Stephen Smith and Mark Woolrich, from whom I had the opportunity to learn so much. I am grateful for their continuous support, patience, and kindness for the last three years, and that's all I could have ever wished for.

I am incredibly thankful to all my colleagues at OHBA and FMRIB. I feel fortunate to have had the opportunity to work with such a unique and talented bunch of individuals who will continue to inspire me. Finally, and most importantly, I would like to thank my parents and siblings for always believing in me; it would not have been possible without your continuous support. Special thanks to my kind and loving partner, who always motivated and guided me throughout this journey.

DECLARATION

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, other than me.

The only material taken for this thesis is from my first-authors published work that are listed below:

1) Pervaiz U, Vidaurre D, Woolrich MW, Smith SM. Optimising network modelling methods for fMRI. Neuroimage. 2020 May 1;211:116604.

2) Pervaiz U, Vidaurre D, Gohil C, Smith SM, Woolrich MW. Multi-dynamic Modelling Reveals Strongly Time-varying Resting fMRI Correlations. Accepted in Medical Image Analysis. 2021 August 1.

Name: Usama Pervaiz *Date: 27-09-2021*

Abstract

The activity of functional brain networks is responsible for the emergence of time-varying cognition and behaviour. Accordingly, functional connectivity (FC) in fMRI has been shown to be predictive of behavioural traits and psychiatric and neurological conditions.

Many studies estimate FC by calculating the average correlation across an entire fMRI scanning session data (static FC). A major goal of such neuroimaging studies is to develop predictive models to analyze the relationship between whole-brain FC patterns and behavioural traits. However, there is no single widely accepted standard pipeline for analyzing static FC. We proposed six steps for designing FC-based predictive models. This standardized framework for estimating **static FC** should accelerate both congruence and scientific progress within the neuroscientific community.

However, there is also increasing interest in studying the time-varying nature of FC in fMRI (time-varying FC). Typically, methods that measure time-varying FC have several limitations that bias the estimation of time-varying FC, to appear more stable over time than it actually is. We designed a new method, MAGE, that offers a potential explanation and a solution for the homogeneity of time-varying FC seen in existing approaches. This new framework for estimating **time-varying FC** showed that FC is not static over time and fluctuates on a sub-minute time-scale during resting-state fMRI.

This work also investigates the clinical applications of both static and time-varying FC. We found that MAGE estimated time-varying FC can provide a more precise understanding of how brain networks are related to cognitive abilities, both in health and disease. Lastly, we demonstrated that MAGE can reliably capture the time-varying changes in FC across a range of behavioral and cognitive tasks, overcoming the limitations of most popular existing methods.

TABLE OF CONTENTS

1 Introduction **1**

1.1 Brain at Rest 1

1.2 Functional Brain Imaging: Towards a New Future 3

1.2.1 Resting-State fMRI 3

1.3 Resting State Networks (RSNs) 4

1.4 Functional Connectivity in fMRI 6

1.4.1 Functional Connectivity versus Effective Connectivity . . 7

1.4.2 Static versus Time-varying Functional Connectivity 8

1.5 Static Functional Connectivity Relevance in Health and Disease . 9

1.5.1 Neurological Conditions 9

1.5.2 Lifestyle and behaviour 10

1.6 Clinical and non-clinical relevance of Time-varying Functional Connectivity 11

1.6.1 Revisiting TVFC 11

1.6.2 Ongoing "task-related" Cognition 12

1.6.3 Ongoing "spontaneous" Cognition 12
1.6.4 Neurological Conditions 13
1.7 Thesis Objectives 14
1.7.1 Methods Development for Static Functional Connectivity Analysis 14
1.7.2 Methods Development for Time-varying Functional Connectivity Analysis 15
1.7.3 Clinical Applications of Static and Time-varying Functional Connectivity 16

2 General Materials **17**
2.1 Resting-state fMRI datasets 18
2.1.1 UK Biobank 18
2.1.2 Human Connectome Project 20
2.1.3 Autism brain imaging data exchange (ABIDE) 22
2.1.4 Addiction Connectome Preprocessed Initiative (ACPI) . . 23
2.2 Task-fMRI Datasets 24
2.2.1 Cognitive-task data-set 24
2.2.2 Movies-task data-set 26

3 Optimising Static Network Modelling Methods for fMRI **28**
3.1 Overview 28
3.2 Introduction 29
3.3 Methods 34
3.3.1 Finding the best parcellation approach 34
3.3.2 Finding the best correlation-based FC estimate method . . 37
3.3.3 Finding the best classifier/predictor 45
3.3.4 Training, validation and testing 50

3.3.5 Statistical testing . . . 51
3.3.6 Imaging Data . . . 52
3.3.7 Non-Imaging Data . . . 53
3.4 Results . . . 54
3.4.1 Parcellation and FC Estimation . . . 54
3.4.2 Tangent Space Projection and Shrinkage . . . 56
3.4.3 Deconfounding . . . 60
3.4.4 Predictors/Classifiers Performance . . . 62
3.4.5 Top Performing Configurations . . . 64
3.4.6 Joint Results . . . 67
3.4.7 Deconfounding revisited . . . 71
3.4.8 Comparison with other commonly used methodologies . . . 72
3.5 Discussion . . . 73
3.6 Conclusion . . . 84

4 Optimising Time-varying Network Modelling methods for fMRI 86
4.1 Overview . . . 86
4.2 Introduction . . . 87
4.3 Methods . . . 90
4.3.1 Multi-dynamic Model . . . 90
4.3.2 What do we mean by functional connectivity? . . . 95
4.3.3 Multi-dynamic Adversarial Generator-Encoder (MAGE) . . . 95
4.3.4 Dual-Estimation of the group-level MAGE . . . 103
4.3.5 Hyper-parameters tuning and cross-validation . . . 104
4.4 Results . . . 105
4.4.1 SAGE Performance on HMM-like Categorical Simulations 105
4.4.2 MAGE can infer dynamic linear mixtures of FC states . . . 108

4.4.3 Failing to model distinct mean activity dynamics can result in homogeneous estimation of FC over time 111
4.4.4 MAGE can infer dynamics in the FC that are distinct from dynamics in the means of brain activity 112
4.4.5 Multi-dynamic approach reveals stronger changes in FC over time in resting fMRI 113
4.4.6 Multi-dynamic TVFC predicts behavioural variability better than the single-dynamic approaches 122
4.5 Discussion 123
4.6 Conclusion 128

5 Clinical Applications of Static and Time-varying Functional Connectivity 129
5.1 Introduction 129
5.2 Static Functional Connectivity in the Clinical Realm 130
5.2.1 Neurological Conditions and Drug Addiction 130
5.2.2 Subject's Identifiability using Functional Connectome . . 134
5.2.3 Spotlight on the Mental Health 135
5.3 Contributions of Time-varying Functional Connectivity to Human Behaviour 142
5.3.1 Ongoing Task-related Cognition 142
5.3.2 Ongoing Spontaneous Cognition 148
5.3.3 Age-related Changes in Resting-state Functional Connectivity 150
5.4 Static FC is Worse than Time-varying FC at Predicting Behaviour 152
5.5 Discussion 154

6 General Discussion and Conclusion 156

6.0.1 Methodological Contributions 156
6.0.2 Achievements and Limitations in Clinical Applications . . 158
6.0.3 Reproducibility and Generalizability of Proposed Machine Learning Algorithms 160
6.0.4 Future of Functional Connectivity 161
S1 Supplementary Material of Chapter 3 163
S2 Supplementary Material of Chapter 4 180
S2.1 Method's Supplementary 180
S2.2 Result's Supplementary 188
S3 Supplementary Figures for Chapter 5 199
References 201

LIST OF ABBREVIATIONS

FC ..... Functional Connectivity

TVFC ..... Time-varying Functional Connectivity

MRI ..... Magnetic Resonance Imaging

fMRI ..... functional Magnetic Resonance Imaging

rfMRI ..... resting-state functional Magnetic Resonance Imaging

tfMRI ..... task functional Magnetic Resonance Imaging

HCP ..... Human Connectome Project

ABIDE ..... Autism Brain Imaging Data Exchange

ACPI ..... Addiction Connectome Preprocessed Initiative

ADHD ..... Attention Deficit Hyperactivity Disorder

MTA ..... Multimodal Treatment Study of ADHD

MS ..... Multiple Sclerosis

ASD . . . . . Autism Spectrum Disorder

ALS . . . . . Amyotrophic Lateral Sclerosis

FLIRT . . . . . FMRIB's Linear Image Registration Tool

BBR . . . . . Boundary-Based Registration

FIX . . . . . FMRIB's ICA-based X-noisefier

C-PAC . . . . . Pre-processed Connectome Project pipeline

MAGE . . . . . Multi-dynamic Adversarial Generator-Encoder

HMM . . . . . Hidden Markov Model

GAN . . . . . Generative Adversarial Network

CHAPTER 1

INTRODUCTION

1.1 Brain at Rest

Understanding the functioning of the human brain is one of the greatest intellectual quests of all times. It is arguably the most fascinating organ in the human body, but unfortunately, is still the least understood. In the last few decades, we are starting to comprehend that the human brain is always active on some level regardless of whether we are engaged in some task or not. However, it is worth noting that idea of the brain being active even when not engaged in an explicit task is quite old, as noted by *Senenca* in $\sim$ 60 A.D., "The fact that the body is lying down is no reason for supposing that the mind is at peace. Rest is... far from restful" (Campbell et al., 2004).

Though *Seneca* and others wondered about the "active" state of the human brain in an unstimulated "rest" condition, neuroscientists spent most of the 20^{th} century studying human brain activity in response to predefined external tasks. Crucially,

these studies paved the way towards a greater understanding of the function of brain regions and effectively allowed us to characterize different brain regions based on their engagement across external tasks. However, studying the brain only during predefined tasks has several limitations. Practically, we are exploring the possibility that the brain only becomes active or reacts when it is explicitly compelled to engage in a task. We know that is not the whole story, as we can all relate to experiences of mind wandering when we are less focused on the external environment and heavily occupied with thoughts that have nothing to do with external sensory information.

The motivation to understand spontaneous brain activity when the brain is unstimulated leads scientists to a challenging quest. In the past two decades, scientists have been increasingly focused on measuring the brain in the absence of a stimulus or a task. These studies have found that the resting brain is engaged in a substantial activity and that the brain activity at rest is not random. For example, it was first noted by (Biswal, Zerrin Yetkin, Haughton, & Hyde, 1995) that the resting-state activity of different brain regions is far from accidental, and the fluctuations across these regions exhibit a high degree of similarity. This study was conducted with functional magnetic resonance imaging (fMRI). From there onwards, a significant amount of research undertaken to understand spontaneous resting-state brain activity has relied on fMRI. In the following section, we will take a brief look into the background of fMRI.

1.2 Functional Brain Imaging: Towards a New Future

fMRI is a widely used neuroimaging technique to record brain activity non-invasively in humans. So, what does fMRI measure? fMRI provides an indirect estimate of brain activity. Compared to the standard electrophysiological techniques (i.e., EEG, MEG), fMRI does not measure the brain's electrical activity, but it indirectly captures the underlying neural activity of the brain. All the neural processes in the brain require energy; for example, when a specific brain region is "activated" in response to some cognitive task, it requires an increase in the supply of oxygen to process that task. The magnetic susceptibility of oxygenated versus deoxygenated blood is different, and fMRI utilizes this property to measure the blood-oxygen-level-dependent (BOLD) signal. fMRI has many strengths, but one of its major strengths is its excellent spatial resolution compared to electrophysiological data. This allows us to measure brain activity with incredible spatial precision.

On the contrary, the principal weakness of fMRI is its low temporal resolution as it is limited by the hemodynamic response time. Nevertheless, by developing appropriate analysis methods, we can cater to such temporal limitations to some extent. fMRI, especially resting-state fMRI, is a powerful resource to improve our understanding of brain function, both in health and disease. All the research presented in the thesis is conducted on human participants' fMRI data.

1.2.1 Resting-State fMRI

There are several motivations to working with resting-state fMRI data. The first reason is that it is easy to design such experiments with fewer constraints and parameters. Secondly, it allows us to study a broad range of subjects to understand

a wide spectrum of neurological conditions that may not be possible using task fMRI. For example, it is not possible to examine patients in an unconscious state using task-based experiments.

During the last decade, one significant factor that has led to the broader acceptance of resting-state fMRI is identifying brain activity patterns during rest that resemble those seen involved in a task. (S. M. Smith et al., 2009) demonstrated that the set of activity patterns seen in resting-state data closely matches the activity patterns under different task conditions. Another valuable piece of work was by (Damoiseaux et al., 2006), which showed that across several healthy subjects in resting-state, the activity patterns were found to be spatially consistent. It was a surprising finding as different people scanned in a resting state are most probably involved in a diverse set of thoughts.

1.3 Resting State Networks (RSNs)

(Biswal et al., 1995) performed an unusual experiment where authors advised their fMRI test subjects to do nothing. It was expected that the brain activity at rest would be random, but it was found that activity of different regions in the sensorimotor brain system was synchronous and organized. It was the first milestone towards a more comprehensive understanding of resting-state activity. This work initially did not get major attention until (Raichle et al., 2001) using PET imaging data identified one fundamental resting state network (RSN) [1], commonly known today as the Default Mode Network (DMN). The authors noticed that some brain regions (DMN) were active during the rest, but such activity decreased when subjects were asked to engage in various cognitive tasks. This sparked significant interest, and later work consistently showed that DMN is activated when individu-

[1] RSNs are defined as the sets of brain areas to represent specific frameworks of brain function.

als are focused on their internal mental-state processes such as self-awareness and memory (Buckner, Andrews-Hanna, & Schacter, 2008; Raichle, 2015). Later on, (Greicius, Krasnow, Reiss, & Menon, 2003) confirmed these findings and identified the DMN in resting-state fMRI. After that, several other RSNs were identified (De Luca, Beckmann, De Stefano, Matthews, & Smith, 2006; Fox, Corbetta, Snyder, Vincent, & Raichle, 2006; Seeley, Menon, et al., 2007). The surprising finding was that similar networks were detected in both task and rest.

(S. M. Smith et al., 2009) using a data-driven approach, identified multiple independent networks that were consistent across both resting-state and task fMRI studies. Precisely, this study demonstrated that at least ten networks were unambiguously paired between the resting-state and task fMRI data sets (as illustrated in Figure 1.1). These networks corresponded to the visual, DMN, sensorimotor, auditory, executive control, and frontoparietal areas. This result implied that the resting brain also actively uses the same functional networks when involved in external tasks. This helped the field truly appreciate the utility of the resting-state fMRI, raised many questions, and motivated researchers to continue additional investigation.

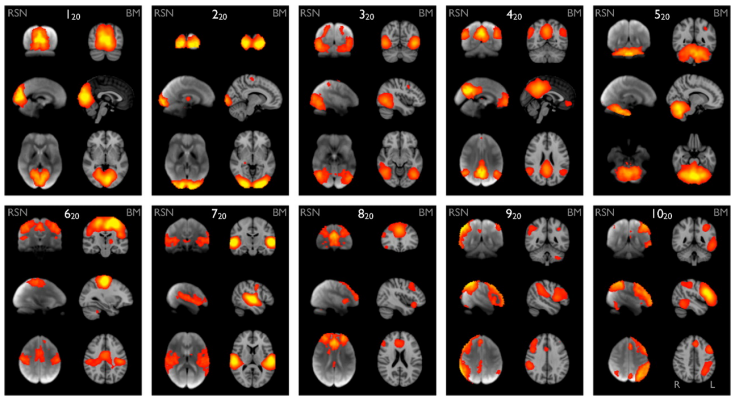


Figure 1.1: The image taken from (S. M. Smith et al., 2009) shows the ten well-matched pairs of networks from the BrainMap activation database and the 36-subject resting fMRI dataset. The shown networks were paired using spatial cross-correlation with a mean correlation of 0.53.

1.4 Functional Connectivity in fMRI

Once we have identified these different functional regions, the neuroimaging field moved on to study the integration of these brain regions (K. J. Friston, 2011). One common way to learn such integration is in terms of functional connectivity analysis (S. M. Smith et al., 2011). Functional connectivity is defined as the statistical association between time series of anatomically distinct brain regions (Rogers, Morgan, Newton, & Gore, 2007). In other words, it is a measure of the statistical dependencies among remote neurophysiological events (K. J. Friston, 2011). To simplify, if two brain regions have correlated BOLD signals, it means they are functionally connected.

The first step of such connectivity analysis is identifying a set of functional nodes (also known as the regions of interest) (S. M. Smith, Vidaurre, et al., 2013). As the fMRI scan comprises thousands of voxels at each time-point per subject, this poses extreme challenges to working with raw voxel-wise time-series data. To avoid working in high dimensional space (and to increase interpretability and ef-

fective SNR), the standard procedure entails determining nodes defined over a group of voxels sharing similar time courses. Once the nodes are identified, the statistical association between these nodes can be quantified with measures of statistical dependencies, such as correlations (S. M. Smith, Vidaurre, et al., 2013). We can apply this kind of functional connectivity analysis both to resting-state and task fMRI studies.

1.4.1 Functional Connectivity versus Effective Connectivity

Functional connectivity informs us about the strength of correlation between two regions, but it does not tell about the causality (the direction of information flow). Ideally, one should generally attempt to estimate both strengths and the direction of information flow (causality [2]); this could be more meaningful than correlation-based functional connectivity estimates. Such connectivity analysis is commonly referred to as effective connectivity (K. J. Friston, 2011; Valdes-Sosa, Roebroeck, Daunizeau, & Friston, 2011), and it is important to distinguish it from functional connectivity. Effective connectivity has advantages but comes with challenges; it is hard to estimate robustly on resting data, and such methods can only cope with fewer nodes (S. M. Smith et al., 2011). Then, we have functional connectivity, which comes with the benefits of being simpler, well-conditioned, and being able to handle more nodes. Ideally, we would like to infer the more informative effective connectivity. Still, functional connectivity is a pragmatic choice considering the current limitations of effective connectivity. We have illustrated different network modelling methods in Figure 1.2.

[2]Causality is defined in terms of a cause preceding and increasing the probability of the effect - a brief overview of causality is given in (S. M. Smith, 2012).

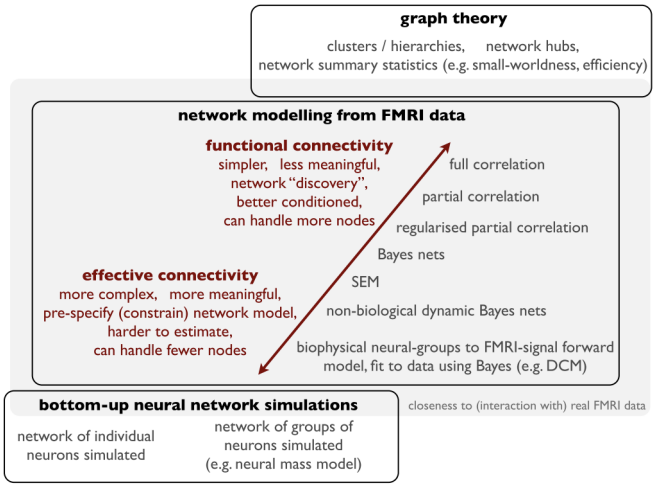


Figure 1.2: The image taken from (S. M. Smith, 2012) shows the relationships between different network modelling methods for fMRI analysis. These methods range from neural-level simulations to abstract graph theoretic summaries of a connectivity matrix. Between these two extremes, we also have several popular network modelling methods (applied to real fMRI data) that can mainly be categorized into the study of functional vs effective connectivity (an over-simplistic division). In this work, we aimed to identify an optimal pipeline to estimate and analyze functional connectivity.

1.4.2 Static versus Time-varying Functional Connectivity

Functional connectivity is dynamic as it can be observed to change over time (Hutchison et al., 2013; Allen et al., 2014). Contrasting with this metric is static connectivity, an average of all time-varying connectivity states, and is less noisy. Static (or time-averaged) connectivity is also simpler to estimate and interpret and is, at this point in more common usage [3]. However, there is now increasing interest in studying the time-varying nature of functional connectivity in fMRI [4]. The methodological research conducted in this thesis is based on exploring both the static and time-varying functional connectivity estimates [5].

[3] The detailed introduction of static functional connectivity is presented in Chapter 3.

[4] The detailed introduction of time-varying functional connectivity is presented in Chapter 4.

[5] Chapter 3 and Chapter 4 attempts to propose a comprehensive pipeline to analyze static and time-varying functional connectivity estimates.

We will refer to functional connectivity as *FC* and time-varying functional connectivity as *TVFC*. In the next few sections, we establish the significance of FC, both static and time-varying.

1.5 Static Functional Connectivity Relevance in Health and Disease

The patterns of FC are sensitive to individual differences in health and disease, and emerging work suggests they may have powerful utility as clinical biomarkers. In this section, we highlight the applications of FC in clinical research settings.

1.5.1 Neurological Conditions

There could be several ways in which one can conduct resting-state functional analysis in the clinical research settings. For now, the most common way is to compare FC patterns between groups of healthy subjects versus those with a neurological condition. Several studies have shown the utility of FC in identifying neurological diseases. For example, it has been shown that there are decreased correlations and anti-correlations within the DMN for people who have Alzheimer's disease (Li et al., 2002; Greicius, Srivastava, Reiss, & Menon, 2004; Supekar, Menon, Rubin, Musen, & Greicius, 2008). Similar changes in patterns of FC have been identified for people suffering from mild cognitive impairment (Sorg et al., 2007).

Moreover, (Seeley, Allman, et al., 2007) has shown that there is a decrease in correlation in the salience network for people with dementia (the salience network includes regions in the dorsal anterior cingulate and bilateral insular cortices). (Lowe et al., 2002) has shown decreased correlation within the somatomotor net-

work for subjects with Multiple Sclerosis (MS). Numerous research studies have shown a decrease or increase in correlation within the DMN, or between DMN and other regions for subjects with schizophrenia, (Liang et al., 2006; Bluhm et al., 2007; Salvador et al., 2007). For subjects with Attention Deficit Hyperactivity Disorder (ADHD), several studies have shown reduced connectivity within the DMN and increased connectivity in the salience network (Zhu et al., 2005; Cao et al., 2006; Tian et al., 2006). Also, decreased connectivity within the DMN is observed in subjects with Autism Spectrum Disorder (ASD) (Cherkassky, Kana, Keller, & Just, 2006; Monk et al., 2009).

Numerous other studies highlight the significance of FC in a wide range of neurological conditions, such as Amyotrophic Lateral Sclerosis (ALS), Tourette syndrome, coma state, and others (Fox & Greicius, 2010). This emphasizes that it is of value to design methods for accurate estimation and analysis of FC.

1.5.2 Lifestyle and behaviour

Patterns in FC are also associated with a wide range of lifestyle factors, mental health conditions, and cognition. It has been shown that FC patterns change along with healthy aging. For example, several studies have shown decreased connectivity within several RSNs with aging (Andrews-Hanna et al., 2007; Damoiseaux et al., 2008), relating this to age-associated cognitive decline.

On the other hand, FC patterns have been associated with mood disorders (Liu et al., 2015). For subjects with anxiety disorder, it has been observed they have increased connectivity between the amygdala and frontoparietal network and decreased connectivity between the amygdala and salience network (Etkin, Prater, Schatzberg, Menon, & Greicius, 2009). For subjects with depression and bipolar, decreased corticolimbic activity has been observed (Anand, Li, Wang, Lowe, &

Dzemidzic, 2009). Lastly, patterns of FC have also been associated with working memory, fluid intelligence, attention, health, income, life satisfaction-related behavioural measures (S. M. Smith et al., 2015).

These studies provide compelling evidence that the functional architecture of the brain captured by FC is indeed a sensitive biomarker to measure individual differences that exist in cognitive function and abilities.

1.6 Clinical and non-clinical relevance of Time-varying Functional Connectivity

1.6.1 Revisiting TVFC

FC has been undoubtedly proven effective as providing markers associated with neurological conditions in the last two decades. It is worth repeating that FC estimates provide us the average connectivity information across the whole scanning session while assuming that connectivity remains static across time. It does not give us any information about the varying connectivity dynamics, and yet we believe that underlying this static (time-averaged) connectivity state are the transient connectivity states that inherently fluctuate across time. FC could be thought of as the simplistic representation of brain networks. However, this representation might be considered too simplistic (D. Lurie et al., 2018). On the other hand, TVFC, though hard to estimate and analyze, provides a much more detailed description of brain network dynamics. There is immense meaningful information present in the temporal characteristics of these brain networks that cannot be estimated by solely looking into static FC (D. J. Lurie et al., 2020).

In the sections below, we establish the behavioural and cognitive relevance of

TVFC during both rest and cognitive tasks.

1.6.2 Ongoing "task-related" Cognition

A growing area of interest is to understand task-related changes in patterns of FC. Recently, many studies have focused on understanding FC changes during different cognitive states primarily imposed by the external experimental paradigm. The existing methods to estimate TVFC have several limitations, but we are still able to demonstrate that the TVFC measures capture the time-varying changes in FC across a range of behavioural tasks (Cohen, 2018; Gonzalez-Castillo & Bandettini, 2018). Moreover, we can demonstrate that we can use the TVFC patterns to predict which task a person is engaged in (Xie et al., 2019). These studies can only capture long or medium time-scale fluctuations in FC and not the instantaneous or short time-scale changes, essentially because of limitations in the techniques used to compute TVFC. Nevertheless, such task experiments also give us a fantastic opportunity to test the efficacy of TVFC methods and demonstrate how accurately our TVFC methods can decode these experimentally imposed conditions.

1.6.3 Ongoing "spontaneous" Cognition

The absence of external tasks does not imply the absence of ongoing cognition. As discussed previously, there is already significant evidence of fluctuations in FC during externally cued cognitive tasks. On the other hand, it is less clear how strongly fluctuations in FC will describe the "spontaneous" fluctuations in the mental state. Some studies have shown that the fluctuations in arousal and vigilance are related to the fluctuations in FC (Shine et al., 2016; Wang, Ong, Patanaik, Zhou, & Chee, 2016). Though arousal and vigilance are not pure resting-state conditions, they are extremely close to wakeful rest. Some investigations have shown within-session fluctuations in FC during resting-state fMRI (D. J. Lurie et

al., 2020; Vidaurre, Smith, & Woolrich, 2017). On the other hand, several studies report that there is insufficient evidence of within-session fluctuations in FC in resting fMRI data (Laumann et al., 2017a). While it is possible that close to homogeneous FC in resting-state fMRI is a genuine phenomenon, it is also possible that it is caused by limitations in the methods used to compute TVFC.

1.6.4 Neurological Conditions

Some studies have recently shown that individual differences in TVFC might be a more sensitive biomarker to behavioural and cognitive traits as compared to static FC (Liégeois et al., 2019), e.g., it was recently found that some aspects of intelligence could only be explained by TVFC (Vidaurre, Llera, Smith, & Woolrich, 2021).

Moreover, several research studies have also shown the utility of TVFC in identifying many psychiatric and neurological conditions. For example, MS patients showed reduced TVFC between the visual and cognitive networks as compared to the control group (d'Ambrosio et al., 2020). Higher TVFC strength was shown between sensorimotor, parietal, and precuneus areas in schizophrenia patients as compared to the control group (Abrol, Rashid, Rachakonda, Damaraju, & Calhoun, 2017). Lower network connections (Fiedler value) were observed in Parkinson's disease patients as compared to the healthy group (Cai et al., 2018). Lower TVFC between sensorimotor and visual-related cortices in stroke patients as compared to the healthy group was reported (J. Chen et al., 2018). ASD adolescents showed they had lowered TVFC between the posterior central cortex, ventromedial prefrontal cortex, and anterior insula compared with typically developing adolescents (Guo et al., 2019). Reduced TVFC was observed between the cerebellar networks and frontoparietal network in epilepsy patients as compared to the healthy subjects (Klugah-Brown et al., 2019). Subjects suffering from major de-

pressive disorder showed higher TVFC strength between the superior frontal and middle frontal gyrus (Zhi et al., 2018).

1.7 Thesis Objectives

In **Chapter 1**, we give an overview of FC and describe its importance in cognitive functioning and also a variety of neurological dysfunctions. **Chapter 2** is general introduction to materials. We will briefly discuss several resting-state and task datasets used during the work.

1.7.1 Methods Development for Static Functional Connectivity Analysis

There is currently no general agreement on a single standard pipeline for estimating and using FC. We aim to help with this situation by answering several significant questions, 1) How should we parcellate the brain? 2) How should we estimate correlation-based FC? 3) How can we preserve the geometry and shape of correlation-based FC elements? 4) How should we remove the effect of confounding variables in resting-state fMRI? 5) Which classifier/predictor provides an optimal prediction of non-imaging variables using these FC estimates?

Chapter 3 attempts to address the above questions to propose a comprehensive pipeline that considers several central methodological issues when estimating and utilizing static FC.

1.7.2 Methods Development for Time-varying Functional Connectivity Analysis

As discussed previously, there is a lack of evidence of within-session fluctuations in FC in resting-state fMRI data. Numerous studies have reported that TVFC appears stable and homogeneous across time during resting-state. Typically, methods that measure TVFC, such as sliding window approaches, do not separately model when changes occur in the mean activity levels from when changes happen in the FC. This might be the reason behind biased estimation of TVFC, appearing more stable over time than it is.

There are also other limitations when it comes to the estimation of TVFC using existing methods. For example, one major limitation is that some of the popular existing methods to estimate TVFC only capture the short-range temporal dynamics of FC. Also, these current approaches to estimate TVFC compel different FC states to act in a mutually exclusive way (e.g., multiple states can not be active at one time-point).

We aim to ameliorate these problems by answering several significant questions, 1) Should we capture the long-range temporal dynamics of FC while estimating TVFC and how? 2) Should we model the temporal fluctuations in FC independently from fluctuations in the mean of the activity? 3) Should we allow FC states to be non-mutually exclusive (e.g., multiple states can be active at one time-point)?

Keeping these questions in mind, **Chapter 4** proposes an alternative approach that models changes in the mean brain activity and the FC as being able to occur at different times to each other. We refer to this method as the Multi-dynamic Adversarial Generator Encoder (MAGE) model, which includes a model of the network dynamics that captures long-range time dependencies. MAGE models

fMRI data as a set of reoccurring brain states, and importantly, these states do not have to be binary and mutually exclusive.

1.7.3 Clinical Applications of Static and Time-varying Functional Connectivity

Chapter 5 focuses on different clinical applications of FC. We discuss its relevance for mood disorders characterization and several other clinical applications (e.g., autism diagnosis and nicotine addiction). We also demonstrate the practical applications of TVFC in understanding spontaneous and task-related cognitive processes. Lastly, we show some of the advantages of TVFC over static FC.

Chapter 6 covers general discussion and conclusions. We discuss the future directions while studying the clinical relevance of FC. We also talk about the prospects of machine learning in fMRI in particular and health care in general. Lastly, we summarise the main findings from previous chapters while discussing the limitations and future directions.

CHAPTER 2

GENERAL MATERIALS

This chapter details the different materials and resources employed during this work. We have utilized two large-scale cohorts, UK Biobank and the Human Connectome Project, focused on understanding the relation between brain and behaviour in the general population. On top of that, we have used two other cohorts, Autism Brain Imaging Data Exchange and Addiction Connectome Preprocessed Initiative, which specifically focused on people with neurological dysfunctions. For the studies mentioned earlier, we have categorically focused on the brain dynamics that occur in a resting or task-negative state when no explicit task is performed.

Another motivation of this work is to understand changes in brain activity during different cognitive states imposed by the external paradigm. In terms of FC, we know there exists some level of spatial overlap between rest and task-derived FC networks. We interrogated whether changes in connectivity strength during different task experiments are powerful enough to predict which task a person is

engaged in? Considering that, we focused on understanding the brain changes that occurred because of externally imposed task experiments. We have focused on two additional cohorts; the first cohort comprises participants who underwent separate task-based (visual, motor, and visuomotor) and task-free (resting) steady-state scans. The second task-based cohort measured brain activity from participants while they were watching different movies.

The resting-state and task data-sets will improve our understanding of brain dynamics in a task-negative or task-positive state.

2.1 Resting-state fMRI datasets

This section is specifically focused on explaining the different resting-state fMRI data sets used for this work. We explain the background and motivation for each data set, then we outline critical fMRI acquisition details, and lastly, we define the pre-processing steps employed for each data set.

2.1.1 UK Biobank

UK Biobank is a prospective epidemiological cohort comprising 500,000 people aged between 45 and 73 years recruited from the UK's general population between 2006 and 2010. In 2014 UK Biobank commenced the imaging sub-study, whereby 100,000 of the original cohort were invited back for comprehensive brain, heart, and body imaging. Approximately 50,000 brain images have now been made available along with the outputs of standardized, automated pipelines (Alfaro-Almagro et al., 2018). For each individual, the raw and pre-processed images, along with approximately 2500 imaging derived phenotypes (IDP's), are now available through the UK Biobank portal.

The UK Biobank brain imaging protocol is described elsewhere (Miller et al., 2016). Multiple modalities (T1 weighted structural, T2 weighted FLAIR, susceptibility-weighted imaging (SWI), diffusion-weighted imaging (DWI), resting fMRI, and task fMRI) were acquired across three dedicated UK Biobank imaging centre equipped with identical scanners and head coils (3T Siemens Skyra, software VD13). For this thesis, we have used the resting fMRI data of 13,301 subjects from UK Biobank. UK biobank utilizes Echo-planar imaging (EPI) based fMRI acquisition that uses a multislice acceleration factor of 8. The acquisition details can be found in (*UK Bibank Preprocessing Pipeline*, 2020), but we have summarised some main parameters in Table 2.1.

Table 2.1: UK Biobank resting-state fMRI acquisition details

Parameters	Values
Resolution	2.4x2.4x2.4 mm
Duration	6 minutes (490 timepoints)
Repetition time	0.735 s
EPI	x8 multislice acceleration factor

Preprocessing

The pre-processing pipeline for UKB is described in (*UK Bibank Preprocessing Pipeline*, 2020) although we can summarize the main steps as, 1) motion correction using MCFLIRT (Jenkinson, Bannister, Brady, & Smith, 2002), 2) highpass temporal filtering, 3) alignment of fMRI data to structural data using FLIRT (FMRIB's Linear Image Registration Tool) Boundary-Based Registration (BBR) tool (Jenkinson & Smith, 2001; Greve & Fischl, 2009), 4) removal of structural artifacts with ICA (Independent Component Analysis and FMRIB's ICA-based X-noisefier(FIX) (Beckmann & Smith, 2004) (Salimi-Khorshidi et al., 2014); FIX was hand-trained on 40 Biobank rfMRI datasets, with classification accuracy of

99.1/100.0% for non-artifact components and 98.1/98.3% for artifact components, 4) brain parcellation using the techniques that we cover in Chapter 3.

After applying a parcellation scheme, we can extract the mean time series of each parcel from the pre-processed time series data. Hence we extract a time-series signal for each subject, $X \in (R)^{T \times d}$, where T corresponds to the number of time-points and d corresponds to the number of parcels for each subject. Based on that, we can compute the FC estimates, both static and time-varying (we cover this in detail in Chapter 3).

The presented work in the upcoming chapters has extensively used the resting-state fMRI data of UK Biobank. In particular, we use the UK Biobank data to design and optimize the modelling methods for static and time-varying FC estimation in Chapter 3 and Chapter 4 respectively. Moreover, we also employ the UK Biobank data in Chapter 5 to highlight the clinical applications of FC.

2.1.2 Human Connectome Project

The Human Connectome Project (HCP) consortium aimed to characterize human brain connectivity and function in a population of 1200 healthy adults and to enable meaningful comparisons between the brain circuits and behaviour not only across the whole groups but also at the individual level (Van Essen et al., 2013). The subjects enrolled in HCP are adult twins and their non-twin siblings, and their age range is 22 to 35 years. The MR brain imaging protocol of HCP includes four imaging modalities: structural MRI, resting-state fMRI, task fMRI, and diffusion MRI (dMRI). For HCP, resting-state fMRI data were acquired in four runs of approximately 15 minutes each, two runs in one session and two in another session, with eyes open with a relaxed fixation on a projected bright cross-hair on a dark background. The detailed acquisition details can be found in (Glasser et al., 2013),

but we have summarised some primary parameters in Table 2.2.

Table 2.2: HCP resting fMRI acquisition details

Parameters	Values
Resolution	2x2x2 mm
Duration	57.32 minutes (4800 timepoints)
Repetition time	0.72 s
EPI	x8 multislice acceleration factor

Preprocessing

HCP provides the data in both volumetric and surface (grayordinates) spaces. HCP has two pre-processing pipelines, *fMRIVolume* and *fMRISurface* to process volumetric and surface data respectively. The first functional pipeline, *fMRIVolume*, as described in (Glasser et al., 2013; *HCP Preprocessing Pipeline*, 2017) can be summarized as, 1) gradient distortion correction, 2) motion correction using MCFLIRT (Jenkinson et al., 2002), 3) registering the fMRI data to the structural space using FLIRT BBR tool (Jenkinson & Smith, 2001; Greve & Fischl, 2009), 4) bias field reduction and intensity normalization to a global mean. After pre-processing the functional volumetric data, the second pipeline, *fMRISurface*, is applied to convert the time-series from the volumetric space to the grayordinates space. Once the volume to surface mapping is achieved, a novel surface-based smoothing is applied (Glasser et al., 2013). The output of *fMRISurface* is a CIFTI dense time-series that can be used for further resting-state fMRI analysis.

For HCP data, we use the minimally-preprocessed released data in grayordinates, and for UKB data, we used the pre-processed released data in standard volumetric space. UKB has not yet been processed with a pipeline like HCP, but future UKB data releases will include versions of the data transformed into grayordinate space. For now, this helps to span the space of approaches commonly taken (some

volumetric, some surface-based)

Like UK Biobank, the presented work in later chapters has also extensively used the resting-state fMRI data of the HCP. In particular, we use the HCP data to design and optimize the modelling methods for static and time-varying FC estimation in Chapter 3 and Chapter 4 respectively. HCP does not provide as extensive phenotypic data as UK Biobank does, so we largely employ the UK Biobank data to investigate clinical applications.

2.1.3 Autism brain imaging data exchange (ABIDE)

Autism Brain Imaging Data Exchange (ABIDE) aggregated the resting-state fMRI data-sets from multiple imaging sites of individuals with Autism Spectrum Disorder (ASD) and age-matched typical controls (Di Martino et al., 2014). The ABIDE consortium provides the resting-state fMRI data for ages of participants range from 7 to 64 years with a median age of 14.7 years across groups. The MRI imaging protocol is different depending on each imaging site and can be accessed through (*ABIDE Imaging Protocol*, 2017). One thing worth noting is that the ABIDE data is comparatively noisier than HCP and UKB data; the number of time points varies between 115 and 295. ABIDE data is available for 871 subjects (autism and control), quite close to the number of subjects in the HCP dataset.

Preprocessing

We used the pre-processed resting-state fMRI data from the Pre-processed Connectome Project pipeline (C-PAC) project (Craddock et al., 2013). C-PAC workflows are mostly built from AFNI (Cox, 1996) and FSL tools. The C-PAC functional pre-processing pipeline can be summarized as, 1) slice time correction using AFNI's 3dTshift, 2) motion correction to the mean fMRI using AFNI's 3dvolreg

(Cox & Jesmanowicz, 1999), 3) skull-stripping using AFNI's 3dAutomask, 4) mean intensity normalization to a global mean, 5) registering of fMRI to structural image using BBR tool. The C-PAC pipeline also provides the option of global signal regression and band-pass filtering (0.01-0.1Hz), but we did not incorporate these steps in the implemented pre-processing pipeline.

We use the ABIDE data in Chapter 3 to optimize the modelling methods for FC analysis. Moreover, we also use the ABIDE data in Chapter 5 to understand whether FC can predict individuals accurately with or without ASD.

2.1.4 Addiction Connectome Preprocessed Initiative (ACPI)

We used the resting-state fMRI data of subjects with Attention Deficit Hyperactivity Disorder (ADHD). The principal aim was to study the effect of cannabis use among adults with a childhood diagnosis of ADHD. The project was originally an extension of the long-standing Multimodal Treatment Study of ADHD (MTA), one of the most extensive studies of ADHD ever conducted. ACPI provides datasets from a subset of MTA participants who were studied with multimodal brain imaging. The acquisition details can be found in (*ACPI Imaging Protocol*, 2016), but we have summarised some main parameters in Table 2.3. One thing worth noting is that the number of time points in the ACPI dataset for each participant is 179. The total number of subjects is 126, far less than the UKB, HCP, and ABIDE datasets.

Table 2.3: ACPI resting-state fMRI acquisition details

Parameters	Values
Resolution	3.4x3.4x4 mm
Duration	6.04 minutes (179 timepoints)
Repetition time	2 s
EPI	x4 multislice acceleration factor

Preprocessing

The data was again processed using C-PAC. The pipeline employed is similar to the one used for ACPI pre-processing (we did not employ global signal regression and band-pass filtering). The detailed pre-processing strategy is explained in (*ACPI Preprocessing Protocol*, 2016).

We use the ACPI data in Chapter 3 to optimize the modelling methods for static FC analysis. Moreover, we also use the ACPI data in Chapter 5 to understand whether inter-individual differences in FC can help us differentiate the healthy subjects from those who suffer from nicotine addiction.

2.2 Task-fMRI Datasets

Task fMRI data sets allow the identification of patterns evoked by tasks performed during the scanning session. This section is specifically focused on explaining the different task fMRI data-sets used for this work. We explain the background, motivation, fMRI acquisition details, and pre-processing steps employed for each data set.

2.2.1 Cognitive-task data-set

This task fMRI data was collected as part of a previous study from 15 healthy volunteers (7 females, 8 males, age = 27.3 +/- 4.4 years, all right-handed) without any previous neurological disorders (Kieliba, Madugula, Filippini, Duff, & Makin, 2019; Duff, Makin, Cottaar, Smith, & Woolrich, 2018). Participants performed four different tasks: rest, motor only, visual only, and simultaneous (but independent) visual and motor tasks, and each task was repeated four times. The motor task involved continuous finger-tapping against the thumb using the right

hand. The vision task consisted of videos of colorful abstract shapes in motion. During the combined motor-visual task, participants performed the finger-tapping task while simultaneously watching the videos. Functional data were acquired using a Siemens Vario 3T scanner, with a high-resolution multiband (factor 6) with the parameters explained in Table 2.4.

Table 2.4: Cognitive task data-set fMRI acquisition details

Parameters	Values
Resolution	3x3x3 mm
Duration	9.25 minutes (185 timepoints)
Repetition time	3 s
EPI	x6 multislice acceleration factor

Preprocessing

The fMRI data were processed using FSL FEAT (S. M. Smith, 2004). Preprocessing steps included: 1) motion and field map correction (Jenkinson et al., 2002), and brain extraction (S. M. Smith, 2002), 2) data was cleaned using FIX to account for any non-neuronal contribution, 3) fMRI data were aligned to structural data using BBR, 4) pre-whitening of time-courses were performed using FILM (FMRIB's Improved Linear Model) (S. M. Smith, 2004), 5) intensity normalization of the images to the global mean, and 6) brain parcellation using the techniques that we cover in Chapter 3.

We use this data set to understand task-related changes in patterns of TVFC. Specifically, in Chapter 5, we use this data to evaluate whether changes in connectivity strength can decode the task performed by an individual. This data set also helped us to determine the performance of our proposed method (MAGE as outlined in Chapter 1) that can be employed to estimate TVFC.

2.2.2 Movies-task data-set

We used the fMRI data from 44 human subjects (male and female) who were presented with 16 three-minute stories (Baldassano, Hasson, & Norman, 2018). The movies are comprised of four schematic events drawn from two different scripts; eating at a restaurant or going through the airport. The stories within each script were similar in their high-level flow of events but differed in their low-level peculiarities. Functional images were collected using a 3T Siemens Prisma scanner. The images were acquired using a high-resolution multiband (factor 4) with the parameters given in Table 2.5.

Table 2.5: Movies task data-set fMRI acquisition details

Parameters	Values
Resolution	2x2x2 mm
Duration	$\sim$ 22 minutes ($\sim$900 timepoints)
Repetition time	1.5 s
EPI	x4 multislice acceleration factor

Preprocessing

We used the FreeSurfer Functional Analysis Stream (FsFast) to pre-process the fMRI data (Fischl, 2012). The main pre-processing steps include 1) alignment of the functional images to the structure, 2) motion and B0 correction, 3) global signal regression. Based on prior work on the representation of high-level, cross-modal situation models (Zadbood, Chen, Leong, Norman, & Hasson, 2017), the writers of the study concentrated on regions within the DMN. They selected the networks 15, 16, and 17 that made up the entire DMN from YEO 17-network parcellation and then merged spatially contiguous parcels. This analysis yielded six ROIs: angular gyrus (1868 vertices), superior temporal sulcus (2118 vertices), superior frontal gyrus (2461 vertices), medial prefrontal cortex (2069 vertices),

parahippocsampal cortex (882 vertices), and posterior medial cortex (2495 vertices).

We use this data-set to understand the participants' brain activity response to these movies and whether that response is consistent across participants or not. Particularly, we use this data in Chapter 5 to gauge differences in inter-individual brain activity when they are viewing similar movies versus different movies.

CHAPTER 3

OPTIMISING STATIC NETWORK MODELLING METHODS FOR FMRI

3.1 Overview

There is no single widely-accepted standard pipeline for analyzing static FC. The common procedure for designing FC based predictive models entails three main steps: parcellating the brain, estimating the interaction between defined parcels, and lastly, using these integrated associations between brain parcels as features fed to a classifier for predicting non-imaging variables e.g., behavioural traits, demographics, emotional measures, etc. There are also additional considerations when using correlation-based measures of FC, resulting in three supplementary steps: utilising Riemannian geometry tangent space parameterization to preserve the geometry of FC; penalizing the connectivity estimates with shrinkage approaches to handle challenges related to short time-series (and noisy) data; and removing confounding variables from brain-behaviour data. These six steps are contingent

on each-other, and to optimise a general framework one should ideally examine these various methods simultaneously. In this work, we investigated strengths and short-comings, both independently and jointly, of the following measures: parcellation techniques of four kinds (categorized further depending upon number of parcels), five measures of FC, the decision of staying in the ambient space of connectivity matrices or in tangent space, the choice of applying shrinkage estimators, six alternative techniques for handling confounds and finally four novel classifiers/predictors. For performance evaluation, we have selected two of the largest datasets, UK Biobank and the Human Connectome Project resting state fMRI data, and have run more than 9000 different pipeline variants on a total of $\sim$14000 individuals to determine the optimum pipeline.

3.2 Introduction

It is valuable to understand large-scale networks of brain activity. As discussed in Chapter 1, most of the literature is focused on recognizing brain networks that are spatially and temporally linked to a given task (Fox et al., 2005). In this chapter, we aim to explore brain networks measured from resting state fMRI. In contrast to task-based fMRI, during rfMRI the brain is not occupied by a pre-defined task. However, as discussed before in Chapter 1, studies have shown that the brain activity at rest is not random (De Luca et al., 2006; Tavor et al., 2016, 2016), but is hierarchically organized in time and significantly correlated with behaviour (Vidaurre et al., 2017).

The analysis of rfMRI data largely relies on estimates of FC. FC is dynamic as it can be observed to change over time (Vidaurre et al., 2017; Hutchison et al., 2013; Allen et al., 2014). Contrasting with this metric is static connectivity, which is an average of all dynamic connectivity states, and is thereby less noisy. The

methodological research conducted in this chapter is purely based on exploring static FC.

As the resting fMRI scan is composed of thousands of voxels at each time-point per subject, the standard processing procedure entails the determination of parcels which are defined over a group of voxels sharing similar timecourses. However, the method by which the brain is segmented into these different regions, or parcels, before estimating FC is far from standardized (Eickhoff, Thirion, Varoquaux, & Bzdok, 2015). Some of the existing parcellation techniques are based on mapping anatomical or functional atlases onto an individual's brain (Shen, Tokoglu, Papademetris, & Constable, 2013; Desikan et al., 2006; Fischl et al., 2004). Other brain parcellation techniques are more data driven and attempt to derive parcels based upon common features within the data (Honnorat et al., 2015; Parisot, Glocker, Schirmer, & Rueckert, 2016; Parisot, Arslan, Passerat-Palmbach, Wells III, & Rueckert, 2016). After parcellating the brain, the next step is estimating FC. FC is conventionally estimated through taking a full (Pearson) correlation between brain parcels over time (Bastos & Schoffelen, 2016). The use of full correlation in deriving these estimates however, has some drawbacks as this method does not distinguish direct vs. indirect pathways by which two brain regions may be connected (S. M. Smith et al., 2011). Additionally, the utilization of this method results in numerous other limitations related to low SNR, primarily because of short scanning sessions. This work provides a detailed account of several FC estimation techniques that may mitigate these challenges, which consequently can result in enhanced sensitivity in signal detection.

After derivation of brain parcels and FC estimates, there remains further decisions of how to optimally utilize these in conjunction with other data, such as behavioural metrics. The commonly followed procedure is to directly use elements of correlation-based FC estimates as features for regression or classi-

fication algorithms. This direct usage however has fallen subject to criticism, as correlation-based FC does not naturally form a Euclidean space (Varoquaux, Baronnet, Kleinschmidt, Fillard, & Thirion, 2010; Barachant, Bonnet, Congedo, & Jutten, 2013). In this work, we evaluate an arguably more principled approach to represent correlation-based FC using Riemannian geometry (Varoquaux et al., 2010; Barachant et al., 2013). This method is founded upon mathematical principles which involve manipulating correlation-based estimates in the tangential space of Riemannian manifold (Ng, Varoquaux, Poline, Greicius, & Thirion, 2015) where Euclidean geometry can be applied.

Additionally, there is also a concern of noise causing an increase in inter- and intra-subject variance in FC estimates. One method to alleviate this is to apply conventional regularisation techniques in the ambient space when FC is estimated. The other approach is to apply covariance shrinkage estimators (Y. Chen, Wiesel, Eldar, & Hero, 2010) in the tangential space of Riemannian manifold. In this work, we have evaluated the effects of covariance shrinkage techniques both in the ambient space and the tangent space (isotropic (Ledoit & Wolf, 2004) and non-isotropic covariance shrinkage techniques (Rahim, Thirion, & Varoquaux, 2019)) on the prediction of individual demographic and behavioural phenotypes. Moreover, we have assessed the performance of shrinkage approaches in following framework, (i) no shrinkage in the ambient space or the tangent space, (ii) shrinkage in the ambient space but no shrinkage in the tangent space, (iii) no shrinkage in the ambient space but shrinkage in the tangent space, and (iv) shrinkage approaches both in the ambient space and the tangent space.

To predict individual phenotypes from resting fMRI data, the elements of the FC matrix, where each element represents the strength of connection between two parcels, can be used as input features for a machine learning classifier (or continuous variable predictor). In this work, we compare the performance of a selection

of existing state of the art methods with several proposed deep learning based architectures. We also demonstrate a baseline application for using neural networks in brain FC modelling

There is currently no general agreement on a single common pipeline for estimating and using FC. We aim to help with this situation by answering several major questions:

1. How should we parcellate the brain? Should this technique be data driven or based upon hard parcellation? How many parcels should we derive?
2. How should we estimate correlation-based FC? Should we remain with full, or partial correlation, or do we have better choices?
3. Should we preserve the geometry and shape of correlation-based FC elements within a Riemannian framework, and what would be the optimal way to do this?
4. Should we utilize covariance shrinkage estimators? Should shrinkage occur in ambient space of covariances or in tangent space?
5. How should we remove the effect of confounding variables in rfMRI?
6. Which classifier/predictor provides optimal prediction of age, sex and other non-imaging variables using these FC estimates? Is a deep learning approach of value?

The questions raised above may be treated as six steps to derive a framework for estimation and accurate usage of FC estimates. This chapter will attempt to address the above questions in order to propose a comprehensive pipeline.

For performance evaluation of the different methods, we have chosen two rich datasets which contain both resting fMRI and phenotypic data. The first is 13301

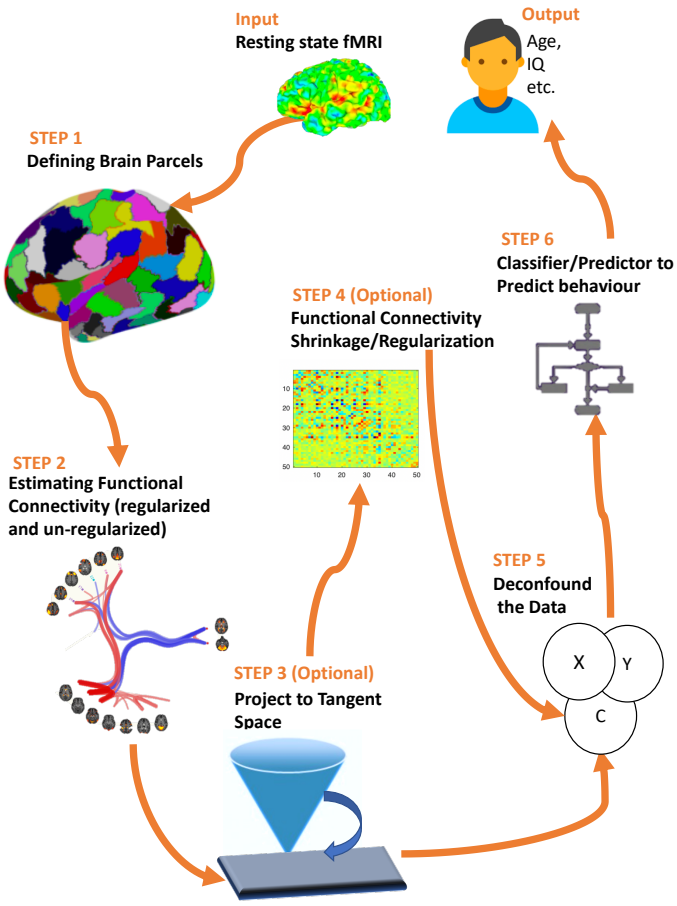


Figure 3.1: Flow chart summarizing the six major steps of the methodology framework for resting-state fMRI analysis.

subjects from UK Biobank (UKB) and other is 1003 healthy subjects from the Human Connectome project (HCP). Our assessment criteria are based on the prediction accuracy of age, sex, fluid intelligence and neuroticism score, using correlation-based FC estimates. To establish the most optimum and reliable pipeline, we have run more than 9000 combinations of different steps of methodology detailed below. This computation has involved thousands of hours of computing on both CPUs and GPUs. The overall methodology is summarized in Figure 3.1.

3.3 Methods

3.3.1 Finding the best parcellation approach

The standard procedure for parcellation entails the grouping of voxels sharing similar time-courses into parcels. A parcel can be any spatial region of the brain, associated with a single time course (found by averaging the spontaneous fluctuation over all voxels within the parcel). For our analyses, we considered the following spatial properties in parcels, 1) parcels may or may not be allowed to overlap, 2) parcels can have both positive and negative spatial weight, or be binary masks 3) parcels may be composed from multiple disconnected regions.

Following the above criteria, we estimated parcels (spatial maps and associated time-series) across the brain using four different methods. In the literature, parcels can also be referred as nodes, modes, networks or region of interests, but here for simplicity and continuity, we will only refer to them as parcels. We only consider rfMRI-driven parcellations, as there is plenty of existing evidence that these outperform parcellations derived from pre-defined structural atlases (Sala-Llonch, Smith, Woolrich, & Duff, 2019; Thirion, Varoquaux, Dohmatob, & Poline, 2014).

Data-driven Parcellation: SHEN functional Parcels

The simplest way to identify parcels is to extract time courses from pre-defined, labelled regions, which often come from anatomical atlases. For resting state studies, a preferred method is to use a functional brain atlas based on correlated BOLD signal, rather than anatomical distinctions (S. M. Smith et al., 2011). In this study, we applied a functional brain atlas known as the SHEN parcellation (Shen et al., 2013) which covers the cortex, subcortex and cerebellum. To derive the SHEN atlas, a spectral clustering method was applied which ensured the functional ho-

mogeneity within each parcel (Shen et al., 2013). SHEN parcellation is based on resting fMRI data of 79 healthy normal volunteers and provides more coherent resting state time-course estimates when compared with anatomical atlases. The total number of parcels in SHEN parcellation is 268. SHEN is a volumetric parcellation, and for both UKB and HCP data, we used the preprocessed released data in standard volumetric space.

Data-driven Parcellation: YEO functional Parcels

We applied another functional brain atlas known as the YEO parcellation (Schaefer et al., 2018). The YEO parcellation is based on resting fMRI data from 1489 subjects which were registered using surface-based alignment. To derive the YEO atlas, a gradient weighted Markov random field approach was employed which integrates local gradient and global similarity approaches. This ensured the parcels were functionally and connectionally homogeneous, and also agreed with the boundaries of certain cortical areas defined using histology and visuotopic fMRI. The YEO parcellation is available at multiple resolution levels, and we have applied the YEO parcellation where the total number of parcels is 100. The YEO parcellation is available both in volumetric and grayordinate space. For HCP data, we use the minimally-preprocessed released data in grayordinates, and for UKB data, we used the preprocessed released data in standard volumetric space. YEO parcellation is often known as the "Schaefer" parcellation, from the name of the first author (Schaefer et al., 2018).

Data Driven Soft Parcellation: Spatially Independent Parcels

We also applied a popular data-driven parcellation scheme called Group Independent Component Analysis, or GroupICA (S. M. Smith, Hyvärinen, Varoquaux, Miller, & Beckmann, 2014; Beckmann & Smith, 2004) to our data. There are

two popular variants of GroupICA, spatial ICA (sICA), which identifies spatially independent components, and temporal ICA (tICA) (S. M. Smith et al., 2012), which put restrictions on temporal dynamics. The parcels identified by tICA are coerced to have orthogonal time-series which therefore precludes further network analyses of the kind covered in this work. Therefore, we opted for sICA and applied [1] sICA by concatenating the time courses from all subjects. The ICA parcels can be spatially overlapping and non-contiguous. The dimensionality (d) of sICA corresponds to the number of desired parcels. For HCP data, we applied sICA at d = 15, 50 and 200, and for UKB, d = 25 and 100. With UKB data, we disregarded four components from ICA 25, and 45 components from ICA 100 as they were judged to be artefacts [2]. This left us with d = 21, and 55 for UKB[3]. For HCP data we use the minimally-preprocessed released data, in grayordinates (cortical surface vertices and subcortical voxels), with ICA FIX denoising having been applied (Salimi-Khorshidi et al., 2014). For UKB data, we used the preprocessed released data in standard volumetric space after ICA FIX denoising.

Data Driven Parcellation: Probabilistic Functional Parcels

While ICA based approaches provide a choice between spatial or temporal independence, these approaches can be problematic as the brain is probably not perfectly segregated either in space or time, and they typically do not have any

[1] To apply sICA, we first generated the group-PCA (Principal Component Analysis) output using MIGP (MELODIC's Incremental group-PCA) (S. M. Smith et al., 2014) using data from all subjects. This comprises the top few thousands (1000-5000) weighted spatial eigenvectors from a group-averaged PCA, which is a very close approximation to the original data (fully concatenating all subjects' time-series). Lastly, group sICA is run on the output of group-PCA.

[2] Regarding the selection of artefactual components; this was part of the central processing done previously on behalf of UKB (Miller et al., 2016). The decisions were made on the basis of spatial layout of components, and average temporal power spectrum.

[3] For UKB, the ICA dimensionality cannot go up further (e.g., d = 200), because for larger dimensionalities, the additional components appear as artefactual; the number of non-artefact components barely grows at all (more than the existing 55). We believe this is because of the limitations of volumetric analysis possibly combined with the 6-minute limited duration of the acquisition.

modelling of between-subject variability. We applied another framework used for identifying large-scale probabilistic functional modes (PROFUMO) (Harrison et al., 2015, 2020) to identify parcels which are allowed to be correlated with each other in space and/or time, and which explicitly models between-subject variability in the parcels. In contrast to the above methods, therefore, PROFUMO does not require parcels to be orthogonal or non-overlapping. We applied PROFUMO on the HCP for d = 50, but we are not yet able to apply PROFUMO on the UKB data due to prohibitive computational cost with the larger cohort.

3.3.2 Finding the best correlation-based FC estimate method

We extracted a timeseries signal for each subject, $X \in (\mathbb{R})^{T \times d}$, where T corresponds to the number of timepoints and d correponds to the number of parcels for each subject. FC estimates are also referred to in the literature as parcellated connectomes or netmats, but here we will refer to these estimates as FC. Table 3.1 lists symbols for the different correlation-based FC estimates we evaluated.

Term	Symbol
Empirical covariance	$\check{\Sigma}$
Full correlation	$\breve{\Sigma}$
Precision matrix	$\hat{\Omega}$
Tikhonov precision matrix	$\bar{\Omega}$
Oracle approximate shrinkage precision matrix	$\dot{\Omega}$
Partial correlation	ρ
Any of above FC matrix	C
Any of above FC matrix in tangent space	$\vec{C}$

Table 3.1: The symbols used for various FC estimates in this work.

Estimating the FC

Sample Covariance: FC is calculated by estimating the empirical (sample) covariance matrix of timeseries signal for each subject. The range of elements of the empirical covariance matrix, $\check{\Sigma}$, lies between $-\infty$ and $+\infty$ and can be defined as illustrated in Equation 3.1:

$$\check{\Sigma} = \frac{1}{T-1}(X - \bar{X})^t(X - \bar{X}) \tag{3.1}$$

where, $X \in (\mathbb{R})^{T \times d}$ corresponds to timeseries signal for each subject, T corresponds to the number of timepoints, d correponds to the number of parcels for each subject and $\bar{X}$ corresponds to the average estimate of X.

Full Correlation: The full correlation coefficient, or Pearson's correlation coefficient can be defined in terms of covariance. Mathematically, full correlation $\breve{\Sigma}$ is a standardized form of covariance. The value of correlation takes place between -1 and +1. Conversely, the value of covariance lies between $-\infty$ and $+\infty$.

Partial Correlation: Although commonly used, using full correlation to derive connectivity estimates is problematic, as it does not distinguish whether two parcels of brain are directly connected or indirectly connected through another parcel. To mitigate this, we use partial correlation (S. M. Smith et al., 2011), which is correlation between the time series of two parcels after adjusting for the time series of all other network parcels. Partial correlation is calculated from the inverse of the covariance matrix, also known as the precision matrix. We calculated inverse covariance using cholesky decomposition (Krishnamoorthy & Menon, 2013) as illustrated in Equation 3.2 . For this derivation, we first factorized $\check{\Sigma}$ into an upper triangular U that satisfies $\check{\Sigma} = U' * U$ and then calculated the precision matrix $\check{\Sigma}^{-1} = \hat{\Omega} = w_{\text{ij(MxM)}}$. To convert precision into partial corre-

lation ρ_{ij}, there is a sign flip of the non-diagonal elements and also normalization as shown in Equation 3.3 where i and j are nodes.

$$\hat{\Omega} = \breve{\Sigma}^{-1} = U^{-1}(U^{-1})^t \tag{3.2}$$

$$\rho_{ij} = -\frac{w_{ij}}{\sqrt{w_{ii}w_{jj}}} \tag{3.3}$$

Tikhonov Partial Correlation: As partial correlation involves the calculation of the inverse of covariance matrix, this method becomes problematic when there are not considerably more timepoints than nodes. Several approaches based on regularization have been proposed in the literature to address this unstable matrix inversion problem. We used the Tikhonov regularization (also referred as L_2 ridge regression) (S. M. Smith et al., 2015; Colclough et al., 2018) as shown in Equation 3.4. This equation involves the addition of a regularization term $\Gamma = \alpha I$, a multiple of the identity matrix. The scalar α is optimized here by minimizing the root mean square distance [4] between regularised subject precision matrices ($\bar{\Omega}$) and the group average of the unregularised subject precision matrices ($\hat{\Omega}$). We optimised a single α for all the subjects.

$$\bar{\Omega} = (\breve{\Sigma} + \Gamma)^{-1} \tag{3.4}$$

Sparse Group Gaussian Graphical Modelling (SGGGM) Partial Correlation: Tikhnov regularization allows for a stable inversion for the covariance matrix;

[4] We have N subjects, and for each subject, we have d parcels. In the case of optimisation of the regularising parameters, we first calculated the difference between the regularised subject precision matrices ($d \times d \times N$) and the group average of the unregularised subject precision matrices ($d \times d$). Then, we calculated the sum of the resulting matrices (after squaring matrix elements), and obtained a $d \times d$ resulting matrix. Then, we summed up the elements in the upper triangular of the resulting matrix, and lastly computed the square root.

however, it uniformly shrinks off-diagonal elements. Sparse Group Graphical Gaussian Modelling (SGGGM) (Ng, Varoquaux, Poline, & Thirion, 2013) learns group-level connectivity in a data-driven way using an L_1 prior and regularization of each subject's connectivity towards the group mean using an L_2 prior. The advantage of SGGGM is that it supports intra-subject inverse covariance being similar, but not exactly the same as the group inverse covariance estimate in a similar manner to (Colclough et al., 2018). The inverse covariance ($\dot{\Omega}$) fed into SGGGM, is calculated using the regularization technique known as oracle approximating shrinkage (OAS), as described in the original paper (Ng et al., 2013), and is well regularized. SGGGM is posed as a regularized consensus optimization problem which can be solved using an alternating direction method of multipliers (ADMM) (Boyd, Parikh, & Chu, 2011). In Equation 3.5, $\|\|.$ corresponds to the Frobenius norm penalty, applied to the difference between the subjects' matrices and the group, where $\dot{\Omega}^S$ and $\dot{\Omega}^G$ denote the subject and group level matrices, respectively. $\| \dot{\Omega}^G \|_1$ is used for enforcing sparsity on the group prior and $tr(\check{\Sigma}^S\dot{\Omega}^S) - logdet(\dot{\Omega}^S)$ is the representation of each subject's own observation. Parameter λ_1 controls sparsity and λ_2 regulates the group consistency. We optimized these hyper-parameters (λ_1 and λ_2) by minimizing the root mean square distance between regularised subject precision matrices estimated using SGGGM and the group average of the unregularised subject precision matrices ($\hat{\Omega}$).

$$\underbrace{min}_{\dot{\Omega}^G>0,\dot{\Omega}^S>0} \quad \lambda_1 \| \dot{\Omega}^G \|_1 + [tr(\check{\Sigma}^S\dot{\Omega}^S) - logdet(\dot{\Omega}^S)] + \lambda_2 \| \dot{\Omega}^S - \dot{\Omega}^G \|^2. \quad (3.5)$$

Tangent Space Parameterization

Covariance matrices [5] by construction are symmetric and live in space of positive definite (PD) matrices. The PD matrix is a symmetric matrix whose eigenvalues are all positive. The standard practice is to directly vectorize (unwrap) these covariance matrices and then feed them into a machine learning predictor/classifier, however there are a few downfalls to this practice. One issue is that PD matrices do not naturally form a Euclidean space, meaning mathematical operations like subtraction do not apply, if geometry is to be preserved (and the PD nature preserved). For example, the difference of two PD matrices does not correspond to the PD covariance matrix of a signal. $C_1, C_2 \in Sym^+{}_n \not\Rightarrow C_1 - C_2 \in Sym^+{}_n$ (Varoquaux et al., 2010). Another problem is that in the space of the positive definite cone, where covariance matrices exist, elements of covariance are inter-correlated. This goes against the underlying assumption for some predictors/classifiers, which assume features are uncorrelated. Lastly, these vectorized connectivity estimates do not follow a normal distribution, which violates the assumptions of some methods, like Linear Discriminant Analysis (LDA).

If we follow the geometry of PD matrices, we note that the space of PD matrices forms a differentiable manifold $\boldsymbol{M}$. PD matrices therefore cannot directly be treated as an element of Euclidean space; therefore, to correctly apply mathematical formulations on these matrices, we utilise Riemannian geometry (Pennec, Fillard, & Ayache, 2006). For each covariance matrix C in manifold $\boldsymbol{M}$, we can define a scalar product in the associated tangent space (T) as $\boldsymbol{MT}_C$. This tangent space is homomorphic to the manifold, and therefore we can approximate the Riemannian distance in the manifold by using Euclidean distance in the tangent space. This allows us to apply simple Euclidean mathematical operations in the tangent

[5] In a broad sense, we have used the word covariance for different variants of the FC estimates. This should not be confused with empirical covariance, $\check{\Sigma}$.

space without breaking the geometry of connectivity matrices. Moreover, many popular classification/prediction algorithms (for example, Neural Network, LDA and SVM) cannot be directly implemented to act in a Riemannian manifold. As the projected connectivity matrices in the tangent space are treated as Euclidean objects, such classification algorithms can be readily applied in the associated tangent space.

Another important consideration for the tangent space parametrization is that the projection requires a reference point in the manifold (this is the point where the tangent plane touches the manifold) that should be close to the subject's projected covariance matrix. If we have a different reference point for each subject, each subject's covariance matrix would be projected to a different tangent plane. To ensure that all projected covariance matrices lie in the same tangent plane, we must find a reference covariance matrix close to all subjects' covariance estimates. This reference covariance matrix could be an average of the whole set of covariance matrices (Barachant, Bonnet, Congedo, & Jutten, 2011) and is referred to here as C_G. Following the logarithmic map for projecting any covariance matrix to its tangent plane (Barachant et al., 2013), any covariance matrix C can be projected to tangent space as shown in Equation 3.6, where log_m is matrix logarithm . Once the covariance matrices are in tangent space, they are no longer linked by the PD constraint, and hence these uncorrelated features are more useful for classifiers/predictors.

$$\vec{C} = log_m(C_G^{-1/2} C C_G^{-1/2}) \tag{3.6}$$

Reference Point in Tangent Space: C_G symbolizes the group reference estimate which may be estimated in different ways. The choice of reference PD matrix would normally be a group mean estimate as stated in Table 3.2. Euclidean mean

is the average of all covariance matrices whereas harmonic mean is the reciprocal of the arithmetic mean of the reciprocals. The Euclidean and harmonic means are calculated in the manifold space. Log Euclidean is calculated by first applying matrix logarithm, then computing the mean, and lastly applying the matrix exponential to bring back the mean to the manifold. The Riemannian mean is introduced in (Moakher, 2005) and can be computed in three iterative steps: (1) projecting connectivity estimates to the tangent space (initialized by using the arithmetic mean, C_e), (2) calculating the arithmetic mean in the tangent space, $\vec{C_e}$, and (3) mapping the arithmetic mean back to the manifold i.e., $C_e = C_e exp_m(\vec{C_e}){C_e}^T$. These three steps are repeated until convergence ($||\vec{C_e}||_F < \xi$) (Barachant et al., 2013). The Kullback mean is the geometric mean between harmonic mean and arithmetic mean. It is not evident in the literature which mean would lead to better representation of the manifold and therefore, we considered all these alternatives for reference point estimation.

Reference Mean	**Equation**
Euclidean	$C_e = \frac{1}{N}\sum_i C_i$
Harmonic	$C_h = (\frac{1}{N}\sum_i {C_i}^{-1})^{-1}$
Log Euclidean	$C_{le} = \exp_m(\frac{1}{N}\sum_i \log_m C_i)$
Riemannian Mean	$C_r = \arg\min(\sum_i \delta_R(C_e C_i)^2)$
Kullback	$C_k = {C_e}^{1/2}\left({C_e}^{-1/2} C_h {C_e}^{-1/2}\right)^{\alpha} {C_e}^{1/2}$

Table 3.2: Reference Mean Estimation. C_i denotes the FC matrix for subject i.

Shrinkage in Tangent Space

To further reduce estimation variability, we can apply regularization in the tangent space. The Ledoit-Wolf estimator shrinks the covariance towards a target matrix (H) in the ambient space as $C_{LW} = (1-\lambda)C + \lambda H$. In (Ledoit & Wolf, 2004),

the authors proposed shrinkage towards an identity matrix. A better alternative is to shrink towards the group reference estimate (as estimated in Table 3.2) (Rahim et al., 2019). The Ledoit-Wolf shrinkage can also be applied in tangent space as shown in Equation 3.7.

$$\vec{C}_{LW} = (1 - \lambda)\vec{C} + \lambda C_G \tag{3.7}$$

This shrinkage is isotropic and therefore does not take into consideration the distribution of the population. Recently, another method called PoSCE was proposed (Rahim et al., 2019) in which the population prior is not based on just the mean, but uses information from the probabilistic distribution of covariances. The prior dispersion is calculated as the mean outer product of the parametrized tangent space connectivity estimates as $\Lambda_o = \frac{1}{N-1}\sum_{i=1}^{N}\left\langle \vec{C}_i, \vec{C}_i \right\rangle$ and $\Lambda = \lambda I$ is the covariance of the likelihood where λ is the shrinkage control parameter. The shrunk covariance matrices in the tangent space can then be calculated as in Equation 3.8.

$$\vec{C}_{PoSCE} = (\Lambda^{-1} + \Lambda_o^{-1})^{-1}\Lambda^{-1}\vec{C} \tag{3.8}$$

These isotropic and non-isotropic shrinkage techniques could ideally be applied to any projected connectivity estimate (covariance, full correlation, partial correlation, Tikhonov partial correlation or SGGGM partial correlation) in the tangent space. Theoretically, Tikhonov partial correlation and SGGGM connectivity estimates are already well-regularized so they may not require any further shrinkage in the tangent space, but we have tested all possible variants in this study (regularization in the ambient space and then further regularization in the tangent space). Moreover, we have introduced concepts of three spaces as detailed below:

- **Ambient Space**: The estimated FC estimates (Section 3.3.2) were not pro-

jected into the tangent space. The classifier/predictor was directly applied to connectivity estimates in the ambient space.

- **Tangent Space**: The various FC estimates (covariance, full correlation, partial correlation, Tikhonov partial correlation or SGGGM partial correlation as defined in Section 3.3.2) were projected to tangent space following Section 3.3.2, but no shrinkage was applied in the tangent space. The classifier/predictor was applied on the parameterized tangent space connectivity estimates.

- **Tangent_Shrinkage Space**: The shrunk tangent space in which FC estimates were projected to tangent space and then shrinkage was applied in the tangent space (Section 3.3.2). In this case, the classifier/predictor was applied on the parameterized shrunk tangent space connectivity estimates.

3.3.3 Finding the best classifier/predictor

The last step is to feed these FC matrices to predictor/classifier[6] as features to predict non-imaging variables.

Elastic Net: The first classifier/predictor we tested is Elastic Net (Zou & Hastie, 2005), which is a regularized regression method that combines the penalties of lasso (L_1) and ridge (L_2) methods. L_1 encourages a sparser model (more zeros) but fails to accomplish group variable selection (where strongly correlated features are selected or discarded together) (Vidaurre, Bielza, & Larrañaga, 2013). Alternatively, L_2 encourages the grouping effect and removes the limitation on the number of chosen variables but it does not achieve sparsity. L_2 is a reasonable

[6] *Predictor* is used sometimes as synonymous for input variable, and *classifier* is mostly used for categorical variables prediction. We have used *predictor/classifier* combination in an effort to make sure that the audience is not confused, and this implies prediction of both categorical and continuous non-imaging variables.

approach to handle the ill-posed problems that result from extremely correlated features. Elastic net aims to overcome the limitations of both L_1 and L_2 penalties. The data matrix X is $s \times p$, where s is the number of subjects and p is the number of features, and y is the response vector, such as age or sex. The FC matrices are symmetric and only values above the diagonal need to be retained and vectorized, and hence p = d(d-1)/2, where d is number of parcels. We then applied univariate pre-feature selection, and only retained the top 50% features (based on correlation with the target variable). In Equation 3.9, $\|_1$ corresponds to the L_1 norm penalty and the quadratic part (the second term) is the is L_2 norm.

$$\hat{\beta} = \underset{\beta}{argmin} \| \ y - X\beta \ \|^2 + \lambda_2 \| \ \beta \ \|^2 + \lambda_1 \ \| \ \beta \ \|_1 \tag{3.9}$$

BrainNetCNN: The second classifier/predictor, BrainNetCNN, is specifically developed to leverage the topological structure of networks (Kawahara et al., 2017). It consists of edge-to-edge (E2E), edge-to-node (E2N) and node-to-graph (N2G) convolutional filters, which are all designed to analyze FC to predict the behavioural information. The main contribution of BrainNetCNN is using a cross shape filter (the dimension of some of the filters are set to be $d \times 1$ and others $1 \times d$, whereas, the overall size of input connectivity matrix is $d \times d$) instead of the traditional box shape filter. There are primarily three main layers used in this network, E2E, E2N and N2G layers. The E2E layer considers the weights of all the edges which share a node together. Intuitively, the E2E layer resembles a convolutional filter, as for an edge (j,k) in a FC matrix, it combines the signal with the signal from direct neighbours (edges connected to either node j or node k), but applies a cross shape filter instead of a box shape filter. The output of the E2E layer is fed into the E2N layer. The E2N layer is equivalent to convolving the FC matrix with a spatial convolutional 1D row filter and a 1D column filter (and adding their

results). It gives a unique output value for each node, j, as it takes the average of incoming and outgoing weights of each edge associated to node j. Finally, the N2G graph layer is used, which is similar to the fully connected layer. The N2G layer reduces the spatial dimensionality, and outputs a single scalar for weighted combination of nodes per feature map. The output features from the N2G layer are then fed into a predictor to provide final prediction (non-imaging variables). The architecture of the BrainNetCNN is shown in Figure S2 (sub-figure A).

Recurrent Convolutional Neural Network: A Convolutional neural network (CNN) is a feed-forward network inspired by the brain's visual system; however, in comparing it with the brain, CNN lacks the recurrent connections which are abundant in the brain. Inspired by (Liang & Hu, 2015), we have designed a recurrent convolutional neural network (RCNN) by incorporating recurrent connections in each convolutional layer. The activities of RCNN evolve over time (the activity of each unit is modulated by the activities of its neighboring units), which provide the advantage of an increased receptive field, and results in a greater awareness of contextual information. We have implemented a 2D RCNN to extract features from FC matrices. The most important module of RCNN is the recurrent convolution layer (RCL). The net input z_{ijk} at unit (i,j) on kth feature map is given by

$$z_{ijk} = (w_k^f)^T u^{(i,j)}(t) + (w_k^r)^T x^{(i,j)}(t-1) + b_k \tag{3.10}$$

In Equation 3.10, $u^{(i,j)}(t)$ is the feed forward input, as in a CNN, while (w_k^f) is the vectorized feed forward weights. The second term of the equation arises because of the recurrent connections and $x^{(i,j)}(t-1)$ represents the recurrent input and $(w_k^r)^T$ is associated with recurrent weights. In other words, $u^{(i,j)}(t)$ and $x^{(i,j)}(t-1)$ are centered at (i, j), and are the vectorized square patches of the feature maps

in the present and foregoing layer respectively. The last term of the equation, (b_k), is the bias. z_{ijk} will pass through the rectified linear activation function $f(z_{ijk})$ and then local response normalization (LRN) will be applied on it, which imitates lateral inhibition in cortex. In neurobiology, lateral inhibition refers to the capacity of an excited neuron to reduce the activity of its neighbors. Similarly here, feature maps will challenge each other for maximum representation. The architecture of 2D RCNN is shown in Figure S1. Equation 3.10 describes the behaviour of one RCL (recurrent block) in Figure S1. Unfolding the RCL for t = 0 will result in a purely feed-forward network. In Figure S1, we have unfolded each RCL for t = 3, leading to a feed-forward network with largest depth of t+1 = 4 and smallest depth of t = 1. Each RCL block has convolution, ReLU, batch normalization and addition layers (collectively named as a residual block). Moreover, each residual block receives input directly from feed-forward and recurrent connections. The recurrent input evolves over iterations but the feed-forward input remains the same for all iterations. The overall 2D RCNN is composed of four RCL blocks, and between each RCL block, only feed forward connections are used. The output from the last RCL is fed into fully connected layers.

Graph Convolutional Neural Network: The standard convolution is limited to data on a Euclidean grid; however, the graph convolutional neural network (GraphCNN) extends beyond traditional CNNs to handle data that is supported on a graph. GraphCNN exploits the Laplacian of the graph and is built on spectral filters, graph coarsening, and efficient pooling, which are all based on established tools in graph signal processing (Shuman, Narang, Frossard, Ortega, & Vandergheynst, 2013). We followed the approach of (Defferrard, Bresson, & Vandergheynst, 2016) and constructed the graph $G = (V, E, A)$ where $d_x = |V|$ is the set of vertices, E is the edges and A is the adjacency matrix. GraphCNN takes input as a feature matrix $X = [x_1, ..., x_n]^T \in \mathbb{R}^{s \times d_x}$, where s is number of

nodes and d is number of features for each node. In our case, each subject is represented by a node and corresponding features are computed by vectorizing the FC matrix. The other input to the GraphCNN is an adjacency matrix, $A \in \mathbb{R}^{d_x \times d_x}$ which is essentially a graph between features. We defined the adjacency matrix based on the similarity estimate between features. For constructing the graph, the Euclidean distance $d_{ij} = ||x_i - x_j||_2$ was used as a distance function, a Gaussian kernel $a_{ij} = \exp(-d_{ij}^2/2\sigma^2)$ was applied and a *k* nearest neighbours graph was formulated. The architecture of GraphCNN is shown in Figure S2 (sub-figure B).

Deconfounding the Data

Many participant-related confounds exist in resting state data, which can corrupt associations between the imaging data and non-imaging data. Our main goal was to predict non-imaging variables based only on *pure* FC, hence we found it was crucial to minimize the impact of confounding variables. The basic confounds we controlled for were age, sex, ethnicity, weight, average head motion, brain size, intercranial volume, variables (x, y, z, table) related to bed position in scanner, imaging centre and confounds modelling slow date-related drift. We also included some non-linear effects, i.e., age^2, $sex \times age$, $sex \times age^2$. We evaluated the effect of various deconfounding schemes on X, a data matrix (i.e, subjects' FC), and Y, a response vector (e.g, fluid intelligence). For X there are three possible strategies, (i) no deconfounding (X0), (ii) confound regression from data that is not contained within the cross-validation loop (X1), (iii) fold-wise confound regression (X2), following (Snoek, Miletić, & Scholte, 2019). For Y, there are two options, (i) no deconfounding (Y0), or (ii) fold-wise confound regression (Y1). For example, deconfounded $\tilde{X}$ is calculated as $\tilde{X} = X - C\beta$ where C are confounds and β are regression weights. The β can be calculated as $\beta = C^+X$, where C^+ is the pseudoinverse of the confounds.

3.3.4 Training, validation and testing

For the Elastic Net, k fold nested cross-validation was performed (k=5). The subjects from both HCP and UKB data were divided into k folds. For each folds, a set was selected as the outer test set and the remaining k-1 sets were used as the outer training set. Then, an inner loop is used to tune the hyper-parameters via k-fold validation on the training set. The optimised model (using tuned hyper-parameters) was then evaluated on the test set and this procedure was repeated for all k test folds. In summary, the outer loop is used for model evaluation and the inner loop for the model selection phase. The important hyper-parameter to tune for Elastic Net is the weights on the ridge and lasso penalty. Lastly, we considered the family structure for HCP data and the family members were kept in the same folds. A summary figure explaining the process of nested cross-validation is shown in Figure S3.

In case of CNNs, there are a number of hyper-parameters to tune, e.g., learning rate, number of hidden units, width of convolutional kernel, etc. For our benchmark study, we have hundreds of different variants of FC estimates as input, e.g., one possible variant of FC is estimated by following steps: step-1 (ICA parcellation), step-2 (SGGGM), step-3 (tangent space), step-4 (shrinkage in tangent space), step-5 (X1Y1 as deconfounding). This computationally makes it impossible to tune hyper parameters separately for each input. We randomly selected a few input variants ($\sim$ 20) and tune the hyper-parameters for these selected inputs (for all non-imaging variables prediction). This gave us the estimate of hyper-parameter values for which CNNs perform optimally (convergence and relative higher prediction accuracy). We then fixed these hyper-parameters values for all non-imaging variables, prediction for both UKB and HCP datasets for all FC estimates. In Chapter 5, we performed external validation of the pipeline, and these

hyper-parameters were still fixed on un-seen new datasets. This approach has the disadvantage that it can negatively affect CNN performance but it does avoid over-fitting concerns, and also tests the generalisability of the CNN architecture. The fine-tuning was performed within the training-validation-testing framework. The tuned hyper parameters for CNNs are reported in Table S1.

We took advantage of large N of UKB (and HCP) and did not apply the inner cross-validation framework for CNN evaluation, but used the repeated training-validation-test splits. The subjects were split into training (72%), validation (8%) and testing (20%) sets. The training, validation and testing split was repeated 5 times; therefore, each model was trained five times and all subjects were tested once. The validation set here was used to assess performance of CNN, e.g., early stopping (when error on validation set grows but can select the previous optimum model) and model preparation (deep features selection) etc. It should be noted that any pre-processing of connectivity estimates (e.g., normalization) is performed within the cross-validation/training-validation-testing folds, and all CNN based results reported in this work are on the testing data.

3.3.5 Statistical testing

Paired t-test: A paired t-test is performed to determine whether there is statistical evidence that the mean difference between paired observations (e.g., no shrinkage vs shrinkage approaches) on a particular outcome (i.e., fluid intelligence score prediction) is significantly different from zero. We have used symbols to report the test results; ns (not significant) when $P > 0.05$, * when $P \leq 0.05$, ** when $P \leq 0.01$ and *** when $P \leq 0.001$.

Fisher transformation: To generate confidence intervals on the prediction correlations (i.e., continuous outputs e.g., age), we computed the Fisher transformation.

If the prediction correlation is r, then Fisher transformation F(r) is $arctanh(r)$. F(r) approximately follows a normal distribution, with mean = F(p) = $arctanh(p)$, and standard error, $SE = \frac{1}{\sqrt{n-3}}$, where n is sample size (i.e., number of subjects) and p is the true correlation coefficient. Then z-score is calculated as $z = \frac{x-mean}{SE} = [F(r) - F(p)]\sqrt{n-3}$. The confidence interval (CI) for p can be calculated as in Equation 3.11. The inverse Fisher transform is then calculated to bring the CI to the r units.

$$100(1-\alpha)\%CI : arctanh(p) \in [arctanh(r) \pm z_{\alpha/2}SE] \tag{3.11}$$

Wilson test: To draw the confidence intervals on the prediction accuracies (i.e., discrete outputs such as sex), we performed the Wilson test (suitable for the binomial distribution). The Wilson score interval is explained in (Bender, 2001).

Repeated k-fold cross validation: We repeated 100 times the nested 5-fold cross validation on the top performing pipelines. We used the 99% confidence limits on these 100 prediction values to generate the confidence intervals. The rationale is to probe well the sampling variability and to test reproducibility of the top performing pipelines.

3.3.6 Imaging Data

The rfMRI data of 13301 subjects from UKB and 1003 subjects from the HCP were used in this analysis. The pre-processing pipelines for UKB and HCP are described in Chapter 2. The length of the scanning session for UKB is 6 minutes, and for HCP is 1 hour (4 sessions of 15 mins each). For UKB, we have 490 timepoints and for HCP, we used data from all 4 sessions which gave us 4800 time points in total (1200 timepoints in each session).

3.3.7 Non-Imaging Data

Here, we have chosen four non-imaging variables: age, sex, fluid intelligence score and neuroticism score (the latter only for UKB), for following reasons:

Age: Improving prediction of age from brain data is of interest, to further clinical understanding of how age links to neurodevelopment and degeneration. For instance, comparing the discrepancy between an individual's chronological age and the age predicted from neuroimaging data can serve as a biomarker for several age-linked brain diseases (Abbott, 2011). Age can be accurately predicted from structural MRI data (Cole et al., 2017); however, prediction of age from FC estimates is not as accurate. Nevertheless, this represents a non-imaging variable that can be predicted reasonably strongly (as opposed to fluid intelligence and neuroticism).

Sex: Sex classification through FC is comparatively accurate, and so this variable provides a valuable evaluation parameter to compare benchmark performance of different models.

Fluid Intelligence: Fluid intelligence refers to the ability to reason and to solve new problems independently of previously acquired knowledge (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008). It is highly related to professional and educational success and represents overall performance across a broad range of cognitive abilities. A number of recent studies have predicted fluid intelligence from FC (He et al., 2018), but due to measurement variability, correlations remain rather weak. However, more precise and novel methods for the estimation and use of FC could lead to better predictions, and provide a more precise understanding of how brain networks relate to cognitive abilities.

Neuroticism: Mental health is being assessed in a number of ways within UKB.

Participants answer a 12-item neuroticism questionnaire as part of their baseline assessment. We used the algorithm employed in (*Neuroticism Score*, 2018) to calculate the score of neuroticism for each subject.

For the purpose of this work, we have shown the prediction accuracy (for categorical variables) and prediction correlation (for continuous variables). There could be an argument in favor of variance (or coefficient of determination) over correlation. However, correlation is commonly used and well-understood by the readers and that was our rationale behind selecting correlation (as we want this pipeline work to be directly comparable to existing literature work). Also, for the reported results in this work, correlation or coefficient of determination carried exactly the same information.

3.4 Results

3.4.1 Parcellation and FC Estimation

Figure 3.2 shows the prediction accuracy/correlation for non-imaging measures from F estimates across different parcellation schemes (Section 3.3.1) and FC estimation methods (Section 3.3.2). The results in Figure 3.2 were calculated by directly vectorizing the connectivity estimates without projection into tangent space. Elastic Net was used as a predictor/classifier with nested 5-fold cross-validation. The results illustrated in Figure 3.2 are a reduction of the full set of results, chosen to highlight the performance of methods employed in step 1 (brain parcellation) and step 2 (FC estimation) using the relatively robust Elastic Net as the predictor/classifier. The results shown in Figure 3.2 were calculated after removing the effect of confounds from data (deconfounding), and Figure S4 shows prediction estimates without confound removal.

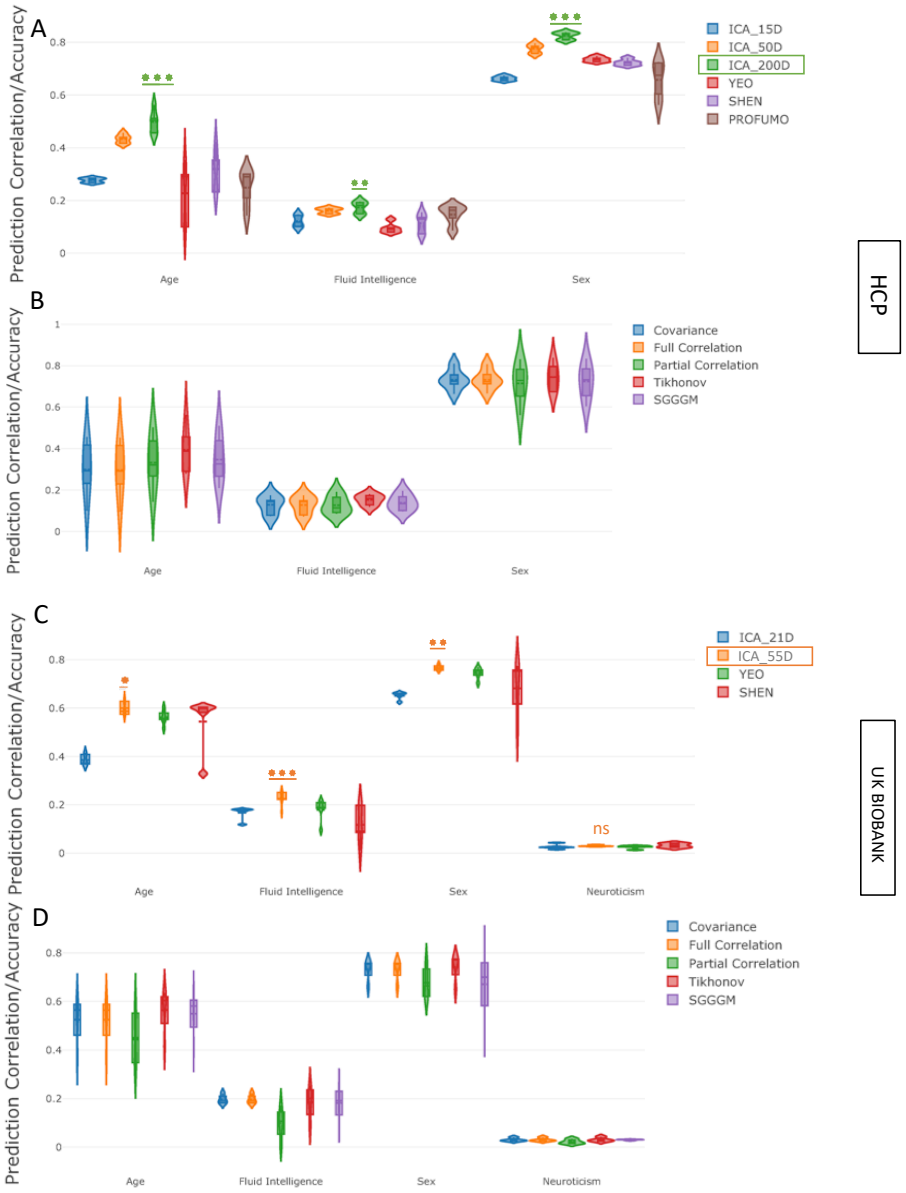


Figure 3.2: The impact of various parcellation strategies and FC estimation methods on prediction power for non-imaging variables **after deconfounding**. **[A,B] (HCP Data), [C,D] (UKB Data):** The violin plots in **A** and **C** show the prediction variability over 5 measures of FC estimates and in **B** and **D** show the prediction variability the over different parcellation schemes. For HCP, the ICA based parcellation schemes are ICA_15D, ICA_50D and ICA_200D, and for UKB are ICA_21D and ICA_55D, where D = the number of parcels. For both HCP and UKB, SHEN parcellation was 268D, YEO was 100D, and PROFUMO was 50D (for HCP only). The stars refer to comparison against the next-best method.

3.4.2 Tangent Space Projection and Shrinkage

A reference point is required when projecting FC estimates into the tangent space. This reference mean binds each of the connectivity estimates into the same tangential plane. The results in Figure S5 show the prediction correlation for fluid intelligence using different reference means. For each mean, FC estimates from all parcellation schemes (Section 3.3.1) were estimated by using all FC estimation methods (Section 3.3.2). The predictor/classifier employed here for fluid intelligence prediction was Elastic Net using nested 5 fold cross validation. For simplicity, only fluid intelligence prediction scores are displayed in Figure S5, as the prediction correlations for other non-imaging variables follows the same trend across various reference mean estimation techniques. The violin plots in Figure S5 show the prediction variability over 4 different parcellation schemes and 5 measures of FC estimates in the tangent space. The results shown in Figure S5 were calculated after deconfounding, and Figure S6 shows prediction estimates without deconfounding.

Figure 3.3 (sub-figures A, C, E and G) illustrates how the prediction accuracy was modified by projecting connectivity estimates into the tangent plane. Again, we varied the parcellation schemes and FC estimation techniques, but fixed Elastic Net as the predictor/classifier. The results illustrated in Figure 3.3 (sub-figures A,C,E and G) are a reduction of the full set of results, chosen to highlight the performance of tangent space parameterization (Section 3.3.2). Figure 3.3 (sub-figures B,D,F and H) compares the effect of applying shrinkage in the tangent space (Section 3.3.2). The shrinkage applied was non-isotropic shrinkage towards the population dispersion (PoSCE). The results in Figure 3.3 (sub-figures B,D,F and H) are based on FC estimates that have been projected into the tangent plane (Section 3.3.2). The violin plots in Figure 3.3 (sub-figures B,D,F and

H) show the prediction variability over 4 different parcellation schemes and 5 measures of FC estimates in the normal tangent space (no shrinkage) versus tangent_shrinkage space (PoSCE), (Section 3.3.2). Similarly to the previously displayed results, Elastic Net was used as the predictor/classifier for prediction of non-imaging measures. Essentially, Figure 3.3 shows side-by-side the effect of tangent space projection (from ambient space to tangent space) and tangent space shrinkage (staying within tangent space). The length of the scanning session in HCP is 60 minutes, Figure 3.3 also shows the prediction performance on the original 60 minutes scan, and on cut-down data (first 15 minutes and first 5 minutes of the scan for each subject). Moreover, the results shown in Figure 3.3 were calculated after deconfounding, and Figure S7 shows similar prediction estimates without deconfounding (for the full length scanning session only).

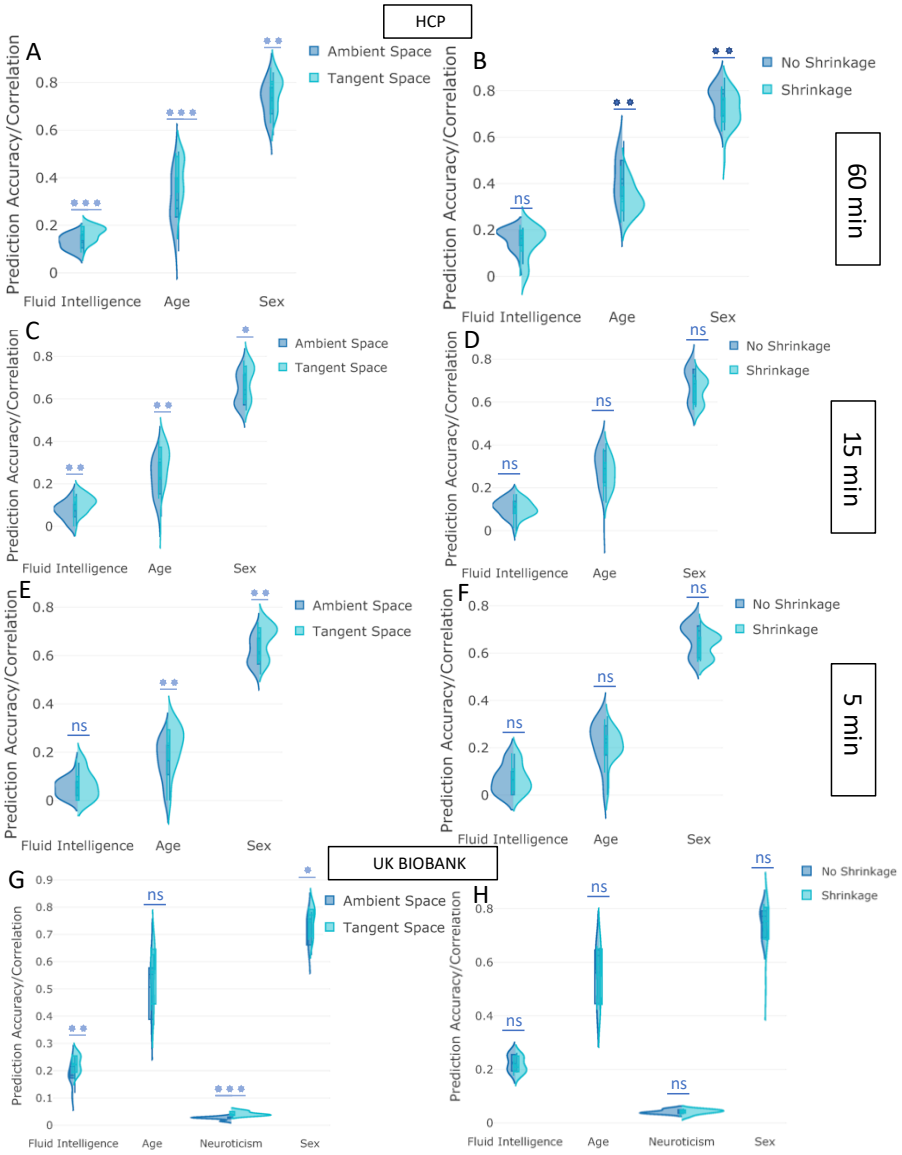


Figure 3.3: The impact of projecting to tangent space and applying shrinkage in tangent space on prediction power **after deconfounding**. **[A-F] (HCP Data), [G,H] (UKB):** The y-axis depicts the prediction accuracy/correlation for different behavioural measures. "Tangent Space" means that tangent space projection was applied on FC estimates (originally in the "Ambient Space"). The "Shrinkage" strategy means that non-isotropic PoSCE shrinkage was applied to connectivity estimates in tangent space before feeding to the predictor/classifier. "No Shrinkage" means that projected FC estimates in tangent space were directly fed to the predictor/classifier, and did not undergo PoSCE shrinkage. The violin plots show the prediction variability over 4 different parcellation schemes and 5 measures of FC estimates.

The Bland-Altman plot is shown in Figure 3.4 to describe agreement between shrinkage versus no shrinkage approaches. The Figure 3.3 (B and H sub-figures) and Figure 3.4 are precisely two different ways of visualizing the similar results. Figure S8 compares the performance of isotropic (Ledoit-Wolf) versus non-isotropic (PoSCE) shrinkage in the tangent space and Figure S9 compares the performance of isotropic (Ledoit-Wolf) versus no shrinkage strategy (projected FC estimates in the tangent space were directly fed to the predictor/classifier, and did not undergo any shrinkage). All the results displayed in Figure 3.3, 3.4, S8 and S9 were calculated after deconfounding. The violin plots in Figure S10 show the effect of varying the dimensionality of parcellation on shrinkage approaches for 5 measures of FC estimates, e.g., it compared the effect of tangent space shrinkage for higher dimensional ICA vs low dimensional ICA.

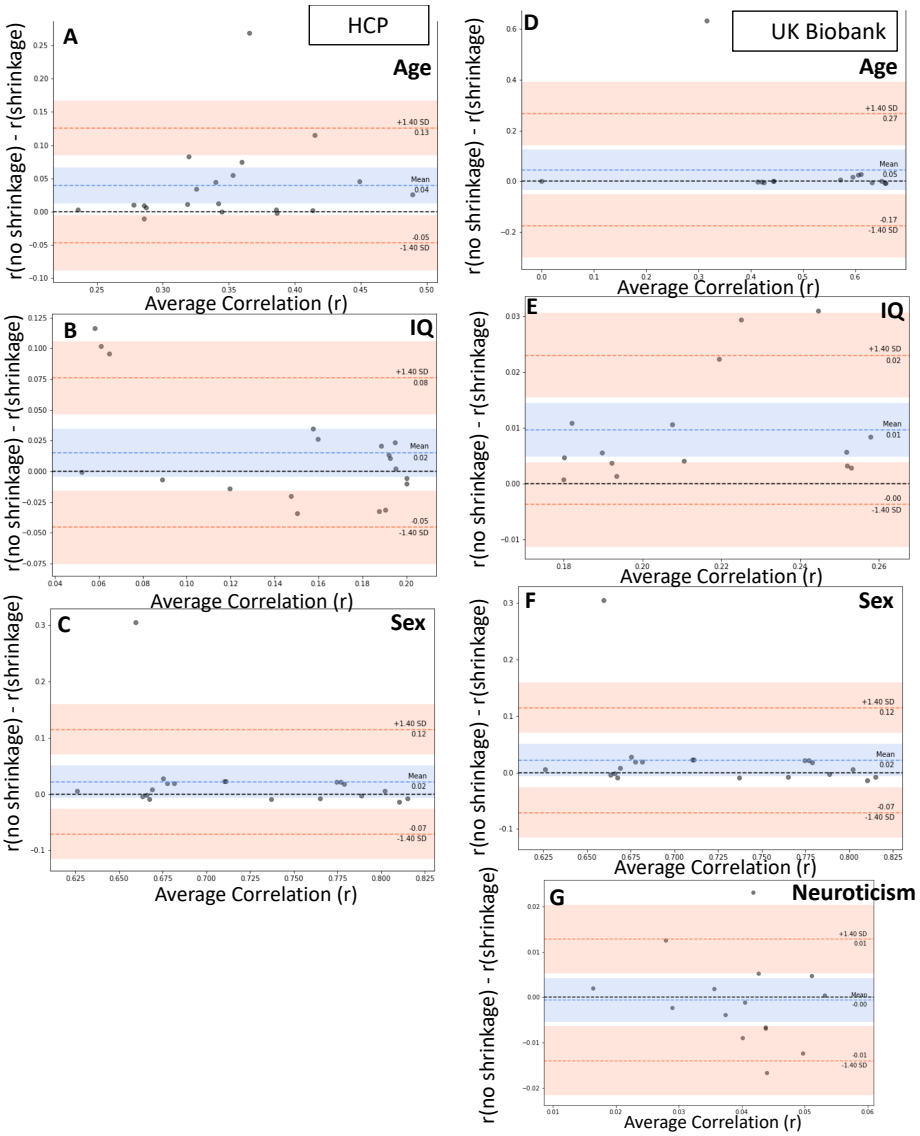


Figure 3.4: The impact of applying shrinkage in tangent space on prediction power **after deconfounding**. **[A,B,C] (HCP Data), [D,E,F,G] (UKB Data):** The y-axis depicts the difference of two methods (no shrinkage - shrinkage) and x-axis depicts the mean of methods ((no shrinkage + shrinkage) /2).

3.4.3 Deconfounding

We compared various deconfounding strategies (Section 3.3.3), by varying ICA dimensionality (Section 3.3.1) and using Elastic Net as the predictor/classifier.

Figure 3.5 shows the prediction correlations of fluid intelligence from FC estimates for each deconfounding strategy. The violin plots in Figure 3.5 shows the prediction variability over each specific sICA dimensionality, 5 measures of FC estimates and 3 different spaces (ambient space, tangent space and tangent_shrinkage space). Again, the results illustrated in Figure 3.5 are a reduction of the full set of results, chosen to highlight the performance of different deconfounding strategies. (Section 3.3.3)

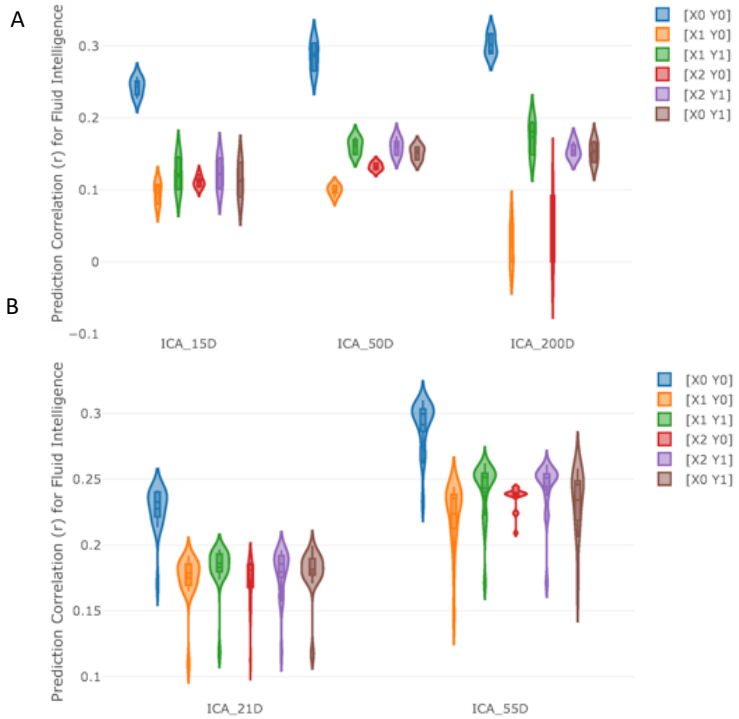


Figure 3.5: The impact of deconfounding strategies on the prediction power. **[A] (HCP), [B] (UKB)** This figure shows the prediction correlation for fluid intelligence score for various ICA based parcellation techniques across the 6 possible deconfounding strategies that were tested. The violin plots show the prediction variability over each specific sICA dimensionality, 5 measures of FC estimates and 3 different spaces (ambient space, tangent space and tangent_shrinkage space). X is an input matrix (i.e, subjects' FC), and Y, a response vector (e.g., fluid intelligence). The details of different deconfounding strategies are explained in Section 3.3.3.

3.4.4 Predictors/Classifiers Performance

Figure 3.6 compares the performance of different predictors (Section 3.3.3) on the prediction between FC and non-imaging variables. Inputs included FC estimates (Section 3.3.2) both in ambient space (without tangent space projection) and tangent space (Section 3.3.2). Estimates with tangent space parameterization included estimates with and without PoSCE shrinkage (tangent space and tangent_shrinkage space). Therefore, the violin plots in Figure 3.6 show the prediction variability over 4 different parcellation schemes, 5 measures of FC estimates and 3 different spaces (ambient space, tangent space and tangent_shrinkage space). Again, the results illustrated in Figure 3.6 are a reduction of the full set of results, chosen to highlight the performance of different classifiers/predictors (Section 3.3.3). The results in Figure 3.6 compared the performance of Elastic Net, 2D RCNN and BrainNetCNN (Section 3.3.3). We included GraphCNN initially, but dropped it due to consistently poor performance compared with the other models. Figure S11 shows the comparative performance of GraphCNN and Elastic Net. The subset of results chosen for Figure S11 were the ones where GraphCNN exhibited the best performance, although still underperforming in comparison to Elastic Net.

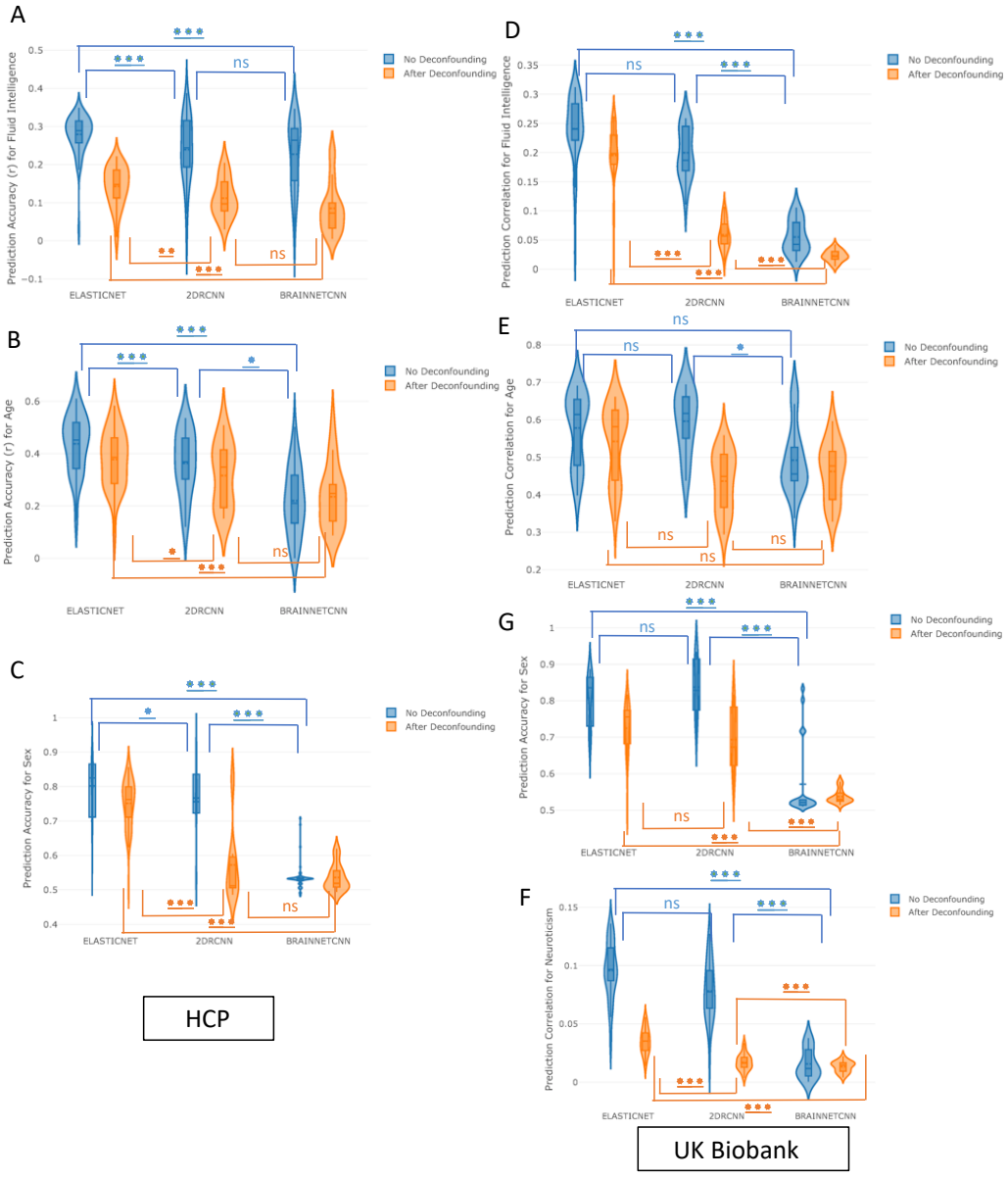


Figure 3.6: The performance of different classifiers/predictors on prediction power with and without deconfounding. **[A,B,C] (HCP Data), [D,E,F,G] (UKB Data):** The prediction accuracy/correlation for different non-imaging measures is depicted on the y-axis. The violin plots show the prediction variability over 4 different parcellation schemes, 5 measures of FC estimates and 3 different spaces (ambient space, tangent space and tangent_shrinkage space).

3.4.5 Top Performing Configurations

Figures 3.7 and 3.8 show the top ten methods in terms of predictive power, with strategies from different steps identified. The results shown in Figure 3.7 and 3.8 were calculated after deconfounding, and Figures S12 and S13 show prediction estimates without deconfounding. The error bars in Figures 3.7, 3.8, S12 and S13 were generated using the Fisher transformation (for continuous output) and Wilson test (for discrete output) (Section 3.3.5). The highlighted red blocks show the recommended pipelines (rationale explained in Section 3.5), and the red dotted lines highlight the point when the error bar of pipeline after the dotted line is out of range from the error bar of the top (first) pipeline.

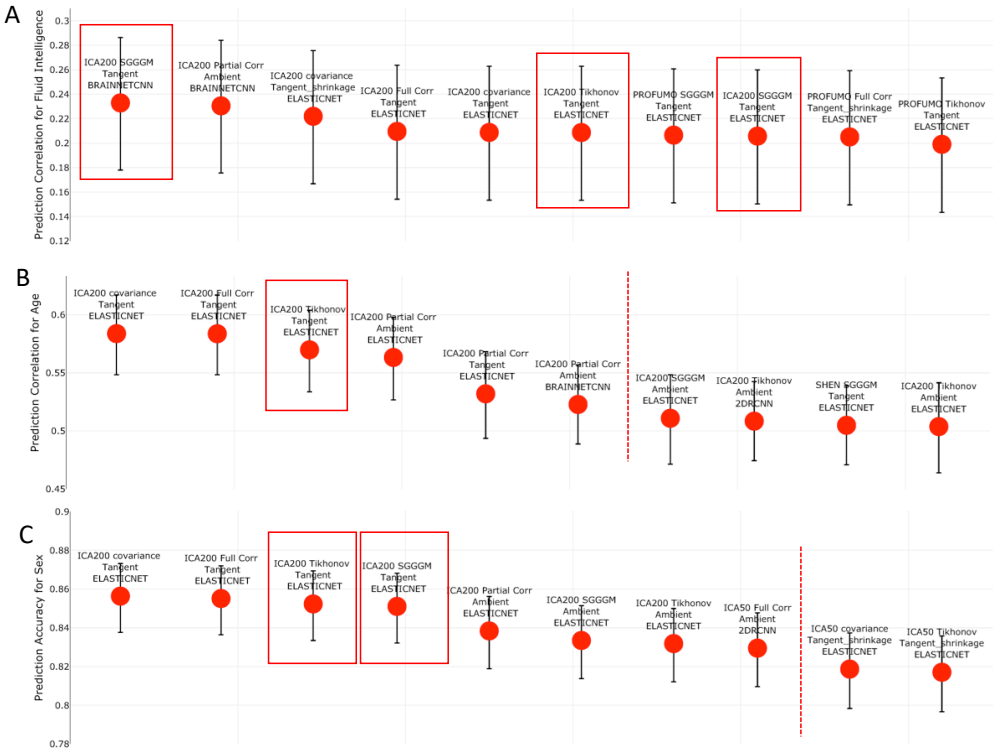


Figure 3.7: The top performing ten configurations for the prediction of each non-imaging variable by dataset **after deconfounding**. **[A,B,C] (HCP Data)** Each data point represents a different configuration strategy that may vary in terms of parcellation strategy, the FC estimation method, whether tangent space parameterization was employed, whether tangent space regularization was employed, and the predictor that was used. The first word indicates the parcellation strategy, and the second word refers to the FC estimation method. The third word refers to the geometry in which classifier/predictor is applied, ambient referring to non-tangent space and tangent referring to the projected covariance matrices in tangent space. If non-isotropic shrinkage was applied after projecting covariance matrices to tangent space, the fourth word will be "shrinkage". The last word indicates the type of classifier/predictor that was used. The highlighted red blocks show the recommended pipelines (rationale explained in Section 3.5), and red dotted lines highlight the point when the error bar of pipeline after the dotted line is out of range from the error bar of the top (first) pipeline.

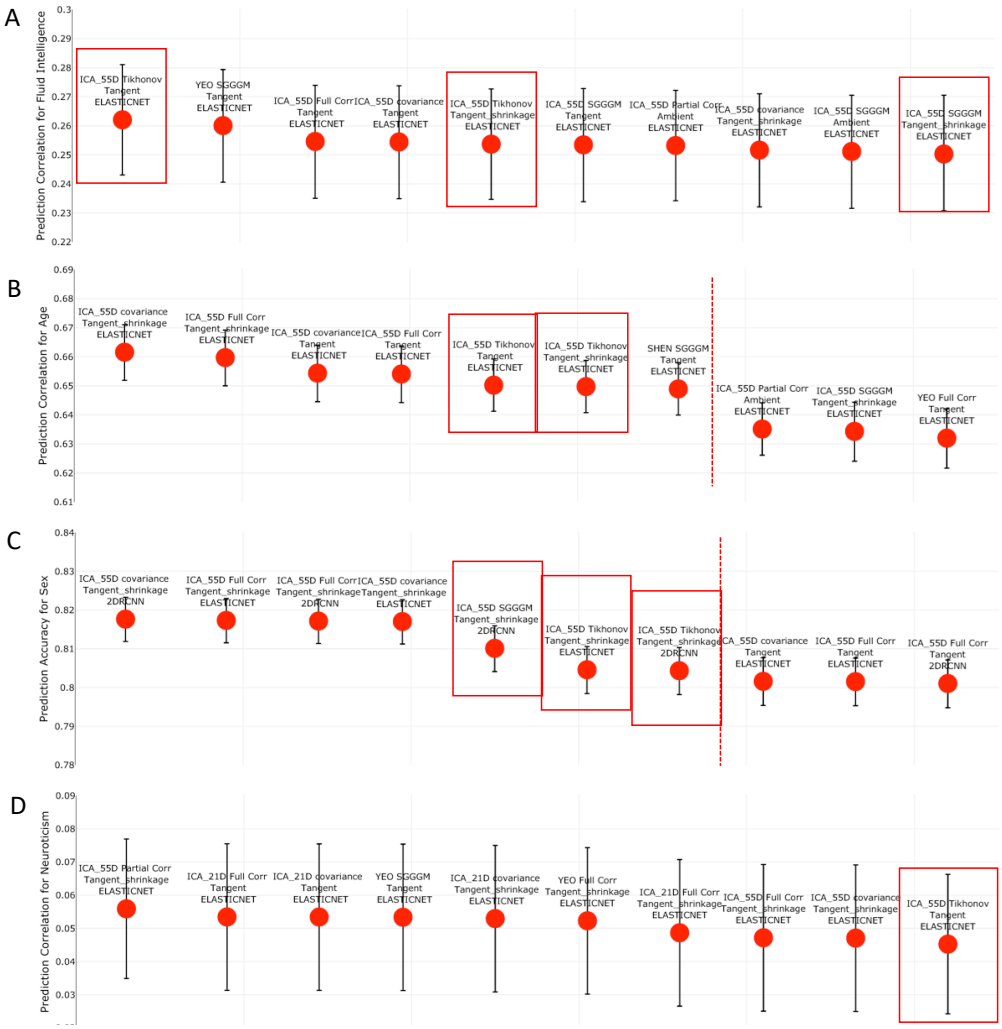


Figure 3.8: The top performing ten configurations for the prediction of each non-imaging variable by dataset **after deconfounding**. **[A,B,C,D] (UKB Data)** Each data point represents a different configuration strategy that may vary in terms of parcellation strategy, the FC estimation method, whether tangent space parameterization was employed, whether tangent space regularization was employed, and the predictor/classifier that was used. The first word indicates the parcellation strategy, and the second word refers to the FC estimation method. The third word refers to the geometry in which classifier is applied, ambient referring to non-tangent space and tangent referring to the projected covariance matrices in tangent space. If non-isotropic shrinkage was applied after projecting covariance matrices to tangent space, the fourth word will be "shrinkage". The last word indicates the type of classifier/predictor that was used. The highlighted red blocks show the recommended pipelines (rationale explained in Section 3.5), and red dotted lines highlight the point when the error bar of pipeline after the dotted line is out of range from the error bar of the top (first) pipeline.

3.4.6 Joint Results

Figures 3.9, 3.10 and 3.11 show the combination of various techniques at different stages of the pipeline for fluid intelligence, sex, and age prediction respectively for both HCP and UKB data after deconfounding. Figures S14, S15, S16 show similar results for fluid intelligence, sex and age prediction respectively but without deconfounding. In each of these figures, a few parameters were fixed across the different pipeline combinations (Euclidean mean as reference mean, non-isotropic shrinkage as shrinkage in tangent space). Figure S17 provides results for the prediction of neuroticism score in UKB data, and results are displayed both before and after controlling for confounds.

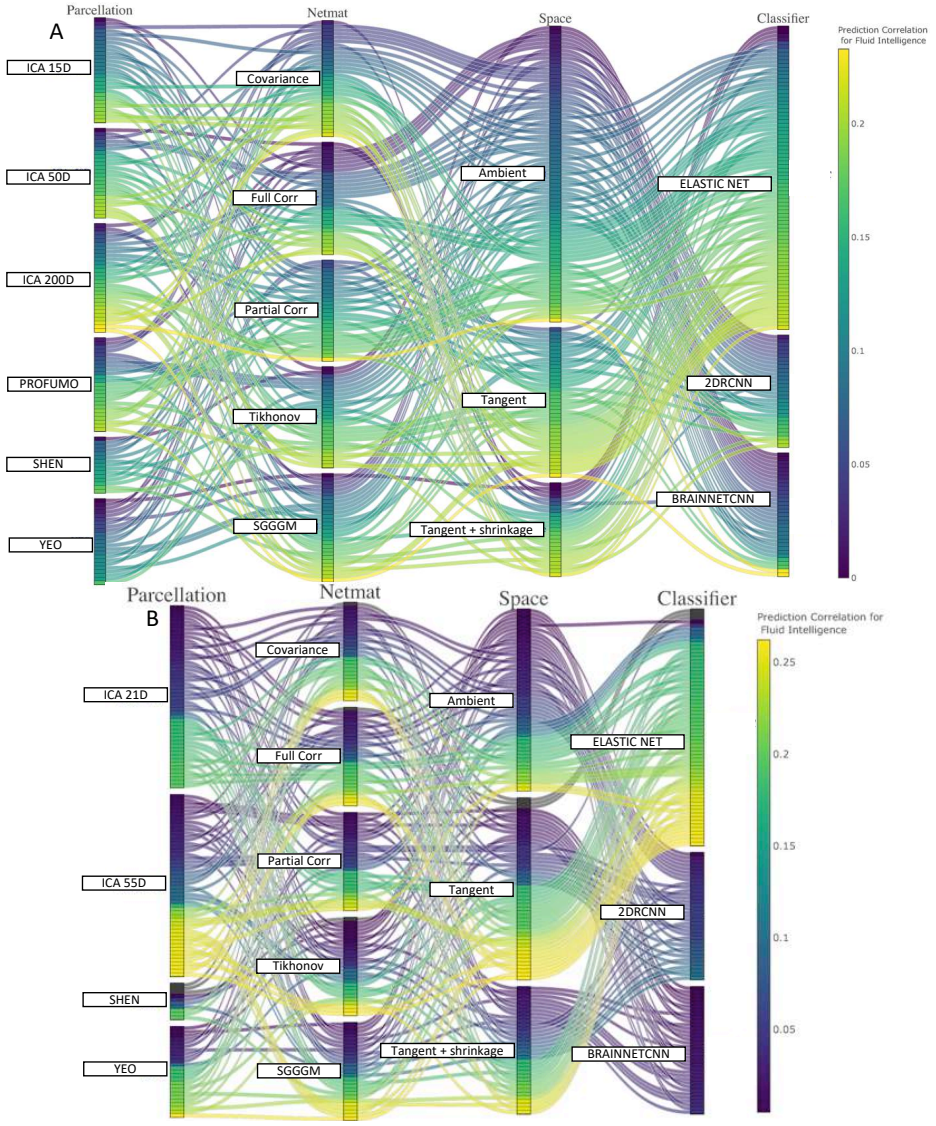


Figure 3.9: This parallel coordinates plot provides a visualization of all possible combinations of options in the pipeline to predict fluid intelligence scores from FC **after deconfounding**. **[A] (HCP Data), [B] (UKB Data):** The lines are color-coded according to their prediction performance. The vertical ordering is chosen with respect to prediction performance for each sub-block (in ascending order of prediction correlation/accuracy). For example, the lower (from bottom) the position of lines in a block, the higher the prediction accuracy of that variant (and vice-versa). With respect to the vertical ordering of sub-blocks (e.g., ambient or tangent space), that is chosen by hand (but it should not make any difference as the crossing of different lines does not explain any meaningful information here).

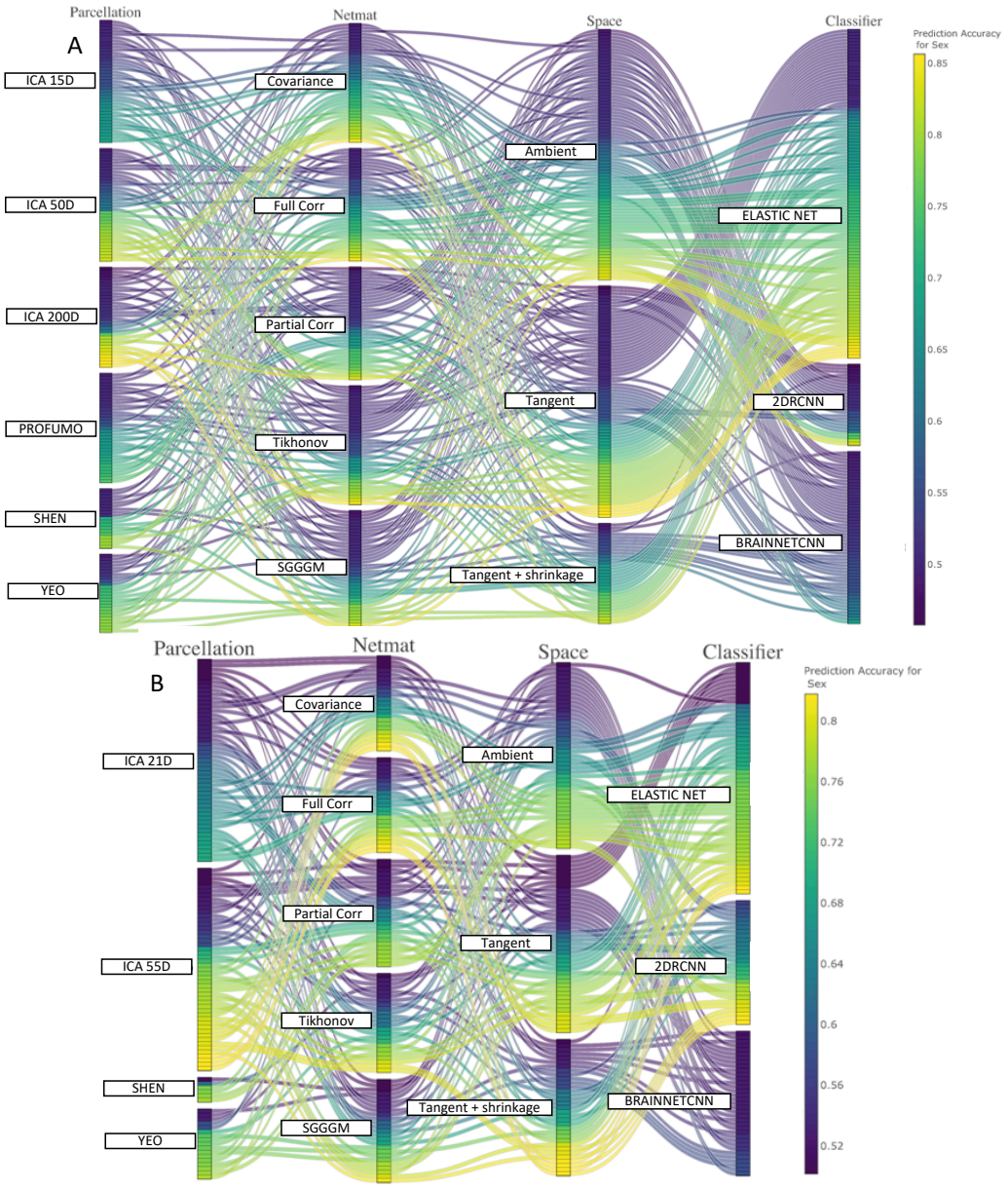


Figure 3.10: This parallel coordinates plot provides a visualization of all possible combinations of options in the pipeline to pipeline to predict sex from FC **after deconfounding**. **[A] (HCP Data), [B] (UKB Data):** The lines are color-coded according to their prediction performance.

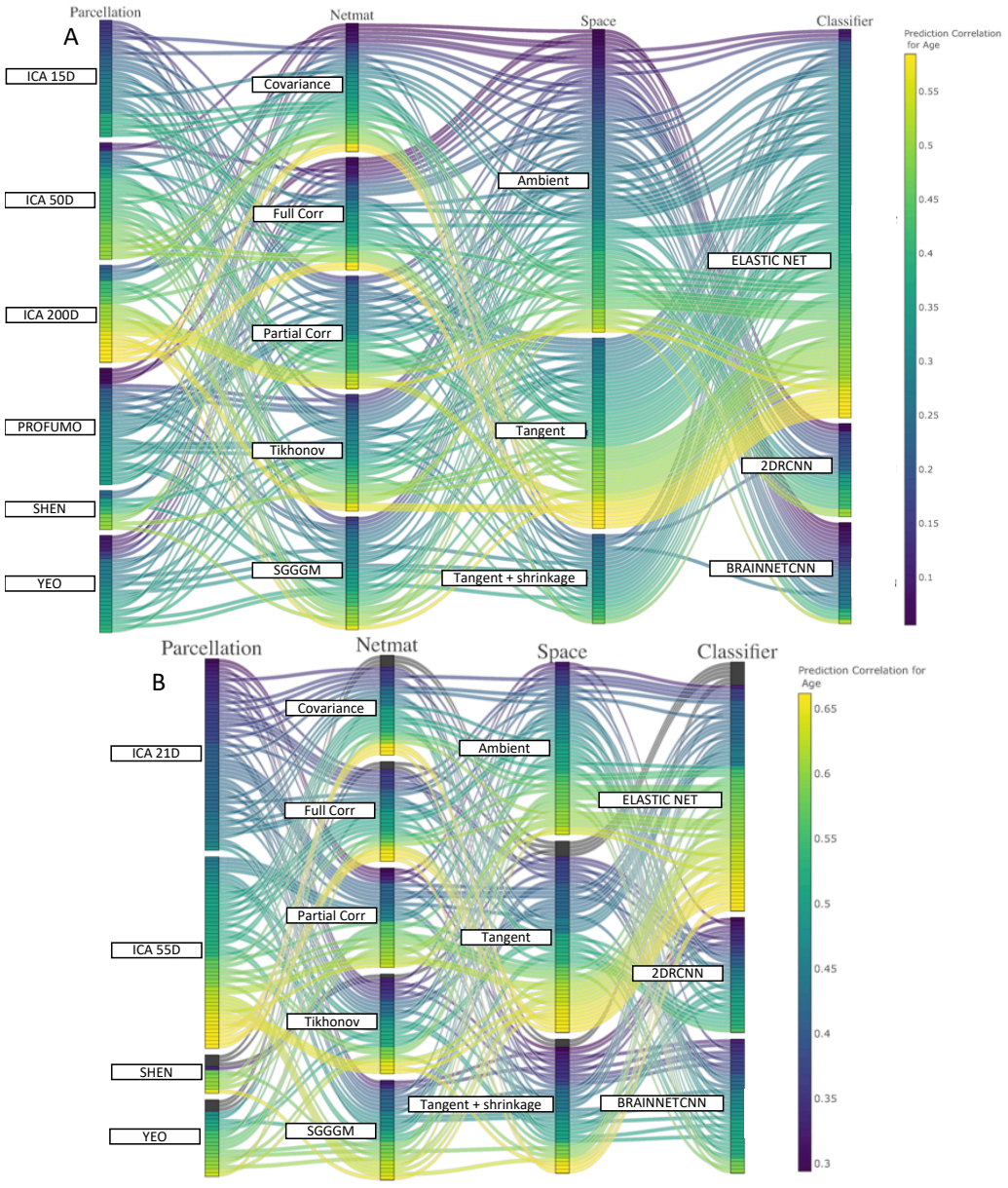


Figure 3.11: This parallel coordinates plot provides a visualization of all possible combinations of options in the pipeline to pipeline to predict age from FC **after deconfounding**. **[A] (HCP Data), [B] (UKB Data):** The lines are color-coded according to their prediction performance.

3.4.7 Deconfounding revisited

All the results illustrated in this work were using X1Y1 or X1Y0 deconfounding strategy (apart from Figure 3.5 where all six deconfounding strategies were employed). We have applied the X1Y1 deconfounding for age, fluid intelligence and neuroticism score prediction and X1Y0 for sex prediction (X1Y1 and X1Y0 are detailed in Section 3.3.3 and the reasoning for this is discussed in Section 3.5). Confounds (particularly head size and motion) are clearly a serious problem if not dealt with for fluid intelligence score prediction. Head size is positively correlated with fluid intelligence and head motion is negatively correlated. For neuroticism score prediction, the confounds which have the most effect on prediction power are age, table position, imaging-centre and date-drift related confounds. This could be because of non-linear interactions between imaging-centre/date-drift related confounds, age, table-position and neuroticism. For age, the crucial confound is head-motion (positively correlated with age), although other than head-motion, we did not find prediction was significantly affected by any confounds. For sex prediction, the crucial confounds are age, height, weight, head motion and head size. For fluid intelligence, neuroticism and age prediction, the deconfounded results are calculated after regressing out all the potential confounding variables. However, in case of sex prediction, for the deconfounded results, we regressed out all the confounds from data except head-size and body density related confounds. The rationale is that sex is a major casual factor for head-size, and head-size causes some confounding effects in the imaging. Regressing out head-size reduces the overall sex prediction accuracy but allows one to focus on the non-headsize sex effects in the imaging. However, one important goal of this work is to optimize FC estimation methods (i.e., allow maximum predictors/classifiers to converge to make performance of different models compa-

rable). Therefore, we didn't include head-size and body density related variables as confounds when predicting sex. Most of the results without removing the effect of confounds (i.e., higher prediction accuracy/correlation as compared to deconfounded results in the main text) are shown in the supplementary material.

3.4.8 Comparison with other commonly used methodologies

Connectome-based predictive modelling: There is another pipeline in the literature known as Connectome-based predictive modelling (CPM) (Shen et al., 2017) that is mainly comprised of three steps: (i) parcellating the brain using the SHEN functional atlas, (ii) estimating connectivity using full correlation, (iii) correlating each edge in the connectivity matrix with behavioural measures, selecting only the most correlated N edges, averaging across these N edges to give a single predictor value for each subject, and then fitting a linear model (e.g., robust regression, polynomial curve fitting) to predict non-imaging variables. We have compared the performance of Elastic Net with CPM in Figure S18.

Random forests: Random forests algorithm are an ensemble learning method and work as a large collection of decorrelated random trees (Liaw, Wiener, et al., 2002). We have compared the prediction performance of Elastic Net and Random Forest for sex prediction, and results are shown in Figure S19.

Dictionary Learning: Massive online dictionary learning (MODL) was used to extract the time-series data (Mensch, Mairal, Thirion, & Varoquaux, 2017) on ABIDE data-set only. These time-series were originally used in (Mensch et al., 2017). We have compared the prediction performance of ICA and Dictionary Learning for a subset of tests, and results are shown in Figure S20.

Higher Dimensional YEO Parcellation: The YEO parcellation is available at multiple resolution levels. For most analysis in this work, we have used d = 100

parcels, but we have compared a subset of methodologies for YEO parcellation with d = 200 and d = 400 parcels. The results are shown in Figure 3.12 (sub-figure D).

3.5 Discussion

A number of research groups are increasingly focussing on resting fMRI connectivity, and developing a plethora of methods (Abraham et al., 2017; Finn et al., 2015). In this work, we aimed to identify an optimal pipeline to estimate and analyze FC, entailing few major steps, and in the following section, we will discuss the overall recommendations (and rationale behind these).

We started this work with the hypothesis that different tasks (non-imaging variables to predict) and different datasets may tell different stories, and for that reason, we carefully chose a few different tasks (age, sex, fluid intelligence, neuroticism and smoking status) and four different datasets. We wanted to evaluate trends across different tasks and across different datasets independently. Once we had considered these independent results, we aimed to identify "one best/recommended" pipeline. The remainder of this section evolves systematically, testing on HCP and UKB datasets, and lastly deducing recommendations which are reliable in all scenarios (consolidated results across all tasks and datasets). Lastly, we illustrate that the recommendations which are based on individual scenarios, and recommendations based on consolidated results are very similar.

The best parcellation strategy: The data-driven parcellation techniques we explored for identification of brain parcels included sICA, PROFUMO, SHEN and YEO. PROFUMO yielded less predictive power than sICA identified parcels, as shown in Figure 3.2. This is consistent with results in (Bijsterbosch et al., 2018), where PROFUMO network matrices tend to be quite empty, in part because of the

fairly strong temporal independence between probabilistic functional modes. Using sICA, we investigated if the number of parcels (ICA dimensionality) affected the predictive power (illustrated in Figure 3.5 and briefly discussed). Our results showed that increasing dimensionality leads to an increase in predictive power. However, higher predictive power using sICA could actually be attributable to cross-subject variations in the spatial configuration of functional brain regions that are represented as changes in FCC (Bijsterbosch et al., 2018). Therefore, although PROFUMO yielded lower predictive power, it still may provide a *purer* estimate, as it should not contain as much potential confounding anatomical information. sICA also outperformed SHEN and YEO parcellations in terms of predictive power as illustrated in Figure 3.2 (sub-figures A and C). SHEN and YEO parcellations can be affected if the reality is that of overlapping functional organisation, due to incorrect assumption of hard parcels leading to mixing of extracted time-series (Bijsterbosch, Beckmann, Woolrich, Smith, & Harrison, 2019); however, further detailed work is require to understand and address this issue. To conclude, in terms of quantitative assessment, sICA outperformed other parcellations.

The best FC estimation method: When FC estimates are left in ambient space, without projection into tangent space, our results revealed estimation using regularised partial correlation estimation yields higher predictive power than does using full correlation/covariance and unregularised partial correlation as illustrated by the mean of the violin plots in sub-figures B and D in Figure 3.2. For example, for age prediction task in UKB, average correlation using covariance, full correlation, unregularised partial correlation, Tikhonov and SGGGM is 0.52, 0.52, 0.45, 0.56, and 0.55 respectively. Using partial correlation results in the removal/weakening of a considerable number of marginal and negative connections. The marginal connections were most likely due to global effects or indirect con-

nections. Hence, partial correlation also enhanced the sparsity in the FC matrix. Partial correlation involves the inversion of covariance matrix, and to stably invert the covariance matrix, regularization should be normally applied to the covariance matrices (Thirion et al., 2014). We employed Tikhonov regularization, as it is an efficient and stable method for estimation of a regularized partial correlation matrix. We found this method resulted in increased predictive power as compared to simple partial correlation (particularly for UKB, as duration of scanning session is shorter than with HCP) as illustrated in Figure 3.2 (sub-figure D). However, we found a more sophisticated, recent regularization method, SGGGM, also provided reasonable predictive power (on par with Tikhonov). Through incorporation of group information when deriving intra-subject connectivity estimates, this method enables more accurately data-adaptive regularisation. This method produced the robust predictions, although optimised tuning of the parameters of SGGGM (such as sparsity level and group consistency) is essential, as the performance of SGGGM is very sensitive to these parameters. Therefore, to find optimal values for these parameters, a comprehensive grid search strategy must be applied (and, for validity, this needs to be done within cross-validation). To conclude, both in terms of qualitative and quantitative assessment, Tikhonov and SGGGM outperformed other connectivity estimation techniques.

Is tangent space parameterization important: To test how tangent space parameterization affects predictive power, we projected different FC estimates into tangent space. As FC matrices being projected in to tangent space should always be positive definitive, estimates containing values outside of the positive definitive manifold ideally require adjustment. FC estimates through covariance or SGGGM are always positive definitive, therefore did not require additional processing. However, partial correlation and Tikhonov connectivity matrices are not always positive definitive. To adjust for these values, we calculated the eigen-

decomposition of the matrix. Any eigenvalues close to zero were adjusted to a fixed, small non-zero value. Our results demonstrated that the performance of most FC estimation techniques were reasonably improved by projecting them into tangent space, as illustrated in Figure 3.3 (subfigures A,C,E and G).These convincing results suggest the importance for future studies to consider the application of Riemannian geometry principles (e.g., tangent space parametrization) to FC estimates. When applied to different FC estimation methods, we found tangent space parametrization increased the predictive ability of almost all estimation techniques. Our results however failed to delineate a single connectivity estimate that was best improved by this parameterization. Projection of these estimates into tangent space required a reference mean, and so we also compared the performance of different reference means. Specifically, we compared the performance of Euclidean, harmonic, log determinant, log Euclidean and Riemannian means as reference means for tangent space parameterization. In terms of prediction accuracy/correlation, our results demonstrate the Euclidean mean tends to have the most stability (less standard deviation), as shown in Figure S5 and S6 (although the differences were not large between any of the reference mean methods). Therefore, we implemented tangent space parameterization using the Euclidean mean as a reference mean for proceeding analyses. The main advantage of tangent space parameterization is that it moved the FC estimates into a space where the following computations (e.g., prediction) could be done in a Euclidean framework with a corresponding boost in efficiency. To conclude, both in terms of qualitative and quantitative assessment, tangent space parameterization improves the predictive power.

Regularization in tangent space: The FC estimates often have high estimation variability. After projecting FC estimates to tangent space, we evaluated three possible regularization strategies, (i) No shrinkage, (ii) applying the non-isotropic

Population Shrinkage Covariance Estimator (PoSCE) shrinkage, (iii) applying isotropic Ledoit-Wolf shrinkage. We found that applying the PoSCE shrinkage technique as compared to no shrinkage slightly improved the performance in terms of prediction accuracy for UKB data as illustrated from top pipelines in Figure 3.8 and Figure S13, but this improvement is not significant, as illustrated in Figure 3.3 (sub-figure H). Moreover, It did not yield meaningful improvements for HCP data, but deteriorated the performance as shown in Figure 3.3 (sub-figure B,D and F). We also tested the effect of tangent space projection and shrinkage on cut-down HCP data (from 60 minutes to 15 and 5 minutes), and found out that it is the tangent space projection which significantly improved the performance as illustrated in Figure 3.3 (sub-figures C and E), but there is no significant performance change using tangent space shrinkage as illustrated in Figure 3.3 (sub-figures D and F). We also tested the effect on PoSCE shrinkage by varying the number of parcels, and did not find any significant improvement as illustrated in Figure S10. However, our analyses also revealed that optimization of the shrinkage parameter in PoSCE was not straightforward. For PoSCE, we chose the shrinkage parameter which resulted in the strongest prediction for the non-imaging variables (within an appropriate cross-validation framework). We then compared non-isotropic PoSCE shrinkage to Ledoit-Wolf isotropic shrinkage, and found that PoSCE did not yield a meaningful advantage compared with Ledoit-Wolf shrinkage (as illustrated in Figure S8). Lastly, we compared the isotropic Ledoit-Wolf shrinkage method with the no shrinkage approach, and found that applying simple Ledoit-Wolf shrinkage improved performance for HCP data, but did not significantly improve performance for UKB data (as illustrated in Figure S9). These results imply that simple isotropic shrinkage is more suitable for less noisy data (e.g., HCP data), and aggressive PoSCE can slightly improve performance for data with comparatively high inter- and intra-subject variance (e.g., UKB data, as duration of scanning

session is shorter than with HCP). To conclude, we are recommending shrinkage in tangent space as an optional step which could be applied if data is noisy.

Deconfounding: Many imaging studies fail to control for confounds in their analyses, which is problematic in providing replicable and interpretable results that are not attributable to extraneous variables. It is correct that in some cases the "confounds" may be useful to keep in the data - that is one reason why we show most results with and without confounds. However, some confounds (e.g., head motion) are generally considered to be an important factor to remove. On the other hand, age is a good example of being a factor that the experimenter may or may not want to remove - although in the context of alzheimer's disease prediction, presumably one would include it as an additional predictor, as opposed to simply ignoring it (or deconfounding it), in order to both obtain optimal prediction and most interpretable prediction parameters. For an example, in UKB fluid intelligence prediction, one could carry out the intermediate option of deconfounding for all confounds except for head motion or age, which would give intermediate prediction correlations (in between the two extremes already presented).

In our analyses, we found the optimal deconfounding strategy was X1Y1 (and X1Y0 in some cases). The reason we recommend X1Y1 over X2Y1 is because X2Y1 combination consists of performing fold-wise confound regression on the X (the FC estimates), and fold-wise deconfounding the target variable Y. However, the deconfounding strategy X1Y1 (deconfounding X once, outside of cross-validation) is more aggressive than X2Y1 (X2Y1 is less aggressive as it used cross-validation to limit the parts of the confounds actually used). For some non-imaging target variables prediction (e.g., sex) there is no chance that confound variables would feed into them (e.g., sex could not be corrupted because of head motion or age). In such cases, it is recommended to deconfound only the imaging variables, X (i.e., FC estimates) and not to deconfound the target variable, Y (i.e.,

sex); therefore, X1Y0 (or X2Y0) deconfounding strategy should be used for such target variables. The pretext of supporting X1Y0 over X2YO is the same (X1Y0 is more aggressive as compared to X2Y0). Our analysis have revealed that X0Y1 (deconfounding only Y), also theoretically yields plausible results; however, for some cases (e.g., sex prediction), it is not applicable. Based on our analysis, we have the following recommendations, (i) deconfound both X and Y when there is a chance that confounds can corrupt both the imaging data and non-imaging data, such as fluid intelligence score or any given mental health variable, and (ii) deconfound only X when there is no chance of confounds feeding into target variable, such as sex prediction (otherwise, this could cause Berkson's paradox (Pearl, 2009) [7].

The best classifier/predictor: Figure 3.6 illustrates the performance of different classifiers/predictors in combination with different parcellation and estimation techniques. Elastic Net is a simple regularized regression technique, whereas the other models tested are based on convolutional neural networks. The prediction of age, fluid intelligence and neuroticism is a continuous prediction task, whereas sex prediction is a binary classification problem. 2D RCNN consistently demonstrated the best performance for sex prediction (e.g., UKB data). 2D RCNN and Elastic Net were comparable for the prediction of fluid intelligence score (e.g., HCP data). For the prediction of neuroticism, our results demonstrated that ElasticNet provided the highest predictive power. Across the prediction of all variables, BrainNetCNN underperformed in comparison to other classifiers/predictors, although it is possible that further optimization of param-

[7]The description of Berkson's paradox is: "If there exist a causal structure $X \rightarrow C \leftarrow Y$, where X and Y are not directly connected. If we try to account for ("condition on") C in a partial correlation analysis (i.e., testing for conditional independence), by regressing C out of X and Y, we induce a negative correlation between X and Y". In the context of the discussion above, if there is no chance of confounds (C), being fed into a target variable (Y), but still by conditioning on C, we might create a spurious association between FC (X) and Y.

eters could lead to higher accuracy. GraphCNN consistently underperformed in comparison with the other models (results presented in Figure S11). It should be considered that we tested these deep learning architectures (2D RCNN, BrainNetCNN, and GraphCNN) for hundreds of different FC estimates, which impose practical limitations on our ability to tune hyper-parameters for each possible configuration. This indirectly suggests that only the most stable architecture will show consistent results across different connectivity estimates. Moreover, this suggests the most robust, off-the-shelf types of neural networks must minimize sensitivity to hyper-parameter tuning, to maximize generalizability and minimize overfitting. Overall, we have demonstrated that a carefully designed 2D Recurrent Convolutional Network or BrainNetCNN (top performing for fluid intelligence prediction in HCP data) can compete with (and in some cases outperform) a simpler regression model. These results contrast with (He et al., 2018), where kernel based regression (non-parametric classical machine learning algorithm (Murphy, 2012)) demonstrated better performance than deep neural networks.

Overall Results (HCP and UKB): In the previous sections, we recommended optimal methodologies from each step of the pipeline. Our recommendation from each step is, (1) sICA for parcellation, (2) Tikhonov or SGGGM for FC estimation, (3) projecting the connectivity estimates to tangent space, (4) optional shrinkage in tangent space, (5) deconfounding the data (X1Y1 or X1Y0) and (6) Elastic Net or 2D RCNN as a predictor/classifier. It is also important to investigate these methods simultaneously (to identify potential interactions between different pipeline decisions). Figures 3.7, 3.8, S12 and S13 illustrate the top performing pipelines out of hundreds of combinations. The top performing methods which repeatedly stand out are in-fact the joint combination of these methodologies; i.e., "ICA SGGGM Tangent Elastic Net" or "ICA Tikhonov Tangent_shrinkage 2DRCNN", etc. It is to be noted that our recommendations are not solely based

on quantitative results but also on the qualitative assessment of each step as explained through-out the Discussions (e.g., one has more parcels than time points, Tikhonov or SGGGM may be able to give stable results). The error bars drawn (using Fisher Transformation/Wilson test) in Figures 3.7, 3.8, S12 and S13 show that for sex and age prediction, these recommended pipelines are significantly different from the other non-recommended pipelines. To make sure that all non-imaging variables prediction are reproducible and significant, we ran repeated cross-validation on the top performing pipelines. With a 99% confidence interval, the fluid intelligence scores prediction correlation is $\sim 0.35 \pm 0.006$, for age it is $\sim 0.58 \pm 0.005$ and for sex it is $\sim 0.92 \pm 0.001$. This highlights that the identified optimum pipelines are significantly different and reproducible.

Combined Results across all tasks and datasets: Lastly, we condensed the results across all four datasets (UKB, HCP, ABIDE and ACPI) and all tasks (non-imaging variables prediction), and have shown the consolidated results in Figure 3.12. The relative prediction scores shown in Figure 3.12 are defined as the prediction score of each pipeline relative to the average across all pipelines. The recommendations which we have found on individual datasets are similar to our conclusions deduced from consolidated results: (1) the mean of relative prediction scores using sICA is 0.07, compared to YEO (0.04), SHEN (0.03) and PROFUMO (-0.03), (2) the mean of relative prediction scores using Tikhonov and SGGGM is 0.02, compared to covariance and Full correlation (0.0), and partial correlation (-0.02), (3) the mean of relative prediction scores using tangent space is 0.05, compared to ambient space (-0.02) and tangent_sharinkage space (-0.04), (4) the mean of relative prediction scores after deconfounding is -0.05, which is negative, but we believe the results without applying deconfounding are inflated due to the shared confounds, and lastly (5) the mean of prediction score using Elastic Net and 2D RCNN is 0.11 and 0.04 respectively, compared to BrainNetCNN

(0.0), and GraphCNN (-0.16). To conclude, there is an approximately "optimal" pipeline (or set of recommendations) which performed reasonably well across all tasks whether analysed individually or in a consolidated fashion.

Dimensionality of Parcellation Revisited: Finally, it is important to consider the impact of the number of parcels across various parcellation methods, although it is hard to fully disentangle different choices of dimensionality from different parcellation methods. We have nevertheless attempted to draw recommendations by comparing YEO parcellation with varying the number of parcels (d = 100, 200, 400) with sICA (d = 15, 50, 200) as shown in Figure 3.12 (sub-figure D). We found that higher dimensional sICA (d=200) outperforms lower dimensional sICA (d=50), which is also illustrated in Figure 3.5, and briefly discussed above the "best parcellation strategy" section. Moreover, lower dimensional sICA (d=50) outperforms both lower and higher dimensional YEO parcellation (d = 100 or 400). Lastly, we found that prediction performance is not always improved by increasing the dimensionality of parcellation method, as better results were obtained by lower dimensional YEO (d=100) compared to higher dimensional YEO (d = 200 or 400). This was also the main rationale behind showing lower dimensional YEO (d=100) results throughout the main text. It is also important to comment on the dimensionality of sICA for UKB data, as in theory higher ICA dimensionality might have been useful for UKB data, but we can not explore high dimensionality with ICA as the number of non-artefact (neuronal signal) components barely grows at all (more than the existing 55). As discussed above, we believe this is because of the limitations of volumetric analysis, possibly combined with the 6-minute limited duration of the acquisition in UKB (as we do not face such limitations in case of HCP).

Comparison with Literature: Somewhat similar evaluations of various methods (in this problem domain) have previously been carried out in (Ng et al., 2015;

Dadi et al., 2019). In (Ng et al., 2015), authors proposed that matrix whitening transform and parallel transport could be utilized to project covariance matrices into a common tangent space and evaluated this method on twenty four healthy subjects. These former results build a foundation for estimation of FC using manifold operations, as higher prediction accuracy was achieved using tangent space parameterization. (Dadi et al., 2019) evaluated different modelling choices and recommended a three stage pipeline: (i) ICA or MODL for brain parcellation, (ii) tangent space embedding for FC estimation and (iii) use of non-sparse models like SVM for non-imaging variables prediction. For brain parcellation, (Dadi et al., 2019) applied a number of parcellation techniques that differed from our pipeline, however also found that sICA yielded the best performance (alongside with MODL). In terms of step-1 (brain parcellation), we compared sICA with MODL, and found that the performance of these is similar (although sICA outperforms for age prediction, as shown in Figure S20). We have not included MODL in main analysis as it is already thoroughly evaluated in (Dadi et al., 2019) (and it remains a valuable tool for brain parcellation).

Lastly, we have applied deep learning architecture such as CNNs to analyze FC whereas (Dadi et al., 2019) utilized traditional machine learning algorithms. Similarly to the last step of our pipeline, (He et al., 2018) compared the performance of three CNNs to kernel regression. The three CNNs tested included a generic fully-connected feedforward network, BrainNetCNN and Graph CNN. This study reported that CNNs did not outperform kernel regression across a wide range of behavioural and demographic measures. However, our results found that our proposed 2D RCNN (architecture illustrated in Figure S1), which has not been previously evaluated against these methods, outperforms the regularized regression and existing CNNs for some tasks (e.g., Figure S15).

3.6 Conclusion

Our results have demonstrated that within data driven parcellation, ICA provides the most predictive power for non-imaging variables. For estimating FC, we have shown that regularized partial correlation like SGGGM or Tikhonov outperforms unregularized partial correlation or full correlation. To correctly apply mathematical formulations on FC estimates, it is recommended to take into account that these measures should ideally be worked with within the Riemannian manifold. To do so, Riemannian space is approximated with an associated tangent space and then FC estimates are mapped to the tangent space. We have reported in this work that this tangent space parameterization results in a significant increase in predictive power of FC estimates. Our results have also demonstrated that additional shrinkage is not necessarily required on well regularized connectivity estimates in tangent space; however, could be applied for noiser data. Lastly, we evaluated various classifiers for prediction of non-imaging variables from connectivity estimates and concluded that a carefully designed deep learning based architecture (2D RCNN) can be a valuable tool for analyzing FC. However, Elastic Net probably performs better at present overall.

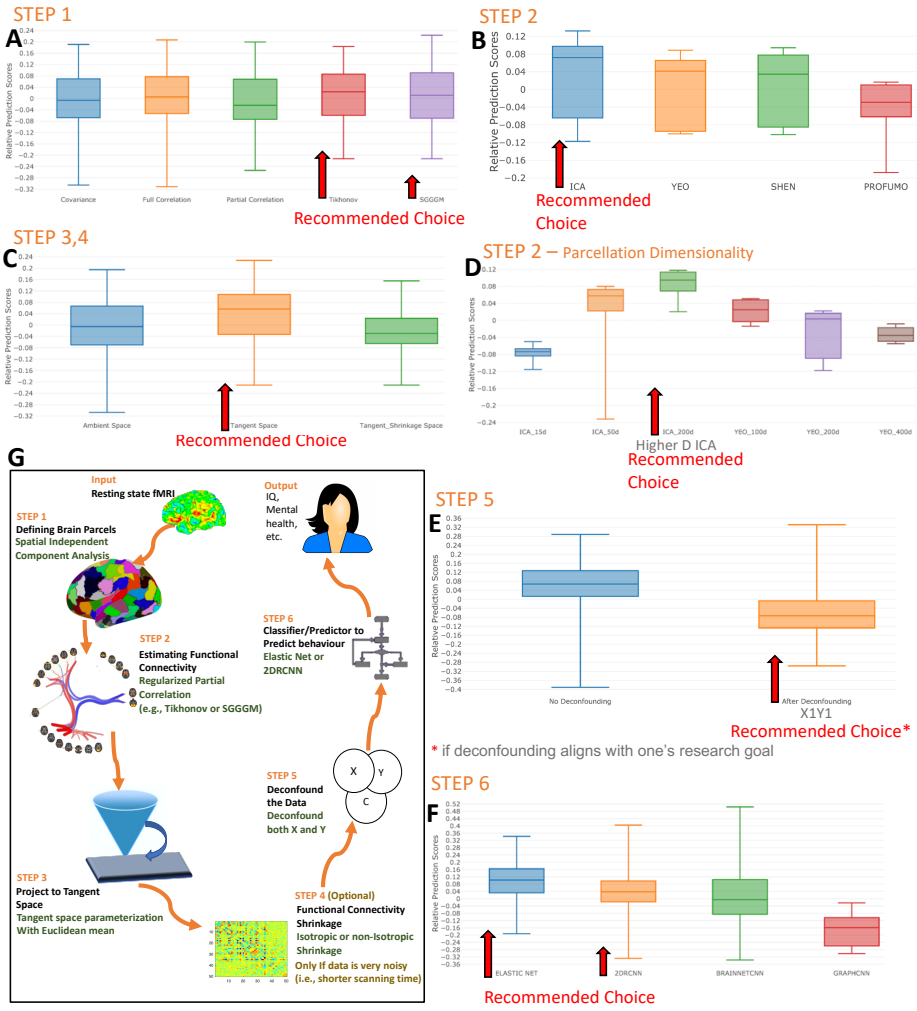


Figure 3.12: The relative impact of different steps of pipeline on the non-imaging variables prediction performance for participants in UKB, HCP, ABIDE, and ACPI datasets. Each sub-figure highlights the recommended options which are selected based on quantitative and qualitative analysis. The recommended optimal methodologies from each step of the pipeline are, (1) sICA for parcellation, (2) Tikhonov or SGGGM for FC estimation, (3) projecting the connectivity estimates to tangent space, (4) optional shrinkage in tangent space, (5) deconfounding the data (both the target and predictor variables) and (6) Elastic Net or 2D RCNN as a predictor/classifier.

CHAPTER 4

OPTIMISING TIME-VARYING NETWORK MODELLING METHODS FOR FMRI

4.1 Overview

Typically, methods that measure TVFC, such as sliding windows approaches, do not separately model when changes occur in the mean activity levels from when changes occur in the FC, therefore conflating these two distinct types of modulation. We show that this can bias the estimation of time-varying FC to appear more stable over time than it actually is. Here, we propose an alternative approach that models changes in the mean brain activity and in the FC as being able to occur at different times to each other. We refer to this method as the Multi-dynamic Adversarial Generator Encoder (MAGE) model, which includes a model of the network dynamics that captures long-range time dependencies, and is estimated on fMRI data using principles of Generative Adversarial Networks. We evaluated the approach across several simulation studies and resting fMRI data from the

Human Connectome Project (1003 subjects), as well as from UK Biobank (13301 subjects). Importantly, we find that separating fluctuations in the mean activity levels from those in the FC reveals much stronger changes in FC over time, and is a better predictor of individual behavioural variability.

4.2 Introduction

Many studies estimate FC by calculating the average correlation across an entire fMRI scanning session data (Chapter 3) (Rogers et al., 2007; O'Reilly, Woolrich, Behrens, Smith, & Johansen-Berg, 2012; S. M. Smith, Beckmann, et al., 2013; S. M. Smith, 2012; Pervaiz, Vidaurre, Woolrich, & Smith, 2020), and these time-averaged measures of FC have been widely associated with cognitive phenotypes and neuropsychiatric disorders as discussed in Chapter 1. There is now increasing interest in studying the time-varying nature of FC in fMRI. Significant within-session fluctuations in FC during rest or task fMRI have been demonstrated in a number of studies (D. J. Lurie et al., 2020; Vidaurre et al., 2017; Hutchison et al., 2013; Allen et al., 2014; Parr, Rees, & Friston, 2018) and are also well-established in electrophysiological data (Baker et al., 2014; Vidaurre, Hunt, et al., 2018; De Pasquale et al., 2010, 2012). Further, a number of research studies have shown the utility of TVFC in identifying many psychiatric and neurological conditions (as explained in Chapter 1; and the dynamics of TVFC have been shown to coactivate with spontaneous replay of recently acquired information (C. Higgins et al., 2021). Moreover, inter-subject differences in TVFC have also been associated with a number of behavioural and cognitive traits (Liégeois et al., 2019; Vidaurre et al., 2017), with some aspects of intelligence only being explained by TVFC (Vidaurre et al., 2021).

Despite these successes in the estimation and use of TVFC, it is striking how

homogeneous FC actually appears to be in fMRI data over time. Indeed, there are a number of studies reporting that there is even insufficient evidence of within-session fluctuations in FC in resting fMRI data (Laumann et al., 2017b; Hindriks et al., 2016; Leonardi & Van De Ville, 2015). While it is possible that close to homogeneous FC in fMRI is a genuine phenomenon, it is also possible that it is instead caused by limitations in the methods used to compute TVFC.

Most previous work is based on the sliding window correlation (SWC) approach (Sakoğlu et al., 2010). There are many variants of the SWC method (Allen et al., 2014; Lindquist, Xu, Nebel, & Caffo, 2014; Liégeois et al., 2016), which are typically limited by using sliding windows of a specified, fixed width. This means that they fail to adapt to the intrinsic time-point by time-point variability of the FC, resulting in changes in FC getting smeared together, thereby making FC look more homogeneous over time. Some of these limitations can be overcome by methods that adapt automatically to time periods of distinct network activity, such as those based on Hidden Markov modelling (HMM) (Vidaurre, Abeysuriya, et al., 2018; Quinn et al., 2018). Even then, a surprising amount of apparent homogeneity in the FC over time persists.

SWC and the HMM have in common that the mean activity and FC estimations are coupled; that is, changes in the mean activity level are by definition accompanied by changes in the FC, when both are modelled. For example, the most common version of the HMM for fMRI characterises each state as a Gaussian distribution with both mean and covariance. Since there is no biological reason for this coupling assumption, in this work we consider the consequences of relaxing that constraint, in fact showing that this change can offer a potential explanation for the homogeneity of TVFC seen in approaches that tie the dynamics of the mean and the FC together.

Our proposed method for uncoupling the dynamics of the mean activity and FC is called the Multi-dynamic Adversarial Generator-Encoder (MAGE). This proposes a multi-dynamic modelling approach that allows the mean activity and FC to fluctuate in time independently from each other. At the same time, MAGE captures long range temporal dynamics of FC and mean activity using recurrent neural networks (in contrast to the Markovian assumption of the HMM) (Gers, Schmidhuber, & Cummins, 2000; Graves & Schmidhuber, 2005). Finally, our proposed model allows multiple brain states to be simultaneously active (in contrast to the categorical "mutual exclusivity" assumption of the standard HMM). To estimate this model on fMRI data, we developed a novel adversarial generator-encoder network framework inspired by the adversarial regularization used in generative adversarial networks (GANs) (Goodfellow et al., 2014; Lowd & Meek, 2005; Makhzani, Shlens, Jaitly, Goodfellow, & Frey, 2015).

In this work, we demonstrate the validity of MAGE on several simulation studies and a cognitive task dataset. We also apply the method to resting-state fMRI studies and deduce several outcomes regarding TVFC temporal dynamics, revealing more distinct patterns of FC over time than are obtained with SWC and the HMM, providing an explanation as to why resting fMRI FC has appeared so temporally stable in previous work (Laumann et al., 2017b; Hindriks et al., 2016; Leonardi & Van De Ville, 2015). Finally, we substantiate our findings by demonstrating improved prediction of behavioural and cognitive traits using MAGE-derived TVFC estimates.

4.3 Methods

4.3.1 Multi-dynamic Model

Our proposed approach assumes that fMRI data is generated by a multivariate Normal distribution process with a time-varying mean, m_t, and covariance, C_t:

$$Y_t = \mathcal{N}(m_t, C_t) \tag{4.1}$$

where Y_t is the value of standardised fMRI data, Y (N x T) at time point t, and N is the number of channels (e.g., brain regions).

In order to separately model the time-varying variances and the correlations, we partition the covariance into a NxN diagonal matrix of standard deviations for each brain region, G_t, and an NxN correlation matrix, F_t:

$$C_t = G_t F_t G_t \tag{4.2}$$

where the correlation matrix, F_t, captures instantaneous correlation between brain regions; which we refer to as instantaneous FC.

Clearly, it will not be possible to estimate the instantaneous correlation and means (and variances) at every timepoint without further constraints. As such, we model regularised versions of them by capturing their repeating spatio-temporal structure using an adaptive temporal prior.

Specifically, we assume that the means and correlations (and variances) are generated from their own linear mixture of a limited, underlying set of dynamic modes or states:

$$m_t = \sum_p \alpha_{tp} \mathcal{S}_p \tag{4.3}$$

$$F_t = \sum_q \beta_{tq} \mathcal{D}_q \tag{4.4}$$

$$G_t = \sum_r \gamma_{tr} \text{diag}(\mathcal{E}_r) \tag{4.5}$$

where α_{tp} is a scalar corresponding to the time course of the p^{th} "mean activity" state (p=1...P), and $\mathcal{S}_p$ is its Nx1 spatial map of mean activity; and β_{tq} is a scalar corresponding to the time course of the q^{th} "correlation" state (q=1...Q), and $\mathcal{D}_q$ is its NxN correlation matrix; and γ_{tr} is a scalar corresponding to the time course of the r^{th} "variance" state (r=1...R), and $\mathcal{E}_r$ is its Nx1 spatial map of the standard deviations. The $\mathcal{D}_q$ correlation matrices capture the state-specific correlation between brain regions; which we refer to as the state-specific FC. As we shall describe in detail later, we assume that both the state time courses [α, β and γ] and the state-specific maps [$\mathcal{S}_p$, $\mathcal{D}_q$, $\mathcal{E}_r$] are unknown and need inferring from the data.

Note that in Equations (4.5S8S9 the correlations and means (and variances) are regularised using their own distinct priors, or dynamics. This is in contrast to most TVFC methods, which are *single-dynamic* approaches in that the dynamics of the means and the correlations (i.e., the FC) are tied together. However, there is no imperative to assume that the dynamics of the means and correlations are the same as each other. As such, the proposed method is designed to explore what happens when we do not tie all of the dynamics together. We refer to this approach as *multi-dynamic*, as illustrated in Figure 4.1[A].

Special case - Sliding Window Correlation

Sliding window correlation (SWC) measures of TVFC capture how the between-brain-region correlation fluctuates over time. For simplicity here (but without loss of generality), we assume non-overlapping sliding windows. This corresponds to Equations (4.5S8S9, where α_{tp}, β_{tq} and γ_{tr} are set to be the same, and are fixed a priori to correspond to the P=Q=R different positions the sliding window can take as it slides through time (e.g., in the case of square windows, α and β and γ have a value of one inside the sliding window, and a value of zero outside of it). Note that P=Q=R needs to typically be very large to span all the data over time; and so SWC approaches are normally followed by a clustering approach, such as k-means, to find repeating correlation patterns (Allen et al., 2014; Tagliazucchi, Von Wegner, Morzelewski, Brodbeck, & Laufs, 2012). Typically, information about how the mean fluctuates over time is ignored. Since α_{tp}, β_{tq} and γ_{tr} are set to be the same, i.e., the means and correlations (and variances) are modelled using the same dynamics, then SWC corresponds to what we refer to as a *single-dynamics* approach (see Figure 4.1[A]).

Special case - Hidden Markov Models

Hidden Markov Models (HMM), with multi-variate Normal observation models, capture how both the mean activity levels and the between-brain-region correlations fluctuate over time. This form of the HMM is the same as SWC in that it corresponds to Equations (4.5S8S9 where the time courses α_{tp}, β_{tq} and γ_{tr} are the same however, α_{tp}, β_{tq} and γ_{tr} are now inferred from the data at the same time as $\mathcal{S}_p$, $\mathcal{D}_q$ and $\mathcal{E}_r$. This means that the approach can adjust to how the transient network states are coming and going for different durations at different points in time. Note that in addition, the HMM assumes that $\alpha_{tp} = \beta_{tq} = \gamma_{tr}$ are binary time courses, defining when states p switches on and off, and that no two states

can be active at the same time (i.e., the states are mutually exclusive). As with SWC, since the means and correlations (and variances) are modelled using the same dynamics, then this version of the HMM corresponds to what we refer to as a *single-dynamics* approach (see Figure 4.1[A]).

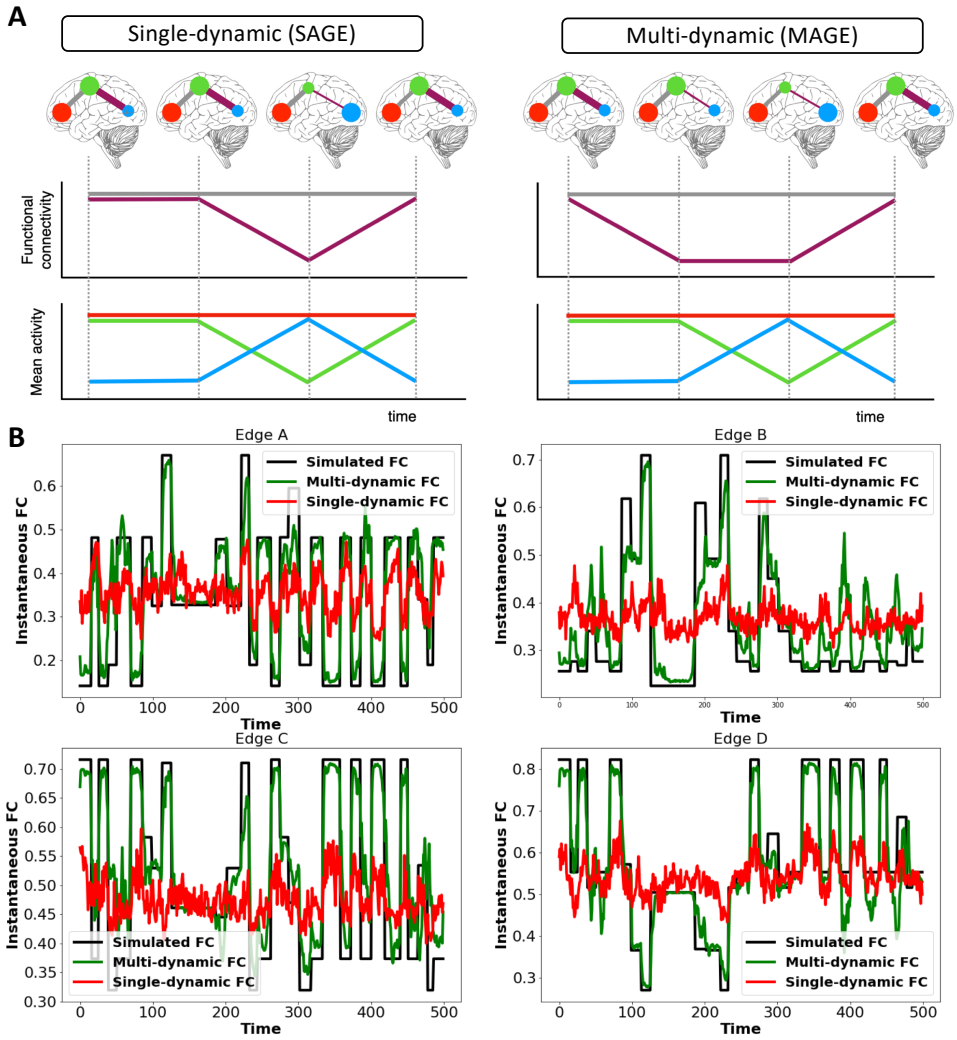


Figure 4.1: Single-dynamic versus multi-dynamic brain networks. **[A]** Top plot shows example brain network dynamics for the two scenarios; line width indicates functional connectivity (FC) strength, and size of node indicates mean activity amplitude. Middle and bottom plots highlight how the single-dynamic case assumes that the means and the FC change together, whereas in the multi-dynamic case they can change at distinct times. **[B]** Simple simulation illustrating the potential benefits of a multi-dynamic approach. The ground truth is that there are different dynamics underlying the time-varying means and FC. The multi-dynamic approach is able to accurately adjust to fluctuating FC (illustrated for four edges of the FC matrix), whereas the single-dynamic approach is not able to accurately capture these fluctuations, producing time-varying, instantaneous FC that erroneously looks much more homogeneous, as is often observed in time-varying FC estimates in real fMRI data. Lastly, the multi-dynamic approach is also able to adjust to fluctuating mean as shown in Figure S25.

4.3.2 What do we mean by functional connectivity?

Typically, correlation-based functional connectivity in fMRI refers to the correlation between two fMRI time series. However, in the model we are proposing, there are multiple entities that could be described as FC, all of which capture some different aspect of correlation between brain regions. First, the model describes the correlation at each time point, i.e., F_t (Equations (4.1, 4.2)). We refer to these as the **instantaneous FCs**. Second, a set of state-specific correlations, i.e., D_q, are mixed together to regularise the instantaneous FC (see Equation S9). We will refer to these as **state-specific FCs**. Finally, we might also consider correlations in the mean activity time courses, which is a single correlation matrix that corresponds to the correlation of m_t over all time, i.e., Corr(m_t). We will refer to this as the **mean activity FC**; this is potentially quite independent of the state-specific FCs.

Our model provides a regularised estimate of the instantaneous FC, and so we can more parsimoniously represent the instantaneous FC using the state-specific FCs. Hence in this work, we focus on using the state-specific FCs as our main TVFC feature of interest, for example when predicting behavioural traits.

4.3.3 Multi-dynamic Adversarial Generator-Encoder (MAGE)

Equations (4.14.24.5S8S9 describe the modelling framework for the multi-dynamic approach we are proposing. In this section, we will describe the more complete generative model, including the generative process assumed for the dynamics of α_{tp}, β_{tq} and γ_{tr}, and how we infer on the model. We refer to the overall approach as the *Multi-dynamic Adversarial Generator-Encoder* (MAGE).

In practice, MAGE assumes that the dynamics of the means and variances are tied together, but crucially the correlation is assumed to fluctuate independently.

This corresponds to Equations (4.5S8S9 with α_{tp} = γ_{tr} (P=R), but with β_{tq} free to be inferred separately [1]. Alongside the state time courses we also infer the state-specific features: $\mathcal{S}_p$, $\mathcal{D}_q$ and $\mathcal{E}_r$. Note that while the dynamics of both the means and the variances are always modelled (and are tied together with the same dynamics), for simplicity we will tend to refer to just the dynamics of the means throughout the work.

In addition, unlike the HMM, MAGE does not assume that state time courses are binary and mutually exclusive. Instead, we assume that they describe time-varying linear mixtures with a sum-to-one constraint (α_{tp}, β_{tq} and γ_{tr} separately sums to one at each time point), which we will refer to as "partial volume" state time-courses. This sum-to-one constraint is an identification condition that is imposed to ensure identifiability (Huang, 2014) on the inferred states and makes the post-model temporal analysis interpretable.

In summary, MAGE describes the data as being generated from a finite set of states. These states are linearly mixed together at each time-point to generate a time-varying description of the data's mean and covariance.

Generative Model

The generative model for the data, Y_t, corresponds to Equation 4.1. Each instance of Y_t is described by a multivariate Normal distribution parameterised by the mean activity at time t, m_t, (Equation S8) and the covariance matrix at time t, C_t (Equation 4.2). We model the state time courses, α_t, γ_t and β_t [2], based on their underlying logits:

[1] The reason MAGE models the variances to be independent of the correlation (i.e., γ_{tr} free to be not equal to β_{tq}) is to allow maximum temporal independence to the correlation. We will expand onto this in Section 4.5 but allowing γ_{tr} = β_{tq} resulted in weaker state-specific FCs dynamics, and hence was not a preferable modelling choice.

[2] We dropped the state sub-indices (p,q,r) in the definition of state time-courses for the sake of simplicity - α_t, γ_t and β_t represents the summations of all states at each time-point

$$\alpha_t = \gamma_t = \zeta(\theta_t^m) \tag{4.6}$$

$$\beta_t = \zeta(\theta_t^c) \tag{4.7}$$

where ζ is the softmax function, ensuring that α and β sum to one (over states) at each time point.

The dynamics in θ_t^m and θ_t^c (and implicitly the dynamics of the state time courses) are captured using a unidirectional sequence-to-sequence long short-term memory (LSTM) model, which uses the history of θ, to predict the value at the next time point. Specifically, it predicts the prior means of the latent, time-varying parameters $[\theta_t^m, \theta_t^c]$.

$$\hat{\mu}_t^{m,\theta} = f(LSTM_{uni}(\theta_{<t}^m, \omega_g^m), \lambda_g^m) \tag{4.8}$$

$$\hat{\mu}_t^{c,\theta} = f(LSTM_{uni}(\theta_{<t}^c, \omega_g^c), \lambda_g^c) \tag{4.9}$$

where $\hat{\mu}_t^{m,\theta}$ and $\hat{\mu}_t^{c,\theta}$ are the prior means of θ_t^m and θ_t^c respectively. $f()$ is the affine transformation and λ_g^m and λ_g^c represents the weights of the affine transformation. ω_g^m and ω_g^c are the generative model LSTM weights. The proposed generative model for the data is illustrated in Figure 4.2 (and the in-depth architecture is illustrated in Figure S22).

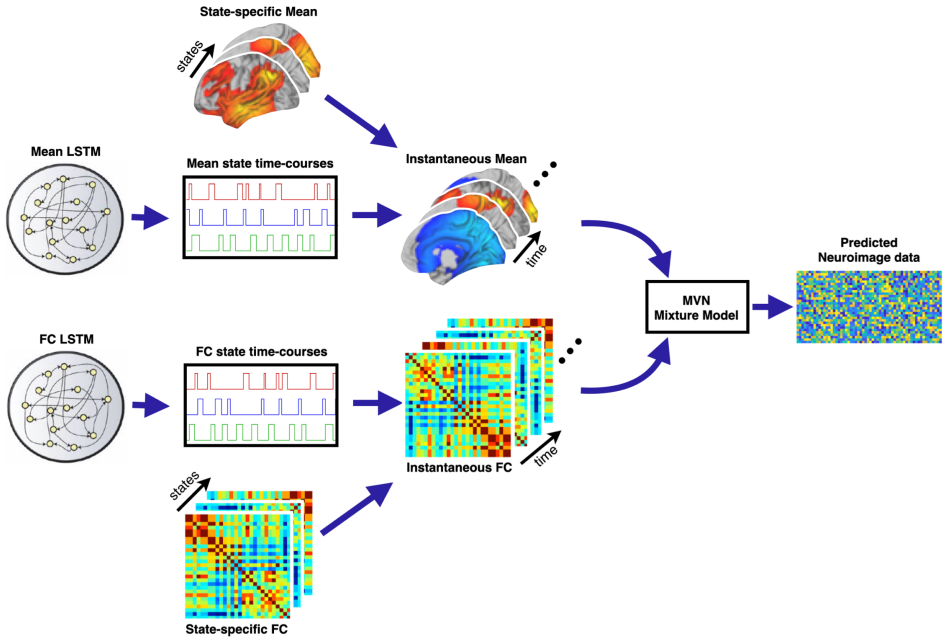


Figure 4.2: Generative model for MAGE. The proposed model generates data by first generating instantaneous means and the instantaneous correlation matrices (i.e., functional connectivity). MAGE also models instantaneous variances, whose dynamics are tied to be the same as the mean (not shown here). The instantaneous mean is modelled using an underlying set of states, for which the state time courses are generated using a long short-term memory (LSTM) model. The instantaneous correlation is also modelled using an underlying set of states, whose state time courses are generated using a completely different LSTM, making the approach multi-dynamic.

Inference

Given fMRI data, we need to infer on the parameters [$\mathcal{S}_p$, $\mathcal{D}_q$, $\mathcal{E}_r$, θ_t^m, θ_t^c, ω_g^m, ω_g^c, λ_g^m, λ_g^c] in the generative model. We infer point estimates with uniform priors for the global parameters in the model, i.e., [$\mathcal{S}_p$, $\mathcal{D}_q$, $\mathcal{E}_r$, ω_g^m, ω_g^c, λ_g^m, λ_g^c]; but for the more challenging inference of the latent, time-varying parameters [θ_t^m, θ_t^c], we want to regularise using priors (which here corresponds to the prior means of [θ_t^m, θ_t^c] as stated in Equations (4.8, 4.9)). This could be attempted using a number of different approaches. For example, Variational Bayes (VB) methods could be used based on the ideas from Variational autoencoders (VAEs) (Kingma

& Welling, 2013), where the generative model acts as a decoder that maps from the latent, time-varying parameters $[\theta_t^m, \theta_t^c]$ to the fMRI data. The loss function in VB inference is based on optimizing the variational free energy (K. Friston, Mattout, Trujillo-Barreto, Ashburner, & Penny, 2007) and consists of two terms - the reconstruction likelihood, and the Kullback-Leibler (KL) divergence that regularises the solution by matching the learned posterior distribution to a prior distribution over the latent variables (i.e., Equations (4.8, 4.9)).

Here, we take an approach based on an adversarial autoencoder (AAE) (Makhzani et al., 2015). Again, the generative model acts as a decoder that maps from the parameters underlying the state time courses, $[\theta_t^m, \theta_t^c]$, to the fMRI data, and hence the reconstruction likelihood term of the AAE is the same as it would be in a VAE. However, the AAE replaces the regularising KL divergence term of the VAE with an adversarial training criterion (Makhzani et al., 2015; Mescheder, Nowozin, & Geiger, 2017). This takes the idea of adversarial loss from GANs to achieve the prior regularization - by forcing the posterior means underlying the state time courses, $[\theta_t^m, \theta_t^c]$, to be close to the specified prior means (i.e., Equations (4.8, 4.9)). To do this, we train a discriminator, separately from the training that corresponds to the generative model parameter inference, to be able to distinguish between the prior and posterior versions of the $[\theta_t^m, \theta_t^c]$. The trained discriminator is then used to enforce closeness between the posterior and prior (alongside the reconstruction likelihood) when performing the model parameter inference.

Encoder Model The encoder model maps from the fMRI data to the posterior estimates of the latent, time-varying parameters $[\theta_t^m, \theta_t^c]$. This can also be referred to as the inference network or inference model and corresponds to the idea of amortized inference in VAEs, whereby the encoder allows an efficient means to âlook-upâ the posterior estimates given the data (Kingma & Welling, 2013).

The authors who proposed AAEs (Makhzani et al., 2015) explored the benefits of being fully Bayesian on the latent variables, by assuming that the posterior is a Gaussian distribution (with the mean and variance predicted by the encoder, in a similar manner to the VAE); but after an extensive hyper-parameter search, they did not find any additional advantages, and only reported results using a deterministic version of the posterior. As such, we learned a single nonlinear mapping that is used to map from the data, Y_t, to the posterior means at any time point:

$$\mu_t^{m,\theta} = f(LSTM_{bi}(Y, \omega_e^m), \lambda_e^m) \tag{4.10}$$

$$\mu_t^{c,\theta} = f(LSTM_{bi}(Y, \omega_e^c), \lambda_e^c) \tag{4.11}$$

where $\mu_t^{m,\theta}$ and $\mu_t^{c,\theta}$ are the posterior means of θ_t^m and θ_t^c respectively. $LSTM_{bi}$ is a sequence-to-sequence, bidirectional LSTM model, and ω_e^m and ω_e^c are the encoder LSTM weights. $f()$ is the affine transformation and λ_e^m and λ_e^c represents the weights of the affine transformation. Note that the approach is still stochastic since we use batching of the data.

Discriminator Model As mentioned above, a discriminator is needed to be able to distinguish between the prior and posterior mean estimates of the state time courses, $[\theta_t^m, \theta_t^c]$. The discriminator is also constructed using a sequence-to-sequence, bidirectional LSTM model and it classifies whether samples are from the posterior or prior estimates of $[\theta_t^m, \theta_t^c]$:

$$D_{\theta_t^m} = \sigma(f(LSTM_{bi}([\zeta(\hat{\mu}_t^{m,\theta}), \zeta(\mu_t^{m,\theta})], \omega_d^m), \lambda_d^m)) \tag{4.12}$$

$$D_{\theta_t^c} = \sigma(f(LSTM_{bi}([\zeta(\hat{\mu}_t^{c,\theta}), \zeta(\mu_t^{c,\theta})], \omega_d^c), \lambda_d^c)) \tag{4.13}$$

where σ is the sigmoid activation to ensure that discriminator output is either 0 or 1, and ω_d^m and ω_d^c are the discriminator LSTM weights. $f()$ is the affine transformation and λ_d^m and λ_d^c represents the weights of the affine transformation. Note that the discriminator takes in as input the full history of both the prior and posterior mean state time courses into its LSTM. Taken together with the fact that we do not factorise the posteriors of $[\theta_t^m, \theta_t^c]$ over time, this means that the trained discriminator will help drive the prior and posterior state time courses to have similar characteristics not just over short, but also over long, time scales.

How it Works

We trained this model as a three-player game played between the generator (i.e., the generative model with parameters $[\mathcal{S}_p, \mathcal{D}_q, \mathcal{E}_r, \theta_t^m, \theta_t^c, \omega_g^m, \omega_g^c, \lambda_g^m, \lambda_g^c]$), encoder (with parameters $[\omega_e^m, \omega_e^c, \lambda_e^m, \lambda_e^c]$), and discriminator (with parameters $[\omega_d^m, \omega_d^c, \lambda_d^m, \lambda_d^c]$). The discriminator model is trained to distinguish prior mean estimates of $[\theta_t^m, \theta_t^c]$ from their posterior mean estimates. The discriminator is then used to enforce closeness between the posterior and prior of those parameters (while also honouring the reconstruction likelihood) when training the encoder and generator. In this game, when the discriminator successfully identifies the prior from posterior, it is rewarded; whereas both encoder and generator models are penalized. In contrast, when the combined efforts of the encoder and generator models fools the discriminator, the discriminator is penalized. Note, that at the same time the encoder and generator models must also look to maximise the reconstruction likelihood.

This three-player game is played with two main loss functions that are listed be-

low:

1. **Reconstruction (Likelihood) loss:** The first loss corresponds to the likelihood - this tells us how well we are explaining our data according to the current estimate of the generative model, and is given by:

$$L = \sum_{t=1}^{T} log\, p(Y_t|\theta_t^m = \mu_t^{m,\theta}, \theta_t^c = \mu_t^{c,\theta}) \tag{4.14}$$

2. **Regularization (Prior) Loss:** The second loss regularises the estimate of the latent, time-varying parameters $[\theta_t^m, \theta_t^c]$ using an adaptive prior - this penalises when the posterior estimates of $[\theta_t^m, \theta_t^c]$ deviate from the prior:

$$R = \sum_{t=1}^{T} [\text{CrossEntropy}\,(\mu_t^{m,\theta}||\hat{\mu}_t^{m,\theta}) + \text{CrossEntropy}\,(\mu_t^{c,\theta}||\hat{\mu}_t^{c,\theta})] \tag{4.15}$$

This corresponds to the cross-entropy between the prior and the posterior of the latent, time-varying parameters $[\theta_t^m, \theta_t^c]$.

The total loss is the summation of these two terms:

$$F = \Lambda L + \frac{1-\Lambda}{2} R \tag{4.16}$$

where we attribute a certain weight to each term using Λ. The expression in Equation 4.16 can be thought of as an approximation of the variational free energy, but where the KL term is replaced by regularisation loss in Equation 4.15. The complete derivation for the regularisation loss in Equation 4.15 is given in Section S2.1. We trained the parameters (i.e., $\mathcal{S}_p$, $\mathcal{D}_q$, $\mathcal{E}_r$, θ_t^m, θ_t^c, ω_g^m, ω_g^c, ω_e^m, ω_e^c, ω_d^m, ω_d^c, λ_g^m, λ_g^c, λ_e^m, λ_e^c, λ_d^m, λ_d^c) using these losses in two phases - the reconstruction

phase and the regularization phase - which are explained below:

- **Reconstruction Phase:** In the reconstruction phase (Equation 4.14), the generator model (with parameters [$\mathcal{S}_p$, $\mathcal{D}_q$, $\mathcal{E}_r$, θ_t^m, θ_t^c, ω_g^m, ω_g^c, λ_g^m, λ_g^c]) and encoder model (with parameters [ω_e^m, ω_e^c, λ_e^m, λ_e^c]) are updated to minimize the reconstruction loss. We kept the discriminator model (with parameters [ω_d^m, ω_d^c, λ_d^m, λ_d^c]) constant during this phase, otherwise the generator and encoder would be trying to hit a moving target and might never converge.

- **Regularization Phase:** In the regularization phase (Equation 4.15), the discriminator model (with parameters [ω_d^m, ω_d^c, λ_d^m, λ_d^c]) is first updated to distinguish prior and posterior mean estimates of [θ_t^m, θ_t^c]. Then, the parameters of the generator model and encoder model are updated to confuse the discriminator model.

Single-dynamic Adversarial Generator-Encoder (SAGE)

There is variant of the proposed model, SAGE, that assumes the time courses α_{tp}, γ_{tr} and β_{tq} are the same. In this regard it is the same as the HMM. However, SAGE maintains the long-range temporal modelling and the partial volume modelling of the network state dynamics that are not part of the standard HMM. Throughout this work, we will compare MAGE with SAGE to highlight the benefits of the multi-dynamic modelling. SAGE is explained in Section S2.1.

4.3.4 Dual-Estimation of the group-level MAGE

We performed a process of dual-estimation (analogous to dual-regression following group-ICA) (Beckmann, Mackay, Filippini, & Smith, 2009) to infer the subject-level state versions of the group-level MAGE. This allows us to obtain a

subject-specific description of time-varying coorelation and time-varying mean. To achieve this, we used the trained parameters inferred from the group-level MAGE (α_{tp}, β_{tq}, $\mathcal{S}_p$, $\mathcal{D}_q$ and $\mathcal{E}_r$). Then we re-estimated the C_t and m_t for each subject by holding constant the α and β parameters, and only retraining $\mathcal{S}_p$, $\mathcal{D}_q$ and $\mathcal{E}_r$.

4.3.5 Hyper-parameters tuning and cross-validation

The hyper-parameters for MAGE were chosen using extensive trial and error by monitoring the validation loss. Some of the hyper-parameters are detailed as follows: 1) The sequence/window length, W, (number of data time-points) was set to 100 and batch-size, B, (number of grouped sequences) was set to 32 - this implies that each batch has a dimension of B*W*N. 2) a small learning rate of 0.0001 was used with large momentum value of 0.9. 3) Nesterov-accelerated Adaptive Moment Estimation (NAdam) optimizer was applied for training the data. 4) the number of units (the dimension of the inner cells in LSTM) for each LSTM layer was set between 8 to 64 (depending on the number of subjects). 5) Λ was optimised depending on the data-set (mostly constrained between 0.85 to 0.95 (I. Higgins et al., 2016)). To ensure that the reported MAGE results are reproducible, we performed cross-validations on different non-overlapping sets of subjects. We calculated the similarity of the MAGE estimated state-specific FC correlation and mean activity estimates across these different runs to highlight that the MAGE estimated states are replicable.

4.4 Results

4.4.1 SAGE Performance on HMM-like Categorical Simulations

We generated a multivariate normal distribution based on the assumptions that the state time courses are categorical and then simulated a multi-channel data-set from the distribution. For the simulation of the state time courses, we used Markov and semi-Markov chains, where transitions are dictated by a transition probability matrix (the probability of transition from one state to another). The main difference between the semi-Markov chain and the Markov chain is that the state lifetimes (i.e. how long one state will stay active before transitioning to a new state) in the semi-Markov chain are not governed by a Markovian dependency. Instead, they are, for example, sampled from a distribution of state lifetimes. The rationale behind the Markov chain based simulations is to match the HMM assumptions, and to demonstrate that SAGE performs well even in those scenarios. Moreover, the reason for including the semi-Markov chain based simulations is to determine SAGE performance when the state time-courses have long-range temporal dependencies (i.e., violate the HMM Markovian assumption).

We have illustrated the performance of SAGE (a single-dynamic counterpart of SAGE) across several categorical scenarios in Figure 4.3 (and compared the performance with the HMM [3] and SWC [4]). These simulated scenarios have certain hyper-parameters in common, e.g., 1) The number of channels, N, in the data. 2) The number of modelled states, P and Q and R. 3) The percentage of observational error, E. 4) Duration of the state lifetimes, k, for semi-Markov chains

[3] We have used the independent implementation of HMM that is available here (*HMM Implementation*, 2017).

[4] HMM and SWC approaches are described in Section S2.1.

(sampled from a gamma distribution where k represents the mean of the distribution). 5) The number of batches, B, used for training the data. The default values (unless specified otherwise) in these simulations are: N = 20 channels, P = Q = R = 6 states, E = 0.04% and B = 200 batches. For each of the simulation configurations, we generated fifty different simulations, in which we varied the observation model (randomly generated mean and covariance) and the state time courses (with randomly generated transition probability).

The performance is assessed by measuring the accuracy between the simulated versus predicted states using Markov chains in Figure 4.3[A-C]. In Figure 4.3[A], N is varied from 20 to 100, whereas, in Figure 4.3[B], P, Q and R are varied from 2 to 16 and in Figure 4.3[C], E is varied from 0.02% to 0.2%. Each box plot shows the prediction variability over 50 different simulations, and SAGE performance is compared with HMM and SWC in all the above-mentioned scenarios. In Figure 4.3[D,E], we conducted the stability analysis of the proposed SAGE model based on two factors: 1) How the performance is changed as we increase the lifetime values of state time courses, and 2) How much data is required to train the model to obtain a satisfactory performance? For Figure 4.3[D], we simulated semi-Markov chains that are characterized by k (varied from 1 to 400 time points).

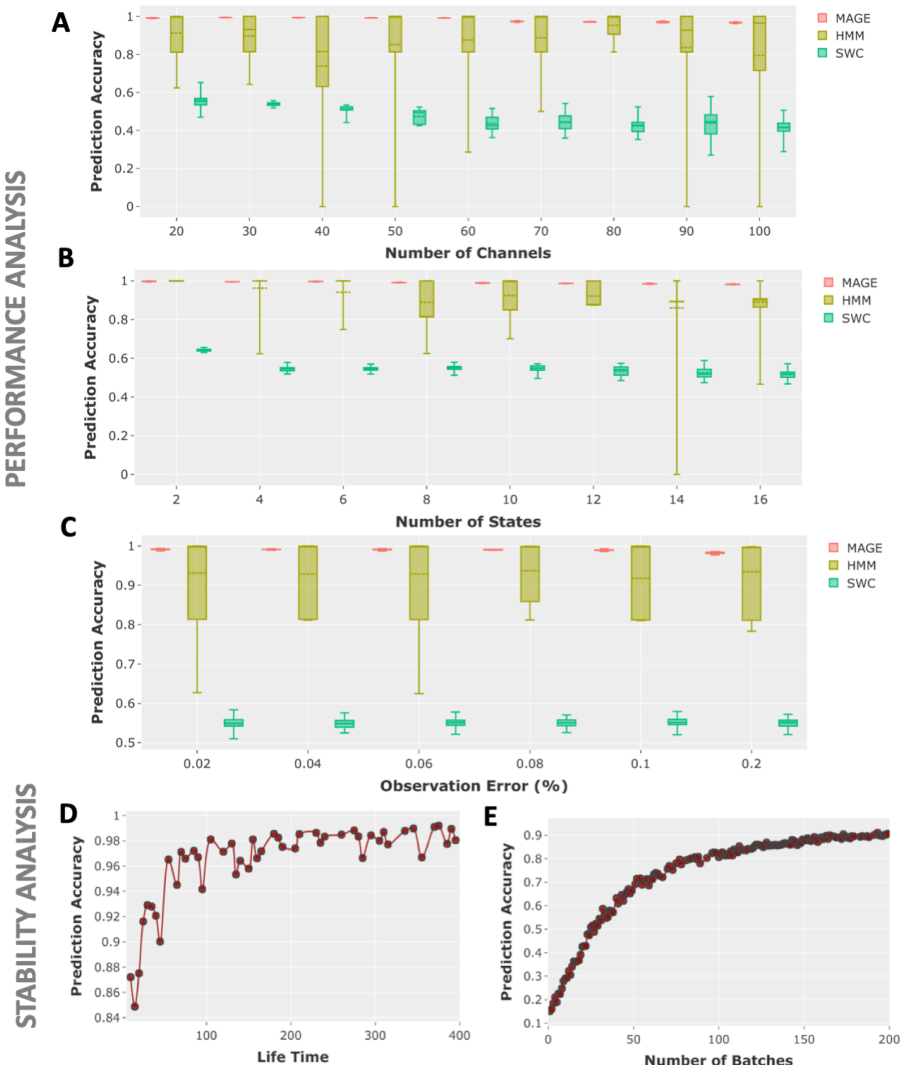


Figure 4.3: The performance and stability analysis of SAGE on the categorical simulations. SAGE, Sliding Window Correlation (SWC) and the Hidden Markov Model (HMM) were compared using same simulation parameters. In **[A-C]**, the impact of varying the simulation parameters on the prediction accuracy of the SAGE, HMM and SWC was assessed using Markov chains. Each violin plot shows the prediction variability over 50 different simulations. Specifically, we have illustrated the performance as we varied number of channels in **[A]**, the number of states in **[B]**, and observational error percentage in **[C]**. In **[D-E]**, we conducted a stability analysis to see the impact on performance of SAGE as we increased the ground truth lifetime of the simulated states using semi-Markov chains **[D]** or as we increased the number of batches **[E]**.

4.4.2 MAGE can infer dynamic linear mixtures of FC states

We start by considering the performance of our proposed method, MAGE, when we set aside the proposed multi-dynamic modelling. This version of MAGE models single-dynamics (where the state time courses for the mean activity and the correlations are the same, i.e., $\alpha_{tp} = \beta_{tq}$) and is referred to as SAGE. SAGE otherwise maintains all of the other features of MAGE, including the long-range temporal modelling and the partial volume modelling of the state dynamics.

We first demonstrate SAGE's ability to infer state time courses that are dynamic linear mixtures (i.e., partial volume mixtures) with more than one underlying state active at each time point, t. Note that this is in contrast to categorical simulations (i.e., HMM assumptions), where the ground-truth state time-courses, α are either 1 or 0 (see Figure 4.3 for simulations demonstrating SAGE's performance on categorical data). In the partial volume simulations, the ground-truth state time-courses, α could be any continuous number from 0 to 1 for any state, but where the sum of α for all states, α_{jt} at any time-point, t, should be 1.

Figure 4.4[A], shows illustrative examples where the number of simulated states is two and three respectively. Either one, two, or three can be active at any time point. The simulated ground-truth in Figure 4.4[A] is *single-dynamic* and the simulated state time courses (demonstrated in Figure 4.4[A]) are common to both the time-varying mean and correlations. We compared SAGE with the HMM to see how they recover the underlying simulated state time courses and the observation model parameters. The HMM failed to infer these state time courses accurately, as HMM is built around the assumption that only one state is active at any time point. On the other hand, SAGE does a good job of inferring these partial volume state time courses. To quantitatively assess the performance of these models, we calculated the Riemannian distance between the simulated covariances and the inferred

covariances for each time point. The lower the Riemannian distance between the inferred and simulated covariance, the better the performance of the model. We can see that SAGE can accurately infer the underlying time-varying covariance, but this is not the case for the HMM.

Figure 4.4[B,C] shows the performance of the approaches across many partial volume scenarios. We varied the number of overlapping states (either 2, 3, or 4), and then compared the performance of SAGE against the HMM. For each overlapping state configuration, we generated 50 different simulations in which we varied the state time courses (% of overlap), and the observation model itself. We illustrated the Riemannian distance between the inferred and simulated covariances. We also showed the prediction correlation (not accuracy, as state time courses, are no longer categorical) between the inferred and the simulated state time courses.

Lastly, we demonstrate the performance of MAGE for scenarios where the ground-truth is such that FC is static (i.e., no fluctuations in edges of the FC matrix over time). We show that MAGE does not infer time-varying correlations when we infer more than one state for such simulations (see Figure S23). This highlights that MAGE only infers time-varying FC when the underlying FC is time-varying.

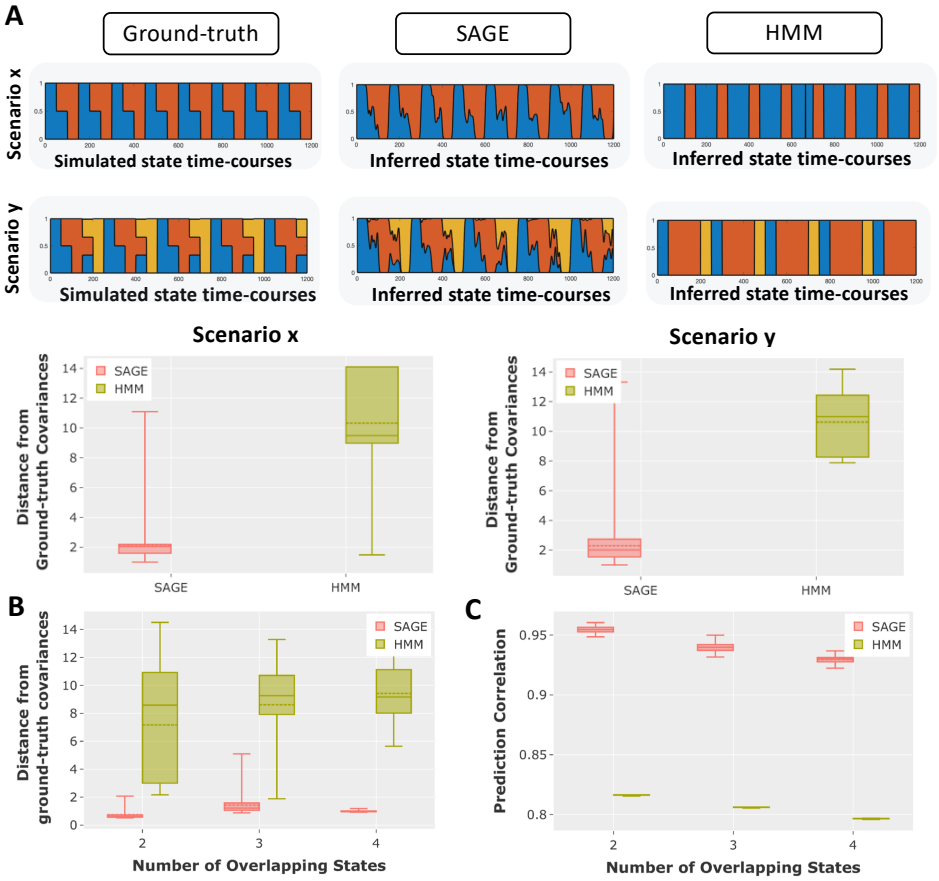


Figure 4.4: Comparison of the performance of the single-dynamic approach SAGE versus HMM, using simulations of dynamic linear mixtures of FC states (i.e., partial volume) and single-dynamics dynamics (i.e., where the state dynamics of the means and correlations are tied together). **[A]** Two illustrative partial volume examples, comparing SAGE performance with HMM. In scenario x, two states can be simultaneously active, whereas, in scenario y, up to three states can be simultaneously active. Note that these state-time courses were common to both the time-varying mean and correlations (i.e., instantaneous FC). The bottom row of **[A]** shows the Riemannian distance between the inferred and simulated covariances for both HMM and SAGE. **[B-C]** Comparison of the performance of SAGE with HMM on a number of partial volume simulations. Each box plot shows the distribution of prediction performance results over 50 different simulations. Riemannian distance between inferred and simulated covariances was used as a measure of success in **[B]**, and prediction correlation between the inferred and simulated state time-courses in **[C]**.

4.4.3 Failing to model distinct mean activity dynamics can result in homogeneous estimation of FC over time

Commonly-used SWC and HMM approaches are what we refer to as single-dynamic models, as they tend to tie the dynamics of the correlations (i.e., the FC) to be the same as the dynamics of the means of the brain activity. However, if the underlying truth is that the dynamics of the means and correlations are actually distinct from each other, then using single-dynamic approaches to infer TVFC can have negative consequences. This is because forcing α_{tp} to be the same as β_{tq}, when the ground truth is that they are not the same, causes the dynamics used for estimating the state-specific FC to be incorrect. This then leads to a smearing of the FC estimation over time, making it look more homogeneous than it really is.

We illustrate the mechanism of the multi-dynamic approach using a simple scenario in Figure 4.1[A]. Furthermore, in Figure 4.1[B], we demonstrate the potential benefits of the multi-dynamic approach on a simulation [5] in which the ground truth is such that the correlation dynamics and the mean activity dynamics are not tied together. Specifically, Figure 4.1[B] shows the instantaneous FC estimated using SAGE or MAGE when we simulate using equations (4.14.24.5S8S9 with α_{tp} not equal to β_{tq}.

When we infer TVFC using SAGE, the single-dynamic version of MAGE (with α_{tp} set to be the same as β_{tq}), there are substantially reduced fluctuations in the inferred instantaneous FC over time (i.e., FC is mapped as being near-homogeneous/static) as is often observed in TVFC estimates in real fMRI data. In contrast, the multi-dynamic version of MAGE (with α_{tp} free to be not equal to β_{tq}) is better able to capture the true fluctuations in instantaneous FC over time.

[5] For this particular simulation, the number of channels in the data is 20, the number of modelled states (P=Q=R) is 12, and multi-dynamics state time courses for the means and the correlations are simulated using two different hidden semi-Markov model chains.

4.4.4 MAGE can infer dynamics in the FC that are distinct from dynamics in the means of brain activity

We next sought to see if MAGE could identify multi-dynamics in a range of scenarios. For this, we simulated distinct dynamics (i.e., multi-dynamics) for the mean and correlations (i.e., the FC) by simulating two different state time courses, α_{tp} and β_{tq}, using two different hidden semi-Markov model chains, whose state lifetime values were sampled from two different gamma distributions such that α_{tp} was not equal to β_{tq}. These state time courses were then used in equations (4.14.24.5S8S9 to simulate the data.

We compared the performance of the multi-dynamic approach (MAGE) with the single-dynamic approach (SAGE) in Figure 4.5[A,B]. As illustrated in Figure 4.5[A], the lifetimes of both the mean and correlation states in the simulation were sampled from the gamma distribution. However, we fixed the distribution of lifetimes of the correlation state time courses by setting the gamma distribution parameter to k=30 (where, k represents the mean of the distribution), whereas, for the mean state time courses, we varied k from 5 to 80, so that the mean state time courses spanned a variety of timescales in different scenarios. In Figure 4.5[B], it is the other way around, where we fixed the lifetime distribution of the mean state time courses (k =30), but varied the lifetime distribution for the correlation state time-courses (again from k = 5 to 80). In Figure 4.5[A,B], we have plotted the prediction accuracy for the state time courses of correlation and mean. MAGE does a better job in predicting the underlying simulated labels for state time-courses as compared to SAGE. Note that we did not draw any comparison with the HMM here, as the HMM is also a single-dynamic approach, and the HMM is not able to infer the mean and correlation in such a multi-dynamic fashion (and hence the results would be expected to be closest to SAGE).

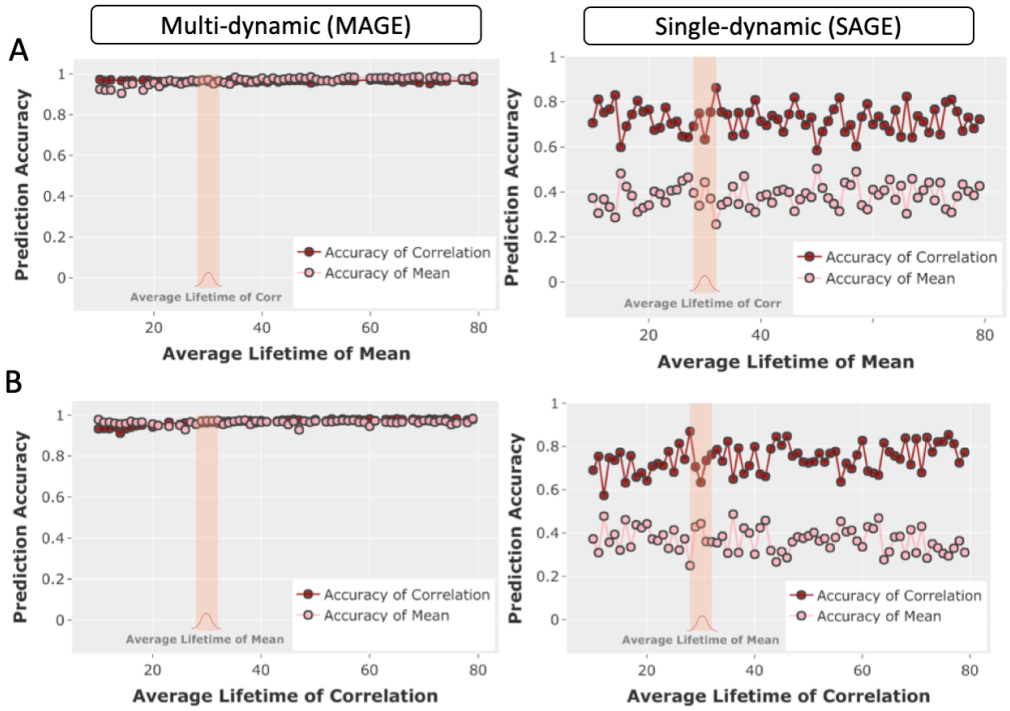


Figure 4.5: Comparison of the performance of the multi-dynamic approach (MAGE) and the single-dynamics approach (SAGE) on multi-dynamic simulations where multi-dynamics are the ground truth (i.e., where the state dynamics of the means and correlations are free to fluctuate independently). **[A,B]** The prediction performance of multi-dynamic and single-dynamic approaches are compared across a range of simulations. The x-axis shows the lifetime parameter (sampled from a gamma distribution) for the mean and correlation state time-courses, whereas the y-axis shows the prediction accuracy of the inferred state time-courses. Specifically, in **[A]**, we fixed the lifetime distribution of the correlation state time-courses (k = 30, as shown by the shaded bar) but varied the lifetime distribution for the mean state time-courses (from k = 5 to 80), and it is the other way around in **[B]**.

4.4.5 Multi-dynamic approach reveals stronger changes in FC over time in resting fMRI

The simulations shown in Figure 4.1 suggest that multi-dynamic approaches may be better able to identify TVFC than single-dynamic approaches, which may erroneously infer FC as being too homogeneous over time. We looked to test this idea

by applying the single-dynamic (SAGE) and multi-dynamic (MAGE) approaches to resting-state fMRI data. In particular, we can analyse how the temporal dynamics of mean activity levels and FC differ (i.e., α_{tp}, β_{tq}) when modelled independently and see if this has any effect on the inferred state-specific FC maps. Lastly, we compared MAGE results with the HMM and SWC approaches, to further substantiate our findings.

We used rfMRI data from 13,301 subjects from the UK Biobank (UKB) and 1003 subjects from the Human Connectome Project (HCP). We applied a popular data-driven parcellation scheme called spatial Group Independent Component Analysis (groupICA) to our data to identify the parcels/channels (N) that form the data to be input into MAGE.

Temporal and Spatial Dynamics We first compare the single-dynamic (SAGE) and multi-dynamic (MAGE) approaches using the HCP data, where the only difference between the two methods is whether or not we model independent state dynamics for the state means and FC (i.e., correlations). In Figure 4.6[A], we show the state-specific mean activity and FC correlation maps for SAGE and MAGE approaches for the HCP data. In total, we modelled P=Q=R=12 states, but we have only displayed the observation model for four of the 12 states for ease of visualisation (the results for all 12 states are illustrated in Figure S26, Figure S27, Figure S28 and Figure S29). As described earlier, MAGE also models time-varying variances, whose dynamics are tied to be the same as the mean. However, we have not displayed the state-specific spatial maps for variance as MAGE infers that there was little between-state variability in the state-specific variances (i.e., the variances remained static across states).

Figure 4.6[A] shows that the state-specific mean activity maps vary reasonably well across states using both SAGE and MAGE. However, with SAGE, it can be

seen that the state-specific FC maps (the rank-one decomposition of the estimated state-specific correlation maps are shown to make these 2D FC maps easier to view) of all the states are predominantly similar, always resembling a somatomotor state. We also attempted another modelling option with SAGE in which we explicitly forced the mean activity to be zero[6] (i.e., $m_t = 0$ in Equation 4.1). This approach also resulted in homogeneous state-specific FCs as shown in Figure S30. In contrast, for the MAGE approach, we can see that the state-specific FC maps correspond to distinct brain networks, with the state-specific FCs looking more distinct from each other than in the SAGE results. To illustrate this quantitatively, we show the pairwise between-state Riemannian distance in Figure 4.6[B]. Specifically, we calculated these pairwise distances over the actual 2D correlation matrices (i.e., FCs). It is clear that these between-state pairwise distances are higher for the state-specific FC maps when inferred using the multi-dynamic MAGE approach, revealing stronger changes in FC over time.

In Figure 4.6[D], we compare summary measures of the mean activity and FC state dynamics. Since MAGE models a linear mixture (i.e., a partial volume) at each time point, it is not straightforward to compute summary statistics normally used on binary state time courses. In order to allow us to do this we threshold the continuous partial volume state time courses (α_{tp} and β_{tq}) at 0.3. Firstly, we show the state switching rate for both models, which is a measure of the frequency of state fluctuation for each subject, and it can be seen that FC-driven state time-courses, β_{tq}, have the lowest switching rate. Then, we show the state lifetimes for both models, which correspond to the number of seconds per state visit. We can see that the FC-driven state time-courses have the highest values for state lifetimes (2 to 5 seconds) as compare to lifetimes of mean state time-courses (2 to 3

[6]For other such examples, see (Vidaurre et al., 2021) for modelling only the covariance and (Charquero-Ballester et al., 2020) for modelling only the mean activity.

seconds). In Figure 4.6[E], we show an example segment of the state time course (of a randomly chosen subject) showing the nature of the mixing of the inferred 12 states and the degree of overlap between the mean activity and the FC state time-courses. Furthermore, in Figure 4.6[C], we quantitatively compare the extent of similarity between the mean activity and the FC state time courses. We plot the mean activity and the FC state time courses correlation matrices as histograms for the null and the HCP dataset. There is a significant non-zero correlation as compared to the equivalent null scenario, but there is clearly still a distinct difference between them (and hence the difference in the performance of SAGE vs MAGE). In short, the means and FCs have some shared dynamics but also some unique/independent dynamics.

It has been previously proposed that the observed estimates of TVFC must be compared against "static" null hypotheses to establish claims of non-stationary (D. J. Lurie et al., 2020). Considering that, we generated null data using autoregressive randomization (ARR) following (Liegeois, Laumann, Snyder, Zhou, & Yeo, 2017). Specifically, a 10-th order Gaussian autoregressive (AR) model was applied to generate null data. For each participant and each pair of brain regions, we compared the observed between-timepoint variances of the instantaneous FCs inferred by MAGE, which can be calculated from the FC state dynamics and state-specific FCs, from the real HCP data against the null data. This analysis resulted in one p-value for each edge of the FC matrix (these are uncorrected p-values). On average, across HCP participants (compared against null data), 95% of edges were significant ($p \leq 0.001$). Furthermore, in Figure S24, we show the histogram of null variance values compared with the real HCP data variance values (this highlights that between-brain region variability over time is significantly higher in the real HCP data as compared to null data).

We find that we get similar results to the HCP dataset when we use MAGE on the

resting fMRI data from the UKB dataset. Here, we also compare the performance of MAGE with other established methods (HMM and SWC) on the UKB dataset. Figure 4.7 shows the state-specific FC maps from MAGE, HMM and SWC, with the HMM and SWC approaches showing similar behaviour to that of SAGE in Figure 4.6. It is also clear that there are smaller between-state differences in FC across time for both the HMM and SWC approaches compared to MAGE (as reflected by the between-state Riemannian distance plots in Figure 4.7[D]), consistent again with the idea that MAGE reveals stronger changes in FC over time due to its ability to model multi-dynamics.

To ensure that the MAGE estimated FC and mean activity states are reproducible, we ran MAGE six times on different non-overlapping subsets of UKB data (2k subjects for each run). In Figure 4.8[A,B], we compare the similarity between MAGE estimated state-specific spatial maps across the six runs. Specifically, in Figure 4.8[A], we show the average correlation plot for the state-specific FC estimates across repeated runs, and in Figure 4.8[B] for the state-specific mean activity estimates, both for a fixed number of states (P=Q=12). This shows that the state-specific FCs are very reproducible and that the state-specific mean activity estimates are moderately reproducible but not to the same extent. We then replicated these reproducibility findings for a wide range of numbers of states (1 to 24) for the FC and mean activity, as shown in Figure 4.8[C], also demonstrating that P does not have to be equal to Q. We show in Figure 4.8[C] that the FC are mostly highly and mean activity are mostly moderately reproducible independent of the number of modelled states. Specifically, mean activity and functional connectivity states are more reproducible if the number of modelled states is more than approximately six.

It has been previously reported in (Vidaurre et al., 2017) that the brain network activity is organized into a hierarchy of two distinct meta-states, with one meta-state

representing hisgher-order cognition, and the other representing the sensorimotor systems. We have taken a similar approach as (Vidaurre et al., 2017), and computed [P x d] or [Q x d] matrices containing the fractional occupancy (time spent in each state) for each subject, where d is the number of subjects. We then use these to compute the [P x P] or [Q x Q] correlation matrices over subjects, which are shown in Figure 4.8[D]. As in (Vidaurre et al., 2017), Figure 4.8[D] shows that a hierarchical meta-state structure is present. Here, it is apparent in both in the functional connectivity and the mean activity driven state time-courses, although it is more apparent in the mean activity state time-courses.

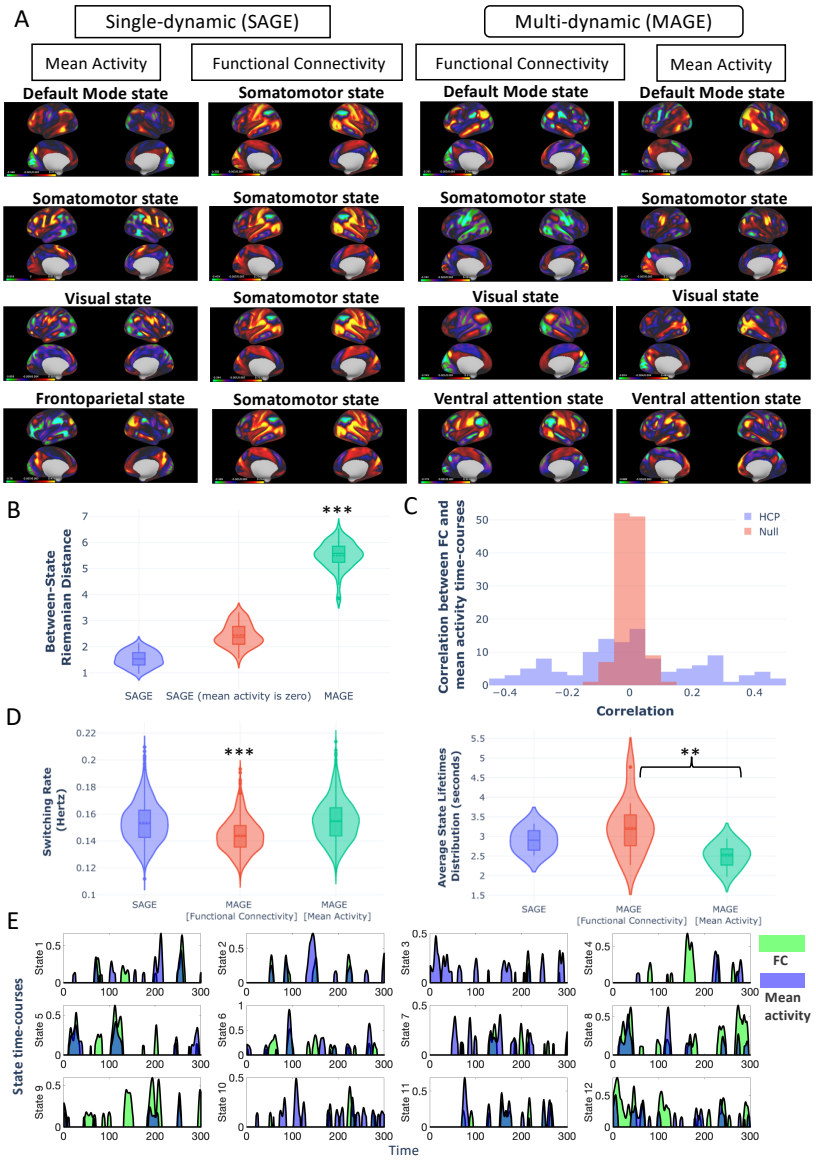


Figure 4.6: **[HCP data] [A-B]** The multi-dynamic (MAGE) approach reveals stronger changes in state-specific FC maps compared to the single-dynamic (SAGE) approach. Using the HCP data, it is evident that qualitatively and quantitatively, MAGE-derived FC shows stronger changes over time. **[A]** State-specific mean activity and FC correlation maps for four of the modelled states for both approaches are shown. For the state-specific mean activity maps, the average amplitude of each node/channel is shown and for the state-specific FC maps, rank-one decomposition (first Eigenvector) of the estimated FC correlation maps are shown. **[B]** The pair-wise Riemannian distance of the state-specific FC correlation matrices between 12 FC states is shown for both MAGE and SAGE (with and without modelling the mean activity, see (Vidaurre et al., 2021) as another example of modelling only the correlation, i.e., $m_t = 0$ in Equation 4.1). **[C]** Correlations between the mean-derived state time-courses and the FC-derived state time-courses plotted as histogram over all pairs of states (and equivalent null scenario). The non-zero correlations indicate that the mean activity and FC have some shared dynamics but also have unique temporal dynamics. **[D]** Summary measures of the state dynamics for both approaches. It can be seen that FC state dynamics show the lowest switching rate (in Hertz) and the highest lifetimes values (in seconds) as compared to other approaches. **(E)** Example segment of a state time course of a randomly chosen subject is shown to qualitatively highlight the extent of similarity (and dissimilarity) between the mean activity and the FC state time courses. These also show how states in MAGE can overlap in time.

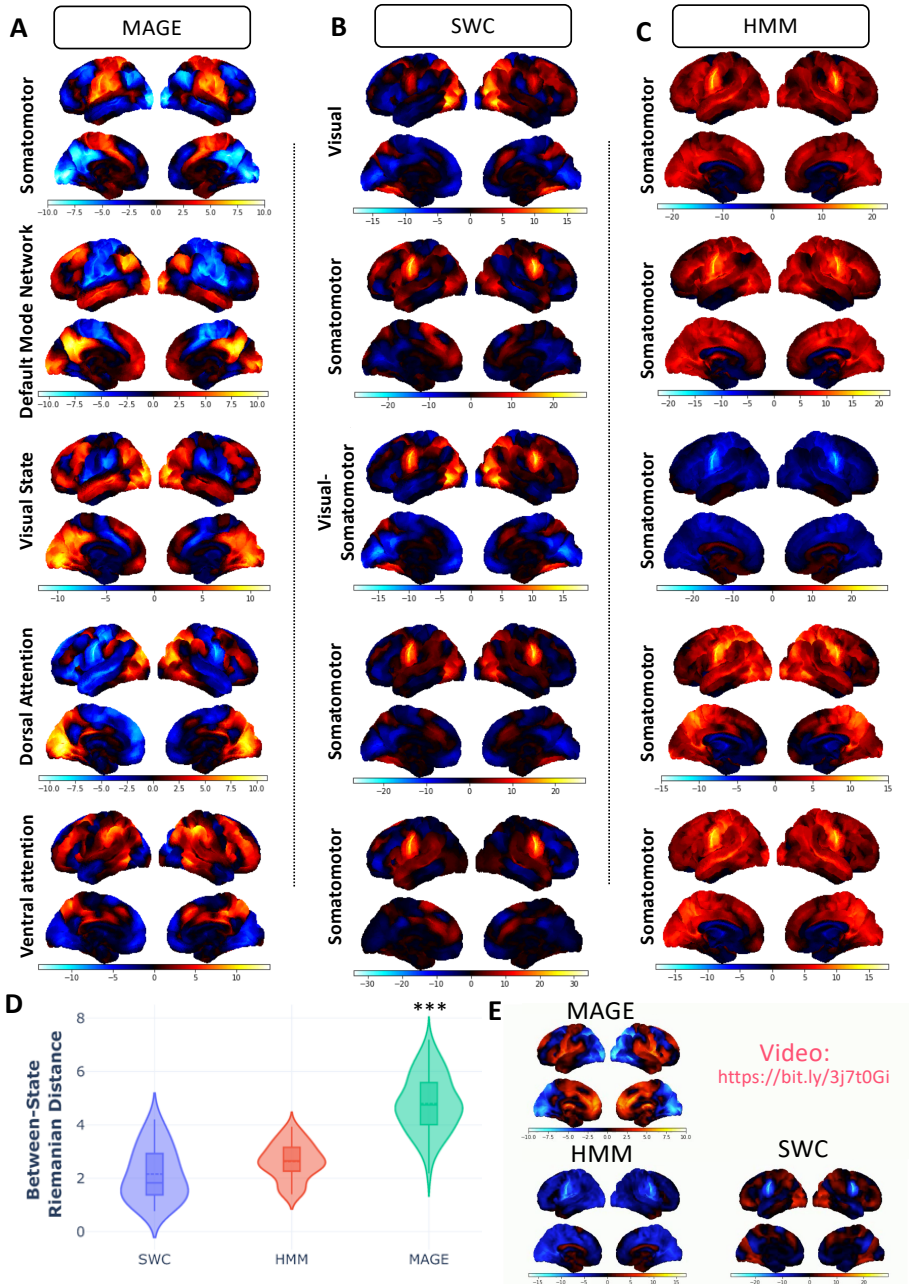


Figure 4.7: **[UKB data]** Demonstration that the between-state spatial variability in state-specific FC maps is significantly higher when modelled using the MAGE approach on UKB data compared with the HMM or SWC approaches. **[A-C]** The pair-wise Riemannian distance of the state-specific FC correlation matrices between 12 FC states is shown for MAGE, HMM and SWC, alongside the FC correlation maps for 5 of the identified states. The state-specific FC maps are shown as rank-one decomposition (first Eigenvector) of the estimated state-specific FC correlation matrices. The channels with positive values (red regions) show they are positively correlated with other regions, whereas channels with negative values (blue regions) show they are negatively correlated with other regions. The state-specific FC maps for all 12 states using SWC, HMM and MAGE are illustrated in Figure S31, Figure S32, and Figure S33 respectively. **[D]** The between-state Riemannian distance for the modelled 12 states, shown for MAGE, SWC and HMM. **[E]** Video for the time-point by time-point instantaneous FC estimation for a randomly chosen UKB subject is shown for the MAGE, SWC and HMM approaches.

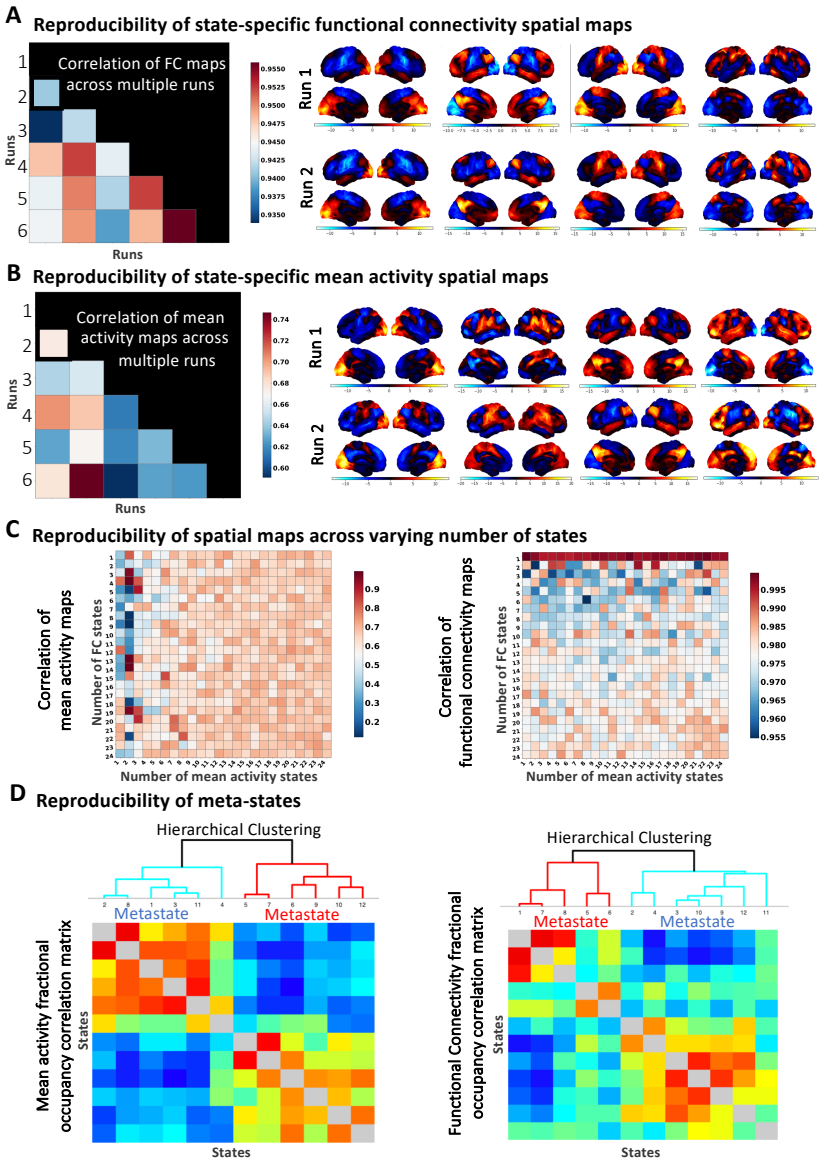


Figure 4.8: **[UKB data]** MAGE estimated FC and mean activity spatial maps are reproducible across various runs on different non-overlapping sets of subjects. **[A]** shows the average correlation between the state-specific FCs across six repeated runs. Moreover, for runs 1 and 2, FC correlation maps are shown for 4 of the identified states (out of the 12 inferred states), demonstrating that the FC states are consistent across repeated runs. On the other hand, **[B]** shows the average correlation plot for the state-specific mean activity estimates across six repeated runs, alongside the mean activity spatial maps (for runs 1 and 2 only). The analysis in **[A]** and **[B]** is performed for a fixed number of states (P = Q = 12), whereas, in **[C]**, we evaluated the reproducibility of state-specific FC and mean activity estimates by varying the total number of modelled states (P, Q). Specifically, **[C]** shows the correlation plots for the state-specific FCs and state-specific mean activity estimates, where we varied the total number of inferred states (P and Q) from 1 to 24. These results suggest that inferred states for the FC are highly and mean activity are moderately reproducible independent of the number of modelled states (Note the difference in the ranges of the colormaps). **[D]** We illustrated that hierarchical meta-state structure exists both in the functional connectivity and the mean activity states, but is more well-defined in the mean activity states (matrices computed as $\upsilon\upsilon'$, where υ is the [number of states by number of subjects] fractional occupancy matrix).

4.4.6 Multi-dynamic TVFC predicts behavioural variability better than the single-dynamic approaches

We next used the UKB data to see if the multi-dynamic MAGE approach provides spatial network features (i.e., the state-specific FCs) that can be used to predict individual behavioural traits, and we compare its performance to the HMM and SWC approaches. Alongside the task fMRI results, this would provide further evidence that the state dynamics being inferred by MAGE are cognitively meaningful.

We applied the dual-estimation method to obtain subject-level estimates of TVFC. This means that for each subject, we can obtain a set of state-specific FC maps (Q * N * N), where Q is the number of correlation (FC) states and N is the number of channels/brain regions. Then, we can use these subject-level, state-specific FC estimates to predict selected behavioural variables. The classifier/predictor we employed is Elastic Net. The data matrix fed into the classifier is d*[Q*f], where d is the number of subjects, Q is the number of states and f is the number of features. The state-specific FC matrices are symmetric and only values above the diagonal need to be retained and vectorized, and hence f = N(N-1)/2. Lastly, we concatenated the state-by-state subject-level, state-specific FC estimates, resulting in Q*f features for each subject. We performed this prediction analysis for UKB data with d = 13301, Q = 12, and N = 25.

We used four non-imaging variables to compare the performance of multi-dynamic (MAGE) approach with HMM and SWC. The details for the SWC and HMM approaches are explained in the Section S2.1. These variables are age, sex, fluid intelligence score, and neuroticism score. In Figure 4.9, we display the prediction results (nested 10-fold cross-validated) from the subject-level FC estimates. We can see that the FC obtained using the MAGE approach is a more accurate

predictor of behaviour as compared to the other approaches

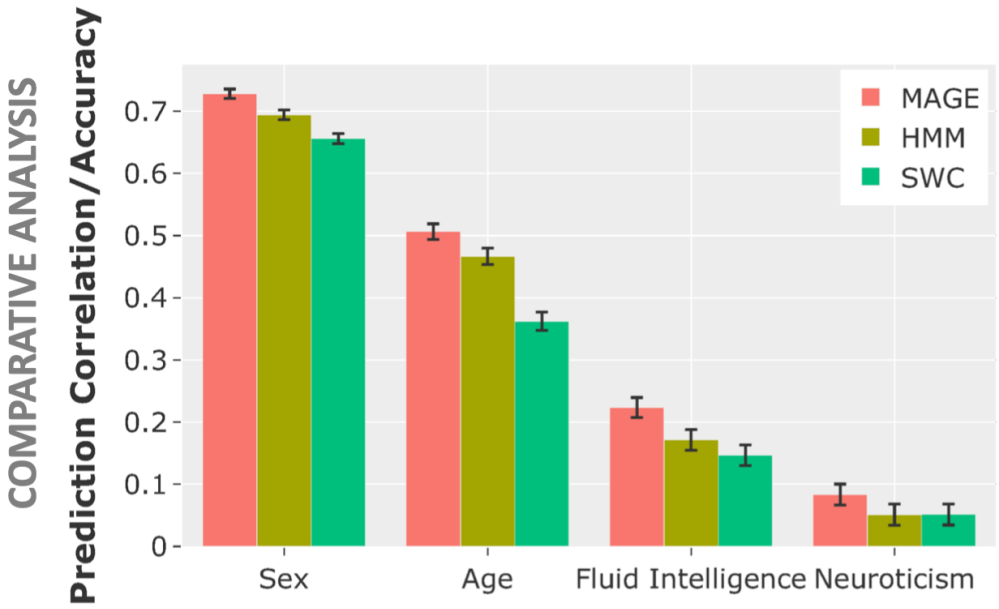


Figure 4.9: Multi-time-scale (MAGE) derived state-specific FCs are used to predict non-imaging variables in the UKB data. The prediction performance of MAGE-derived state-specific FCs compared with the HMM and SWC approaches. The prediction accuracy/correlation is estimated using the combined features by concatenating the state-specific FC estimates from all states. Y-axis shows the prediction correlation (age, fluid intelligence, neuroticism) and prediction accuracy (sex), and the x-axis shows the predicted variable. To draw the confidence intervals on the prediction accuracies for sex variable (discrete output), we performed the Wilson test (suitable for the binomial distribution). Lastly, to generate confidence intervals on the prediction correlations (i.e., continuous outputs e.g., age, fluid intelligence and neuroticism), we computed the Fisher transformation.

4.5 Discussion

We have proposed a novel algorithm, MAGE, that reliably infers the time-point by time-point estimates of FC. Notably, MAGE frees up the dynamics of the FC to fluctuate independently from the dynamics of the means and variances of the brain activity. We found that this multi-dynamic approach reveals much stronger

changes in FC over time than existing single-dynamic approaches (e.g., HMM, SWC), and is a better predictor of individual behavioural variability. This provides a potential explanation and solution as to why time-varying FC has appeared to look so stable in resting fMRI data in previous work (Hindriks et al., 2016; Liegeois et al., 2017).

MAGE inferred multi-dynamic linear mixtures of FC states. We assessed the performance of MAGE on simulations where there can be more than one underlying state active at each time point (Figure 4.4). We demonstrated that categorical models, such as the HMM, under-performed as expected when applied to such simulations. In contrast, MAGE remained accurate across a range of these simulations. MAGE also shows improved performance in simulations designed to be suitable for categorical models (e.g., HMM), i.e., by having only one state active at each time point (mutual exclusivity) (Figure 4.3). This suggests that MAGE performance remained stable when the state time-courses have long-range temporal dependencies (i.e., violate the HMM Markovian assumption) (Figure 4.3[D]). Lastly, we assessed the performance on the multi-dynamic simulations and demonstrated that MAGE accurately identified unique dynamics of the correlations when the ground truth was that the correlations and mean activity dynamics were distinct from each other (Figure 4.5).

Why do we model the mean of the brain activity separately from the functional connectivity? Our assumption in Equation 4.1 is that the mean and the FC (i.e., the correlation matrix) continuously vary over time. From that perspective, there is no imperative to assume that the dynamics of those two entities are the same as each other. As such, MAGE is designed to explore what happens when we do not tie the dynamics together. By allowing the state dynamics of the mean and FC to be independent of each other, MAGE also allows them to potentially operate on different time scales. However, in real fMRI data, we found that the dy-

namics of the mean states and the FC states ended up on fairly similar time-scales (between 2 to 5 seconds for FC states and 2 to 3 seconds for mean activity states, as shown in Figure 4.6). It is also interesting to consider the extent to which the state dynamics of the mean and FC get inferred as being coupled together, despite being free to be inferred as completely independent. Figure 4.6[C] demonstrated that there was indeed some significant non-zero correlation between the two types of dynamics. However, this correspondence between the dynamics of the mean and the dynamics of the FC was not absolute. In other words, the fluctuations in the FC clearly have some unique temporal structure that is not shared with the fluctuations in the mean of the activity. Indeed, the benefit of capturing these unique fluctuations was demonstrated by the improved performance of MAGE (versus SAGE) in inferring stronger fluctuations in FC over time and in predicting behavioural traits.

Why does MAGE reveal stronger functional connectivity dynamics in resting-state fMRI? Our results demonstrated that MAGE revealed stronger changes in FC over time than single-dynamic approaches, where changes in FC over time correspond to between-state differences in the state-specific FCs (e.g., HMM, SWC, SAGE, see Figure 4.6 and Figure 4.7). The lack of change in FC over time in single-dynamic approaches has been observed in other, recent studies (Laumann et al., 2017b; Hindriks et al., 2016; Leonardi & Van De Ville, 2015). Indeed, in some cases, it has been observed that the reported variability in the FC from one state to another state (or against null models) is statistically insignificant (Liegeois et al., 2017). So why does MAGE show comparatively stronger changes in FC over time? We can gain insight into the multi-dynamic model being used in MAGE, by considering a dynamic version of structural equation modelling (SEM), as explained in Section S2.1. This shows that if the underlying truth is that there are distinct dynamics for the external driving inputs and connectivity, then distinct dy-

namics are needed for the mean activity and FC modelled by MAGE. If we were to ignore this, and use a single dynamic approach (e.g., HMM, SAGE, SWC) then we will infer common state dynamics that are a compromise. In turn, use of these incorrect state dynamics will tend to cause the inferred FC patterns to be temporally smeared versions of the true FC patterns, causing them to look more homogeneous over time. We demonstrated this phenomenon with a simple example (Figure 4.1[B]), in which we simulated a ground truth where the FC dynamics and the mean activity dynamics were not tied together. Using SAGE on this data substantially reduced the inferred change in instantaneous FC over time, whereas MAGE better captured the true variability in instantaneous FC across time. This work provides a potential explanation as to why FC in resting fMRI has previously appeared to look so homogeneous over time (Laumann et al., 2017b; Hindriks et al., 2016; Liegeois et al., 2017).

Reproducibility of MAGE states. The presented results in Figure 4.8 demonstrated that the MAGE estimated FC states are highly and mean activity states are moderately reproducible across re-runs on non-overlapping subsets of the data (shown quantitatively and qualitatively in Figure 4.8[A,B]). Also, as is typical with other unsupervised learning methods (e.g., ICA, HMM), the number of states to be modelled is an empirical parameter in our method and it needs to be chosen before running the model. In most of our presented work, we modelled anywhere from 4 to 16 states, though we had experimented with up-to 64 states in the resting-state and task data (and the results remained sensible). In Figure 4.8[C], we found out the estimated states are more consistent across re-runs when the number of modelled states is more than six. Considering that, we suggest that the number of modelled states to be more than approximately six for such dynamic modelling (though we do not have a strong recommendation for an exact number of states).

What do we mean by TVFC? We can gain insight into the multi-dynamic model being proposed, by considering a dynamic version of structural equation modelling (SEM) as explained in Section S2.1. The SEM perspective suggests that the instantaneous FC is the closest representation that we have to the time-varying connectivity in a dynamic SEM (Equation S3). In MAGE, we have assumed that the instantaneous FC, C_t, is the TVFC measure of most interest. MAGE provides a regularised estimate of the instantaneous FC, and so in this work we have represented the instantaneous FC parsimoniously using the state-specific FCs. MAGE also provides other potential features for use in prediction, for example, the mean-activity FC (i.e., correlations of the mean activity, m_t, over all time) or the state-specific mean activity and the mean-derived dynamics.

Variance of the brain activity. We found through resting-state and task data-sets results (Figure 4.6 and Figure 4.7) that there was not much between-state variability in the variance of the activity, and the variance remained static over the scanning session. Nevertheless, we explored another configuration where dynamics of the variance of the activity was modelled with independent state dynamics from the mean activity and the FC, but that also did not result in any meaningful variability in the variance of the activity (though it is computationally more expensive to learn the variance state time courses separately). We also attempted to model state dynamics of the variance of the activity tied to the state dynamics of the FC fluctuations, but independently from the mean activity; however, this resulted in comparatively less between-state variability in the FC. In summary, we found out there was little between-state variability in the state-specific variances irrespective of the modelling choice.

Multi-dynamic TVFC predicted behavioural variability better than existing methods. We evaluated whether the inter-individual differences inferred by MAGE are meaningful. We demonstrated in Figure 4.9 that the MAGE-estimated,

state-specific FCs did a better job at predicting these non-imaging variables as compared to SWC and HMM approaches. For example, age was predicted with an average correlation of 0.50 using MAGE as compared to 0.46 using HMM, and 0.36 using SWC. This provides evidence that the TVFC estimated using the MAGE is a more accurate and meaningful representation of underlying functional networks and is indeed a powerful biomarker for cognitive traits.

4.6 Conclusion

In summary, we proposed a multi-dynamic approach (MAGE) that models temporal fluctuations in FC independently from fluctuations in the mean of the activity. Multi-dynamic modelling provided an explanation and a solution as to why resting fMRI FC has previously looked so stable and homogeneous. We found out that that separating fluctuations in the mean activity levels from those in the FC reveals much stronger changes in FC over time. Lastly, MAGE estimated time-varying FC is a better predictor of behavioural variability in the resting-state fMRI data than established methods.

CHAPTER 5

CLINICAL APPLICATIONS OF STATIC AND TIME-VARYING FUNCTIONAL CONNECTIVITY

5.1 Introduction

The previous two chapters described new methods for estimating both static functional connectivity (FC) and time-varying functional connectivity (TVFC). We are interested in seeing if we can use these new measures to better understand how brain networks are related to cognitive abilities, both in health and disease. In this chapter, we will investigate the clinical applications of both static FC and TVFC in resting-state and task fMRI.

In Chapters 3 and Chapters 4, we primarily evaluated our proposed methods to improve the prediction performance of four non-imaging variables: age, sex, fluid intelligence score, and neuroticism score. These non-imaging variables acted as a valuable evaluation parameter for comparing the benchmark performance of designed models. In this chapter, we look to go beyond the prediction of subject-

specific measures of non-clinical variables: firstly, we investigate our ability to predict measures in clinical applications using our optimised static FC measures; and secondly, we investigate our ability to predict within-subject task and cognitive state using our new TVFC measures.

5.2 Static Functional Connectivity in the Clinical Realm

The important finding of Chapter 3 is that we have proposed an *almost* optimum pipeline for relating whole-brain functional connectivity patterns to non-imaging measures such as neurological and behavioural characteristics. We want to employ this optimum static FC pipeline to answer the following questions: Can the alteration in static FC patterns: 1) Help us diagnose and understand mood disorders? 2) Provide potential biomarkers for predicting neurological dysfunctions? 3) Uniquely identify a given individual from the set of profiles?

5.2.1 Neurological Conditions and Drug Addiction

We expect neurological disorders such as Autism Spectrum Disorder (ASD) and addictions (e.g., the long-term use of nicotine) to disrupt the brain's typical functioning. It seems reasonable to assume that this disruption will be apparent in the brain's functional connectivity. We therefore investigated whether ASD and long-term use of nicotine manifests as irregularities in brain FC patterns at rest, and whether these irregularities are sufficiently robust to be identified by our proposed approaches.

Diagnosis of Autism Spectrum Disorder (ASD)

We used the ABIDE dataset (Di Martino et al., 2014) that is available for 871 subjects (autism and control), as described in Chapter 2 [1]. We picked the best pipeline configurations for static FC estimation based on the generally top-performing techniques as demonstrated in Chapter 3; i.e., "*parcellation (ICA), functional connectivity (Tikhonov or SGGGM), tangent space parameterization (yes), and tangent space shrinkage (optional)*". With regards to the recommended predictor/classifier, Elastic Net tends to be more stable, but CNN-based architectures can out-perform Elastic Net sometimes (as illustrated in Chapter 3); therefore, we tested all the predictors/classifiers for this task.

We predicted whether inter-individual differences in static FC patterns could be reliably associated with ASD using the optimum modelling techniques. Figure 5.1 illustrates the recommended pipeline's complete set of results for ASD diagnosis (and age prediction) for ABIDE participants [2]. It is to be noted that the prediction of ASD is a challenging task, and that is why we added age as a control measure in the results (this will ensure that the pipeline is able to perform the comparatively easy task of age prediction).

[1] Chapter 2 (Section 2.1.3) details the ABIDE pre-processing pipeline.

[2] The reported results are nested cross-validated as described by criteria in Chapter 3 (Section 3.3.4).

	Brain Parcellation	Functional Connectivity Estimation	Tangent Space Projection	Shrinkage In Tangent Space	Classifier/ Predictor	Prediction Correlation Age (r) ABIDE	Prediction Accuracy Autism ABIDE
Recommended Pipeline (1)	ICA_50D	SGGGM	Yes	Yes	ELASTIC NET	0.652	0.55
	ICA_50D	SGGGM	Yes	Yes	2DRCNN	0.784	0.422
	ICA_50D	SGGGM	Yes	Yes	BRAINNETCNN	0.715	0.46
Recommended Pipeline (2)	ICA_50D	Tikhonov	Yes	No	ELASTIC NET	0.643	0.5019
	ICA_50D	Tikhonov	Yes	No	2DRCNN	0.780	0.466
	ICA_50D	Tikhonov	Yes	No	BRAINNETCNN	0.792	0.462

Figure 5.1: [ABIDE] This figure illustrates the prediction performance results of age and autism diagnosis using static FC estimates. The prediction is performed using the best-performing pipelines as identified in Chapter 3: a "*parcellation (ICA), functional connectivity (Tikhonov or SGGGM), tangent space parameterization (yes), and tangent space shrinkage (optional)*" pipeline.

The prediction accuracy for ASD is close to random (random accuracy here is 50%). This is not surprising as many earlier studies cannot predict autism with statistical significance when using data from multiple sites. For example, this review work (Hull et al., 2017) highlights the concern regarding test-retest reliability of machine-learning classifiers when applied to novel ASD data sets. Another work (Nielsen et al., 2013), when used the entire ABIDE consortium data, was not able to achieve more than 55 to 60.0% accuracy for ASD diagnosis. Our best-recommended pipelines also illustrated inconsistent results for autism prediction. In contrast, the same pipelines performed reasonably well for the age prediction task on the same data. In particular, a *"parcellation (ICA), functional connectivity (SGGGM), tangent space parameterization (yes), tangent space shrinkage (yes), and classifier (2DRCNN)"* pipeline can predict age with a correlation of 0.78. This raises the question of whether disruption caused by ASD manifests as a change in FC or not? Our results for ASD diagnosis using FC are not convincing and will discuss this further in Chapter 6.

Diagnosis of Nicotine Addiction

We used the ACPI fMRI data (*ACPI Imaging Protocol*, 2016) from a multimodal treatment study of Attention Deficit Hyperactivity Disorder (ADHD). The data was pre-processed using the C-PAC pipeline as explained in Chapter 2 (Section 2.1.4), and the total number of subjects is 126. We again used the recommended pipelines (shortlisted from results of Chapter 3) and employed different predictors/classifiers to predict whether an individual is experiencing nicotine addiction or not. Figure 5.2 illustrates the recommended pipeline's complete set of results for nicotine addiction diagnosis (and sex prediction) for ACPI participants. Again, it is worth clarifying that predicting nicotine addiction is a challenging task, which is why we added sex prediction as a control measure.

	Brain Parcellation	Functional Connectivity Estimation	Tangent Space Projection	Shrinkage In Tangent Space	Classifier/ Predictor	Prediction Accuracy sex ACPI	Prediction Accuracy Nicotine addiction ACPI
Recommended Pipeline (1)	ICA_50D	SGGGM	Yes	Yes	ELASTIC NET	0.681	0.588
	ICA_50D	SGGGM	Yes	Yes	2DRCNN	0.795	0.641
	ICA_50D	SGGGM	Yes	Yes	BRAINNETCNN	0.789	0.575
Recommended Pipeline (2)	ICA_50D	Tikhonov	Yes	No	ELASTIC NET	0.679	0.549
	ICA_50D	Tikhonov	Yes	No	2DRCNN	0.791	0.648
	ICA_50D	Tikhonov	Yes	No	BRAINNETCNN	0.719	0.614

Figure 5.2: [ACPI] This figure illustrates the prediction performance results of nicotine addiction diagnosis and sex prediction using static FC estimates. The prediction is performed using the best-performing pipelines as identified in Chapter 3: a *"parcellation (ICA), functional connectivity (Tikhonov or SGGGM), tangent space parameterization (yes), and tangent space shrinkage (optional)"* pipeline.

Figure 5.2 shows that we can predict nicotine addiction with an accuracy of 64% using an optimized set of methods proposed in Chapter 3 (i.e. *"parcellation (ICA), functional connectivity (SGGGM), tangent space parameterization (yes), tangent space shrinkage (yes), and classifier (2DRCNN)"* pipeline). This is consistent with the literature as (Fedota & Stein, 2015) found that nicotine exposure decreases the

ability of network communication across the brain, and acute nicotine exposure affects several resting-state networks such as salience, default, and executive control networks.

It has been shown previously that nicotine addiction interacts with neural systems to modulate drug-related circuitry and therefore affects cognition (Volkow, Fowler, Wang, & Goldstein, 2002). There is some existing research that demonstrates that the connectivity measures within the executive control and frontoparietal networks are particularly informative in understanding nicotine addiction (Pariyadath, Stein, & Ross, 2014). (Pariyadath et al., 2014) used the whole-brain FC patterns to predict nicotine addiction with an accuracy of 52.4%. Our results for predicting nicotine addiction (64% accuracy) are arguably more convincing, and we discuss this further in Chapter 6.

5.2.2 Subject's Identifiability using Functional Connectome

(Finn et al., 2015) previously used HCP data from 126 subjects to demonstrate that the FC estimated from one session can be used to uniquely recognize a given person from the group of profiles taken from a following session (both sessions only a day apart). This has been referred to as neuronal finger-printing.

We aimed to carry out a similar study, where we identify individuals from their FC. The main difference between this work and (Finn et al., 2015) is that we used the UK Biobank data (3728 subjects), in which the multiple scanning sessions in UK Biobank are on average two years apart (as opposed to a day apart in (Finn et al., 2015)). The question is that whether an individual's intrinsic functional architecture is reliable enough to be identified even after a few years?

We applied the same procedure as (Finn et al., 2015) and used a set of FC matrices from the first visit for the database, and FC matrices from the second visit acquired

after a few years as the target set. To identify an individual correctly from the target set, we computed the correlations between the query matrix and all the FC matrices in the database. If the predicted identity matched the true identity, this was classified as a hit and was given a score of 1, and if it could not match the true identity, it was classified as a miss and given a score of 0. Based on this analysis, we can compute the percentage of subjects that are correctly identified.

Using the UK Biobank longitudinal data, we were only able to identify 9.03% of subjects (out of 3728 subjects) when we used full correlation as a measure of FC estimation [3]. This was further improved up to 13.11% when we used partial correlation [4] instead of full correlation to measure FC estimation. Lastly, we used a more principled approach to represent FC using Riemannian geometry and projected correlation-based estimates in the tangential space of the Riemannian manifold (as proposed in Chapter 3). This showed a significant performance improvement, and we were able to identify 38.33% of subjects correctly. However, this measure is not as high as the $\sim$ 80% performance reported by (Finn et al., 2015).

One possible explanation is that an individual's intrinsic functional architecture can remain stable across a couple of days but significantly varies after a couple of years. Additional research is needed to properly comprehend the concept of finger-printing across days and months, years, and decades.

5.2.3 Spotlight on the Mental Health

Biomedical research in mental health has long been frustrated by the lack of an apparent mapping between recognized diseases and biological substrates (e.g.,

[3] FC is calculated by estimating the full correlation coefficient of timeseries signal for each subject as explained in Chapter 3 (Section 3.3.2).

[4] Partial correlation is calculated from the inverse of the covariance matrix as described in Chapter 3 (Section 3.3.2).

genes, molecular signatures, imaging biomarkers). For example, there is considerable shared heritability between all psychiatric diseases (Anttila et al., 2018). Whilst it is acknowledged that the traditional diagnostic boundaries are imperfect, it seems likely that the difficulty in mapping between biology and disease is driven by individual variability; i.e., different individuals have different biological underpinnings that have either caused or increased their vulnerability to a particular disease or disorder.

This represents a significant challenge for the development of new therapies. Clinical trials of novel drugs are doomed to modest results (at best) if the drug is hitting the molecular target for only a proportion of trial participants. This is consistent with evidence from clinical practice, where there is a "trial and error" approach for matching the right therapy to each patient; and it is also consistent with evidence from clinical research, where associations between diagnosis groups with genes, molecules, imaging measurements are often marginal (in terms of statistical significance), non-specific, and hard to replicate.

The problems associated with the traditional categorical approach to mental health diagnoses are long acknowledged. For example, the National Institute of Mental Health's (NIMH) Research Domain Criteria (RDoC) initiative seeks to develop an alternative classification system for psychopathology involving dimensional conceptions of clinical phenomena (negative valence, positive valence, cognitive systems, social processes and arousal/regulatory systems) referenced to multimodal psychobiological constructs (genes, molecules, cells, circuits, physiology, behaviour, self-report and paradigms). While this initiative potentially gets us closer to biology, the jury is still out in terms of how RDoC can be applied for clinical trials or in the clinic. In particular, the lack of mapping between traditional diagnoses and RDoC concepts and the many ways in which RDoC elements can be operationalized both create challenges for the emergence of a coherent noso-

logical system for psychopathology (Patrick & Hajcak, 2016). Nevertheless, the RDoC concept of using complex phenotypes to provide a *prism* through which individuals with distinct *biotypes* can be discovered remains promising for stratifying patients in the context of experimental medicine and clinical trials.

Here, we apply the RDoC approach to the mapping between biology and diagnosis of mood disorder by taking advantage of the large N of the UK Biobank, which represents an unparalleled resource for exploring complex phenotypes associated with diseases and mental health disorders. We report here on the UK Biobank participants who have undergone neuroimaging and completed the mood disorder touchscreen assessment at baseline (2006-10) and when they attend their imaging visit (2014-18).

Mood Measures

Mental health has been assessed in several ways within the UK biobank. The original 500K individuals answered a small number of MH questions (N=6), as well as a 12-item neuroticism questionnaire as part of their baseline 'touchscreen' assessment. These were augmented partway through the baselining timeframe, with the last 172K individuals responding to an additional 17 questions relating to mood disorder (D. J. Smith et al., 2013). 157K individuals completed a more wide-ranging mental health assessment via an online follow-up questionnaire; but there is a relatively small overlap between these individuals and those who have undergone imaging, so these data were not analyzed here. The baseline and mood disorder questionnaires were repeated at the imaging visit and are therefore available for (almost) all participants in whom we have brain imaging data.

Classification of probable mood disorder

(D. J. Smith et al., 2013) defined an algorithm to categorize individuals who have undergone the mood disorder questionnaire as having a single probable episode of major depression, recurrent major depression (moderate and severe), bipolar disorder (types I and II), or no history of mood disorder. This questionnaire formed part of the 'touchscreen assessment' for 172K of the original 500K UKB participants and was repeated at the imaging visit. We reproduced the algorithm, which is available here (*MentalStatesDerivation*, 2018) for the data collected at the imaging visit to characterize probable mood disorder within the 21K individuals. Reproducing the algorithm described by (D. J. Smith et al., 2013) in the 21K imaging participants produced 1033 participants with probable single episode depression, 2791 participants with probable recurrent depression, 108 participants with probable bipolar disorder, and 15267 participants with no evidence of mood disorder.

UK Biobank provided variables

1. **Mood swings**
2. **Miserableness**
3. **Irritability**
4. **Sensitivity / hurt feelings**
5. **Fed-up feelings**
6. **Nervous feelings**
7. **Worrier / anxious feelings**
8. **Tense / 'highly strung**
9. **Worry too long after embarrassment**
10. **Suffer from 'nerves'**
11. **Loneliness, isolation**
12. **Guilty feelings**
13. Risk taking
14. Happiness
15. Work/job satisfaction
16. Health satisfaction
17. Family relationship satisfaction
18. Friendships satisfaction
19. Financial situation satisfaction
20. Frequency of depressed mood in last 2 weeks
21. Frequency of unenthusiasm in last 2 weeks
22. Frequency of tenseness in last 2 weeks
23. Frequency of tiredness / lethargy in last 2 weeks
24. Seen doctor (GP) for nerves, anxiety, tension or depression
25. Seen a psychiatrist for nerves, anxiety, tension or depression
26. Ever depressed for a whole week
27. Longest period of depression
28. Number of depression episodes
29. Unenthusiastic/disinterested for a whole week
30. Longest period of unenthusiasm/ disinterest
31. Number of unenthusiastic/ disinterested episodes
32. Ever manic/hyper for 2 days
33. Highly irritable for 2 days
34. Longest manic/irritable episode
35. Severity of manic/irritable episodes

Derived Variables

36. Bipolar and major depression status

The score ranges from 0-5:
0 = No Bipolar or Depression
1 = Single Probable major depression episode
2 = Probable Recurrent major depression (moderate)
3 = Probable Recurrent major depression (severe)
4 = Bipolar I Disorder
5 = Bipolar II Disorder

37. Neuroticism score

The score ranges from 1 – 12, corresponding to number of distinct neurotic traits present

The score depends on number of 'yes' answers to questions from 1 to 12 (highlighted in left block), like 'Do you have mood Swing' or 'Do you feel lonely'

Figure 5.3: The reported variables for mental health diagnosis in UK Biobank. We reproduced the algorithm proposed by (D. J. Smith et al., 2013) to calculate derived variables (e.g., depression and neuroticism) from a set of provided variables.

Pairwise associations between brain imaging and mental health variables

We have two categories of variables; one consists of brain imaging variables (including the static FC and amplitude information, derived from methods proposed in Chapter 3). The other category comprises mental health variables. We pairwise correlated all imaging variables to all mental health variables and calculated 92000 pairwise correlations. We corrected for multiple comparison corrections using false discovery rate (FDR) and Bonferroni correction (considered more conservative than FDR) to find the significant associations. Out of these 95000 correlations, 4704 correlations survived the FDR ($p < 5.83 * 10^{-4}$) and only 234 correlations passed the Bonferroni correction ($p < 3.26 * 10^{-9}$).

It is to be noted that the main confounds are age, sex, age-sex interactions, head-size, head-motion, squared, and quantile-normalized versions of some of the above. However, it is debatable what to list as a confound, and it highly depends on the study's objectives. Some studies are interested in seeing the effect of age on psychiatric illness, so we provided an additional configuration in which we did not remove age-related confounds from the data.

In Figure 5.4, we showed the pairwise associations between brain imaging and mental health variables. As shown in Figure 5.4, amongst all the imaging variables, the edges of the static FC matrix are the ones most significantly correlated with the mental health variables. For example, risk-taking behaviour is linked with the connection of the precentral gyrus (primary motor cortex) to the postcentral gyrus (primary somatosensory cortex). Also, the status of major depression is associated with the primary auditory cortex connection to the supramarginal gyrus and angular gyrus. Lastly, neurotic behaviour is related to the primary somatosensory and primary motor cortex connection to the superior frontal gyrus and para cingulate gyrus.

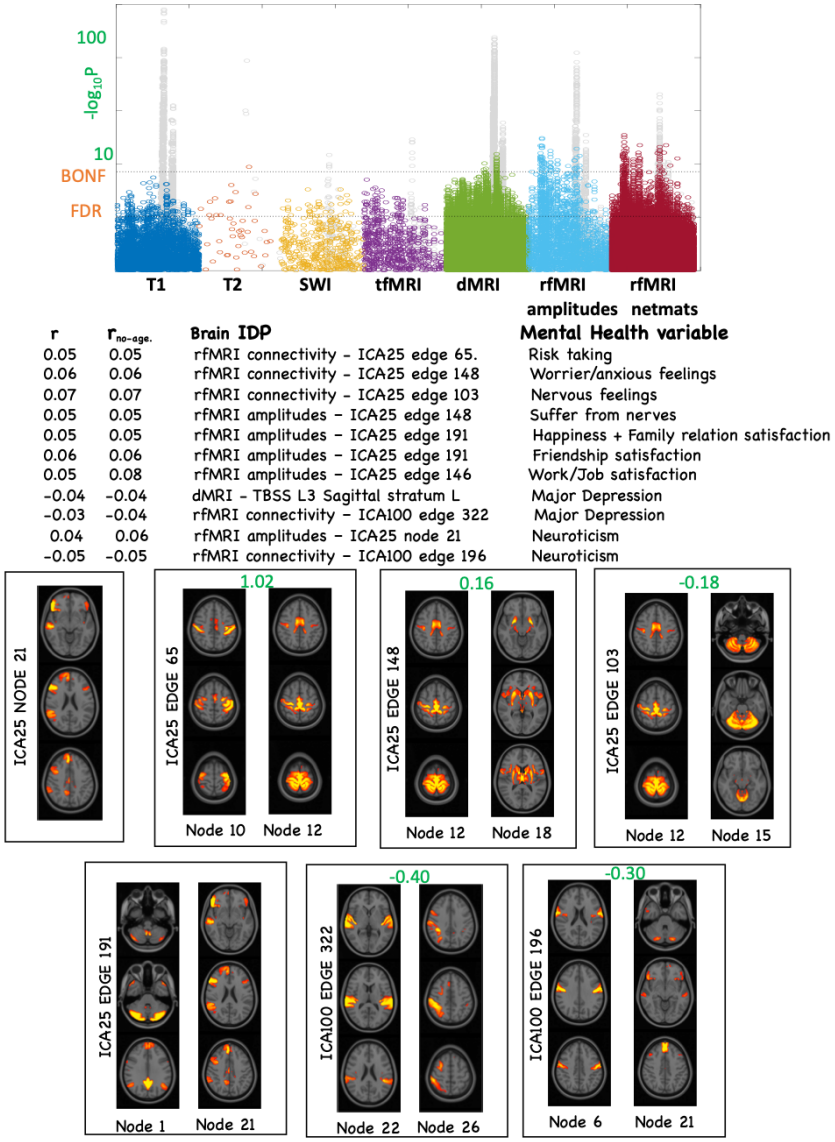


Figure 5.4: [UK Biobank] Manhattan plot illustrates the pairwise correlations between the imaging variables and the mental health variables. Out of the 95000 correlations, only 234 passed the Bonferroni correction; most of them belonged to the resting-state fMRI amplitudes and static FC variables. Lastly, we showed these correlations with (shown in color in top-figure) and without adjusting for age (shown in grey in top-figure), and it is observed that in most cases, adjusting for age decreases the strength of association.

Predicting Personalized Mental Health Status using Static Functional Connectivity

The individual (univariate) associations between imaging and mental health variables in Figure 5.4 provided valuable insight into each variable's contribution, but compromised the predictive power. We are further interested in multivariate associations, and we employed the recommended pipeline from Chapter 3 for the prediction of depression and neuroticism (we only employed Elastic Net as a classifier for the sake of simplicity). The prediction results are illustrated in Figure 5.5.

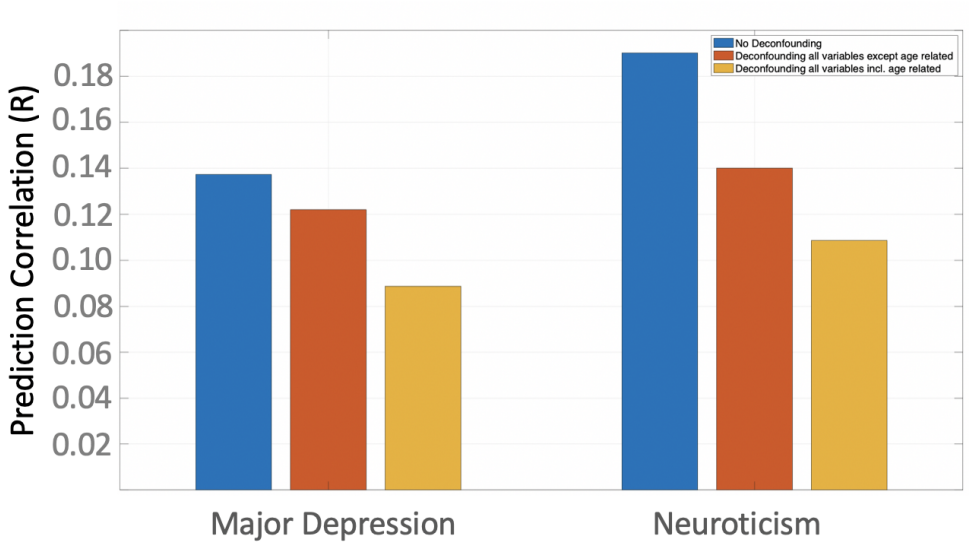


Figure 5.5: [UK Biobank] We used FC estimates to predict mental health variables. The results provided here are based on nested ten-fold cross-validation and illustrate prediction correlation (r) of neuroticism and depression. Moreover, we considered the effects of confounds, and results are presented for three configurations; 1) ignoring the effect of confounds variables, 2) considering all the confounds except age-related, 3) removing the effect of all confounds, including age-related.

As demonstrated in Figure 5.5, the variance explained for neuroticism and depression is around 4% before deconfounding and 1% after deconfounding (reported

results are statistically significant with $p < 0.0001$). The rationale for this low variance is that, as with all large-scale epidemiological cohorts, the data about mental health in the UK Biobank is relatively limited. We were thus unable to explore associations with, for example, age of onset, course of mood symptoms, childhood adversity, etc. Cohorts that include this rich information plus imaging do exist (e.g., (Filippini et al., 2014)), but the size of these cohorts limits the application of the sorts of analyses described here. It is also important to note that the age range of the UK Biobank is not ideal for studying the course of disorders that typically manifest around adolescence/early adulthood.

5.3 Contributions of Time-varying Functional Connectivity to Human Behaviour

In Chapter 4, we proposed MAGE for the reliable estimation of TVFC. We investigate two main applications of TVFC estimated using MAGE. First, we employ MAGE to understand distinct task-related changes in the FC. Second, we use MAGE to understand the fluctuations in FC caused by a spontaneous cognitive process, not imposed by any external stimulus.

5.3.1 Ongoing Task-related Cognition

In this section, we examine within-subject, task-related changes in patterns of FC. For this we use the cognitive and movies-related task fMRI datasets (as introduced in Chapter 2).

Time-varying Functional Connectivity can Show Appropriate Task Dependencies

In order to understand time-related changes in FC by externally imposed tasks, we applied MAGE to cognitive-task fMRI data. The idea is to train MAGE with no knowledge of the task timings (i.e., unsupervised), allowing us to see if the state dynamics being inferred by MAGE are meaningful in the sense that they exhibit dynamics that link to the task. Furthermore, we wanted to compare how the FC varies across the task conditions (and judge whether the inferred spatial maps are meaningful).

This task fMRI data used was collected as part of a previous study as explained in Chapter 2 (Section 2.2.1). Participants performed four different tasks: rest, motor only, visual only, and simultaneous (but independent) visual and motor tasks, and each task was repeated four different times. We applied PROFUMO (Harrison et al., 2015) to identify 30 parcels and their time courses; note that PROFUMO explicitly models between-subject variability in the spatial location and extent of each parcel. These 30 parcel time courses were then fed into MAGE as N=30 channels of data.

Temporal Dynamics In the main results shown in Figure 5.6, we modelled four states using MAGE, so that the number of states is matched to the number of known task conditions. In Figure 5.6[A], we have plotted the task-evoked occupancy, which is the average of the MAGE inferred state time courses across the four repeated runs (as each experimental condition is repeated four times). The task-evoked occupancy shows clearly that the inferred states switch at the same time as the experimental task condition changes. It is worth noting from Figure 5.6[A] that the estimated FC-driven state-time courses, β_{tq}, performed better than the mean activity/variance-driven time-courses, α_{tp}. Moreover, in Fig-

ure S34, we have illustrated the results of task decoding when we modelled more than 4 states. In Figure S34[A], we have modelled 8 states and in Figure S34[B], we have modelled 16 states, and we plotted the relative fractional occupancy (time spent in each state) distribution for each of the modelled states. It is evident that even with the increased number of states, we can still see dynamics that are task dependent and can identify task-specific states (as highlighted by the paired t-test values on the respective states).

Spatial Dynamics In Figure 5.6[B], we show the state-specific FC spatial maps (corresponding to the the first eigenvector of the state-specific correlation matrices) and mean activity spatial maps for each of the four inferred states. The maps are labelled according to the task conditions that their state time courses most correspond to. The state-specific FC maps for MAGE identifies appropriate active regions during each of the tasks. It can be seen during the motor task that the somatomotor area of the brain is active, and during the visual task that the visual cortex is active. During the rest task, the DMN is active, and during the visual-motor task, both the motor and visual cortex are active. The mean activity level maps appear similar and look appropriate with regards to the relevant task condition. During the visual task, the visual cortex of the brain is active, and during the motor task, the motor cortex is active for both approaches. During the rest task, DMN is active, and during the visual-motor task, only the visual cortex is active for the MAGE approach.

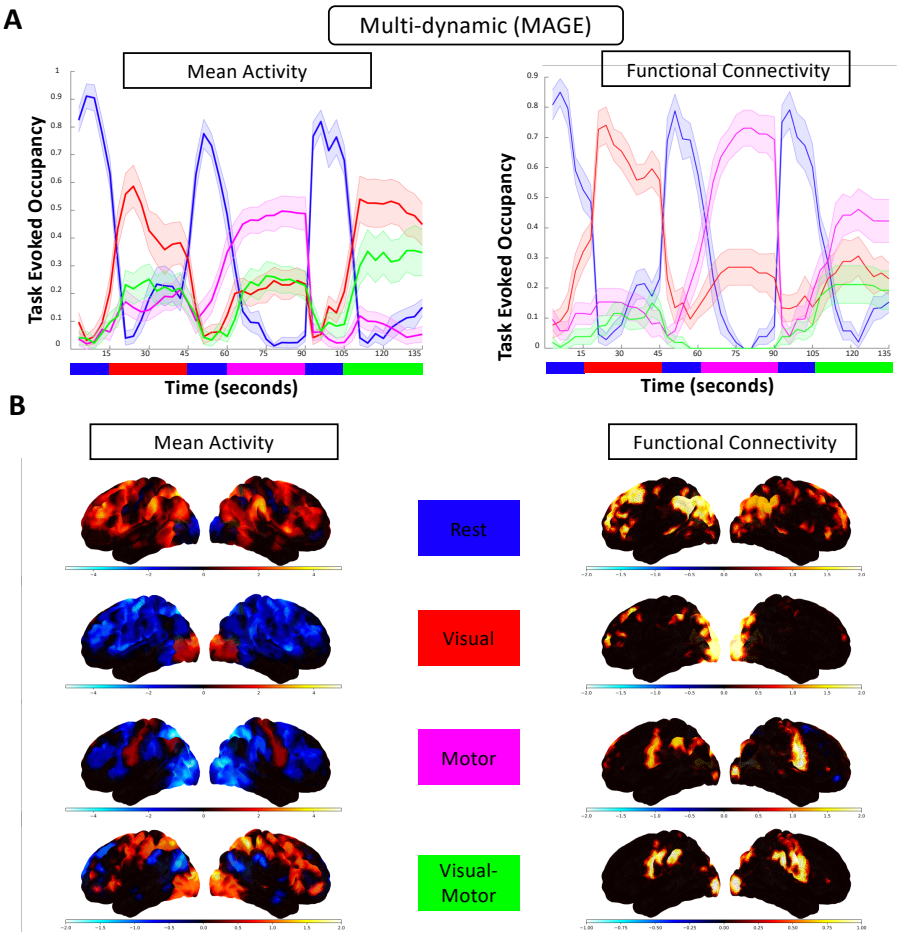


Figure 5.6: [Cognitive-task fMRI data] Time-varying functional connectivity can show appropriate task dependencies. **[A]** There are four different tasks performed by the participants (rest, visual, motor, and visual-motor). Task-evoked occupancy plots of the state time courses from MAGE (learnt with no knowledge of the task timings) across the four tasks, where the x-axis represents the number of time points and the y-axis corresponds to the average activation value for that state. Specifically, the task-evoked plot curves show the mean and standard deviation across four repeated runs of each experimental condition for all 15 subjects. **[B]** The state-specific spatial maps for the functional connectivity (FC) and mean activity, which are labelled according to the task conditions that their state time courses most correspond to. The state-specific spatial maps for variance are not illustrated as there is no between-state variability in the inferred variance states. For the mean activity maps, the average amplitude of each node/channel is shown. The negative mean activity (e.g., deactivation of the visual network during the motor task) refers to decreases in signal change compared to the baseline. For the FC maps, rank-one decompositions (first Eigenvector) of the estimated state-specific FC correlation matrices are shown. The magnitude of the values of the state-specific FC spatial maps (colorbars) correspond to the degree of connectedness of each region with the rest of brain.

The results in Figure 5.6 demonstrates that both the mean activity and FC vary during task conditions, and these variations can robustly decode the imposed task conditions.

Measuring Shared Responses across Subjects using Time-varying Functional Connectivity

Our goal is to predict fluctuations in brain connectivity during complex, dynamic narrative comprehension tasks. We hypothesize that participants' brain connectivity would have a certain level of similarity when involved in similar realistic scenarios (Nastase, Gazzola, Hasson, & Keysers, 2019).

Considering that, we fitted MAGE to the fMRI data from 44 human subjects (male and female) who were presented with 16 three-minute stories (as explained in detail in Chapter 2 (Section 2.2.2)). We focused on different ROIs within the DMN, e.g., parahippocampal cortex, medial prefrontal cortex, posterior medial cortex, and superior frontal gyrus. We applied MAGE individually to these different ROIs and inferred the state time-courses of subjects. This analysis was performed individually on each ROI to ensure that the reported results were in line with previous studies (Baldassano et al., 2018).

We sought to investigate whether the participant's response to these movies was uniform and consistent to a certain extent or not (when participants are watching the same movies). For any two similar stories, we expect the individuals' brain FC to be comparatively more correlated than individuals' brain FC when watching two different movies. For example, if there are two subjects (A and B), we expect their brain FC is more correlated when both subjects A and B are watching the same movie as compared to when subjects A and B are watching two different movies.

Our analysis found that participants' FC-state dynamics (inferred using the multi-dynamic approach MAGE) is consistently more correlated when watching similar movies and clips. These patterns are highly reliable and robust across different ROIs within the DMN. Our findings aligned with (Baldassano et al., 2018) in which it was demonstrated that some ROIs, particularly the posterior medial cortex and medial prefrontal cortex, exhibit robust movie-related patterns that can be identified without explicit ground truth.

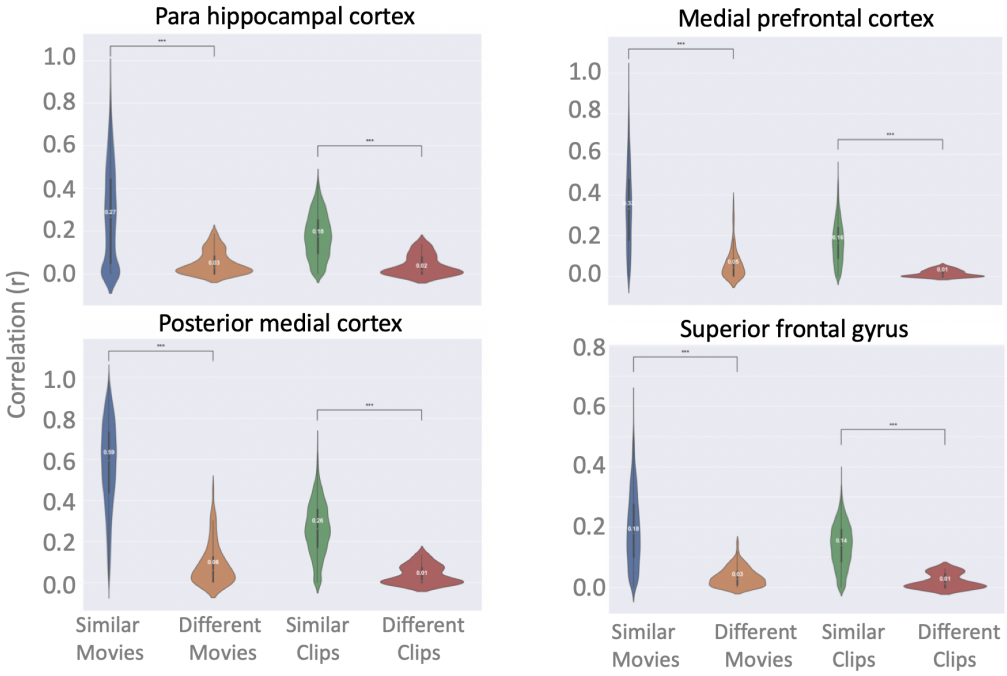


Figure 5.7: [Movies-task fMRI data] We measured shared responses across subjects for movies-task fMRI data. We found that participants, when watched similar movies showed significantly higher event similarity in the FC-state dynamics (inferred using the multi-dynamic approach MAGE) than when they watched different movies.

Figure 5.7 revealed a network of brain regions that are sensitive to high-level details in dynamic naturalistic scenarios. Specifically, it demonstrates that the FC temporal dynamics are consistent across subjects when involved in similar tasks and activities.

5.3.2 Ongoing Spontaneous Cognition

Our results in Chapter 4 demonstrated that there are more significant changes in FC across time using our proposed multi-dynamic approach, MAGE, as compared to existing single-dynamic methods. These results are important as they demonstrated that resting-state FC is not homogeneous across time.

This section aims to specifically investigate the FC-driven state time courses estimated by MAGE, to see how well summary measures of the temporal dynamics can be used to predict behavioural traits (and other non-imaging variables). The summary measure we used was the fractional occupancy (time spent in the different brain states) of the FC-driven states, and we applied Bayesian partial least square (Vidaurre, van Gerven, Bielza, Larrañaga, & Heskes, 2013) combined with permutation testing to predict the behavioural variables.

We applied MAGE on the resting fMRI data from the UK Biobank dataset as described in Chapter 4 (Section 4.4.5) [5]. We divided the behavioral data of UK biobank participants into the following categories: Early life factors, lifestyle and environment (exercise and work), physical measures (general), physical measures (bone density and sizes), physical measures (cardiac and blood vessels), hearing test, blood assays, cognitive phenotypes, health and medical history, genetic measures and mental health self-reported variables. Figure 5.8 shows the variance explained in the non-imaging variables using the MAGE-estimated fractional occupancy.

[5] UK Biobank pre-processing pipeline is described in Chapter 2 (Section 2.1.1)

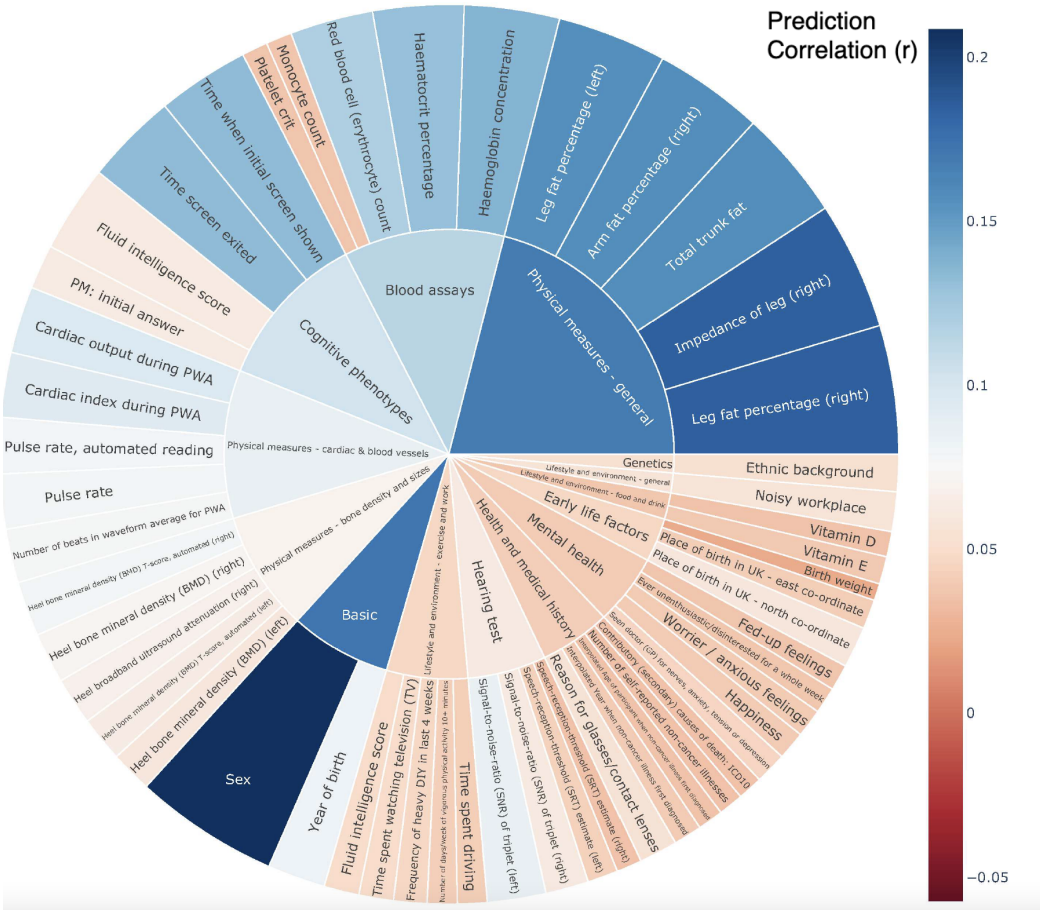


Figure 5.8: [UK Biobank] Inter-individual differences for the time spent in each FC-driven state (fractional occupancy) inferred using the multi-dynamic approach MAGE during resting state fMRI are predictive of behavioral traits. We have illustrated the amount of correlation explained in the behavioural variables using each state's fractional occupancy. The area of the sub-blocks is colour-coded according to their prediction performance.

Figure 5.8 highlights that the brain network dynamics are subject-specific and significantly correlates with cognitive functioning. Specifically, it shows that the differences in FC temporal dynamics can explain individual differences in behaviour (e.g., fluid intelligence score).

5.3.3 Age-related Changes in Resting-state Functional Connectivity

Some recent studies have shown that age-related changes in FC are widely associated with cognitive functioning. The rationale is that the decline in cognitive function associated with aging is because of changes in communication between different brain regions (Zuo et al., 2013). In this section, we aimed to study the whole-brain FC patterns in distinct adult age groups.

We divided the UK Biobank population (imaging and pre-processing protocol is described in Chapter 2 (Section 2.1.1)) into two distinct age groups; first is the younger adult group with participants ages 40 to 45 years, and the other is the older adult group with participants ages 65 to 70 years. We then applied MAGE (similar to pipeline described in Chapter 4 (Section 4.4.5)) to infer TVFC estimates for these two age groups. Some studies that have compared the connectivity between younger adults and older adults have found there exists decreased within-network connections in different RSNs (Grady, Sarraf, Saverino, & Campbell, 2016). Most of these studies were focused on changes within the DMN, but some work has shown that other RSNs such as the sensorimotor or the visual network also exhibited age-related changes (Onoda, Ishihara, & Yamaguchi, 2012).

We kept the objectives of this work simple and solely intended to compare changes in within-network/state connectivity patterns in younger versus older age groups using state-specific network patterns inferred using MAGE. We investigated several RSNs, i.e., DMN, visual network, somatosensory network, frontoparietal network, dorsal and ventral attention networks. In Figure 5.9, we compared the changes in FC within these networks. We found an increase in within-network connectivity in visual, DMN, and frontoparietal networks and a decrease in within-network connectivity in somatosensory, dorsal, and ventral attention networks.

These findings are in line with some of the conclusions drawn from a recent review by (Damoiseaux, 2017).

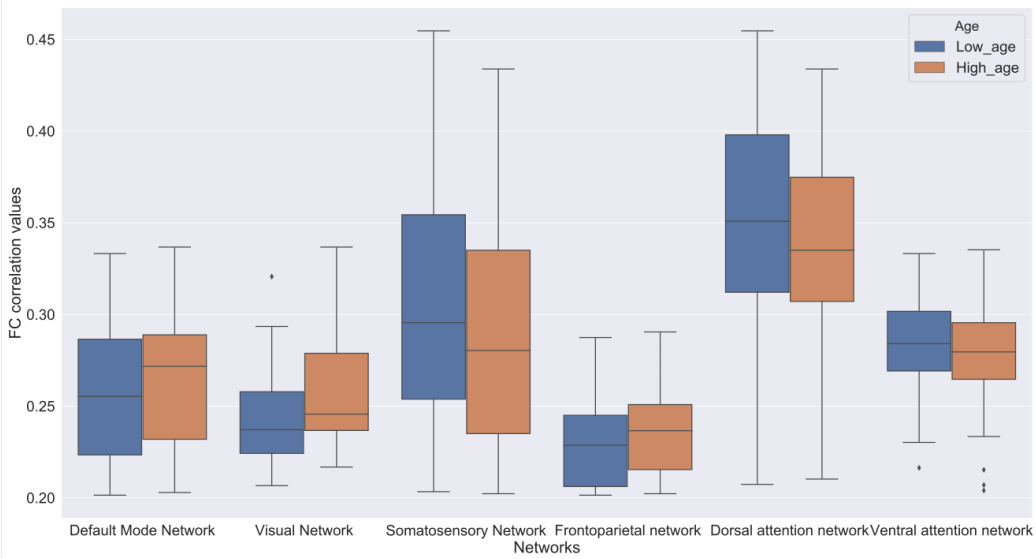


Figure 5.9: [UK Biobank] Age-related changes in the resting-state fMRI data using MAGE estimated time-varying functional connectivity. We compared connectivity patterns in younger versus older age groups and found an increase in within-network connectivity in visual, DMN, and frontoparietal networks and a decrease in within-network connectivity in somatosensory, dorsal, and ventral attention networks in older age groups.

The results in Figure 5.9 show that aging has a significant impact on the within-network connectivity patterns and may provide an insight into the mechanism by which aging can contribute to cognitive decline.

5.4 Static FC is Worse than Time-varying FC at Predicting Behaviour

Static FC can be considered as an average of all state-specific FC states (and the time-varying means) and is conventionally estimated simply by taking a full (Pearson) correlation between brain parcels over the whole scanning time. Unlike TVFC in the form of the state-specific FCs, static FC is a single FC estimate per subject and is already an established predictor of behavioural traits. In Figure 5.10, we compare the combined power of the state-specific FCs [6] versus static FC [7]. This shows that for the mental health and medical-history related variables, predictive performance is improved by using the state-specific FCs in comparison to static FC. We employed Elastic Net for this comparison and the reported results are nested ten-fold cross-validated.

Figure 5.10 compared the predictive power of MAGE-derived TVFC (Chapter 4) versus static FC (using the optimised approach from Chapter 3), and demonstrated the prediction performance on mental-health and medical-history related variables in UK Biobank data. We have demonstrated that the prediction accuracy of these behavioural variables increased when TVFC was used, as compared to static FC e.g., for work/job satisfaction, prediction accuracy was improved by 20.2 % and for depressed for a week, accuracy was improved by 19.5 %, etc.

[6] MAGE-derived state-specific FCs were estimated on the UKB data with the pipeline described in Chapter 4 (Section 4.4.5).

[7] Static FC was calculated on the UK Biobank data by using the proposed pipeline in Chapter 3: "*parcellation (ICA), functional connectivity (Tikhonov), tangent space parameterization (yes)*".

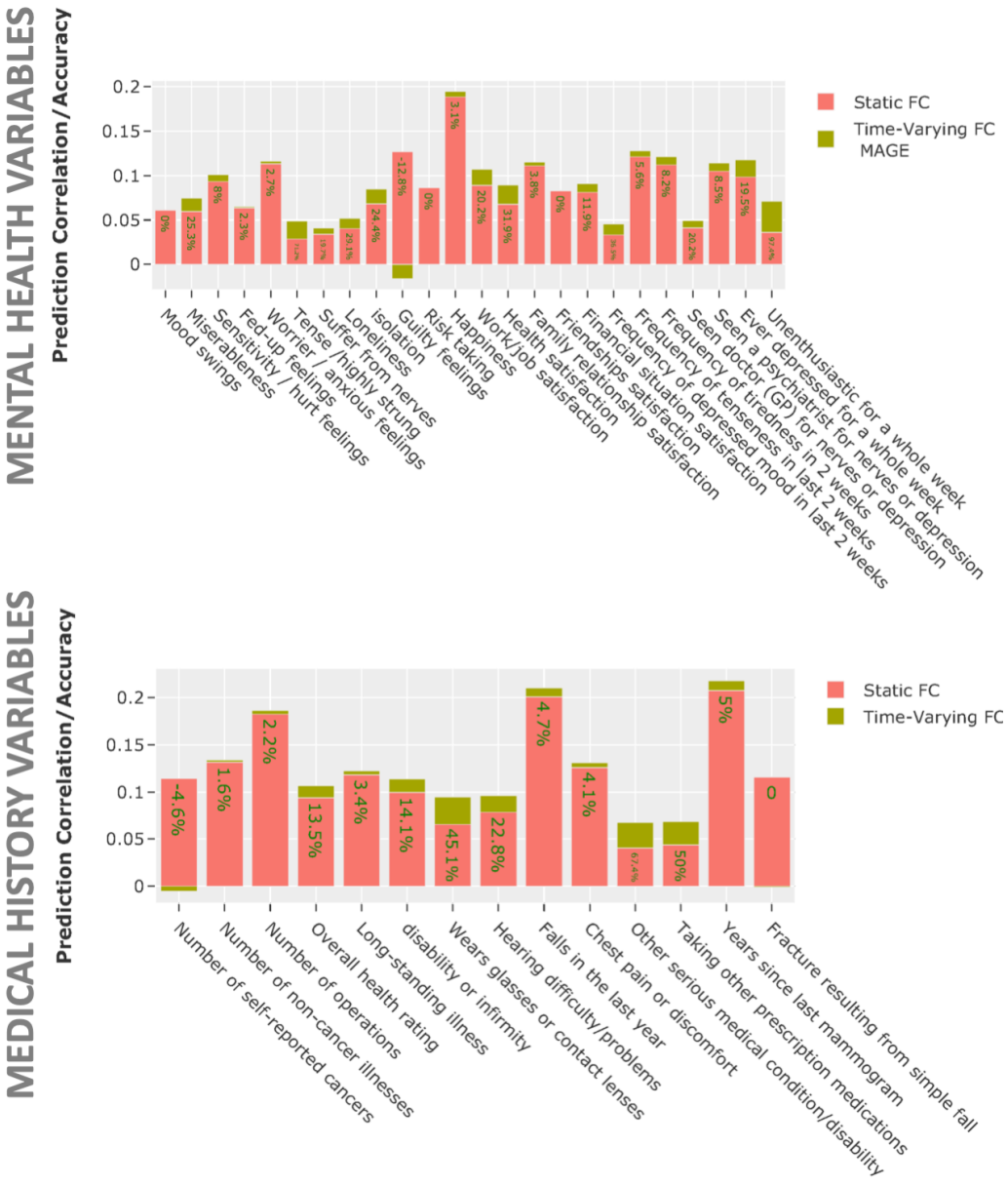


Figure 5.10: **[UK Biobank]** Comparison of the prediction performance of MAGE-derived TVFC with the static FC (using the optimised pipeline from Chapter 3). The bars in red colour show the prediction correlation/accuracy using static FC estimates, whereas bars in green show the increased/decreased prediction correlation/accuracy using TVFC estimates.

This highlights that there are some aspects of behaviour that can be best explained

using TVFC estimates. The fact that the TVFC is more significantly correlated to behavioural variability as compared to static FC highlights that it might be capturing transient neuronal communication fluctuating around a stable neural architecture (Vidaurre et al., 2021).

5.5 Discussion

Static FC: In Section 5.2, we explored the relevance of static FC in explaining individual differences in health and disease. Specifically, we looked into identifying neurological diseases using static FC as a clinical biomarker (Section 5.2.1). We also highlighted the significance of static FC patterns in understanding mood disorders (Section 5.2.3). Lastly, we demonstrated that each individual has a unique FC fingerprint, and FC fingerprints can remain intact (to a certain degree) even after a few years (Section 5.2.2).

Time-varying FC: In Section 5.3, we investigated the utility of MAGE-derived TVFC estimates in understanding within-subject task-related and spontaneous changes in FC. Specifically, we demonstrated that the MAGE-derived TVFC estimates can show appropriate task dependencies when subjects are involved in different cognitive or narrative comprehension tasks (Section 5.3.1). We also looked into FC fluctuations caused by spontaneous cognitive processes and demonstrated that FC temporal dynamics are subject-specific and can explain individuals' behaviour variability to some extent (Section 5.3.2). Lastly, we demonstrated that MAGE estimated spatial network features (i.e., the state-specific FCs) can decode age-related changes in brain connectivity (Section 5.3.3).

Temporal Summary Measures of Time-varying FC: In Section 5.3 and Section 5.4, we mainly focused on the state-specific FCs (spatial network features) as

our main measure of TVFC (e.g., for predicting behavioural traits in Figure 5.10). However, a complete representation of TVFC includes the FC state time courses alongside the state-specific FCs (after all, we combine the state-specific FCs with the FC state time courses to get the instantaneous FC [8]). As such, the prediction may be improved further in the future by including temporal features describing the FC state dynamics. We can already show that inter-individual differences for the time spent in each FC state are linked to behavioural traits, as illustrated in Section 5.3.2 (Figure 5.8). There are other rich temporal features captured by MAGE-estimated TVFC [9], e.g., lifetimes of FC states, interval times of FC states, probability of transition from one FC state to another, etc.). We never fully utilize these other temporal summary measures (i.e., lifetimes, interval times) in our work. However, one suitable approach could be to combine these different *spatio-temporal* measures (e.g., state-specific FCs, fractional occupancy, lifetimes, etc.) to better understand the intraindividual behavioural variability.

Chapter 6 further discusses in detail the achievements and limitations of the presented clinical applications of both static and time-varying FC.

[8] This is shown by Equation S9 in Chapter 4

[9] Detailed description of some of the TVFC summary measures can be found in (*TVFC Summary Measures*, 2017).

CHAPTER 6

GENERAL DISCUSSION AND CONCLUSION

The contributions of the thesis are two ways, one is methodological advancement, and the other is clinical implications. In terms of methods development, we proposed frameworks for estimating both static and time-varying FC. In terms of clinical relevance, we demonstrated the utility of FC in relation to cognitive abilities, both in health and disease. In the sections below, we summarise and discuss both the methodological and clinical contributions of this work.

6.0.1 Methodological Contributions

In Chapter 3, we proposed a comprehensive methodology for designing static FC-based predictive models. We can summarise the significance of Chapter 3 in a few key points:

1. There has been previously no single widely accepted standard pipeline for analyzing FC.

2. We used resting-state fMRI data from four cohorts and ran more than 9000

different pipeline variants on 15000 individuals to determine the optimum pipeline.

3. We proposed six steps for designing connectivity-based predictive models (i.e., predicting non-imaging variables, e.g., behavioural traits, demographics, emotion measures, etc.)

4. The standardized set of methods we proposed, can help accelerate scientific progress within the neuroscientific community.

The framework presented in Chapter 3 considered several central methodological issues when estimating and using static FC. It attempted to establish the optimum and reliable pipeline, paving the way for future studies to accurately analyze the relationship between whole-brain static FC patterns and behavioural traits.

Though the presented work in Chapter 3 is significant, it assumes that the FC is static across time. We were aware of the limitations of that, and we shifted our work focus to studying the time-varying nature of FC in Chapter 4. Considering the limitations in current methods that measure TVFC, we proposed a new model, MAGE, that revealed substantial fluctuations in FC over time, both in task-negative and task-positive states. We can summarise the significance of Chapter 4 in a few key points:

1. MAGE is a multi-dynamic model that models temporal fluctuations in FC independently from fluctuations in the mean of the activity.

2. MAGE revealed more significant changes in FC over time than single-dynamic approaches, such as sliding window correlations.

3. Multi-dynamic modelling provided an explanation and a solution as to why resting fMRI FC has previously looked so stable.

4. MAGE modelled fMRI data as a set of reoccurring brain states, and impor-

tantly, these states do not have to be binary and mutually exclusive (e.g., multiple states can be active at one time-point).

Our presented work in Chapter 4 allowed us to draw several significant conclusions regarding the nature of TVFC in fMRI; the most important one was that we learned MAGE-derived TVFC fluctuates on a sub-minute time-scale and shows strong changes over time during resting fMRI.

6.0.2 Achievements and Limitations in Clinical Applications

Following the development of methods for estimation of both static and time-varying FC, we endeavoured to investigate whether it is probable to use FC as a clinical biomarker, for example, as a predictive tool for cognitive functioning or as a diagnostic tool for neurological and psychiatric disorders.

We employed static FC to predict ASD on data from multiple sites using an optimized pipeline (as recommended in Chapter 3). The results we obtained were not statistically significant for ASD prediction. We are confident that it is not because of methodological limitations as we can predict the age and sex with high accuracy for the same participants. This raises the question of whether there is indeed an alteration in resting-fMRI-derived static FC patterns for subjects with ASD? Studies have shown that ASD is associated with functional and structural deficits in the brain (Hollander et al., 2005; Long, Duan, Mantini, & Chen, 2016) but it remains unclear whether such changes can manifest themselves in the patterns of resting fMRI static FC. One thing that should be indisputable is that static FC solely cannot be predictive of all neurological disorders. Acknowledging that does not negate the potential utility of FC, but helps us to focus on and identify areas of genuine application.

Next, we employed static FC to diagnose subjects experiencing nicotine addic-

tion. Research from preceding studies suggests that nicotine addiction can affect cognitive abilities. We used the same optimized framework (as recommended in Chapter 3) to identify nicotine-addicted subjects. Our results showed convincingly that we can statistically significantly predict subjects suffering from drug abuse.

Similarly, we explored associations between mood characteristics and FC in the UK Biobank. We found several individual associations that met statistical significance, consistent with previous studies using much smaller study populations (Rashid, Damaraju, Pearlson, & Calhoun, 2014; Gruber, Rogowska, & Yurgelun-Todd, 2004). We could also identify subjects with different psychiatric disorders (such as major depression) by taking advantage of the new methodology. These results increase our confidence in the true diagnostic power of static FC. Though we must admit that there is still a long way to go before deploying such methods in clinics, and will only be possible if the proposed methods are reproducible and generalizable across different datasets.

As we already have discussed in Chapter 4, FC can be shown to be fluctuating at a fast time scale, and there exists rich spatiotemporal information in the dynamics of these brain networks. We deployed MAGE, an optimized algorithm to model TVFC (as proposed in Chapter 4) to understand task-induced FC patterns. We demonstrated in Chapter 5 that we could use the TVFC to predict which task a person is engaged in. This is crucial as some former studies argued that FC is sustained during different tasks (Kieliba et al., 2019; Gratton et al., 2018), but we clearly showed that FC varies during different tasks [1].

Another notable result we demonstrated is that FC is not homogeneous over time during resting-state fMRI. We showed that FC largely transitions between sen-

[1]This is inline with some other studies (Di et al., 2015) that also reported reliable task-related FC changes

sory, motor, control, attention, and higher-order default mode networks during wakeful rest. Lastly, our work demonstrated that the time spent in these fluctuating FC states is linked to behavioural traits. This highlights that both the spatial connectivity and temporal dynamics of these spatial connectivity states encode information related to a person's behaviour and lifestyle. Such rich temporal information (e.g., time spent in these different FC states, switching frequency of FC states, probability of transition from one FC state to another, etc.) is not available by solely estimating static FC.

On the debate of static versus time-varying FC, our results recommended that some aspects of mental health, physical well-being, lifestyle, and behaviour can only be best explained using TVFC estimates. We believe that more extended usage of TVFC can significantly expand our knowledge of brain dynamics.

6.0.3 Reproducibility and Generalizability of Proposed Machine Learning Algorithms

It is essential to talk about the limitations of this work. For the scope of this work, it was essential to benchmark the performance of both static and time-varying FC based predictive modelling pipelines, without being fixated on improving prediction performance for one particular variable of interest or one specific data-set. For example, by designing MAGE to be a generative model, unsupervised learning approach - i.e. MAGE is inferred with *no* knowledge of the non-imaging variables and has the ability to adapt its generative model to different datasets - we expect the approach is able to better generalize across any set of non-imaging variables on any fMRI data without the need for fine-tuning. For example, if the goal of this work was only to predict autism in a small study, unintentionally, we could have tuned MAGE's assumptions to specifically enhance its performance

for ASD diagnosis. A downside of this approach is that some of our prediction results are not as high as previously reported. We believe that it is advantageous as it allows the methods to be applicable to a wider variety of data and tasks.

This work has leaned heavily on machine learning, specifically deep learning, which has numerous benefits, such as performing tasks with a level of correctness that was not achievable previously. However, there are a few considerations that are worth mentioning. For example, deploying these models in the healthcare system would need time and resources, and it is no secret that these type of algorithms can be highly susceptible to errors and require continuous monitoring in real-world heath-care settings.

6.0.4 Future of Functional Connectivity

Regarding the future of FC, we envision a massive potential for growth in terms of its association to cognition, both in health and disease. As *Stephen Smith* said in the *The future of FMRI connectivity* (S. M. Smith, 2012), "we will see FMRI connectivity become not just a powerful clinical marker, but a tool for investigating disease mechanism." This thesis provides an improved set of methods, but clearly the methods development will not stop here. Future work will take advantage of ideas from this work and will explore more incredible applications.

We understand that there are tons of opportunities for exciting research in the future because of extensive studies like the UK biobank, where we have massive resources of rich imaging, genotype, and phenotype data. This makes us genuinely enthusiastic about the future of functional connectivity in health care.

Supplementary Material

S1 Supplementary Material of Chapter 3

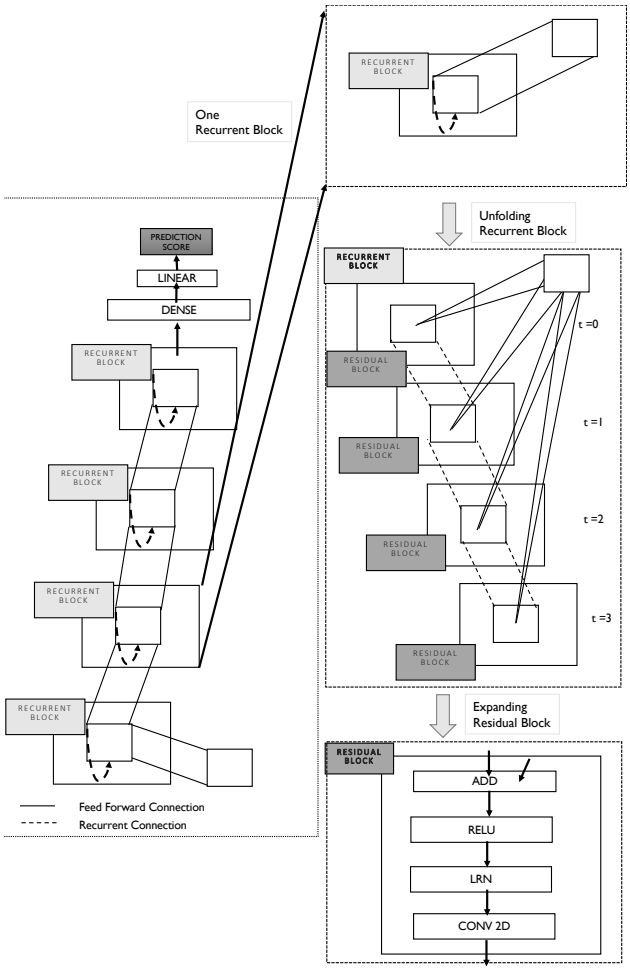


Figure S1: The proposed architecture of 2D RCNN. A recurrent block is unfolded on the right for t = 3 time steps. At t = 0, it is only a feed-forward network with a single residual block. At t =3, it has depth of 4 with additional recurrent connections. Each residual block is further composed of addition, batch normalization, convolution and activation layers.

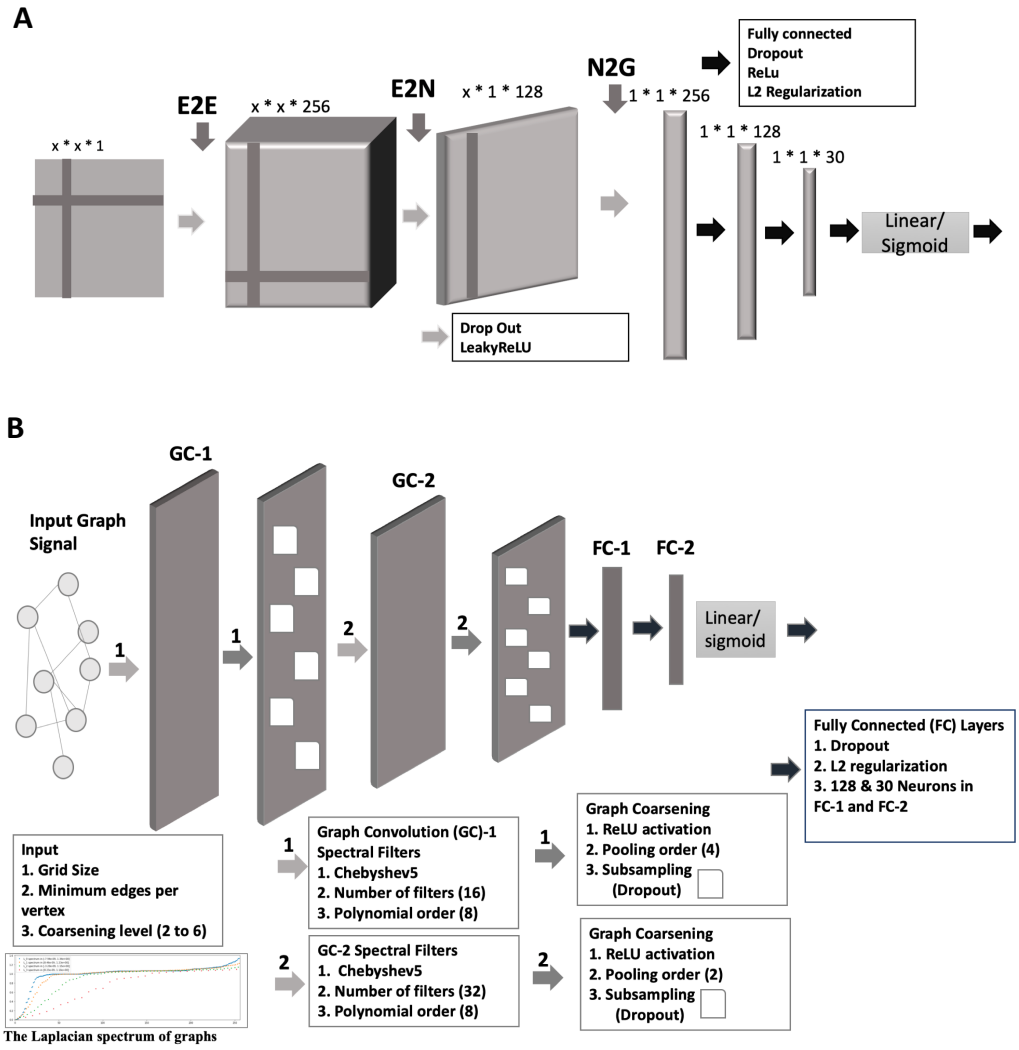


Figure S2: The architectures of DNNs: **(A)** The BrainNetCNN architecture, where the E2E layer considers weights of all neighbouring edges (adjacent brain regions), the E2N layer convolves netmats with 1D convolutional filter producing a single output for each node and finally the N2G layer reduces dimensionality. **(B)** the schematic of GraphCNN, which can be summarized as $H^{(l+1)} = [\tilde{D}^{(-1/2)}\tilde{A}\tilde{D}^{(-1/2)}H^{(l)}W^{(l)}]$, where A is the adjacency matrix, W is weight matrix and D is degree matrix. Input graph signals pass through a set of convolution, pooling and fully connected layers resulting in producing a single output score corresponding to non-imaging variable for each subject's functional connectivity matrix.

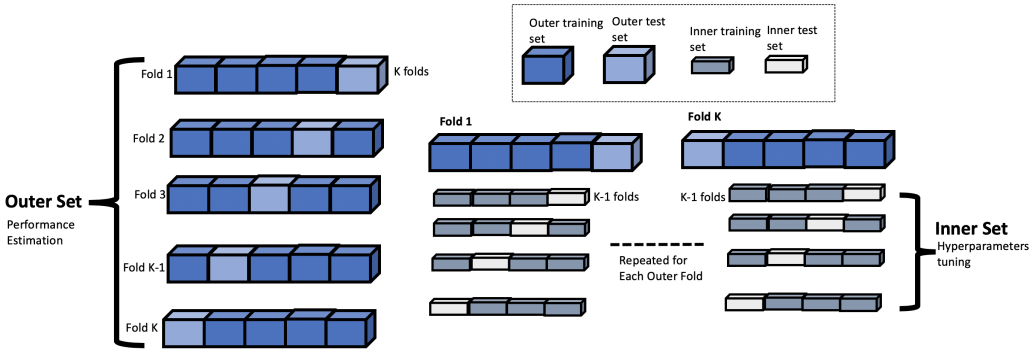


Figure S3: A summary figure explaining the process of nested cross-validation.

BRAINNET CNN	
Parameters	**Fine-tuned Values**
Edge2Edge	256
Edge2Node	128
Node2Graph	256
Fully Connected-1,2,3	256,128, 30
Loss	mean_squared_error/ binary_crossentropy
Optimizer	SGD
Batch Size	4
Learning Rate	0.0000001

2D RCNN	
Parameters	**Fine-tuned Values**
Recurrent Blocks	2-4
Residual Blocks	4
Kernel size	3
Number of kernels	128
Fully Connected-1,2	128, 64
Loss	mean_squared_error/ binary_crossentropy
Optimizer	SGD
Batch Size	128
Learning Rate	0.0001

Table S1: The tuned hyper-parameters for BrainNetCNN and 2DRCNN.

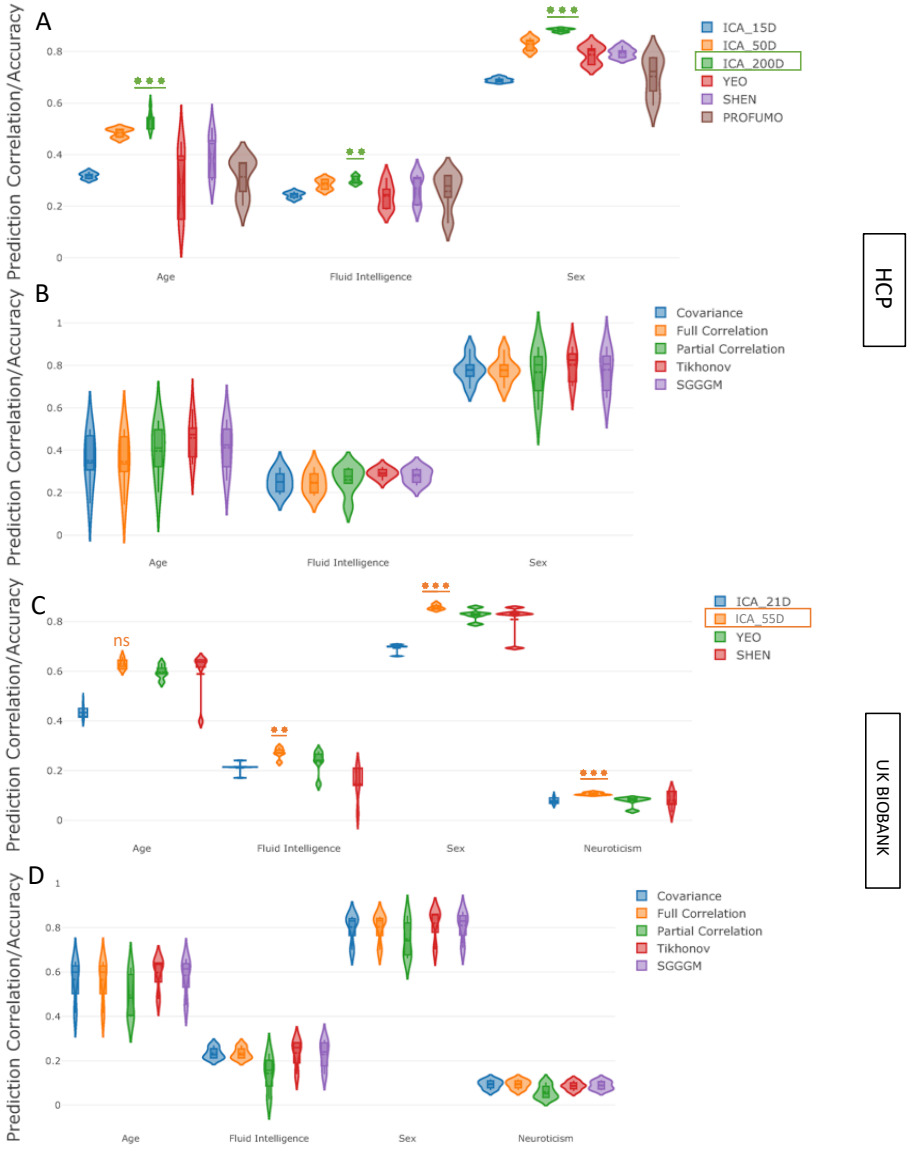


Figure S4: The impact of various parcellation strategies and functional connectivity estimation methods on prediction power for non-imaging variables **without deconfounding**. **[A,B] (HCP Data), [C,D] (UKB Data):** The violin plots in **A** and **C** show the prediction variability over 5 measures of functional connectivity estimates and in **B** and **D** show the prediction variability the over different parcellation schemes. For HCP, the ICA based parcellation schemes are ICA_15D, ICA_50D and ICA_200D, and for UKB are ICA_21D and ICA_55D, where D = the number of parcels. For both HCP and UKB, SHEN parcellation was 268D, YEO was 100D, and PROFUMO was 50D (for HCP only). The stars refer to comparison against the next-best method.

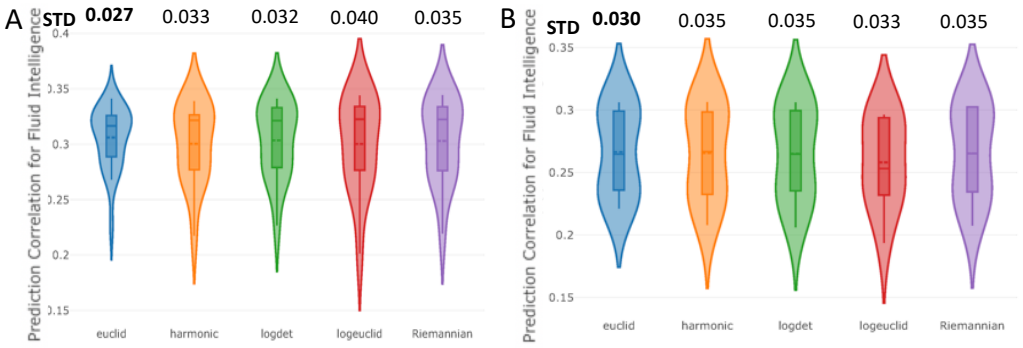


Figure S6: The impact of various reference mean estimation techniques during tangent space parameterization on prediction of non-imaging variables **without deconfounding**. **[A] (HCP Data), [B] (UKB Data)** These figures show the prediction correlation for fluid intelligence scores when functional connectivity estimates are projected into tangent space, using different reference means. The violin plot represents the prediction variability over 3 different parcellation schemes and 5 measures of functional connectivity estimates in the tangent space.

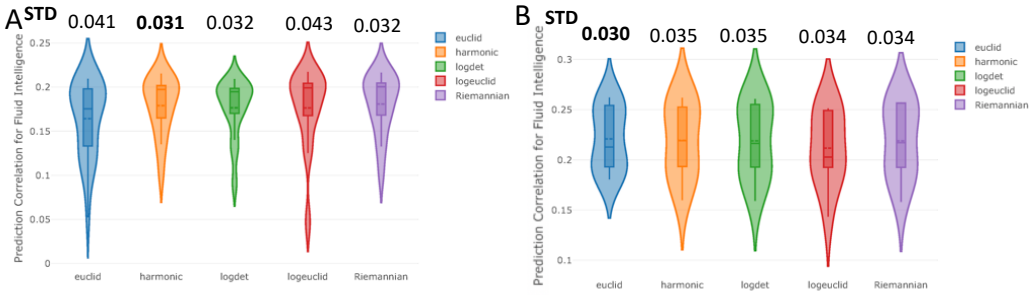


Figure S5: The impact of various reference mean estimation techniques during tangent space parameterization on prediction of non-imaging variables **after deconfounding**. **[A] (HCP Data), [B] (UKB Data)** These figures show the prediction correlation for fluid intelligence scores when functional connectivity estimates are projected into tangent space, using different reference means. The violin plots represent the prediction variability over 4 different parcellation schemes and 5 measures of functional connectivity estimates in the tangent space.

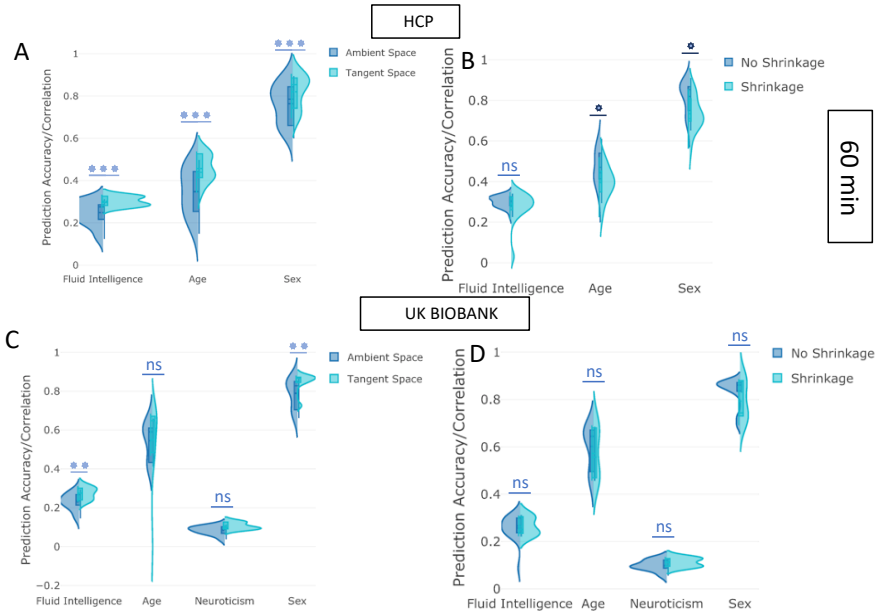


Figure S7: The impact of projecting to tangent space and applying shrinkage in tangent space on prediction power **before deconfounding**. **[A-B] (HCP Data), [C,D] (UKB):** The y-axis depicts the prediction accuracy/correlation for different behavioural measures. "Tangent Space" means that tangent space projection was applied on functional connectivity estimates (originally in the "Ambient Space"). The "Shrinkage" strategy means that non-isotropic PoSCE shrinkage was applied to connectivity estimates in tangent space before feeding to the predictor/classifier. "No Shrinkage" means that projected functional connectivity estimates in tangent space were directly fed to the predictor/classifier, and did not undergo PoSCE shrinkage. The violin plots show the prediction variability over 4 different parcellation schemes and 5 measures of functional connectivity estimates.

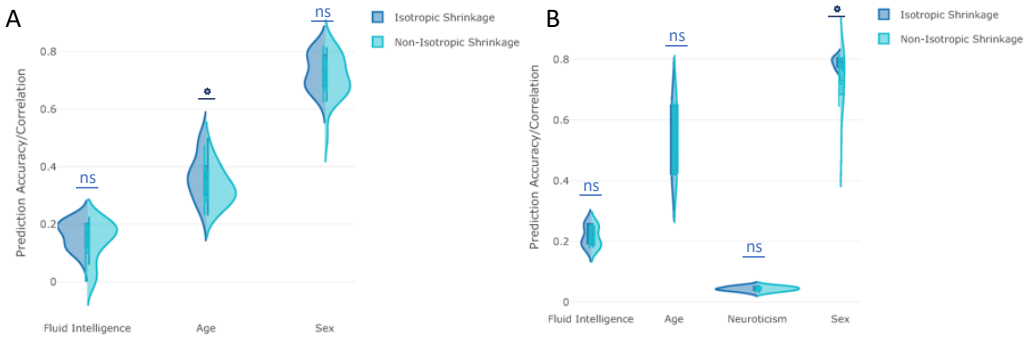


Figure S8: The impact of isotropic versus non-isotropic shrinkage in tangent space on prediction power for non-imaging variables **after deconfounding**. **[A] (HCP Data), [B](UKB)**: The y-axis depicts the prediction accuracy/correlation for different non-imaging measures. "Isotropic Shrinkage" means that Ledoit-Wolf shrinkage was applied to projected functional connectivity estimates in tangent space before they were fed to a predictor/classifier. The "Non-Isotropic Shrinkage" strategy means that PoSCE shrinkage was applied to connectivity estimates in tangent space before they were fed to a predictor/classifier.

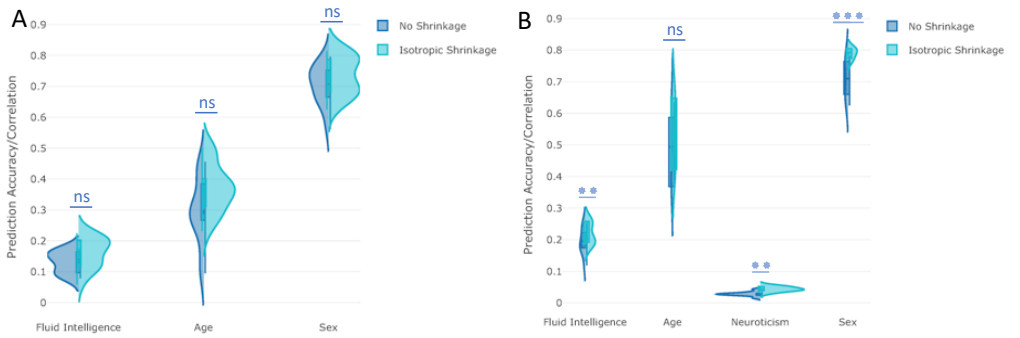


Figure S9: The impact of applying isotropic shrinkage in tangent space on prediction power after deconfounding. **[A] (HCP Data), [B](UKB)**: The y-axis depicts the prediction accuracy/correlation for different non-imaging measures. "Isotropic Shrinkage" means that Ledoit-Wolf shrinkage was applied to projected functional connectivity estimates in tangent space before they were fed to a predictor/classifier. "No Shrinkage" means that projected functional connectivity estimates in tangent space were directly fed to the predictor/classifier, and did not undergo Ledoit-Wolf shrinkage.

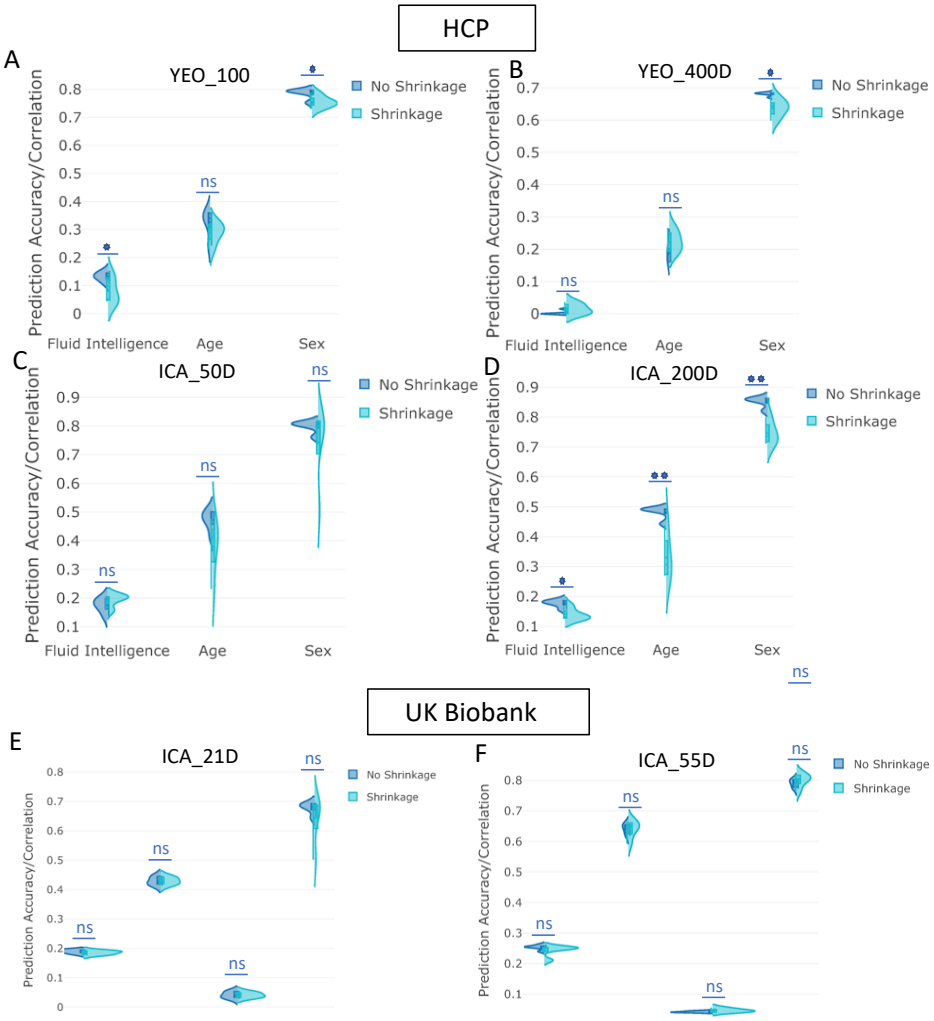


Figure S10: The impact of applying shrinkage in tangent space on prediction power varying the number of parcels. **[A-D] (HCP Data), [E,F](UKB)**: The y-axis depicts the prediction accuracy/correlation for different behavioural measures. The "Shrinkage" strategy means that non-isotropic PoSCE shrinkage was applied to connectivity estimates in tangent space before feeding to the predictor/classifier. "No Shrinkage" means that projected functional connectivity estimates in tangent space were directly fed to the predictor/classifier, and did not undergo PoSCE shrinkage. The "YEO_400D" means that 400D YEO parcellation was applied. The violin plots show the prediction variability over 5 measures of functional connectivity estimates for the mentioned parcellation scheme (e.g., "YEO_400D")

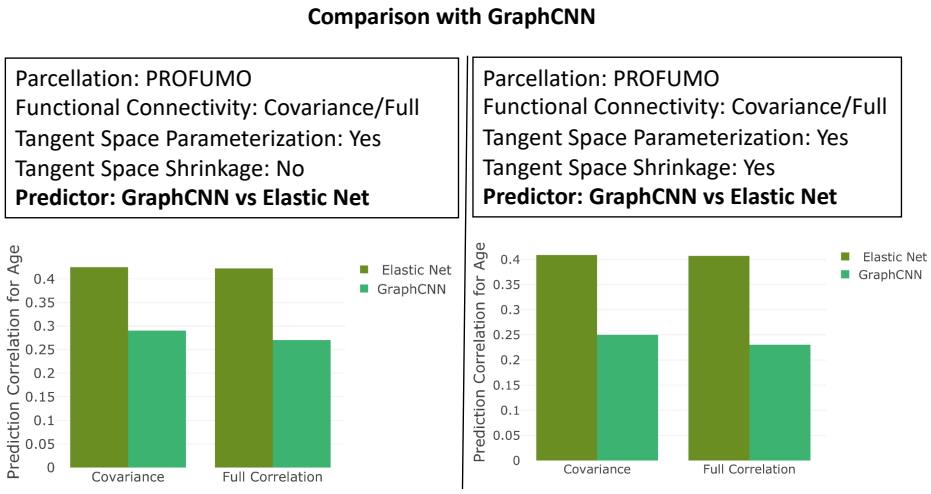


Figure S11: (HCP) The comparison of Elastic Net and GraphCNN prediction performance. We have chosen a subset (e.g., parcellation (PROFUMO), functional connectivity estimation (Full/Covariance)) from all available configurations to illustrate the performance of GraphCNN.

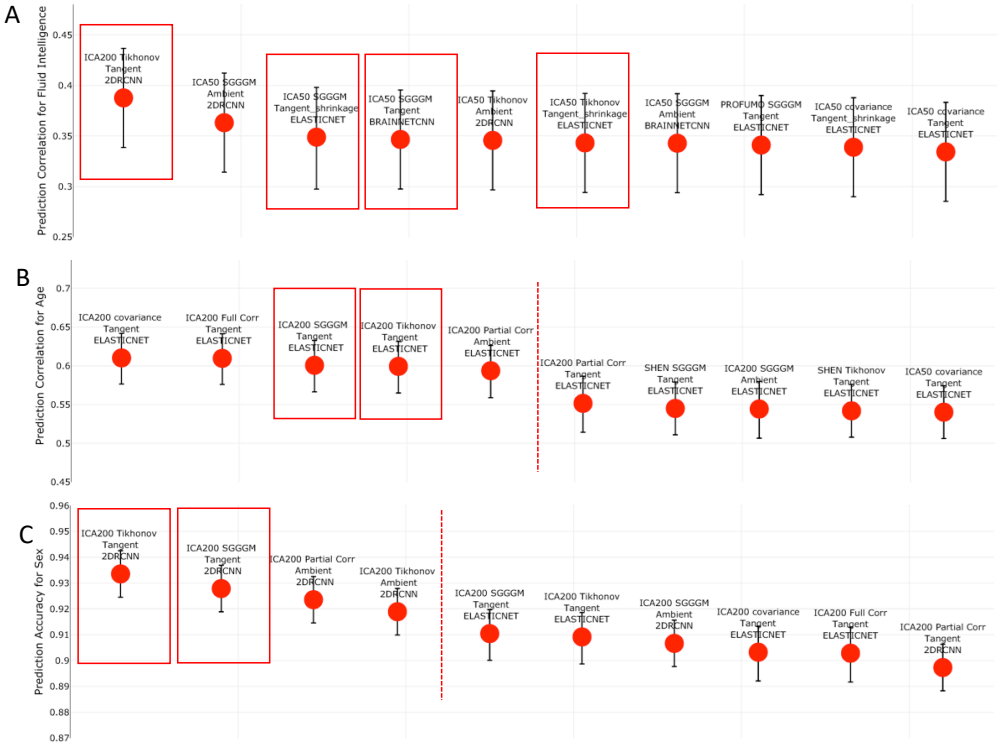


Figure S12: The top performing ten configurations for the prediction of each non-imaging variable by dataset **without deconfounding**. **[A,B,C] (HCP Data)** Each data point represents a different configuration strategy that may vary in terms of parcellation strategy, the functional connectivity estimation method, whether tangent space parameterization was employed, whether tangent space regularization was employed, and the predictor/classifier that was used. The first word indicates the parcellation strategy, and the second word refers to the functional connectivity estimation method. The third word refers to the geometry in which classifier is applied, ambient referring to non-tangent space and tangent referring to the projected covariance matrices in tangent space. If non-isotropic shrinkage was applied after projecting covariance matrices to tangent space, the fourth word will be "shrinkage". The last word indicates the type of classifier/predictor that was used. The highlighted red blocks show the recommended pipelines (rationale explained in Section 4.5), and red dotted lines highlight the point when the error bar of pipeline after the dotted line is out of range from the error bar of the top (first) pipeline.

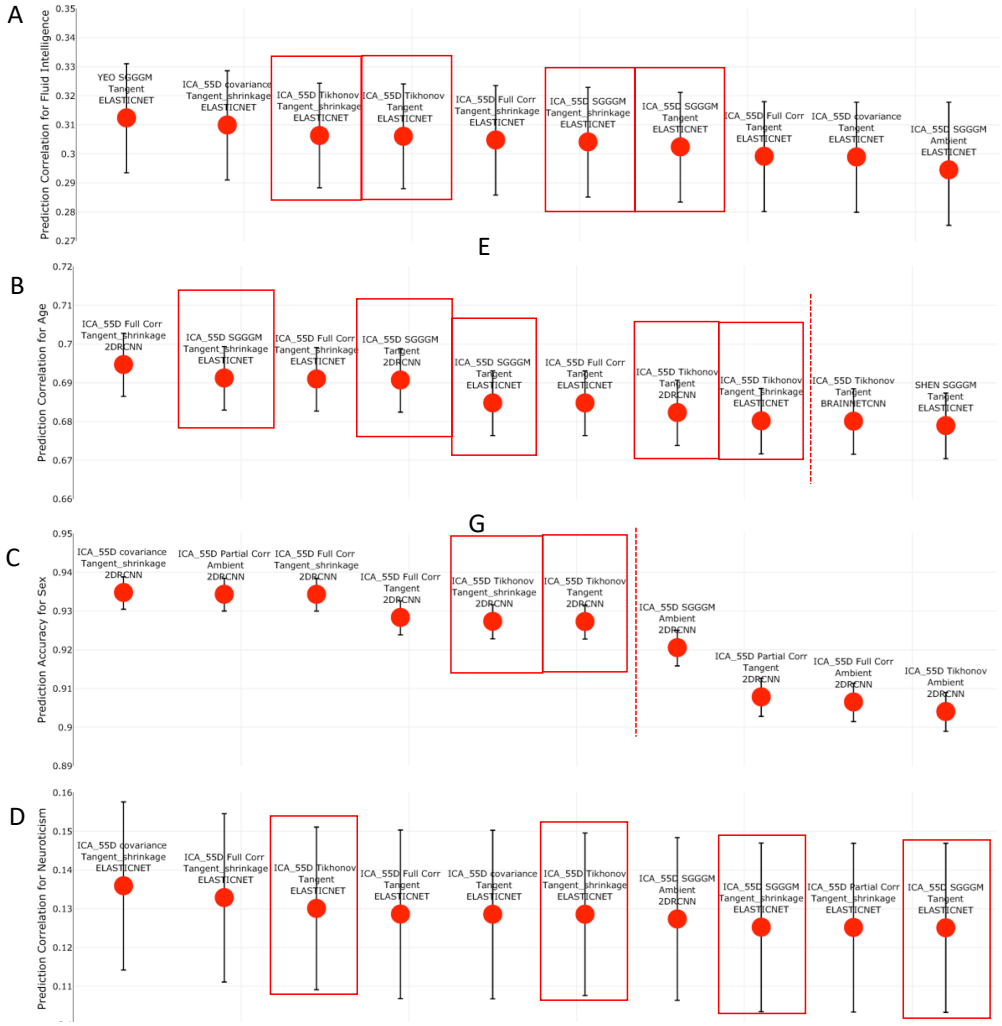


Figure S13: The top performing ten configurations for the prediction of each non-imaging variable by dataset **without deconfounding**. **[A,B,C,D] (UKB Data)** Each data point represents a different configuration strategy that may vary in terms of parcellation strategy, the functional connectivity estimation method, whether tangent space parameterization was employed, whether tangent space regularization was employed, and the predictor/classifier that was used. The first word indicates the parcellation strategy, and the second word refers to the functional connectivity estimation method. The third word refers to the geometry in which classifier/predictor is applied, ambient referring to non-tangent space and tangent referring to the projected covariance matrices in tangent space. If non-isotropic shrinkage was applied after projecting covariance matrices to tangent space, the fourth word will be "shrinkage". The last word indicates the type of classifier/predictor that was used. The highlighted red blocks show the recommended pipelines (rationale explained in Section 4.5), and red dotted lines highlight the point when the error bar of pipeline after the dotted line is out of range from the error bar of the top (first) pipeline.

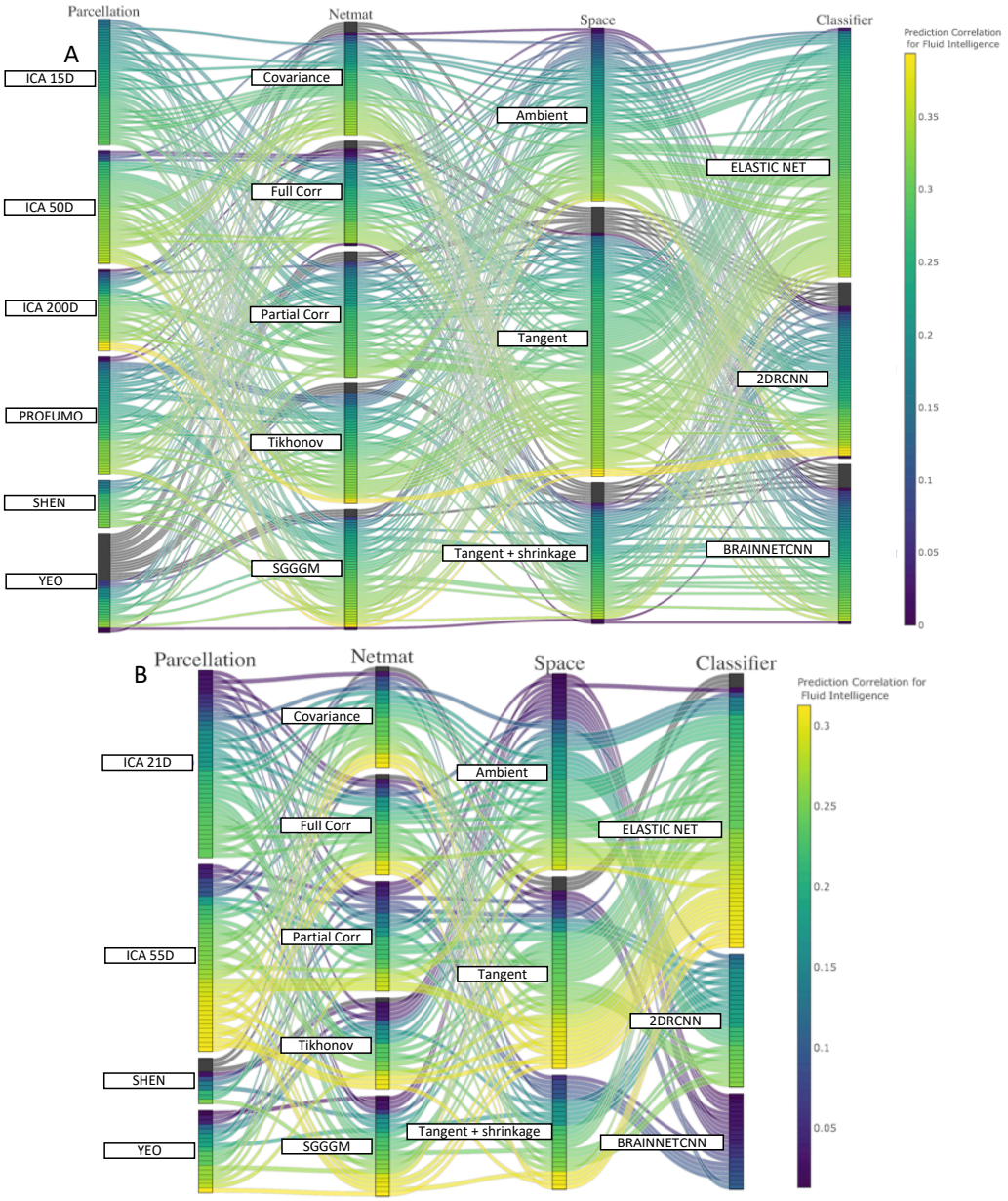


Figure S14: This parallel coordinates plot provides a visualization of all possible combinations of options in the pipeline to predict fluid intelligence scores from functional connectivity **without deconfounding**. **[A] (HCP Data), [B] (UKB Data):** The lines are color-coded according to their prediction performance.

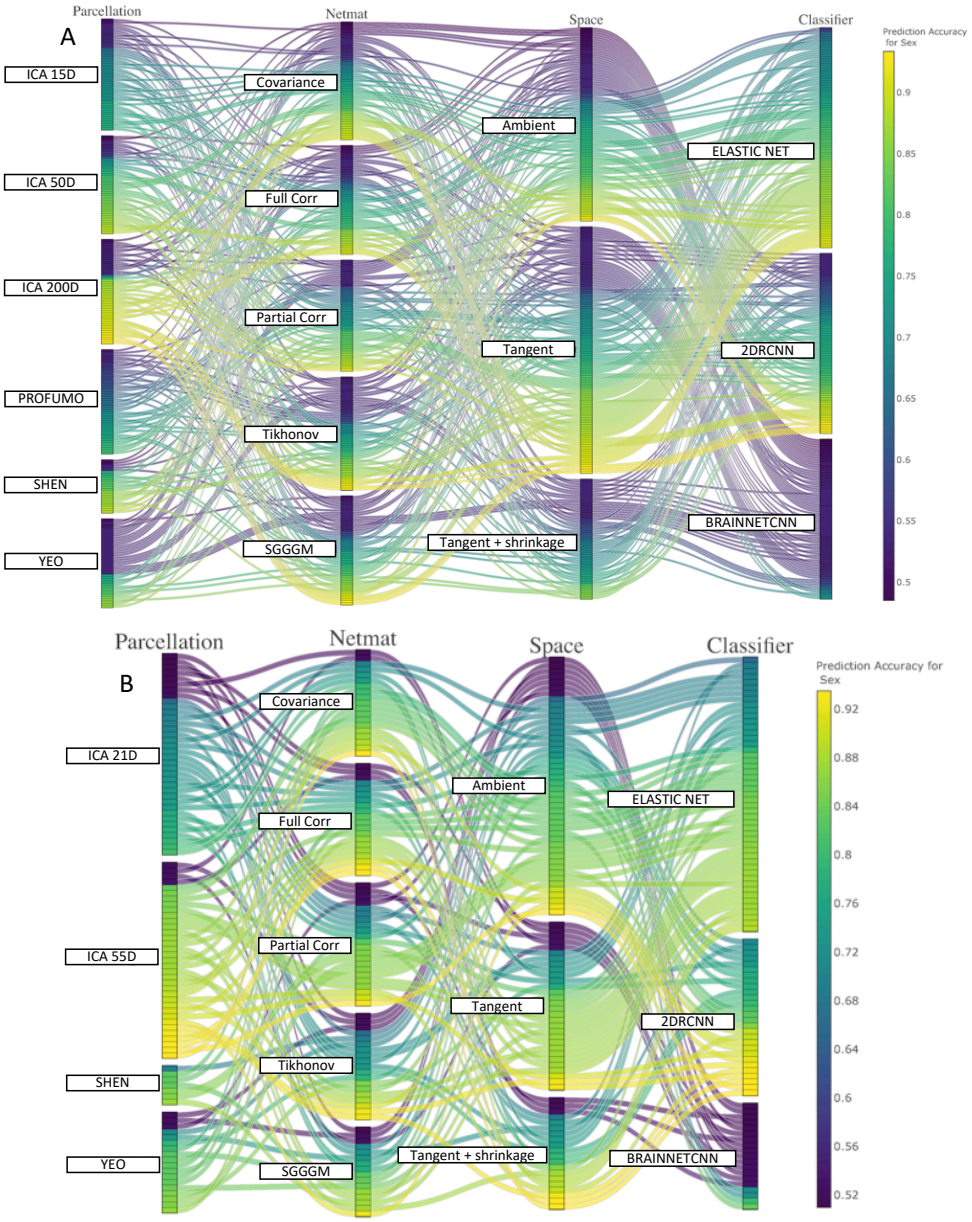


Figure S15: This parallel coordinates plot provides a visualization of all possible combinations of options in the pipeline to pipeline to predict sex from functional connectivity **without deconfounding**. **[A] (HCP Data), [B] (UKB Data):** The lines are color-coded according to their prediction performance.

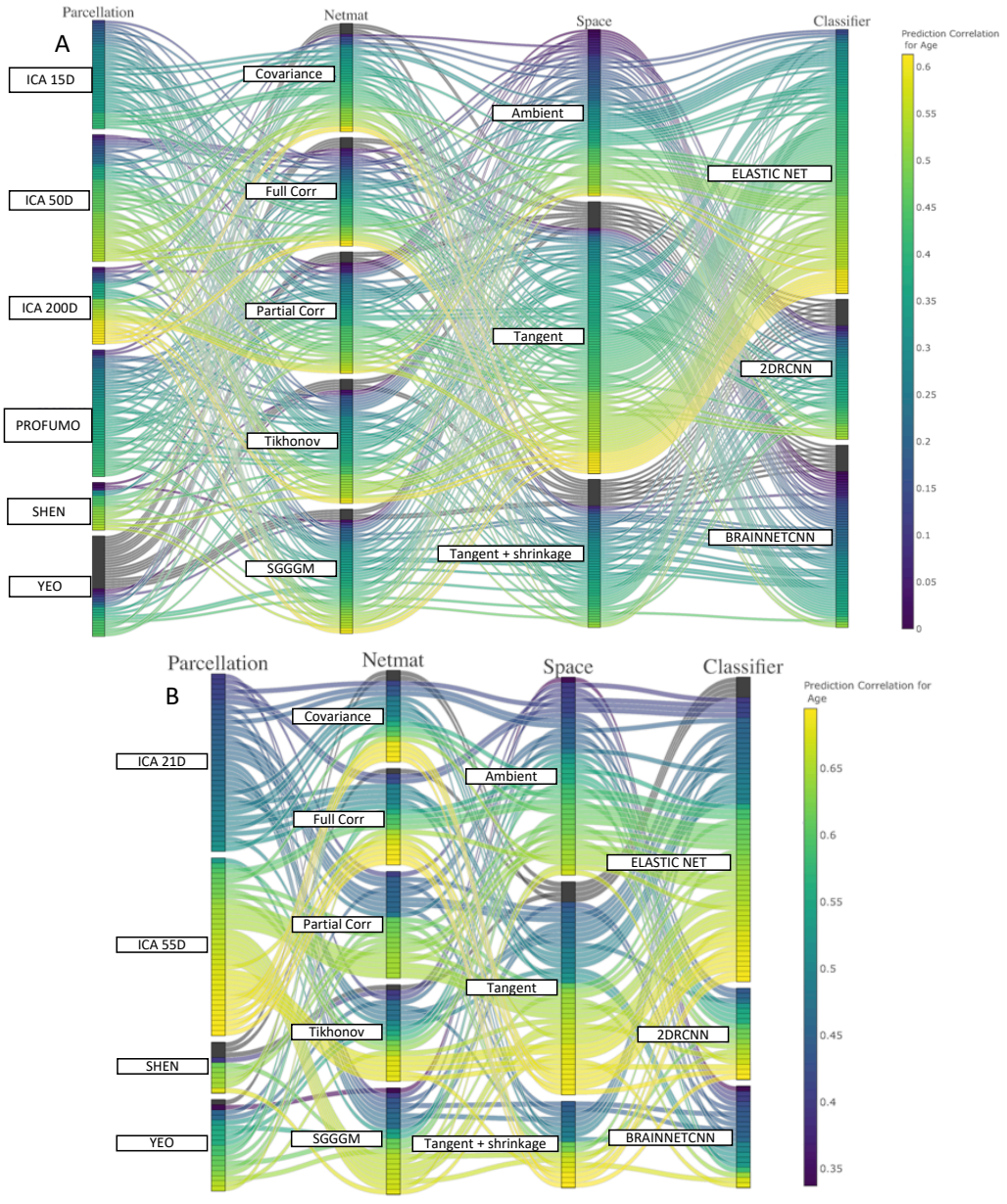


Figure S16: This parallel coordinates plot provides a visualization of all possible combinations of options in the pipeline to pipeline to predict age from functional connectivity **without deconfounding**. **[A] (HCP Data), [B] (UKB Data):** The lines are color-coded according to their prediction performance.

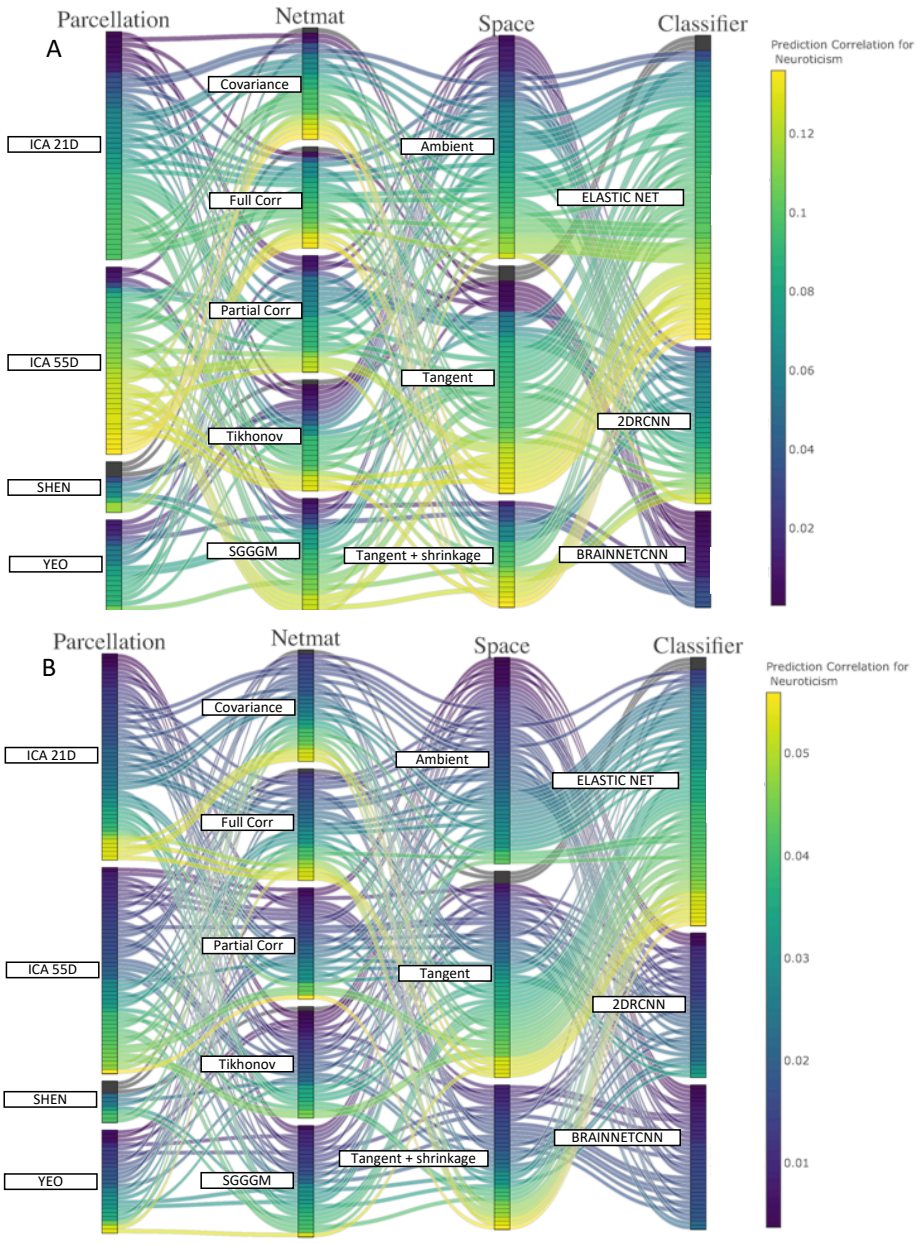


Figure S17: This parallel coordinates plot provides a visualization of all possible combinations of options in the pipeline to predict neuroticism score from functional connectivity. **[A]** shows the result before confounds removal and **[B]** shows the result after regressing out the confounds.

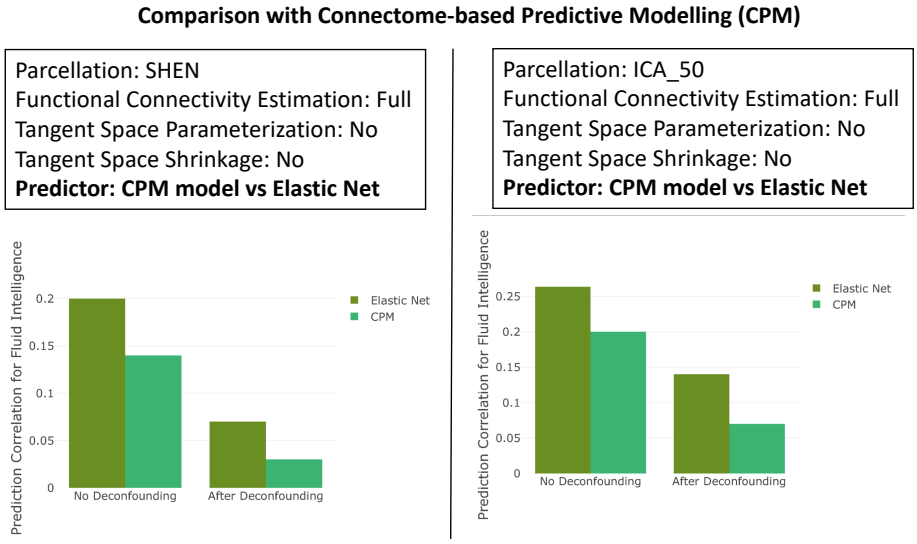


Figure S18: (HCP) The comparison of Elastic Net and CPM prediction performance. CPM predictor/classifier is based on a model that averages connectivity edges from a subset of all edges.

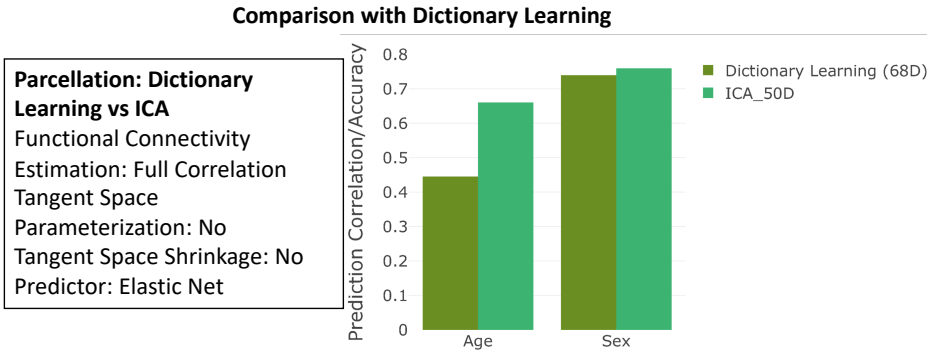


Figure S19: (ABIDE) The comparison of ICA and Dictionary Learning prediction performance. We have chosen a subset (e.g., parcellation (ICA and Dictionary Learning), functional connectivity estimation (Full) from all available configurations to illustrate the comparison.

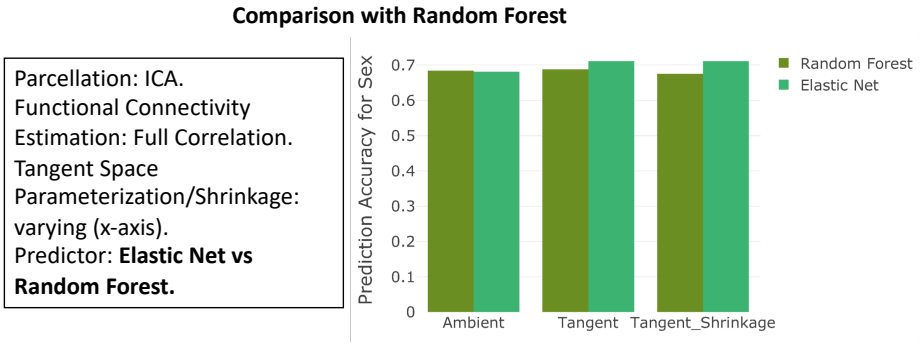


Figure S20: (HCP) The comparison of Elastic Net and Random forest prediction performance. We have chosen a subset (e.g., parcellation (ICA), functional connectivity estimation (Full)) from all available configurations to illustrate the comparison.

S2 Supplementary Material of Chapter 4

S2.1 Method's Supplementary

Dynamic Structural Equation Modelling (SEM) perspective

We can gain insight into the multi-dynamic model being proposed, by considering a dynamic version of structural equation modelling (SEM) (Penny, Stephan, Mechelli, & Friston, 2004):

$$Y_t = A_t Y_t + H\mu_t + e_t \tag{S1}$$

where A_t is an N*N dimensional matrix containing the connectivity information, where A_{ij} denotes a connection from region j to region i and $A_{ii} = 0$ for all regions (in SEM, A_t is referred to as the path coefficient matrix, M). μ_t is an Qx1 vector of Q structured, external inputs that are independent such that $\mu_t \mu_t^\intercal = I$, and H is an NxQ dimensional mixing matrix describing how much of each external input is contributing to each region. Finally, e_t is a random variable modelling the noise in Y_t using a Gaussian distribution with zero mean and an NxN diagonal covariance matrix, i.e., $e_t \sim \mathcal{N}(0, \Sigma)$. Note that the idea is that μ_t would be identifiable through its non-Gaussian, spatio-temporal structure, which is what makes it distinct to e_t. Note that typically in SEM, only the noise, e_t, is modelled and an external input, μ_t, is not (i.e. $\mu_t = 0$). However, other variants of SEM have previously considered modelling structured, external inputs (Shimizu, Hoyer, Hyvärinen, Kerminen, & Jordan, 2006).

It is straightforward[2] to then show that Y_t is Gaussian distributed, with a time-

[2]Such investigations to establish the relationship between SEM and FC has been previously conducted in detailed by (Marrelec & Benali, 2009)

varying mean and covariance with the same form as in Equations (S8, S9), where

$$m_t = (1 - A_t)^{-1} H \, \mu_t \tag{S2}$$

$$F_t = (1 - A_t)^{-1} \left(\Sigma + HH^\intercal\right) \left((1 - A_t)^{-1}\right)^\intercal \tag{S3}$$

In other words, if we assume that in the dynamic SEM the external inputs, μ_t, and the dynamic connectivity, A_t, fluctuate distinctly from each other; then correspondingly m_t and F_t can also fluctuate distinctly from each other (see Equation S2, S3). This is what we are assuming in the *multi-dynamic* approach.

Single-dynamic (SAGE) Model

SAGE model assumes that the mean, variance, and correlation fluctuate on a single (common) time-course. This implies that the time courses α_{tp}, β_{tq} and γ_{tr} are the same in the SAGE approach. The time-varying mean, correlation and variance are modelled as:

$$m_t = \sum_p \alpha_{tp} \mathcal{S}_p \tag{S4}$$

$$F_t = \sum_p \alpha_{tp} \mathcal{D}_p \tag{S5}$$

$$G_t = \sum_p \alpha_{tp} \text{diag}(\mathcal{E}_p) \tag{S6}$$

where, the α_t is modelled by underlying logits θ_t. The loss function for SAGE is:

$$\sum_{t=1}^{T} log\, p(Y_t|\theta_t = \mu_t^\theta) + \sum_{t=1}^{T} \text{CrossEntropy}\,(\mu_t^\theta||\hat{\mu}_t^\theta) \tag{S7}$$

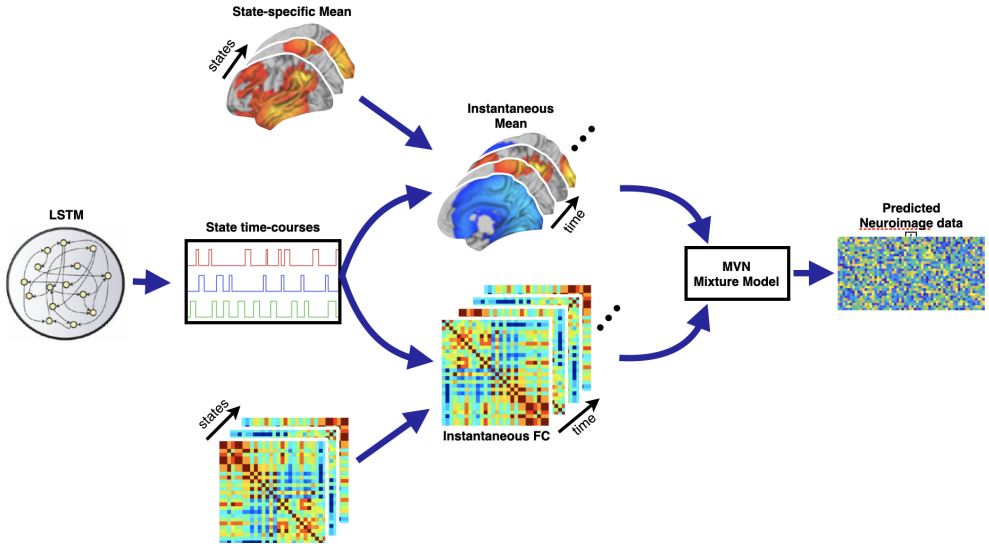


Figure S21: Generative model for SAGE. The instantaneous mean and correlation are modelled using an underlying set of states, for which the state time courses are generated using a long short-term memory (LSTM) model (common for both the mean and correlation), making the approach single-dynamic.

MAGE In-depth architecture

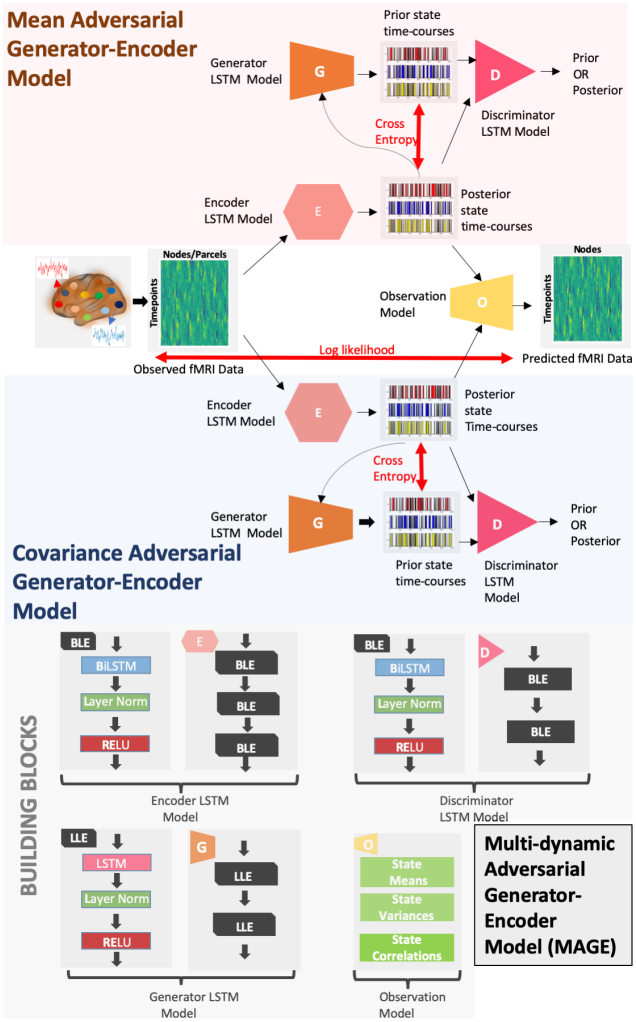


Figure S22: The in-depth architecture of the MAGE is illustrated here. For the multi-time-scale modelling, two sub-blocks are proposed; one for the generation-inference of covariance associated parameters and the other for the mean. Each block is composed of a generator, encoder, and discriminator model. The generator model predicts the prior state time-courses and the encoder model predicts the posterior state time-courses, whereas, the discriminator model enforces the prior state time-courses on the posterior state time-courses. The encoder model is composed of three BLE blocks, whereas each BLE block is composed of a bi-directional LSTM (BiLSTM), layer normalization (Layer Norm), and a rectified linear activation function (ReLU) layer. The discriminator is composed of two BLE blocks whereas, the generator model is composed of two LLE blocks. BLE is an acronym from **B**iLSTM, **L**ayer Norm and **R**e**L**U and LLE is an acronym from **L**STM, **L**ayer Norm and **R**e**L**U. The main difference between LLE and BLE block is the usage of a uni-directional LSTM (referred here as simply LSTM) instead of a bi-directional LSTM.

Sliding Window Correlation (SWC) approach

For the SWC analysis, given a window of length, w, we removed all the frequency components below 1/w. Specifically, we have chosen a window-length of 100 seconds and estimated pair-wise correlations within each window, resulting in one FC correlation matrix per window. We used overlapping windows with an overlap of 25 time-points between two adjacent windows. After performing the SWC analysis, we applied the k-means clustering to the SWC estimated FC matrices. This clustering was applied at the group-level so the identified states are comparable across subjects. We associated each FC matrix with a clustering label/state, whereas, the total number of k-means states, are set to 12 (same as the MAGE and HMM approach). Then, we averaged the FC matrices for each subject as per the clustering labels. This sliding window *plus* clustering analysis provided us with TVFC estimates for each subject where the maximum number of states for each subject is 12 (and a state is not included if it is not active for that subject). Lastly, we concatenated the state-by-state FC matrices for each subject for any post-hoc prediction analysis conducted in the work.

Hidden Markov Modelling (HMM) approach

For the HMM analysis conducted in this work, states are modelled using multivariate Gaussian distributions, each state has a distinct mean and a distinct covariance matrix. In the HMM, the parameters describing the dynamics are categorical state time courses. Specifically, we model the covariance C_t at each time point as a selection of a single covariance matrix from a finite set of covariances, $\mathcal{D}_p$ where p = [1...P]. Similarly, we model the mean m_t at each time point as a selection of a single mean vector from a finite set of mean vectors, $\mathcal{S}_p$.

The mean and covariance at any time-point, t, can be modelled as:

$$m_t = \sum_p \alpha_{tp} \mathcal{S}_p \tag{S8}$$

$$C_t = \sum_p \alpha_{tp} \mathcal{D}_p \tag{S9}$$

where, $\alpha_{tp} \in \{0, 1\}$, indicating one state (out of P possible states) is active ($\alpha_t = 1$) at any time-point, t.

Derivation of the regularization loss function for the SAGE approach

The loss function for the SAGE approach is stated in Equation S7. The second term of the loss function is the regularization term that minimises the cross-entropy between the prior and the posterior of the latent, time-varying state time-courses.

$$\sum_{t=1}^{T} \text{CrossEntropy}\left(\mu_t^\theta || \hat{\mu}_t^\theta\right) \tag{S10}$$

To achieve this, our proposed loss function is composed of two terms, the encoder adversarial loss and the generator adversarial loss.

Encoder regularization loss We trained the discriminator (D) model to distinguish posterior from the prior. Simultaneously, we modified the encoder (E) model to make the discriminator believe that encoder model outputs are samples from the generator (G) model.

The objective of the discriminator model is to correctly classify the μ_t^θ from $\hat{\mu}_t^\theta$. For this, Equation S11 should be maximized (max) and the loss function for the

discriminator model can be written as:

$$\sum_{t=1}^{T} \underset{D}{max}\ [logD(G(\hat{\mu}_t^{\theta}))] + \ [log(1 - D(E(\mu_t^{\theta}))] \tag{S11}$$

On the other hand, we make the encoder model compete against the discriminator model, and it would try to minimize (min) Equation S11 as shown below:

$$\sum_{t=1}^{T} \underset{E}{min}\ [logD(G(\hat{\mu}_t^{\theta})] + \ [log(1 - D(E(\mu_t^{\theta}))] \tag{S12}$$

If we combine the Equation S11 and Equation S12, It can be written as:

$$\sum_{t=1}^{T} \underset{D}{max}\ \underset{E}{min}\ [logD(G(\hat{\mu}_t^{\theta}))] + \ [log(1 - D(E(\mu_t^{\theta})))] \tag{S13}$$

In summary, the encoder regularization loss has two objectives; the objective of the discriminator model is to learn to distinguish posterior from the prior, and the objective of the encoder model is to fool the discriminator model into believing that μ_t^{θ} is not inferred from the encoder model but the generator model.

Generator regularization loss For the generator adversarial loss, we again trained the discriminator model to distinguish posterior from the prior. Unlike the encoder adversarial loss, this time we trained the generator model to make the discriminator believe that generator model outputs are samples from the encoder model.

The objective of the discriminator is:

$$\sum_{t=1}^{T} \underset{D}{max}\ [logD(E(\mu_t^{\theta}))] + \ [log(1 - D(G(\hat{\mu}_t^{\theta}))] \tag{S14}$$

The generator model here would be competing against the discriminator model and would try to minimize Equation S14 as:

$$\sum_{t=1}^{T} \min_{G} [logD(E(\mu_t^{\theta}))] + [log(1 - D(G(\hat{\mu}_t^{\theta}))] \tag{S15}$$

Equation S14 and Equation S15 can be written in condensed form as:

$$\sum_{t=1}^{T} \max_{D} \min_{G} [logD(E(\mu_t^{\theta})] + [log(1 - D(G(\hat{\mu}_t^{\theta}))] \tag{S16}$$

In summary, the generator regularization loss has two objectives, the objective of the discriminator is to learn to distinguish posterior from the prior, and the objective of the generator is to fool the discriminator model into believing that $\hat{\mu}_t^{\theta}$ is not inferred from the generator model but the encoder model.

Total regularization loss The total adversarial loss can be written by combining Equation S13 and Equation S16:

$$\begin{aligned}\sum_{t=1}^{T} \max_{D} \min_{G} \min_{E} [logD(G(\hat{\mu}_t^{\theta})] + [logD(E(\mu_t^{\theta}))] + \\ [log(1 - D(E(\mu_t^{\theta}))] + [log(1 - D(G(\hat{\mu}_t^{\theta}))]\end{aligned} \tag{S17}$$

This derivation can be easily extended to the multi-dynamic (MAGE) approach, where we are modelling θ_t^m and θ_t^c instead of a single θ_t.

S2.2 Result's Supplementary

MAGE performance on static-FC simulations

We demonstrated the performance of MAGE for scenarios where the ground-truth is such that correlation (i.e., the FC) is static (i.e., no fluctuations in the edges of the FC matrix over time). We show in Figure S23 that MAGE does not infer time-varying correlations when we infer more than one state for such simulations. It highlights that MAGE only infers time-varying FC when the underlying FC is time-varying.

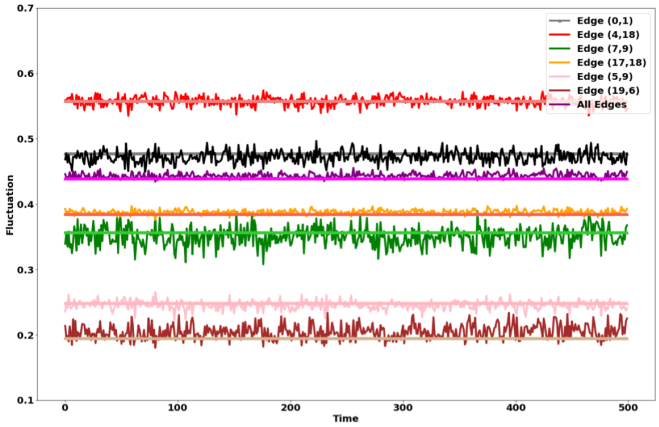


Figure S23: MAGE does not infer time-varying FC when the underlying ground-truth is not time-varying. The ground-truth is that there is only one state active at all the time-points (implies that the FC is not fluctuating). In such a simulation, even when we infer more than one state, MAGE only infers the static FC (closer to the ground-truth). We have illustrated such results for six edges and average across all the edges of the FC matrix. The light-faced straight lines highlight the ground-truth that shows the instantaneous FC is not fluctuating across time for these six edges. The bold-faces lines around the ground-truth highlight the infer instantaneous FC estimates also look homogeneous across time.

MAGE performance on fluctuating mean activity and null (static) HCP data

We generated null data using autoregressive randomization (ARR). Specifically, a 10-th order Gaussian autoregressive (AR) model was applied to generate null data.

For each participant and each pair of brain regions, we compared the observed between-timepoint variances of the instantaneous FCs inferred by MAGE, which can be calculated from the FC state dynamics and state-specific FCs, from the real HCP data against the null data. This analysis resulted in one p-value for each edge of the FC matrix. On average, across HCP participants (compared against null data), 95% of edges were significant ($p \leq 0.001$).

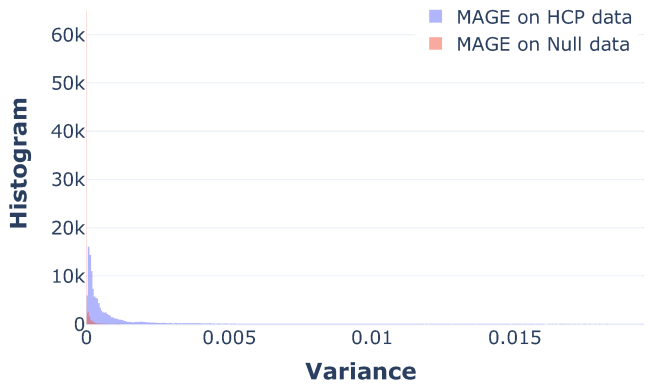


Figure S24: MAGE results on null (static) HCP data. The histogram of null variance estimates is compared with the variance estimates from the real HCP data for each participant and each pair of brain regions. The histogram of the variance estimates highlights that the between-brain region variability over time is significantly higher in the real HCP data as compared to null HCP data.

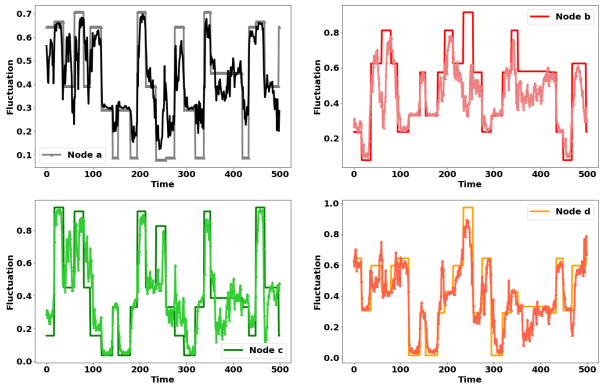


Figure S25: MAGE can adjust to fluctuating mean activity. It is the continuum of the multi-dynamic results presented in Figure 1 of the article. The multi-dynamic approach can adjust to fluctuating FC (as shown in Figure 1) and also accurately adjust to fluctuating mean activity of nodes (as shown here). We illustrated these results for four of the nodes of the FC matrix where we have shown the ground-truth and the inferred estimates.

HCP: MAGE estimated state-specific functional connectivity correlation maps

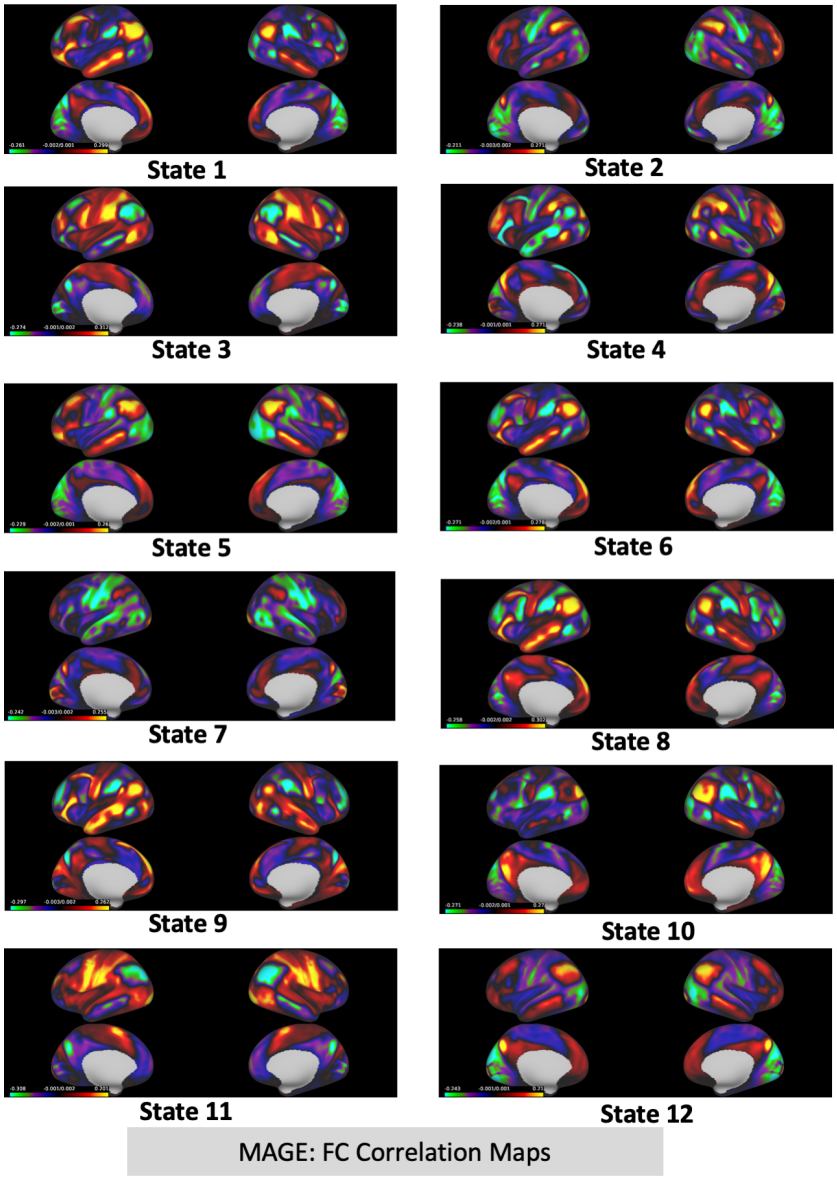


Figure S26: HCP Data: We illustrated the spatial maps for the functional connectivity using the multi-dynamic (MAGE) approach for all the 12 states.

HCP: MAGE estimated state-specific mean activity maps

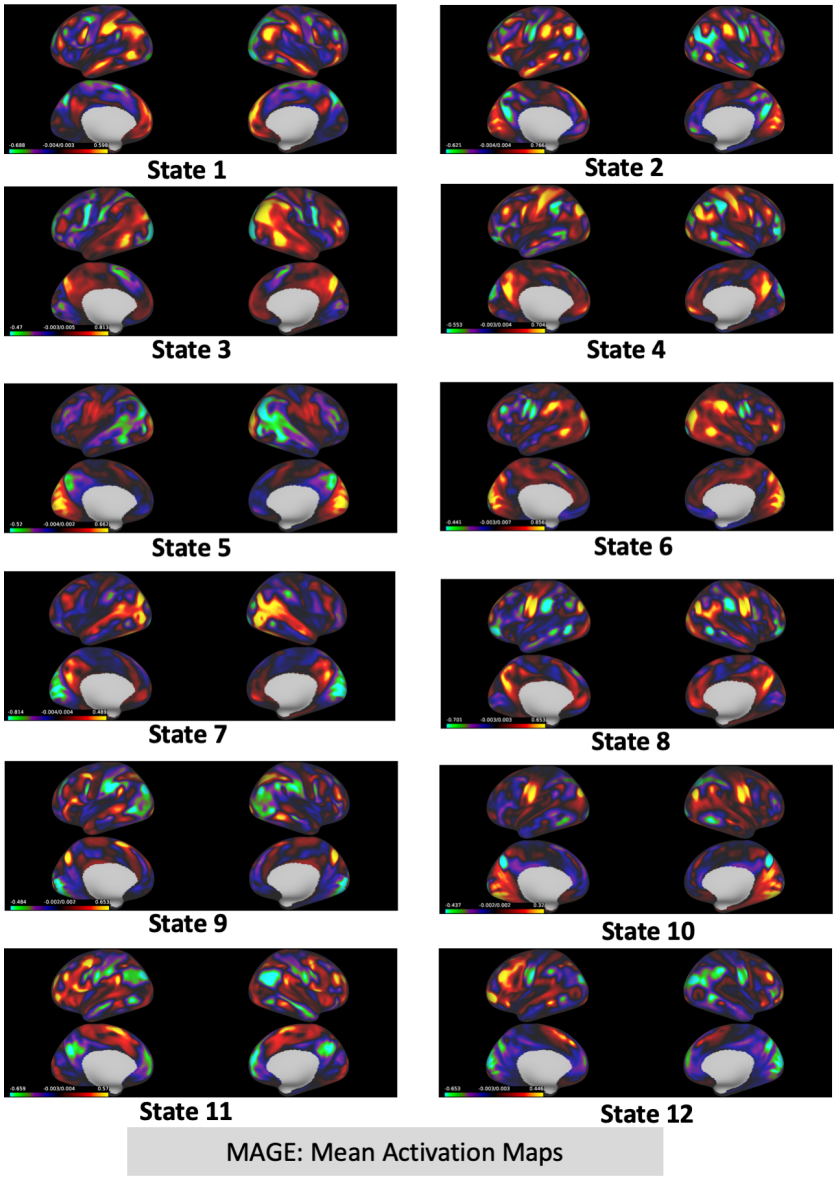


Figure S27: HCP Data: We illustrated the spatial maps for the mean activity levels using the multi-dynamic (MAGE) approach for all the 12 states.

HCP: SAGE estimated state-specific functional connectivity correlation maps

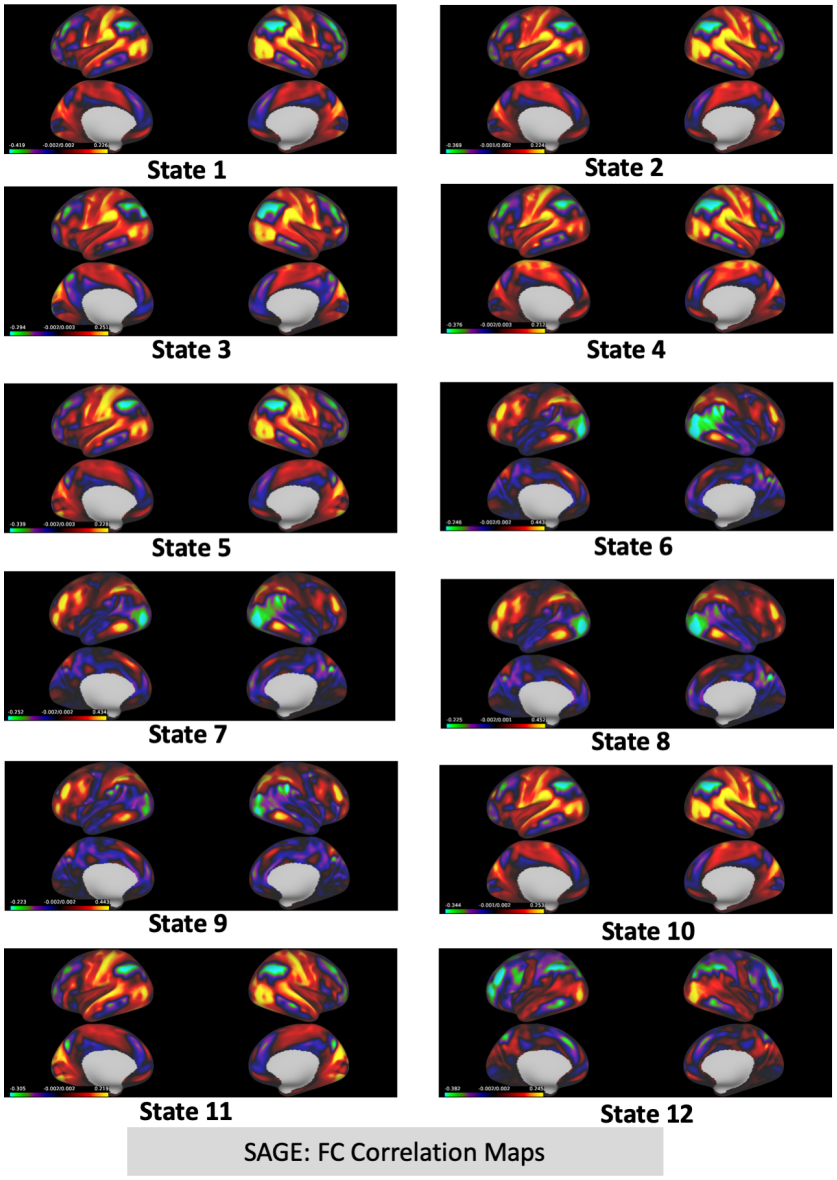


Figure S28: HCP Data: We illustrated the spatial maps for the functional connectivity using the single-dynamic (SAGE) approach for all the 12 states.

HCP: SAGE estimated state-specific mean activity maps

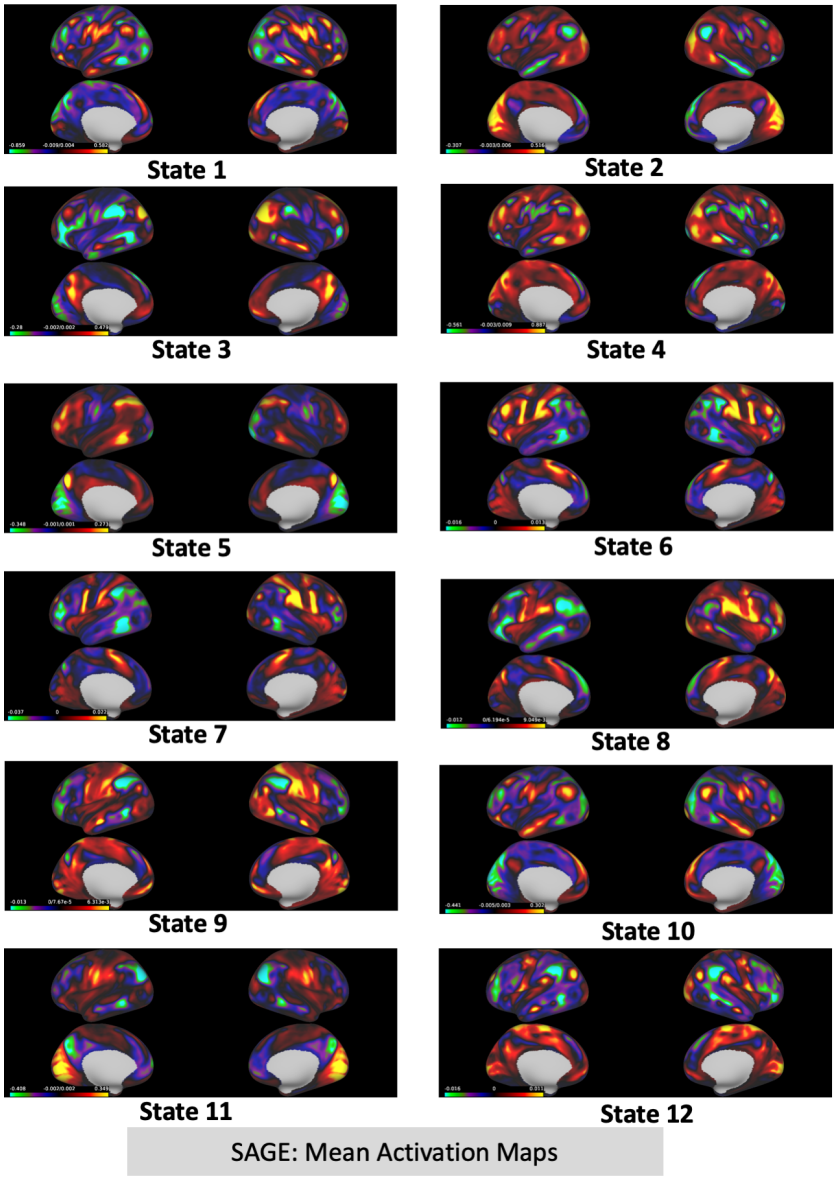


Figure S29: HCP Data: We illustrated the spatial maps for the mean activity levels using the single-dynamic (SAGE) approach for all the 12 states.

HCP: SAGE estimated state-specific functional connectivity correlation maps (mean activity is forced to zero)

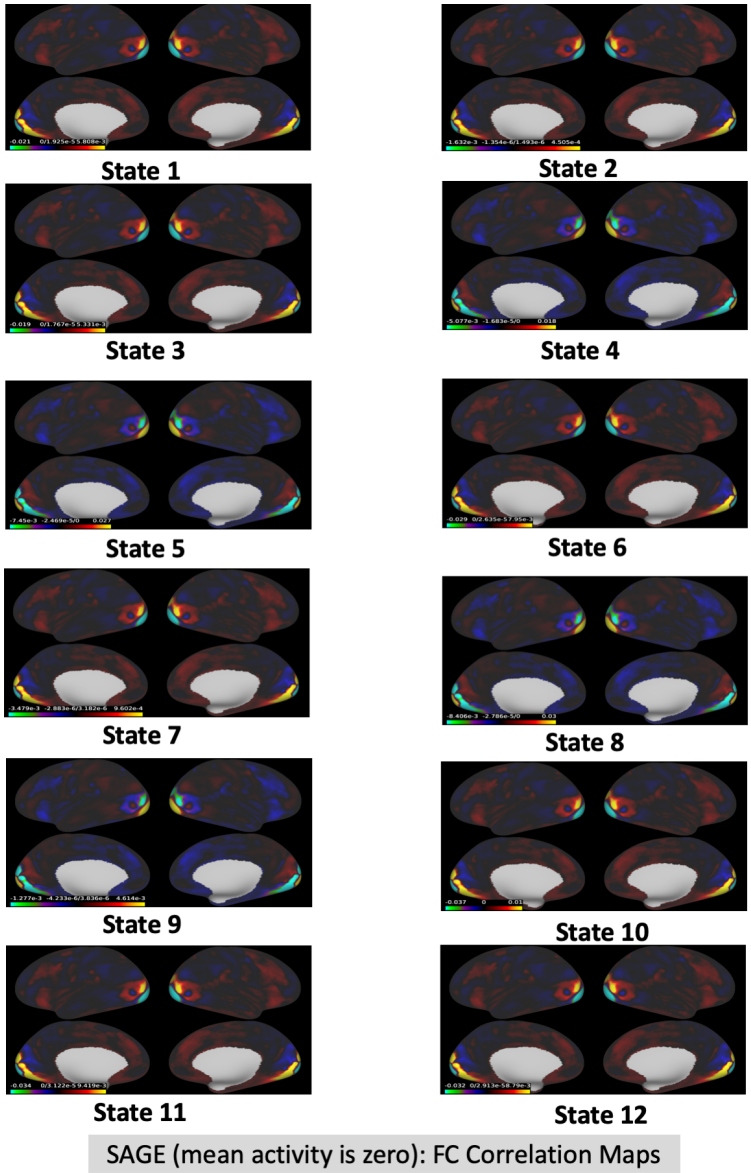


Figure S30: HCP Data: We illustrated the spatial maps for the functional connectivity using the single-dynamic (SAGE) approach for all the 12 states.

UKB: SWC estimated state-specific functional connectivity correlation maps

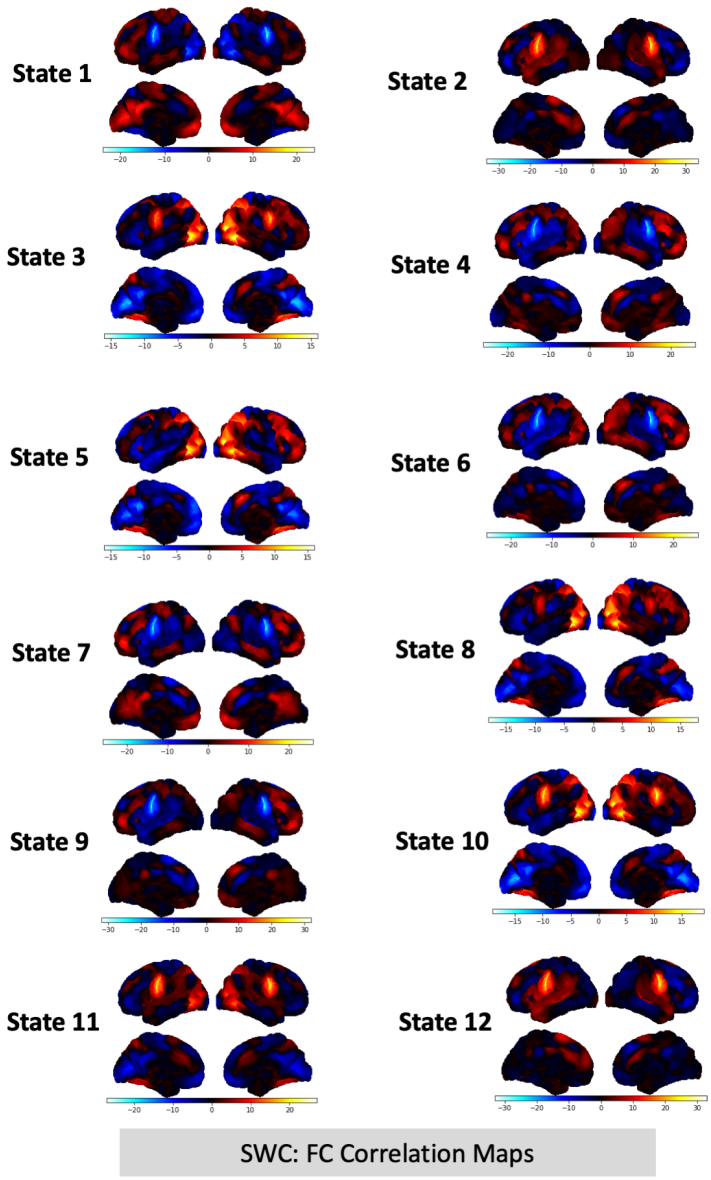


Figure S31: UKB Data: We illustrated the spatial maps for the functional connectivity using the Sliding Window Correlation (SWC) approach for all the 12 states.

UKB: HMM estimated state-specific functional connectivity correlation maps

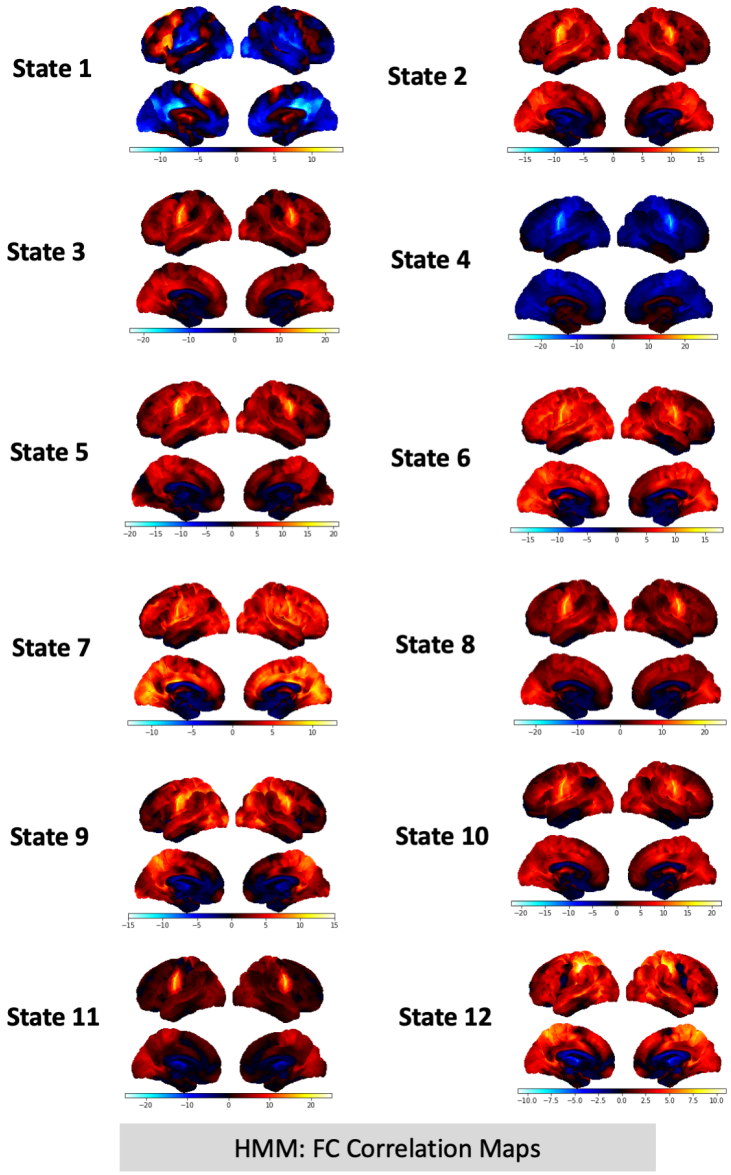


Figure S32: UKB Data: We illustrated the spatial maps for the functional connectivity using the Hidden Markov Models (HMM) approach for all the 12 states.

UKB: MAGE estimated state-specific functional connectivity correlation maps

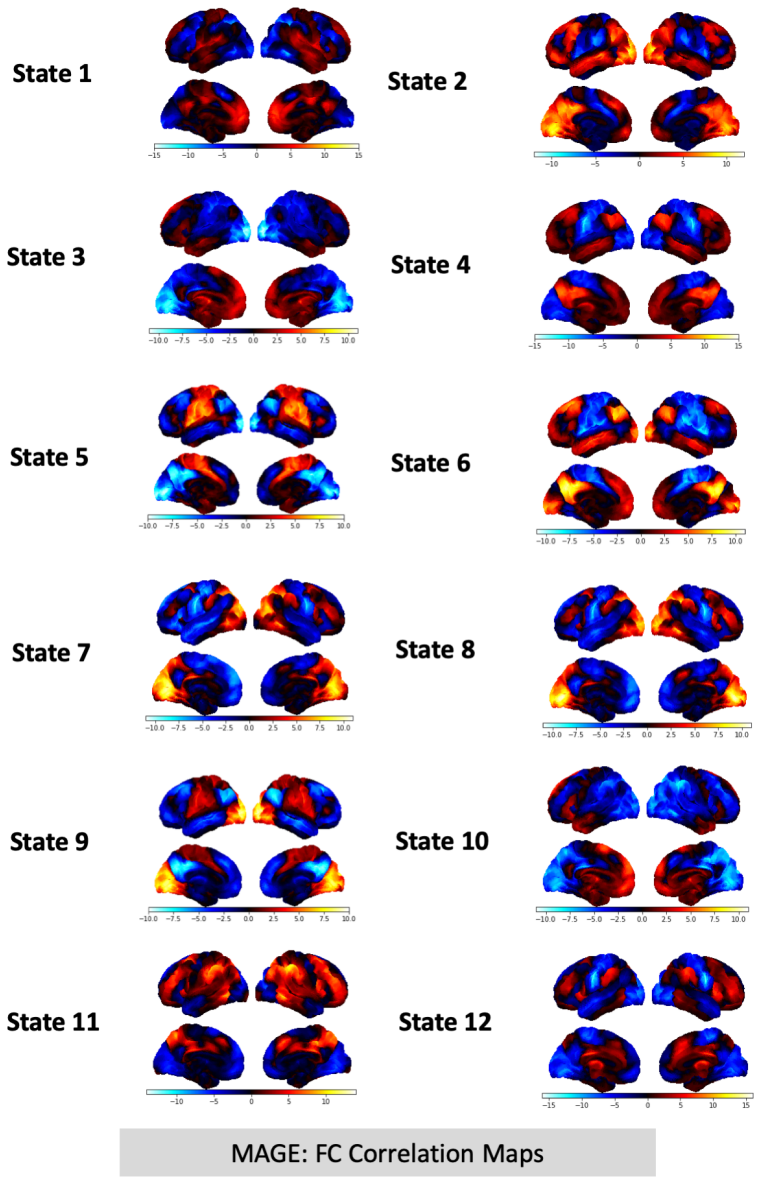


Figure S33: UKB Data: We illustrated the spatial maps for the functional connectivity using the multi-time-scale (MAGE) approach for all the 12 states.

S3 Supplementary Figures for Chapter 5

We have illustrated the results of task decoding when we modelled more than 4 states by linking the fractional occupancy (the time spent in each state) with the experimental task. In Figure S34[A] and Figure S34[B], we have modelled 8 and 16 states respectively. Particularly, we plotted the relative fractional occupancy (mean change in occupancy relative to baseline) distribution for each of the modelled states. It is evident that even with the increased number of states, we can still find task-specific states (as highlighted by the paired t-test values on the respective states).

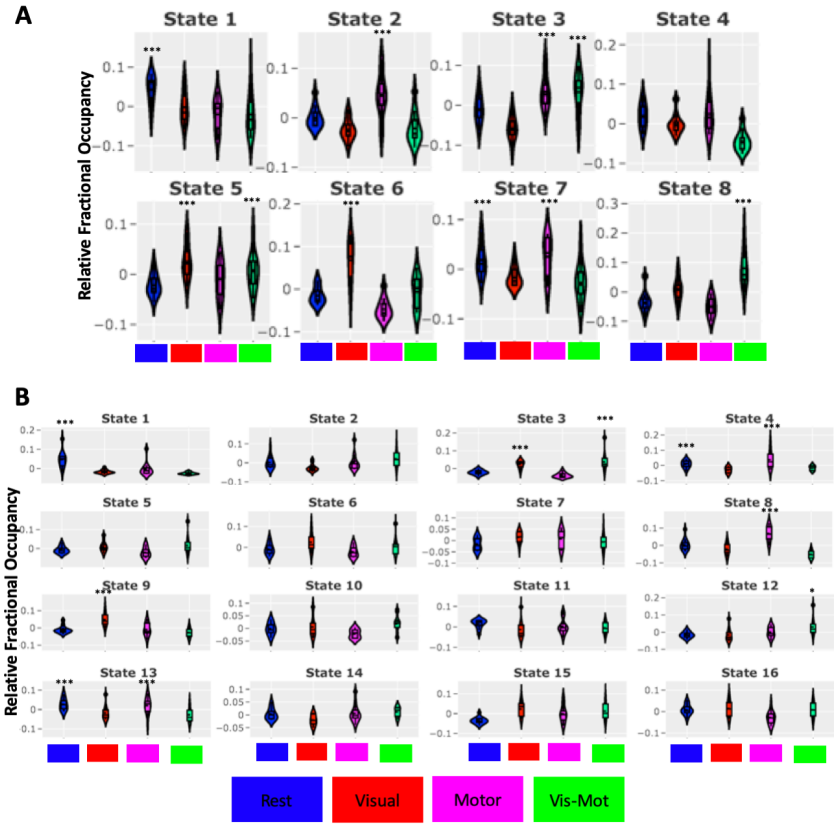


Figure S34: **[A-B]** We illustrated the results of task decoding for the task data set when more than 4 states are modelled. On the y-axis, we have shown the relative fractional occupancy response for each of the modelled states, and the on x-axis, we have highlighted the experimental task at hand. The paired t-test highlights the correspondence of states with the experimental task: **[A]** State 1 is mostly active during the rest, state 2 is active during the motor task, state 3 is active for both motor and visual-motor task, and state 8 is dominantly active for the visual motor task, etc. **[B]** State 1 and State 13 is active during the rest, state 3 and state 9 is active during the visual task, etc.

References

Abbott, A. (2011). Dementia: a problem for our age. *Nature*, *475*(7355), S2–S4.

Abide imaging protocol. (2017). http://fcon_1000.projects.nitrc.org/indi/abide/abide_I.html. (Accessed: 2021-08-30)

Abraham, A., Milham, M. P., Di Martino, A., Craddock, R. C., Samaras, D., Thirion, B., & Varoquaux, G. (2017). Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example. *NeuroImage*, *147*, 736–745.

Abrol, A., Rashid, B., Rachakonda, S., Damaraju, E., & Calhoun, V. D. (2017). Schizophrenia shows disrupted links between brain volume and dynamic functional connectivity. *Frontiers in Neuroscience*, *11*, 624.

Acpi imaging protocol. (2016). https://s3.amazonaws.com/fcp-indi/data/Projects/ACPI/ScanParameters/mta_1_scan_parameters.pdf. (Accessed: 2021-08-30)

Acpi preprocessing protocol. (2016). http://fcon_1000.projects.nitrc.org/indi/ACPI/html/index.html. (Accessed: 2021-08-30)

Alfaro-Almagro, F., Jenkinson, M., Bangerter, N. K., Andersson, J. L., Griffanti, L., Douaud, G., ... others (2018). Image processing and quality control for the first 10,000 brain imaging datasets from uk biobank. *Neuroimage*, *166*, 400–424.

Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., & Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cerebral cortex*, *24*(3), 663–676.

Anand, A., Li, Y., Wang, Y., Lowe, M. J., & Dzemidzic, M. (2009). Resting state corticolimbic connectivity abnormalities in unmedicated bipolar disorder

and unipolar depression. *Psychiatry Research: Neuroimaging*, *171*(3), 189–198.

Andrews-Hanna, J. R., Snyder, A. Z., Vincent, J. L., Lustig, C., Head, D., Raichle, M. E., & Buckner, R. L. (2007). Disruption of large-scale brain systems in advanced aging. *Neuron*, *56*(5), 924–935.

Anttila, V., Bulik-Sullivan, B., Finucane, H. K., Walters, R. K., Bras, J., Duncan, L., ... others (2018). Analysis of shared heritability in common disorders of the brain. *Science*, *360*(6395).

Baker, A. P., Brookes, M. J., Rezek, I. A., Smith, S. M., Behrens, T., Smith, P. J. P., & Woolrich, M. (2014). Fast transient networks in spontaneous human brain activity. *Elife*, *3*, e01867.

Baldassano, C., Hasson, U., & Norman, K. A. (2018). Representation of real-world event schemas during narrative perception. *Journal of Neuroscience*, *38*(45), 9689–9699.

Barachant, A., Bonnet, S., Congedo, M., & Jutten, C. (2011). Multiclass brain–computer interface classification by riemannian geometry. *IEEE Transactions on Biomedical Engineering*, *59*(4), 920–928.

Barachant, A., Bonnet, S., Congedo, M., & Jutten, C. (2013). Classification of covariance matrices using a riemannian-based kernel for bci applications. *Neurocomputing*, *112*, 172–178.

Bastos, A. M., & Schoffelen, J.-M. (2016). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in systems neuroscience*, *9*, 175.

Beckmann, C. F., Mackay, C. E., Filippini, N., & Smith, S. M. (2009). Group comparison of resting-state fmri data using multi-subject ica and dual regression. *Neuroimage*, *47*(Suppl 1), S148.

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic independent component

analysis for functional magnetic resonance imaging. *IEEE transactions on medical imaging*, *23*(2), 137–152.

Bender, R. (2001). Calculating confidence intervals for the number needed to treat. *Controlled clinical trials*, *22*(2), 102–110.

Bijsterbosch, J. D., Beckmann, C. F., Woolrich, M. W., Smith, S. M., & Harrison, S. J. (2019). The relationship between spatial configuration and functional connectivity of brain regions revisited. *Elife*, *8*, e44890.

Bijsterbosch, J. D., Woolrich, M. W., Glasser, M. F., Robinson, E. C., Beckmann, C. F., Van Essen, D. C., ... Smith, S. M. (2018). The relationship between spatial configuration and functional connectivity of brain regions. *Elife*, *7*, e32992.

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic resonance in medicine*, *34*(4), 537–541.

Bluhm, R. L., Miller, J., Lanius, R. A., Osuch, E. A., Boksman, K., Neufeld, R., ... Williamson, P. (2007). Spontaneous low-frequency fluctuations in the bold signal in schizophrenic patients: anomalies in the default network. *Schizophrenia bulletin*, *33*(4), 1004–1012.

Boyd, S., Parikh, N., & Chu, E. (2011). *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease.

Cai, J., Liu, A., Mi, T., Garg, S., Trappe, W., McKeown, M. J., & Wang, Z. J. (2018). Dynamic graph theoretical analysis of functional connectivity in parkinson's disease: The importance of fiedler value. *IEEE journal of biomedical and health informatics*, *23*(4), 1720–1729.

Campbell, R., et al. (2004). *Letters from a stoic: Epistulae morales ad lucilium*. Penguin UK.

Cao, Q., Zang, Y., Sun, L., Sui, M., Long, X., Zou, Q., & Wang, Y. (2006). Abnormal neural activity in children with attention deficit hyperactivity disorder: a resting-state functional magnetic resonance imaging study. *Neuroreport*, *17*(10), 1033–1036.

Charquero-Ballester, M., Kleim, B., Vidaurre, D., Ruff, C., Stark, E., Tuulari, J. J., ... others (2020). Effective psychological treatment for ptsd changes the dynamics of specific large-scale brain networks. *BioRxiv*.

Chen, J., Sun, D., Shi, Y., Jin, W., Wang, Y., Xi, Q., & Ren, C. (2018). Alterations of static functional connectivity and dynamic functional connectivity in motor execution regions after stroke. *Neuroscience letters*, *686*, 112–121.

Chen, Y., Wiesel, A., Eldar, Y. C., & Hero, A. O. (2010). Shrinkage algorithms for mmse covariance estimation. *IEEE Transactions on Signal Processing*, *58*(10), 5016–5029.

Cherkassky, V. L., Kana, R. K., Keller, T. A., & Just, M. A. (2006). Functional connectivity in a baseline resting-state network in autism. *Neuroreport*, *17*(16), 1687–1690.

Cohen, J. R. (2018). The behavioral and cognitive relevance of time-varying, dynamic changes in functional connectivity. *NeuroImage*, *180*, 515–525.

Colclough, G. L., Woolrich, M. W., Harrison, S. J., López, P. A. R., Valdes-Sosa, P. A., & Smith, S. M. (2018). Multi-subject hierarchical inverse covariance modelling improves estimation of functional brain networks. *NeuroImage*, *178*, 370–384.

Cole, J. H., Poudel, R. P., Tsagkrasoulis, D., Caan, M. W., Steves, C., Spector, T. D., & Montana, G. (2017). Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage*,

163, 115–124.

Cox, R. W. (1996). Afni: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research*, *29*(3), 162–173.

Cox, R. W., & Jesmanowicz, A. (1999). Real-time 3d image registration for functional mri. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *42*(6), 1014–1018.

Craddock, C., Sikka, S., Cheung, B., Khanuja, R., Ghosh, S. S., Yan, C., ... others (2013). Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (c-pac). *Front Neuroinform*, *42*, 10–3389.

Dadi, K., Rahim, M., Abraham, A., Chyzhyk, D., Milham, M., Thirion, B., ... others (2019). Benchmarking functional connectome-based predictive models for resting-state fmri. *NeuroImage*, *192*, 115–134.

d'Ambrosio, A., Valsasina, P., Gallo, A., De Stefano, N., Pareto, D., Barkhof, F., ... others (2020). Reduced dynamics of functional connectivity and cognitive impairment in multiple sclerosis. *Multiple Sclerosis Journal*, *26*(4), 476–488.

Damoiseaux, J. S. (2017). Effects of aging on functional and structural brain connectivity. *Neuroimage*, *160*, 32–40.

Damoiseaux, J. S., Beckmann, C., Arigita, E. S., Barkhof, F., Scheltens, P., Stam, C., ... Rombouts, S. (2008). Reduced resting-state brain activity in the âdefault networkâ in normal aging. *Cerebral cortex*, *18*(8), 1856–1864.

Damoiseaux, J. S., Rombouts, S., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., & Beckmann, C. F. (2006). Consistent resting-state networks across healthy subjects. *Proceedings of the national academy of sciences*, *103*(37),

13848–13853.

Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, *29*, 3844–3852.

De Luca, M., Beckmann, C. F., De Stefano, N., Matthews, P. M., & Smith, S. M. (2006). fmri resting state networks define distinct modes of long-distance interactions in the human brain. *Neuroimage*, *29*(4), 1359–1367.

De Pasquale, F., Della Penna, S., Snyder, A. Z., Lewis, C., Mantini, D., Marzetti, L., ... others (2010). Temporal dynamics of spontaneous meg activity in brain networks. *Proceedings of the National Academy of Sciences*, *107*(13), 6040–6045.

De Pasquale, F., Della Penna, S., Snyder, A. Z., Marzetti, L., Pizzella, V., Romani, G. L., & Corbetta, M. (2012). A cortical core for dynamic integration of functional networks in the resting human brain. *Neuron*, *74*(4), 753–764.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... others (2006). An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, *31*(3), 968–980.

Di, X., Fu, Z., Chan, S. C., Hung, Y. S., Biswal, B. B., & Zhang, Z. (2015). Task-related functional connectivity dynamics in a block-designed visual experiment. *Frontiers in human neuroscience*, *9*, 543.

Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., ... others (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, *19*(6), 659–667.

Duff, E. P., Makin, T., Cottaar, M., Smith, S. M., & Woolrich, M. W. (2018). Disambiguating brain functional connectivity. *Neuroimage*, *173*, 540–550.

Eickhoff, S. B., Thirion, B., Varoquaux, G., & Bzdok, D. (2015). Connectivity-based parcellation: Critique and implications. *Human brain mapping*, *36*(12), 4771–4792.

Etkin, A., Prater, K. E., Schatzberg, A. F., Menon, V., & Greicius, M. D. (2009). Disrupted amygdalar subregion functional connectivity and evidence of a compensatory network in generalized anxiety disorder. *Archives of general psychiatry*, *66*(12), 1361–1372.

Fedota, J. R., & Stein, E. A. (2015). Resting-state functional connectivity and nicotine addiction: prospects for biomarker development. *Annals of the New York Academy of Sciences*, *1349*(1), 64.

Filippini, N., Zsoldos, E., Haapakoski, R., Sexton, C. E., Mahmood, A., Allan, C. L., ... others (2014). Study protocol: the whitehall ii imaging sub-study. *BMC psychiatry*, *14*(1), 1–16.

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., ... Constable, R. T. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature neuroscience*, *18*(11), 1664–1671.

Fischl, B. (2012). Freesurfer. *Neuroimage*, *62*(2), 774–781.

Fischl, B., Van Der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., ... others (2004). Automatically parcellating the human cerebral cortex. *Cerebral cortex*, *14*(1), 11–22.

Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L., & Raichle, M. E. (2006). Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proceedings of the National Academy of Sciences*, *103*(26), 10046–10051.

Fox, M. D., & Greicius, M. (2010). Clinical applications of resting state functional connectivity. *Frontiers in systems neuroscience*, *4*, 19.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, *102*(27), 9673–9678.

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., & Penny, W. (2007). Variational free energy and the laplace approximation. *Neuroimage*, *34*(1), 220–234.

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

Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with lstm. *Neural computation*, *12*(10), 2451–2471.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., ... others (2013). The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, *80*, 105–124.

Gonzalez-Castillo, J., & Bandettini, P. A. (2018). Task-based dynamic functional connectivity: Recent findings and open questions. *Neuroimage*, *180*, 526–533.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Generative adversarial networks. *arXiv preprint arXiv:1406.2661*.

Grady, C., Sarraf, S., Saverino, C., & Campbell, K. (2016). Age differences in the functional interactions among the default, frontoparietal control, and dorsal attention networks. *Neurobiology of aging*, *41*, 159–172.

Gratton, C., Laumann, T. O., Nielsen, A. N., Greene, D. J., Gordon, E. M., Gilmore, A. W., ... others (2018). Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron*, *98*(2), 439–452.

Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural networks*, *18*(5-6), 602–610.

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, *100*(1), 253–258.

Greicius, M. D., Srivastava, G., Reiss, A. L., & Menon, V. (2004). Default-mode network activity distinguishes alzheimer's disease from healthy aging: evidence from functional mri. *Proceedings of the National Academy of Sciences*, *101*(13), 4637–4642.

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, *48*(1), 63–72.

Gruber, S. A., Rogowska, J., & Yurgelun-Todd, D. A. (2004). Decreased activation of the anterior cingulate in bipolar patients: an fmri study. *Journal of affective disorders*, *82*(2), 191–201.

Guo, X., Duan, X., Suckling, J., Chen, H., Liao, W., Cui, Q., & Chen, H. (2019). Partially impaired functional connectivity states between right anterior insula and default mode network in autism spectrum disorder. *Human brain mapping*, *40*(4), 1264–1275.

Harrison, S. J., Bijsterbosch, J. D., Segerdahl, A. R., Fitzgibbon, S. P., Farahibozorg, S.-R., Duff, E. P., ... Woolrich, M. W. (2020). Modelling subject variability in the spatial and temporal characteristics of functional modes. *NeuroImage*, *222*, 117226.

Harrison, S. J., Woolrich, M. W., Robinson, E. C., Glasser, M. F., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2015). Large-scale probabilistic functional modes from resting state fmri. *NeuroImage*, *109*, 217–231.

Hcp preprocessing pipeline. (2017). https://www.humanconnectome.org/storage/app/media/documentation/s1200/HCP_S1200_Release_Reference_Manual.pdf. (Accessed: 2021-08-30)

He, T., Kong, R., Holmes, A. J., Nguyen, M., Sabuncu, M. R., Eickhoff, S. B., ... Yeo, B. T. (2018). Do deep neural networks outperform kernel regression for functional connectivity prediction of behavior? *BioRxiv*, 473603.

Higgins, C., Liu, Y., Vidaurre, D., Kurth-Nelson, Z., Dolan, R., Behrens, T., & Woolrich, M. (2021). Replay bursts in humans coincide with activation of the default mode and parietal alpha networks. *Neuron*, *109*(5), 882–893.

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., ... Lerchner, A. (2016). beta-vae: Learning basic visual concepts with a constrained variational framework.

Hindriks, R., Adhikari, M. H., Murayama, Y., Ganzetti, M., Mantini, D., Logothetis, N. K., & Deco, G. (2016). Can sliding-window correlations reveal dynamic functional connectivity in resting-state fmri? *Neuroimage*, *127*, 242–256.

Hmm implementation. (2017). https://github.com/OHBA-analysis/HMM-MAR. (Accessed: 2021-08-30)

Hollander, E., Anagnostou, E., Chaplin, W., Esposito, K., Haznedar, M. M., Licalzi, E., ... Buchsbaum, M. (2005). Striatal volume on magnetic resonance imaging and repetitive behaviors in autism. *Biological psychiatry*, *58*(3), 226–232.

Honnorat, N., Eavani, H., Satterthwaite, T. D., Gur, R. E., Gur, R. C., & Davatzikos, C. (2015). Grasp: geodesic graph-based segmentation with shape priors for the functional parcellation of the cortex. *Neuroimage*, *106*, 207–221.

Huang, G.-H. (2014). Model identifiability. *Wiley StatsRef: Statistics Reference Online*, 1–4.

Hull, J. V., Dokovna, L. B., Jacokes, Z. J., Torgerson, C. M., Irimia, A., & Van Horn, J. D. (2017). Resting-state functional connectivity in autism spectrum disorders: a review. *Frontiers in psychiatry*, *7*, 205.

Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., ... others (2013). Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*, *80*, 360–378.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*(19), 6829–6833.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*(2), 825–841.

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical image analysis*, *5*(2), 143–156.

Kawahara, J., Brown, C. J., Miller, S. P., Booth, B. G., Chau, V., Grunau, R. E., ... Hamarneh, G. (2017). Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, *146*, 1038–1049.

Kieliba, P., Madugula, S., Filippini, N., Duff, E. P., & Makin, T. R. (2019). Large-scale intrinsic connectivity is consistent across varying task demands. *PLoS One*, *14*(4), e0213861.

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Klugah-Brown, B., Luo, C., He, H., Jiang, S., Armah, G. K., Wu, Y., ... Yao, D. (2019). Altered dynamic functional network connectivity in frontal lobe

epilepsy. *Brain topography*, *32*(3), 394–404.

Krishnamoorthy, A., & Menon, D. (2013). Matrix inversion using cholesky decomposition. In *2013 signal processing: Algorithms, architectures, arrangements, and applications (spa)* (pp. 70–72).

Laumann, T. O., Snyder, A. Z., Mitra, A., Gordon, E. M., Gratton, C., Adeyemo, B., ... others (2017a). On the stability of bold fmri correlations. *Cerebral cortex*, *27*(10), 4719–4732.

Laumann, T. O., Snyder, A. Z., Mitra, A., Gordon, E. M., Gratton, C., Adeyemo, B., ... others (2017b). On the stability of bold fmri correlations. *Cerebral cortex*, *27*(10), 4719–4732.

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, *88*(2), 365–411.

Leonardi, N., & Van De Ville, D. (2015). On spurious and real fluctuations of dynamic functional connectivity during rest. *Neuroimage*, *104*, 430–436.

Li, S.-J., Li, Z., Wu, G., Zhang, M.-J., Franczak, M., & Antuono, P. G. (2002). Alzheimer disease: evaluation of a functional mr imaging index as a marker. *Radiology*, *225*(1), 253–259.

Liang, M., & Hu, X. (2015). Recurrent convolutional neural network for object recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 3367–3375).

Liang, M., Zhou, Y., Jiang, T., Liu, Z., Tian, L., Liu, H., & Hao, Y. (2006). Widespread functional disconnectivity in schizophrenia with resting-state functional magnetic resonance imaging. *Neuroreport*, *17*(2), 209–213.

Liaw, A., Wiener, M., et al. (2002). Classification and regression by randomforest. *R news*, *2*(3), 18–22.

Liegeois, R., Laumann, T. O., Snyder, A. Z., Zhou, J., & Yeo, B. T. (2017). In-

terpreting temporal fluctuations in resting-state functional connectivity mri. *Neuroimage*, *163*, 437–455.

Liégeois, R., Li, J., Kong, R., Orban, C., Van De Ville, D., Ge, T., ... Yeo, B. T. (2019). Resting brain dynamics at different timescales capture distinct aspects of human behavior. *Nature communications*, *10*(1), 1–9.

Liégeois, R., Ziegler, E., Phillips, C., Geurts, P., Gómez, F., Bahri, M. A., ... others (2016). Cerebral functional connectivity periodically (de) synchronizes with anatomical constraints. *Brain structure and function*, *221*(6), 2985–2997.

Lindquist, M. A., Xu, Y., Nebel, M. B., & Caffo, B. S. (2014). Evaluating dynamic bivariate correlations in resting-state fmri: a comparison study and a new approach. *NeuroImage*, *101*, 531–546.

Liu, F., Guo, W., Fouche, J.-P., Wang, Y., Wang, W., Ding, J., ... others (2015). Multivariate classification of social anxiety disorder using whole brain functional connectivity. *Brain Structure and Function*, *220*(1), 101–115.

Long, Z., Duan, X., Mantini, D., & Chen, H. (2016). Alteration of functional connectivity in autism spectrum disorder: effect of age and anatomical distance. *Scientific reports*, *6*(1), 1–8.

Lowd, D., & Meek, C. (2005). Adversarial learning. In *Proceedings of the eleventh acm sigkdd international conference on knowledge discovery in data mining* (pp. 641–647).

Lowe, M. J., Phillips, M. D., Lurito, J. T., Mattson, D., Dzemidzic, M., & Mathews, V. P. (2002). Multiple sclerosis: low-frequency temporal blood oxygen level–dependent fluctuations indicate reduced functional connectivityâinitial results. *Radiology*, *224*(1), 184–192.

Lurie, D., Kessler, D., Bassett, D., Betzel, R. F., Breakspear, M., Keilholz, S., ... others (2018). On the nature of time-varying functional connectivity in

resting fmri.

Lurie, D. J., Kessler, D., Bassett, D. S., Betzel, R. F., Breakspear, M., Kheilholz, S., ... others (2020). Questions and controversies in the study of time-varying functional connectivity in resting fmri. *Network Neuroscience*, *4*(1), 30–69.

Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., & Frey, B. (2015). Adversarial autoencoders. *arXiv preprint arXiv:1511.05644*.

Marrelec, G., & Benali, H. (2009). A theoretical investigation of the relationship between structural equation modeling and partial correlation in functional mri effective connectivity. *Computational intelligence and neuroscience*, *2009*.

Mensch, A., Mairal, J., Thirion, B., & Varoquaux, G. (2017). Stochastic subsampling for factorizing huge matrices. *IEEE Transactions on Signal Processing*, *66*(1), 113–128.

Mentalstatesderivation. (2018). https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/MentalStatesDerivation.pdf. (Accessed: 2021-08-30)

Mescheder, L., Nowozin, S., & Geiger, A. (2017). Adversarial variational bayes: Unifying variational autoencoders and generative adversarial networks. In *International conference on machine learning* (pp. 2391–2400).

Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., ... others (2016). Multimodal population brain imaging in the uk biobank prospective epidemiological study. *Nature neuroscience*, *19*(11), 1523–1536.

Moakher, M. (2005). A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, *26*(3), 735–747.

Monk, C. S., Peltier, S. J., Wiggins, J. L., Weng, S.-J., Carrasco, M., Risi, S., & Lord, C. (2009). Abnormalities of intrinsic functional connectivity in autism spectrum disorders. *Neuroimage*, *47*(2), 764–772.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Nastase, S. A., Gazzola, V., Hasson, U., & Keysers, C. (2019). *Measuring shared responses across subjects using intersubject correlation.* Oxford University Press.

Neuroticism score. (2018). https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/MentalStatesDerivation.pdf. (Accessed: 2021-08-30)

Ng, B., Varoquaux, G., Poline, J. B., Greicius, M., & Thirion, B. (2015). Transport on riemannian manifold for connectivity-based brain decoding. *IEEE transactions on medical imaging*, *35*(1), 208–216.

Ng, B., Varoquaux, G., Poline, J. B., & Thirion, B. (2013). A novel sparse group gaussian graphical model for functional connectivity estimation. In *International conference on information processing in medical imaging* (pp. 256–267).

Nielsen, J. A., Zielinski, B. A., Fletcher, P. T., Alexander, A. L., Lange, N., Bigler, E. D., ... Anderson, J. S. (2013). Multisite functional connectivity mri classification of autism: Abide results. *Frontiers in human neuroscience*, *7*, 599.

Onoda, K., Ishihara, M., & Yamaguchi, S. (2012). Decreased functional connectivity by aging is associated with cognitive decline. *Journal of cognitive neuroscience*, *24*(11), 2186–2198.

O'Reilly, J. X., Woolrich, M. W., Behrens, T. E., Smith, S. M., & Johansen-Berg, H. (2012). Tools of the trade: psychophysiological interactions and functional connectivity. *Social cognitive and affective neuroscience*, *7*(5),

604–609.

Parisot, S., Arslan, S., Passerat-Palmbach, J., Wells III, W. M., & Rueckert, D. (2016). Group-wise parcellation of the cortex through multi-scale spectral clustering. *NeuroImage*, *136*, 68–83.

Parisot, S., Glocker, B., Schirmer, M. D., & Rueckert, D. (2016). Grampa: Graph-based multi-modal parcellation of the cortex using fusion moves. In *International conference on medical image computing and computer-assisted intervention* (pp. 148–156).

Pariyadath, V., Stein, E. A., & Ross, T. J. (2014). Machine learning classification of resting state functional connectivity predicts smoking status. *Frontiers in human neuroscience*, *8*, 425.

Parr, T., Rees, G., & Friston, K. J. (2018). Computational neuropsychology and bayesian inference. *Frontiers in human neuroscience*, *12*, 61.

Patrick, C. J., & Hajcak, G. (2016). Rdoc: Translating promise into progress. *Psychophysiology*, *53*(3), 415–424.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics surveys*, *3*, 96–146.

Pennec, X., Fillard, P., & Ayache, N. (2006). A riemannian framework for tensor computing. *International Journal of computer vision*, *66*(1), 41–66.

Penny, W. D., Stephan, K. E., Mechelli, A., & Friston, K. J. (2004). Modelling functional integration: a comparison of structural equation and dynamic causal models. *Neuroimage*, *23*, S264–S274.

Pervaiz, U., Vidaurre, D., Woolrich, M. W., & Smith, S. M. (2020). Optimising network modelling methods for fmri. *Neuroimage*, *211*, 116604.

Quinn, A. J., Vidaurre, D., Abeysuriya, R., Becker, R., Nobre, A. C., & Woolrich, M. W. (2018). Task-evoked dynamic network analysis through hidden markov modeling. *Frontiers in neuroscience*, *12*, 603.

Rahim, M., Thirion, B., & Varoquaux, G. (2019). Population shrinkage of covariance (posce) for better individual brain functional-connectivity estimation. *Medical image analysis*, *54*, 138–148.

Raichle, M. E. (2015). The restless brain: how intrinsic activity organizes brain function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1668), 20140172.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, *98*(2), 676–682.

Rashid, B., Damaraju, E., Pearlson, G. D., & Calhoun, V. D. (2014). Dynamic connectivity states estimated from resting fmri identify differences among schizophrenia, bipolar disorder, and healthy control subjects. *Frontiers in human neuroscience*, *8*, 897.

Rogers, B. P., Morgan, V. L., Newton, A. T., & Gore, J. C. (2007). Assessing functional connectivity in the human brain by fmri. *Magnetic resonance imaging*, *25*(10), 1347–1357.

Sakoğlu, Ü., Pearlson, G. D., Kiehl, K. A., Wang, Y. M., Michael, A. M., & Calhoun, V. D. (2010). A method for evaluating dynamic functional network connectivity and task-modulation: application to schizophrenia. *Magnetic Resonance Materials in Physics, Biology and Medicine*, *23*(5-6), 351–366.

Sala-Llonch, R., Smith, S. M., Woolrich, M., & Duff, E. P. (2019). Spatial parcellations, spectral filtering, and connectivity measures in fmri: Optimizing for discrimination. *Human brain mapping*, *40*(2), 407–419.

Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., & Smith, S. M. (2014). Automatic denoising of functional mri data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage*, *90*, 449–468.

Salvador, R., Martinez, A., Pomarol-Clotet, E., Sarró, S., Suckling, J., & Bullmore, E. (2007). Frequency based mutual information measures between clusters of brain regions in functional magnetic resonance imaging. *Neuroimage*, *35*(1), 83–88.

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

Seeley, W. W., Allman, J. M., Carlin, D. A., Crawford, R. K., Macedo, M. N., Greicius, M. D., ... Miller, B. L. (2007). Divergent social functioning in behavioral variant frontotemporal dementia and alzheimer disease: reciprocal networks and neuronal evolution. *Alzheimer Disease & Associated Disorders*, *21*(4), S50–S57.

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., ... Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience*, *27*(9), 2349–2356.

Shen, X., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X., & Constable, R. T. (2017). Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *nature protocols*, *12*(3), 506–518.

Shen, X., Tokoglu, F., Papademetris, X., & Constable, R. T. (2013). Groupwise whole-brain parcellation from resting-state fmri data for network node identification. *Neuroimage*, *82*, 403–415.

Shimizu, S., Hoyer, P. O., Hyvärinen, A., Kerminen, A., & Jordan, M. (2006). A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, *7*(10).

Shine, J. M., Bissett, P. G., Bell, P. T., Koyejo, O., Balsters, J. H., Gorgolewski, K. J., ... Poldrack, R. A. (2016). The dynamics of functional brain networks: integrated network states during cognitive task performance. *Neuron*, *92*(2), 544–554.

Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., & Vandergheynst, P. (2013). The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE signal processing magazine*, *30*(3), 83–98.

Smith, D. J., Nicholl, B. I., Cullen, B., Martin, D., Ul-Haq, Z., Evans, J., ... others (2013). Prevalence and characteristics of probable major depression and bipolar disorder within uk biobank: cross-sectional study of 172,751 participants. *PloS one*, *8*(11), e75362.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human brain mapping*, *17*(3), 143–155.

Smith, S. M. (2004). Overview of fmri analysis. *The British Journal of Radiology*, *77*(suppl_2), S167–S175.

Smith, S. M. (2012). The future of fmri connectivity. *Neuroimage*, *62*(2), 1257–1266.

Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., ... others (2013). Resting-state fmri in the human connectome project. *Neuroimage*, *80*, 144–168.

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., ... others (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the national academy of sciences*, *106*(31), 13040–13045.

Smith, S. M., Hyvärinen, A., Varoquaux, G., Miller, K. L., & Beckmann, C. F. (2014). Group-pca for very large fmri datasets. *Neuroimage*, *101*, 738–

749.

Smith, S. M., Miller, K. L., Moeller, S., Xu, J., Auerbach, E. J., Woolrich, M. W., ... others (2012). Temporally-independent functional modes of spontaneous brain activity. *Proceedings of the National Academy of Sciences*, *109*(8), 3131–3136.

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., ... Woolrich, M. W. (2011). Network modelling methods for fmri. *Neuroimage*, *54*(2), 875–891.

Smith, S. M., Nichols, T. E., Vidaurre, D., Winkler, A. M., Behrens, T. E., Glasser, M. F., ... Miller, K. L. (2015). A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature neuroscience*, *18*(11), 1565–1567.

Smith, S. M., Vidaurre, D., Beckmann, C. F., Glasser, M. F., Jenkinson, M., Miller, K. L., ... others (2013). Functional connectomics from resting-state fmri. *Trends in cognitive sciences*, *17*(12), 666–682.

Snoek, L., Miletić, S., & Scholte, H. S. (2019). How to control for confounds in decoding analyses of neuroimaging data. *Neuroimage*, *184*, 741–760.

Sorg, C., Riedl, V., Mühlau, M., Calhoun, V. D., Eichele, T., Läer, L., ... others (2007). Selective changes of resting-state networks in individuals at risk for alzheimer's disease. *Proceedings of the National Academy of Sciences*, *104*(47), 18760–18765.

Supekar, K., Menon, V., Rubin, D., Musen, M., & Greicius, M. D. (2008). Network analysis of intrinsic functional brain connectivity in alzheimer's disease. *PLoS computational biology*, *4*(6), e1000100.

Tagliazucchi, E., Von Wegner, F., Morzelewski, A., Brodbeck, V., & Laufs, H. (2012). Dynamic bold functional connectivity in humans and its electrophysiological correlates. *Frontiers in human neuroscience*, *6*, 339.

Tavor, I., Jones, O. P., Mars, R. B., Smith, S., Behrens, T., & Jbabdi, S. (2016). Task-free mri predicts individual differences in brain activity during task performance. *Science*, *352*(6282), 216–220.

Thirion, B., Varoquaux, G., Dohmatob, E., & Poline, J.-B. (2014). Which fmri clustering gives good brain parcellations? *Frontiers in neuroscience*, *8*, 167.

Tian, L., Jiang, T., Wang, Y., Zang, Y., He, Y., Liang, M., ... others (2006). Altered resting-state functional connectivity patterns of anterior cingulate cortex in adolescents with attention deficit hyperactivity disorder. *Neuroscience letters*, *400*(1-2), 39–43.

Tvfc summary measures. (2017). https://github.com/OHBA-analysis/HMM-MAR/wiki/User-Guide. (Accessed: 2021-08-30)

Uk bibank preprocessing pipeline. (2020). https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf. (Accessed: 2021-08-30)

Valdes-Sosa, P. A., Roebroeck, A., Daunizeau, J., & Friston, K. (2011). Effective connectivity: influence, causality and biophysical modeling. *Neuroimage*, *58*(2), 339–361.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., ... others (2013). The wu-minn human connectome project: an overview. *Neuroimage*, *80*, 62–79.

Varoquaux, G., Baronnet, F., Kleinschmidt, A., Fillard, P., & Thirion, B. (2010). Detection of brain functional-connectivity difference in post-stroke patients using group-level covariance modeling. In *International conference on medical image computing and computer-assisted intervention* (pp. 200–208).

Vidaurre, D., Abeysuriya, R., Becker, R., Quinn, A. J., Alfaro-Almagro, F., Smith, S. M., & Woolrich, M. W. (2018). Discovering dynamic brain networks from big data in rest and task. *Neuroimage*, *180*, 646–656.

Vidaurre, D., Bielza, C., & Larrañaga, P. (2013). A survey of l1 regression. *International Statistical Review*, *81*(3), 361–387.

Vidaurre, D., Hunt, L. T., Quinn, A. J., Hunt, B. A., Brookes, M. J., Nobre, A. C., & Woolrich, M. W. (2018). Spontaneous cortical activity transiently organises into frequency specific phase-coupling networks. *Nature communications*, *9*(1), 1–13.

Vidaurre, D., Llera, A., Smith, S. M., & Woolrich, M. W. (2021). Behavioural relevance of spontaneous, transient brain network interactions in fmri. *NeuroImage*, *229*, 117713.

Vidaurre, D., Smith, S. M., & Woolrich, M. W. (2017). Brain network dynamics are hierarchically organized in time. *Proceedings of the National Academy of Sciences*, *114*(48), 12827–12832.

Vidaurre, D., van Gerven, M. A., Bielza, C., Larrañaga, P., & Heskes, T. (2013). Bayesian sparse partial least squares. *Neural computation*, *25*(12), 3318–3339.

Volkow, N. D., Fowler, J. S., Wang, G.-J., & Goldstein, R. Z. (2002). Role of dopamine, the frontal cortex and memory circuits in drug addiction: insight from imaging studies. *Neurobiology of learning and memory*, *78*(3), 610–624.

Wang, C., Ong, J. L., Patanaik, A., Zhou, J., & Chee, M. W. (2016). Spontaneous eyelid closures link vigilance fluctuation with fmri dynamic connectivity states. *Proceedings of the National Academy of Sciences*, *113*(34), 9653–9658.

Xie, H., Zheng, C. Y., Handwerker, D. A., Bandettini, P. A., Calhoun, V. D.,

Mitra, S., & Gonzalez-Castillo, J. (2019). Efficacy of different dynamic functional connectivity methods to capture cognitively relevant information. *NeuroImage*, *188*, 502–514.

Zadbood, A., Chen, J., Leong, Y. C., Norman, K. A., & Hasson, U. (2017). How we transmit memories to other brains: constructing shared neural representations via communication. *Cerebral cortex*, *27*(10), 4988–5000.

Zhi, D., Calhoun, V. D., Lv, L., Ma, X., Ke, Q., Fu, Z., ... others (2018). Aberrant dynamic functional network connectivity and graph properties in major depressive disorder. *Frontiers in psychiatry*, *9*, 339.

Zhu, C., Zang, Y., Liang, M., Tian, L., He, Y., Li, X., ... Jiang, T. (2005). Discriminative analysis of brain function at resting-state for attention-deficit/hyperactivity disorder. In *International conference on medical image computing and computer-assisted intervention* (pp. 468–475).

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), 301–320.

Zuo, X.-N., Xu, T., Jiang, L., Yang, Z., Cao, X.-Y., He, Y., ... Milham, M. P. (2013). Toward reliable characterization of functional homogeneity in the human brain: preprocessing, scan duration, imaging resolution and computational space. *Neuroimage*, *65*, 374–386.