Computational Tools for Neuroscience at Different Scales

Emanuele Sicurella

Supervised by:
Dr. Jiaxiang Zhang

School of Psychology,
Cardiff University

A thesis submitted for the degree of

Doctor of Philosophy

27 November 2023

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Prof. Jiaxiang Zhang, for his expert guidance, patience and friendliness. I particularly appreciated the support I received during a challenging period of my research and life in general. I have learned and grown a lot, both personally and professionally, and for this, I'm deeply grateful. Thank you. I also want to thank my advisors, Prof. Andrew Lawrence, for his valuable feedback and insights and Dr Matthias Gruber, who followed my research progress and provided expert advice.

I am grateful to my lab mates at the CCBrain lab and my other fellow PhD students at CUBRIC for their collaborative spirits and intellectual discussions, which helped me grow as a researcher and as a person. In particular, I want to acknowledge Dr Esin Karahan, Dr Luke Tait and Roguang Si for performing the acquisition and the processing of the data I used in Chapter 3 (DWI, MEG and fMRI, respectively).

I want to acknowledge the generous funding from the European Research Council (716321) and the access to the facilities and the hardware provided by Cardiff University, which was instrumental in my research.

A special mention goes to my parents, Agata and Giovanni, for their continuous support in the last 32 years and a half, even when I did not deserve it. Thank you for everything. I also want to express my gratitude to my cousins Giulia and Adriana (who are more like my sisters), my other relatives and all my friends who always had nice words for me and have believed in me even when I did not.

Lastly, I thank every individual who contributed in ways both big and small to the completion of this thesis. Without your support and encouragement, this achievement would have not been possible.

Grazie.

Abstract

Nowadays, computational tools are essential and ubiquitous in basically every research field. In particular, they enormously helped Psychology and Neuroscience by providing computational models of cognition, description of neurobiological phenomena, data analysis tools and many more. In this thesis, I will propose a classification method based on the scale of the problem under investigation, providing the results of applying three different computational tools for neuroscience at different scales.

First, I perform graph theory-based analysis of whole human structural (DWI) and functional (MEG and fMRI) connectomes using a new approach based on small, induced, connected subgraphs called graphlets. I will show that graphlet provides an elegant and effective way to represent and characterise topological information of brain networks without the need for numerous classical graph-theory measures.

Second, I apply deep learning for parameter recovery of a perceptual decision-making model simulating the evidence accumulation of the LIP area of the brain. Deep learning offers a valuable tool for parameter recovery of more complex biologically plausible models. However, I also stress that successful parameter recovery depends not only on the choice of the tool but also on the careful design of the experiment to avoid the recovery of parameters that have a similar effect on the output, thus making parameter recovery difficult.

Finally, I perform neuronal decoding of spiking neuron activity during a 2D reaching task performed by two monkeys. Velocity decoding performances are generally better than position or acceleration decoding. I also study the effect of PCA on dimensionality reduction of neural data and consequent neuronal decoding, showing that general performances on reduced data are lower except for position decoding from PMd cortex activity of many neurons. Following previous research results, I hypothesise this is caused by the encoding of different processes in the PMd cortex not related to the simple forward motor output.

Overall, this work explores the use of powerful computational tools to solve problems in neuroscience at different scales.

Contents

1 Introduction **1**

1.1 Structure of the thesis 3

2 The model hierarchy of neural networks **5**

2.1 Modelling the brain 5

2.1.1 Single neuron models 7

2.1.1.1 The Integrate-and-Fire Model 9

2.1.1.2 The Hodgkin-Huxley Model 11

2.1.2 Population models and neural mass models 13

2.1.2.1 Dynamical System Theory of Neural Mass Models . . 16

2.1.2.2 The Wilson-Cowan model 17

2.1.3 Modelling perceptual decision-making 19

2.1.3.1 The Wang-Wong model 21

2.2 Graph representations of biological neural networks 23

2.2.1 Graph Theory, Network Science and biological networks . . . 25

2.2.2 Network Neuroscience 31

2.2.3 Novel graph metric: Graphlets and Orbits 33

2.2.3.1 Graphlet Degree Vector 33

2.2.3.2 Graphlet Correlation Matrix 35

2.3 Potential contributions of Machine Learning to Neuroscience 35

2.3.1 Supervised vs. Unsupervised learning 36

2.3.2 Principal Component Analysis (PCA) 38

2.3.3 Deep Learning 39

2.3.4 Deep Learning for parameter recovery 42

2.3.5 Decoding continuous behaviour using Deep Learning 44

2.3.5.1 Latent dynamics analysis of neuronal data 44

3 Graphlet analysis of structural and functional connectomes 46
3.1 Introduction . . . 46
3.2 Methods . . . 48
3.2.1 Neuroimaging . . . 48
3.2.1.1 Magnetic Resonance Imaging . . . 48
3.2.1.2 Diffusion-Weighted Imaging . . . 50
3.2.1.3 Functional MRI . . . 51
3.2.1.4 Magnetoencephalogram . . . 52
3.2.2 Imaging data acquisition and processing . . . 53
3.2.2.1 Participants . . . 53
3.2.2.2 MRI data acquisition . . . 53
3.2.2.3 MEG data acquisition . . . 54
3.2.2.4 Cortical reconstruction . . . 54
3.2.2.5 DWI data preprocessing . . . 55
3.2.2.6 Tractography . . . 56
3.2.2.7 Structural Connectome . . . 57
3.2.2.8 fMRI preprocessing . . . 57
3.2.2.9 fMRI derived functional connectome . . . 57
3.2.2.10 MEG preprocessing . . . 58
3.2.2.11 MEG source reconstruction . . . 58
3.2.2.12 MEG derived functional connectome . . . 59
3.2.2.13 Network binarization . . . 59
3.2.3 Graphlets and Orbits . . . 60
3.2.4 Brain plots on central orbits proportions . . . 60
3.2.5 Associating between modality-specific and theoretical GCMs . 60
3.2.6 Associating between Graphlet and classical graph-based metrics 61
3.3 Results . . . 61
3.3.1 Graphlet correlation matrices from multi-modal whole-brain connectome . . . 61
3.3.2 Spatial variance in graphlet orbit distribution . . . 62
3.3.3 Comparisons with theoretical-derived GCMs . . . 62
3.3.4 Correlation between graphlet and classical graph-based metrics 63
3.4 Discussion . . . 63

4 Deep Learning for parameter recovery of a perceptual decision making model **72**
4.1 Introduction . . . 72
4.1.1 Decision making and sequential sampling models . . . 72
4.1.2 Parameter recovery . . . 73
4.1.2.1 Deep Learning for Parameter Recovery . . . 74
4.1.2.2 Deep Learning as a model of cognition . . . 75
4.2 Methods . . . 75
4.2.1 Neural mass models . . . 75
4.2.2 Data sampling and processing . . . 76
4.2.3 Deep Learning for parameter recovery . . . 80
4.2.4 Neural ensemble . . . 80
4.3 Results . . . 85
4.4 Discussion . . . 86

5 Deep Learning for Neural Decoding of Motor Behaviour **92**
5.1 Introduction . . . 92
5.2 Methods . . . 93
5.2.1 Data collection . . . 93
5.2.2 Data preparation . . . 96
5.2.3 Neural decoding . . . 96
5.3 Results . . . 97
5.4 Discussion . . . 99

6 Discussion **111**
6.1 Summary of the results . . . 111
6.2 Discussion . . . 113
6.3 Limitations and future directions . . . 116

Bibliography **120**

List of Figures

2.1 Anatomy of the neuron 8
2.2 Spike train of the IF neuron 10
2.3 Spiking rates predicted by Mean Field approximation. 14
2.4 Phase portrait of single excitatory population 18
2.5 Phase plane of the Wilson-Cowan model. 19
2.6 Phase plane of the Wang-Wong model. 24
2.7 Example of a graph with some properties outlined. 26
2.8 Graphlet-related methods 34

3.1 Connectivity matrices and GCMs 64
3.2 OCP cortical plots 65
3.3 GCM of ideal models and GCD with real networks 66
3.4 GCD between clusters and model's GCM 67
3.5 Partial correlation coefficients 68

4.1 Original accuracy and reaction time versus coherence plots. 77
4.2 Accuracy and reaction time versus coherence plots 78
4.3 Schematisation of the data sampling and parameter recovery pipeline. 81
4.4 Simplified low diagram of the parameter recovery procedure. 82
4.5 Distribution of model parameters 83
4.6 Schematisation of the CNN + RNN ensemble for RDK discrimination task. 84
4.7 Parameter predicted by the model vs real values for 4 parameters . . 87
4.8 Parameter predicted by the model vs real values for 3 parameters . . 88
4.9 Parameter recovery performances at different number of trials 89
4.10 Accuracy versus coherence plot of a CNN + RNN neural ensemble compared to the Wang-Wong model. 89

5.1 Schematisation of the reaching task performed by a monkey. 94
5.2 Schematisation of neural decoding procedures. 98

5.3 Neurons' connectivity matrices and GCMs. 100
5.4 Position decoding 101
5.5 Velocity decoding 102
5.6 Acceleration decoding 103
5.7 Explained variance of principal components from PCA 104
5.8 Position decoding from 90% PCA reduced neural data 105
5.9 Velocity decoding from 90% PCA reduced neural data 106
5.10 Acceleration decoding from 90% PCA reduced neural data 107
5.11 Decoding performance with different time bins number in input data 108

List of Tables

4.1 Original parameters of the Wang-Wong model with minimum and maximum values for parameter recovery. 76
4.2 Hyperparameter values for optimal performances of the algorithms trained for the 3 cases of parameter recovery. 86

Abbreviations

AEC	Amplitude Envelope Correlation
AI	Artificial Intelligence
ANN	Artificial Neural Network
BCI	Brain Computer Interface
BCT	Brain Connectivity Toolbox
CCA	Canonical Correlation Analysis
CCS	Connectome Computation System
CNN	Convolutional Neural Network
CSF	Cerebro-Spinal Fluid DDM Drift Diffusion Model
DL	Deep Learning
DNN	Deep Neural Network
DWI	Diffusion Weighted Imaging
EOG	Electro-Oculograms
EPI	Echoplanar Imaging
FC	Functional Connectivity
FNN	Forward Neural Network
GCD	Graphlet Correlation Distance
GCM	Graphlet Correlation Matrix
GDV	Graphlet Degree Vector
GM	Grey Matter
HCP	Human Connectome Project
ICA	Independent Component Analysis
LIP	Lateral Intraparietal
M1	Primary Motor Cortex
MEG	Magnetoencephalogram
MLP	Multilayer Perceptron
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error

NaN	Not a Number
NMM	Neural Mass Model
OCP	Orbit Centrality Percentage
PCA	Principal Component Analysis
PLS	Partial Least Squares
PMd	Premotor Dorsal Cortex
RDK	Random Dot Kinematogram
RNN	Recurrent Neural Network
ROI	Region of Interest
SC	Structural Connectivity
SSM	Sequential Sampling Model
WM	White Matter

Chapter 1

Introduction

The human brain is a remarkable machine composed of billions of elements which interact in a highly complex and structured way. Such interactions are what define ourselves, our personalities, fears and desires. It allows us to interact with the external world, make long-term plans and effectively achieve complex goals. For decades, researchers from various fields have tried to demystify how the brain is able to accomplish this extensive set of tasks in such an efficient way. However, a comprehensive understanding of how the brain's structure and functioning translates into such winning problem-solving capabilities still needs to be achieved.

The human brain is one of the best information encoders currently known. Every fraction of a second, the brain receives enormous amounts of information from the external world through sensory inputs. Since our birth, such sensory input is constantly acquired, transformed into some lower-dimensional representation (the "code" of the brain) and elaborated by combining information and generating valuable blueprints for accomplishing complex goals.

The exact mechanism through which the brain performs such information computation is another hurdle that persists in this burgeoning realm of study. With just a few grams of sugar, the brain can quickly perform tasks that even the most sophisticated computer today performs with lower accuracy, such as recognising a person's face or understanding other people's emotions.

Given the high complexity of the human brain and the large amount of data involved in its study, researchers have found the use of computational tools applied to psychology, neuroscience and related disciplines highly beneficial. Computers dramatically helped understand the brain's structure and functioning, facilitating various methodologies. From simulations of individual neurons up to entire networks, advanced data analysis techniques, and the adoption of Artificial Intelligence as a model of brain areas and cognitive processes or as predictive models for psychology

and neuroscience-related tasks. One of the many possible ways to classify different computational tools for studying the human brain is by considering the scale of the specific approach. In the remainder of this chapter, I will follow a bottom-up approach, going from the finer scale to the largest one.

Suppose we want to start limiting ourselves to directing our attention towards the fundamental building block of the brain and the broader nervous system, namely the neuron. In that case, we already encounter a highly complex entity exhibiting correspondingly complex behaviours. A neuron can be represented in many different ways. The choice of representation generally depends on the specific research question we want to answer, as recreating a perfect replica will leave us with a model that is as complex as the system we want to study, thus limiting the knowledge we can extract from it. As we will see, this argument is valid for any model, not only in neuroscience. To study single neurons, we may opt for a biologically plausible model that considers the dynamics of ions and channels. Alternatively, we could simplify it into a mathematical construct that integrates diverse input sources and produces outputs through nonlinear operations.

Proceeding on a wider scale, the next step will consist of putting together several neurons to study how their interaction works. Similarly to the single-neuron case, different models can be proposed depending on what biological detail to include, if any. We could consider, for example, the excitatory and inhibitory connection between neurons or the dynamics of the neurotransmitters near the synapses. As we increase the number of neurons in the model, we should cope with increasing variables, which will require greater computational resources to manage. At some point, we may average the activity of several neurons of the same type (e.g. excitatory or inhibitory) or discard any biological detail and consider just an interaction of nodes connected by a link. This approach will be helpful in studying how local populations of neurons or entire regions specialise themselves to accomplish specific goals.

The latter approach can be gradually extended up to the broadest possible scale, the whole brain [1]. Such description is typical in the approach of Graph Theory or Network Science, which intersects significantly with broader fields like Complex System Theory. Here, the emphasis is typically placed on the connectivity patterns within the network rather than the individual component properties. In this way, the property of a system can be described by specific statistics that can refer to the single node or even to the global system. To name a notable application, node-wise

[1] In reality, an even broader scale may exist if we consider social networks of interacting brains [54].

statistics can be used to understand how structure is related to function. In contrast, global statistics can characterise different groups of brains, for example, to identify neurological diseases or other specific conditions.

Recently, another computational approach has exploded, providing numerous insights about the human brain at the different scales I described. Artificial Intelligence (AI) is one of the most active research fields today, and it is transforming the way we approach the study of the human brain. AI is not only a tool through which we solve complex problems but also a model of cognition which can be both inspired by some functioning of the brain and helpful in understanding some aspects of cognition. I will present an application of AI mainly as a powerful predictive tool while also shortly discussing how it can be used to model a single brain area which accomplishes a specific cognitive task.

In this thesis, I will present a robust set of state-of-the-art computational tools for neuroscience at different scales. Particular attention will be paid to a set of results after applying such tools to study a specific problem. Although this thesis is far from being an exhaustive list of the vast field of computational neuroscience, it will present cutting-edge approaches that will be extensively discussed, including the main limitations and potential future development. After the introduction, I will follow a top-bottom approach to present the results, starting from the broadest possible scale, *i.e.*, the study of the whole brain, down to population models, which use a mean-field approximation, to finally reach the analysis of data coming from single neurons. By employing these sophisticated methodologies, we strive to progress towards the ambitious goal of uncovering the secrets of what many consider the most complex object in the known universe.

1.1 Structure of the thesis

In this thesis, Chapter 2 will provide an overview of some of the most used computational tools for neuroscience at different scales. I will start from the lower level of analysis, presenting some of the most successful single-neuron models. Then, I will discuss how several interacting neurons can be modelled as populations using a mean-field approximation approach, and how population models can explain complex tasks such as perceptual decision-making. A discussion on graph theory and network science will follow, focusing on how to use these tools to study the whole brain, neglecting the single neurons' specific properties, and I will review a novel approach that uses small induced connected subgraphs called graphlets. Finally, I will give a

quick overview of the broad field of Machine Learning, focusing specifically on how it can solve two interesting neuroscience problems: parameter recovery and neural decoding.

Each of the three subsequent chapters starts with a brief introduction of the specific tool application. Then, I will describe in detail the data and the methods used. Finally, the results will be presented and discussed in detail. In particular, Chapter 3 will focus on the graphlet approach to graph data constructed from three different neuroimaging methods: DWI, fMRI and MEG. In Chapter 4, I will use a Deep Learning algorithm for parameter recovery of a simple yet powerful perceptual decision-making model. I will also present a proof-of-concept of a simple neural ensemble to simulate the output of the perceptual decision-making model, which in turn effectively model a specific brain area involved in evidence accumulation during a competing task. In Chapter 5, I will apply a typical neural decoding approach to understand complex data from monkeys performing a 2D reaching task. I will also apply a dimensionality reduction technique to study how the predictive performance changes when using reduced data as input. Overall, I will present a non-exhaustive yet robust set of computational tools for neuroscience at different scales.

In Chapter 6, I will summarise the results, discussing in particular the limitations and possible directions for future research.

Chapter 2

The model hierarchy of neural networks

This chapter aims to provide a high-level overview of some of the most successful computational tools for neuroscience at different scales and present in detail some specific approaches which will be used in later chapters.

2.1 Modelling the brain

The human brain's activity is basically due to neurons, electrically excitable cells that exchange electrical and chemical signals through connections called synapses. The average adult human brain contains around 86 billion neurons, and each of the 86 billion neurons has 7000 synaptic connections to other neurons on average [237]. It has been estimated that the brain of a three-year-old child has about 10^{15} synapses while for an adult, the numbers range from 10^{14} to $5 \cdot 10^{14}$ [50]. These numbers, as well as the large variety of tasks it can perform, should give an overview of the high degree of complexity of the human brain. To understand such a complex system, we may take advantage of using simplified computational modelling [44].

A model can be defined as a simplified description of a system or process which can make predictions. The keyword "simplified" here is highly important because a model which is as complex as the system we want to study does not help better understand such a system. One of the most popular ways to categorise models, in general, is the three-level organisation initially proposed by David Marr to investigate visual systems [144]. Following this definition, a model of a system can be divided into three different levels of abstraction:

1. The *computational* level only models the system's goal without considering any specific detail of the components.

2. The *algorithmic* or *representation level* is a description of a possible implementation of a system.

3. The *implementation level* is an accurate description of how elementary components interact together to generate the system.

To better understand these levels, I will present two examples. These are not necessarily realistic examples, and they serve purely as an illustration of the reasoning approach.

First, suppose we want to create a model of the areas of the brain involved in distinguishing a cat from a dog. A potential top-down approach would start from the computational description (first level). This represents the system's goal, which might be formulated as "quickly identifying discriminating features of an animal which can be assigned to one and only one category with high precision". Moving to the representational level, we have some choices in acquiring and processing the data to identify valuable features. We could first identify elementary features like lines or edges and hierarchically combine them to create more abstract features which will differ between cats and dogs. Eventually, we reach the implementation level where we might, for example, use artificial spiking neurons which form several hierarchically connected layers; the earliest neurons activate preferentially when they detect simple edges or lines and pass this information in the form of activation pattern to the successive layers which combine such information to create more complex features.

The top-down structured flow is one of many possible. As a second example, suppose we want to understand how neurons communicate with each other to generate complex behaviour. Using a CPU, we might create a biologically detailed computational replica of a neuron. Under some controlled conditions, we might study the excitatory and inhibitory connections between two neurons and gradually increase the number of neurons involved (the implementation level). We observe a synchronised behaviour at a certain point where several neurons fire simultaneously. We hypothesise that synchrony is the critical feature through which several neurons represent specific information (the algorithmic or representation level). After combining a different set of connected neurons, we observe that hierarchical synchronisation leads to a system able to solve complex tasks (the computational level). What is just presented is an example of a bottom-up approach where we start looking at the essential components and gradually gain knowledge about larger-scale phenomena.

Depending on the specific research question, a successful model may only require some of the three levels of description. Consider, for example, Artificial Intelligence.

Today, AI algorithms are broadly used to solve tasks with super-human performances (for example, controlling robotic arms to perform even delicate tasks [6], understand questions provided by humans and give detailed elaborate answers [166], autonomous vehicle driving [248], to name a small portion of AI applications). On the other hand, a high effort is being put into making AI smarter to perform tasks that, at the moment, only humans can do effectively: making long-term life plans, writing code to create an app, and publishing an original research paper are just some examples of tasks in which humans still outperform AI [85, 127]. Such algorithms are often not made of essential components like the human brain. Consequently, it is not guaranteed that the representations and algorithms these systems use are comparable to those in the brain, although biology might inspire them [85].

In the following sections, I will start from the lower scale. I will first present two of the most popular single-neuron models that will serve as a starting point for the bottom-up approach to construct computational model from the smallest to the largest possible scale and as an example of how the model changes depending on the biological details that are introduced. Later, I will describe a powerful approach to studying groups of connected IF neurons and how a simple two-population model can replicate the behavioural output of a perceptual decision-making task. I will then describe the graph theory approach to large-scale brain networks, that overlook the fine details of the dynamics of the membrane current to focus on the structure and the information flow of the system we want to study. Finally, I will give a brief overview of Deep Learning and present some algorithms with exciting applications in neuroscience.

2.1.1 Single neuron models

To understand how the brain performs complex cognitive tasks, an intuitive starting point focuses on its essential components: single neurons. A *neuron* is a particular cell that receives, processes, and transmits electrical pulses. Each neuron typically possesses three main components: soma, axon and dendrites [41] (Figure 2.1).

The soma is the primary cell body containing the nucleus and other organelles. Differently from other cells, the neuron can store electric charges via charged ions. This ion concentration usually generates an average membrane resting potential of $-70mV$. The cell is permeable to specific ions thanks to ion channel proteins involved in transporting such ions. The charge distribution may change in specific conditions, and the membrane potential may increase or decrease. The ion transportation mechanisms allow the cell to spontaneously move towards equilibrium by restoring its

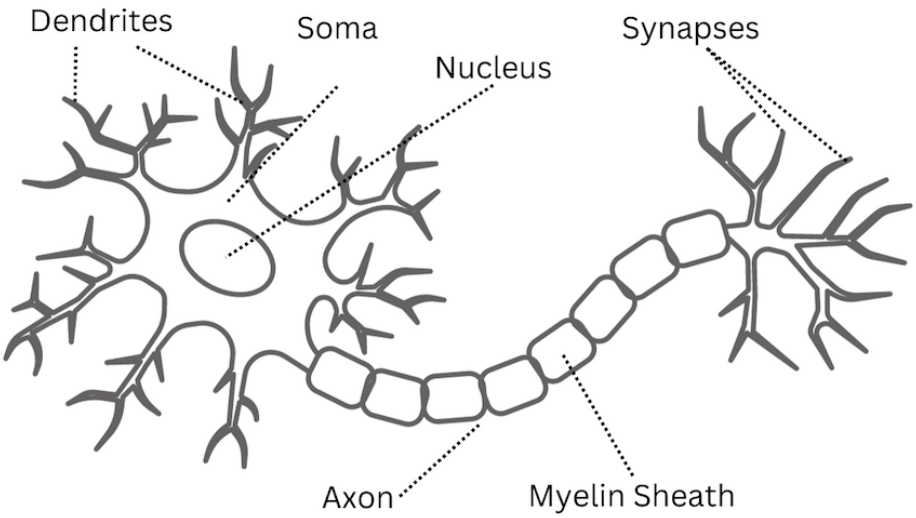

Figure 2.1: Anatomy of the neuron and its main components.

resting potential. However, when the membrane potential reaches a threshold value V_{th} of about $-55mV$, a rapid cascade of events causes a strong depolarisation of the cell followed by a repolarisation [95]. The intense and rapid depolarisation is called a *spike*, an almost instantaneous electrical pulse that can be transmitted to other neurons through the axon. The axon is a relatively long extension of the cell body and can be wrapped in a sheath of myelin, a discontinuous lipid-rich insulating material which serves as a spike propagation facilitator [41]. The ending part of the axon possesses several ramifications, which terminals are called synapses. A synapse is a specialised structure serving as a system between neurons. Such communication typically involves the release of signalling chemical molecules called neurotransmitters. Neurotransmitters are collected by receptors in the dendrites, branched extensions of the soma which, together with the axons, form the extraordinarily complex network of interactions between neurons.

When a spike propagates to other neurons, if certain conditions are met it may cause depolarisation or polarisation of post-synaptic neurons and possibly trigger another spike [77]. In a large ensemble of neurons like the brain, this gives rise to highly complex dynamics, which is believed to encode and elaborate all the information that makes us feel emotions, perform long-term plans or take decisions in front of uncertain situations [188].

A single neuron is an extraordinarily complex unit, and several models have been proposed to represent specific features and behaviours.

2.1.1.1 The Integrate-and-Fire Model

The integrate-and-fire (IF) model is a simple computational model which successfully explains some features of real neurons [41] and will serve as a basis for higher scales analysis. IF neurons are represented by a membrane voltage V, which varies in time due to an input current $I(t)$ following

$$C\frac{dV(t)}{dt} = I(t), \tag{2.1}$$

where C is the capacitance of the membrane. When the voltage reaches a threshold value V_{th}, a spike represented by a delta function occurs, and the voltage is reset to its resting value. Since such models generate spikes with a frequency increasing linearly without bound, a post-spiking refractory period t_{ref} in which another spike cannot occur can be introduced for higher accuracy. One of the significant limitations of this model is that an increase in the voltage, which does not lead to a spike, is retained indefinitely. Conversely, in real neurons, a relaxation to the resting potential is observed. This observed spontaneous relaxation can be addressed by introducing a leaky term in the model.

$$C\frac{dV(t)}{dt} = I(t) - \frac{V(t)}{R}, \tag{2.2}$$

where R is the membrane resistance. This is called the leaky integrate-and-fire model, equivalent to the IF model in the limit $R \to \infty$ (perfect insulating membrane).

If we multiply both sides of Eq. (2.2) by R and introduce a variable $\tau_m = RC$ representing a time, we obtain the standard form of the equation for the leaky IF neuron,

$$\tau_m\frac{dV(t)}{dt} = -[V(t) - V_{rest}] + RI(t), \tag{2.3}$$

where the resting potential V_{rest} is also explicitly introduced such that, at equilibrium, $V(t) = V_{rest}$, and the first term in the second side of Eq. 2.3 is 0.

The solution of Eq. (2.3) with initial value $V(0) = V_{rest} + \Delta V$, considering that for $t > 0$ the input current vanishes $I(t) = 0$, is

$$V(t) - V_{rest} = \Delta V \exp\left(-\frac{t - t_0}{\tau_m}\right) \text{ for } t > t_0. \tag{2.4}$$

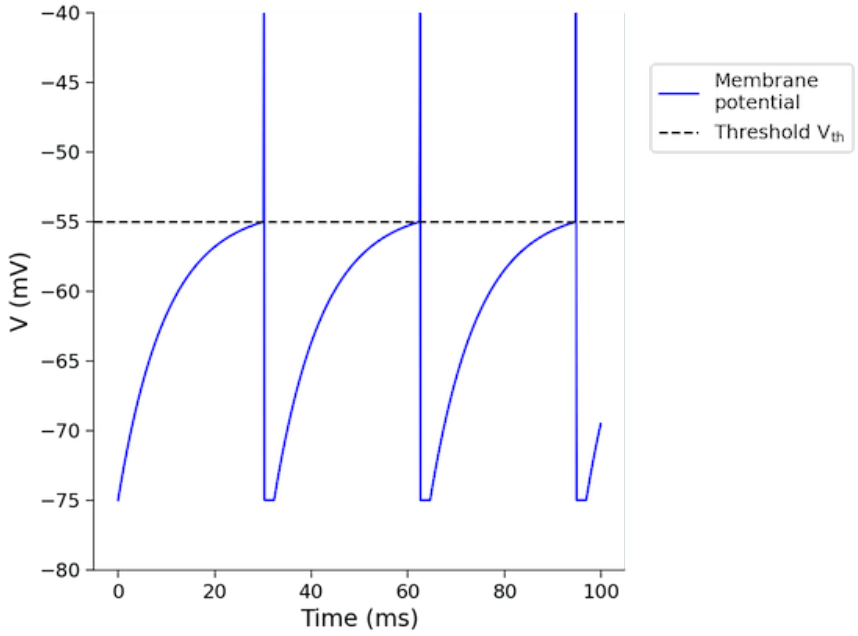

Figure 2.2: Spike train generated by a leaky-IF neuron with a constant current as input. When the potential reaches the threshold value of -55 mV, a spike is generated. The potential is immediately reset to its resting value of -75 mV.

Eq. (2.4) represents an exponential decay to the resting value without inputs.

The leaky IF model includes a threshold potential V_{th} as in the IF model. When this value is reached, a spike is triggered. Mathematically, in the leaky IF model, spikes are represented as points in time using the Dirac δ-function. Suppose the threshold potential is reached at time t^f such that $V(t_f) = V_{th}$, then the spiking event is registered and the potential is reset to a new value $V_r < V_{th} < V_{rest}$ and the dynamics for $t > t^f$ is again described by Eq. (2.3) (Figure 2.2).

The IF model of neurons is a simplified model that does not consider many complex biophysical properties of real neurons, like the role of dendrites or other sub-cellular structures in receiving and processing input signals or the effects of ion channels or neurotransmitters. Moreover, it does not consider essential phenomena like the refractory period or synaptic plasticity. Several extensions of the IF model have been proposed to take into account the refractory period (e.g., the Refractory Exponential IF model [8]) or the synaptic plasticity (e.g., the Spike Timing Dependent Plasticity model [142]).

2.1.1.2 The Hodgkin-Huxley Model

The IF model is a simple yet powerful description of single neurons that is able to describe complex behaviours without introducing many biological details. In the past, another successful single-neuron model has been proposed and it focused more on the biological details of the membrane selective ion channels responsible for the complex electrical activity of neurons. While the IF will be used in later chapter to extend the scale to ensemble of many neurons, it is still interesting to take a short detour and present a model which reproduces biological details.

The first successful attempt to reproduce empirical observations of the membrane depolarisation of neurons has been carried out by Alan Lloyd Hodgkin and Andrew Huxley [95]. The Hodgkin-Huxley model describes the dynamics of neurons starting from the equation of conservation of electric charge,

$$C\frac{dV}{dt} = -F + I, \tag{2.5}$$

where V is the membrane potential (considered spatially constant for simplicity), C is the cell capacitance, F is the membrane current and I is the total current entering the cell from both external or synaptic sources. In this model, the membrane current F is mostly generated by the motion of sodium and potassium ions crossing voltage-dependent channels in the membrane. Minor currents generated by other ions are grouped in a leak term. We can write the equation for F as,

$$F(V, m, h, n) = g_L(V - V_L) + g_K n^4(V - V_K) + g_{Na} h m^3(V - V_{Na}), \tag{2.6}$$

where g_{Na}, g_K and g_L are the sodium, potassium and leaky conductances, respectively, V_{Na}, V_K and V_L are the sodium, potassium and leaky reversal potentials, respectively, and m, h and n are time-and-voltage-dependent conductance variables with values in the interval $[0, 1]$. These variables represent the probability that a specific channel is open at any given time. In particular, m and h control the dynamic of the sodium channel, and they model activation or voltage gates and inactivation of the channel, respectively. The variable n controls the activation of the potassium channel. The temporal evolution of the three gating variables takes the form

$$\frac{dx}{dt} = -\frac{1}{\tau_x(V)}[x - x_0(V)], \tag{2.7}$$

where x can be m, n or h, τ_x is a time constant, and x_0 an asymptotic value for x. In particular, m is close to 0 when the membrane potential is at its resting value,

and it approaches its highest value $m = 1$ asymptotically as the potential grows. The variable n behaves similarly to m but starts to increase at a higher potential. On the other hand, the asymptotic value for h is 0 as it starts being slightly higher than 0.6 at resting and decreases as the potential increases.

The Hodgkin-Huxley model generates spikes resembling the dynamics of real ion channels. Indeed, at equilibrium $V = V_{rest}$ we have m close to zero, $n = 0$ and h slightly above 0.6. Consequently, the voltage gates of sodium and potassium channels are closed while the inactivation gate is open because the concentration of Na^+ ions is higher outside the cell, while the opposite is true for K^+ ions. Following Eq. (2.6), if an input current reaches the membrane, causing an increase in the membrane potential, the conductance of sodium channels increases in correspondence with increasing m (voltage gate opening), and more Na^+ ions enter the cell raising the membrane potential further (depolarisation). If the input current is sufficiently high, the positive feedback loop of Na^+ ions causes a dramatic increase in the concentration of positively charged ions inside the cell, thus generating a spike. At this point, the potassium gate is also activated, but since it has a higher characteristic time, it opens more slowly, and the membrane potential becomes more and more positive. When the reversal potential of sodium V_{Na} is reached, the inactivation variable h stops the flow of Na^+ inside the cell, and the potassium gate controlled by n finally opens, allowing K^+ ions to cross the membrane towards the outside of the cell and decreasing the membrane potential towards its resting value (re-polarisation). The potassium channels close slowly as the characteristic time τ_n is higher. Hence, when the cell reaches the resting potential, the potassium gate is still open, causing K^+ ions to exit the cell and the potential to decrease at lower values until it reaches the reversal potential of the potassium V_K (hyper-polarisation). Eventually, the leaky term slowly restores the potential to its resting value, and the ATP-mediated sodium-potassium pump restores the normal concentration of sodium and potassium ions inside the cell.

Similarly to the IF model, the HH model has some limitations [46]. Firstly, the model is based on data from the giant axon of the squid, and its parameters may not accurately reflect the behaviour of neurons in other species or regions of the nervous system. Moreover, it does not account for other ion channels, receptors, or biophysical factors that may be present in real neurons and influence their behaviour. Again, the model cannot represent more complex behaviours of real neurons, such as bursting and oscillations or synaptic plasticity. In addition to other limitations, it is worth considering that the model is computationally expensive and difficult to simulate large networks of neurons [107]. This is important to note as I will start presenting models

of more interacting neurons in the next section. Several models have been proposed as an extension of the HH model, like the FitzHugh-Nagumo model, to better account for the behaviour of neurons in the presence of synaptic noise [59], the Morris-Lecar model which adds a current generated by calcium ions [153], the Izhikevich model to provide a more computationally efficient simulation of neurons [107].

2.1.2 Population models and neural mass models

In past section I presented two of the most successful single-neuron models, outlining their advantage as well as limitations. However, it is hard to appreciate the brain's computational capabilities by looking at only single neurons. As we will see in deeper detail later, the brain possesses a structured organisation different from a mere collection of randomly connected neurons. It has been shown instead that neurons are organised in richly structured populations (also called pools), which often show a similar activation pattern (synchronisation) [218, 206, 205]. Each population might contain several millions of neurons, and modelling such a high number of neurons with their inputs and outputs is computationally expensive, especially in the case of complex models like the Hodgkin-Huxley neuron presented in 2.1.1.2. What is typically done instead is considering a population model, where all neurons are considered identical, receiving a common input and interacting in a statistically uniform way [67]. This approach is inspired by physics and statistical mechanics where it takes the name of mean field approach [68]. While considering a common input is an easy but effective approximation, defining a population output is more tricky. Here is one possible definition [67]: given a population of N neurons the activity can be written as,

$$r(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \frac{n_{act}(t; t + \Delta t)}{N} = \frac{1}{N} \sum_{j=1}^{N} \sum \delta(t - t_j^{(f)}), \tag{2.8}$$

where $n_{act}(t; t + \Delta t)$ is the number of spikes in a small interval of time Δt, δ is the Dirac delta function, and $t^{(f)}$ represents firing time. Models that study neural populations with a mean-field approach are called Neural Mass Models (NMM), and they aim to predict the temporal evolution of the population activity or mean firing rate $r(t)$ in large and homogeneous populations of spiking neurons [67]. More specifically, NMMs describe how the population-averaged firing rate varies as a function of time and the parameters of the specific network of neurons. Figure 2.3 shows a simple comparison between rates observed from a simulation of 50 IF neurons and rates predicted by Eq. (2.8). It can be seen that an NMM approach is justified by the

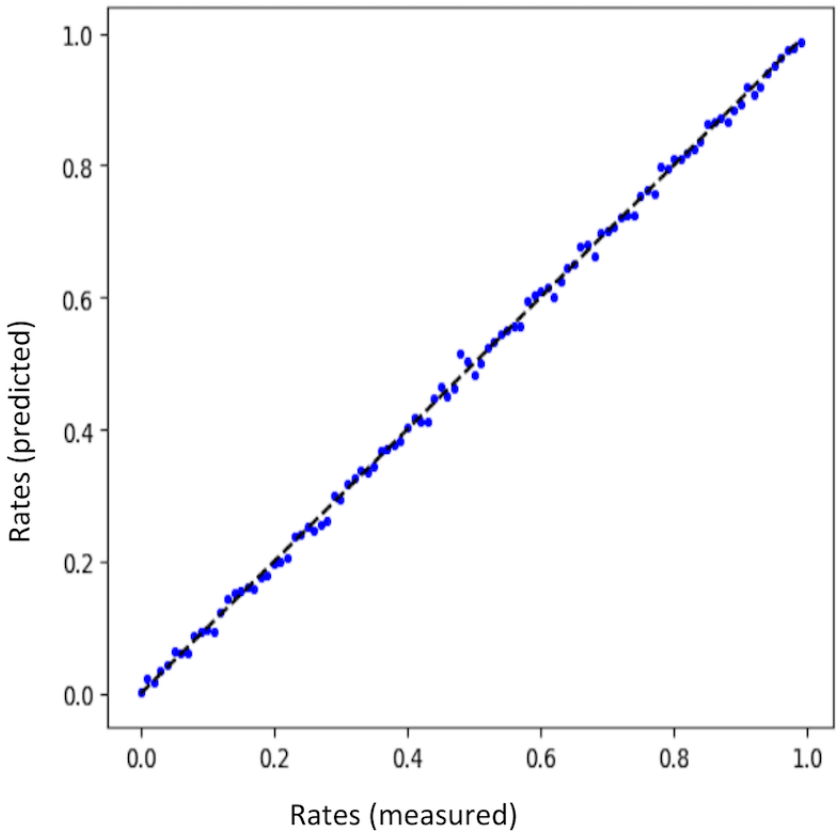

Figure 2.3: Mean field predictions for spiking rates (y-axis) versus rates observed from 50 simulated IF neurons (x-axis). Points are located near the dashed line (identity line), confirming the excellent approximation of the mean-field approach.

excellent approximation of the rates of the averaged neural population on the single neurons' rates. Neural mass models have achieved great success by effectively capturing the collective behaviour of large populations of interconnected neurons, providing valuable insights into brain dynamics and facilitating the understanding of complex neural phenomena [17, 43, 40].

In the simplest case, we may define the differential equation for a feed-forward network, with excitatory neurons only, as

$$\tau \frac{dr}{dt} = -r + F(I_{ext}) \tag{2.9}$$

where $r(t)$ can be defined as in Eq. (2.8) or using other averaging techniques, τ is the characteristic time of the evolution of the average firing rate, I_{ext} represents the external input received by the population, and the transfer function $F(\cdot)$ represents the population activation function in response to all received inputs (also known as

F-I curve where I stands for input) [67]. The term $-r$ is a leaky term indicating the spontaneous decay of the firing rate without external inputs. A popular way to model the transfer function $F(\cdot)$ is by using a sigmoidal function

$$F(x; a, \theta) = \frac{1}{1 + e^{-a(x-\theta)}} - \frac{1}{1 + e^{a\theta}} \tag{2.10}$$

where the argument x represents the input to the population, θ is a threshold indicating how much input is needed for the firing rate to reach its midpoint value 0.5, and a determines the slope (gain) of the rising phase of the F-I curve [24]. The second term is chosen so that $F(0; a, \theta) = 0$. Several alternatives for the transfer function exist, generally being non-linear monotonic functions. Valid examples of different transfer functions are the hyperbolic tangent $tanh(x)$ or the rectified linear function $ReLU(x)$ defined as

$$ReLU(x) = \begin{cases} x & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.11}$$

As we will see in later sections, the rectified linear unit function is one of the most popular functions to introduce non-linearity in Deep Learning architectures.

Eq. (2.9) is interesting because it helps understand how neurons respond to a specific input. However, in biological neural networks, populations of neurons typically have a recurrent component which introduces a more complex behaviour. We may describe the dynamics of a recurrent network or population of neurons as

$$\tau \frac{dr}{dt} = -r + F(w \cdot r + I^{ext}) \tag{2.12}$$

where we added to Eq. (2.9) a parameter w indicating the strength of the connections between neurons in the population multiplied by the firing rate r.

At this point we should have an idea of how to model arbitrarily large neuronal populations. One thing that's important to notice is how the research questions we can answer change with the increasing in the scale we are considering. Single neuron models are very helpful in reproducing the spiking dynamics and other fine phenomena, while when considering neural population we need a different approach to study a more complex dynamics and understand how it relates to higher cognitive functions. For this reason, in the section I will introduce a powerful framework to analyse the dynamics of interacting neuronal populations.

2.1.2.1 Dynamical System Theory of Neural Mass Models

The dynamical systems theory is one of the most used frameworks to study dynamical equations like (2.12). A dynamical system is a system that evolves in time [108]. We will focus on continuous dynamical systems, i.e. systems described by differential equations, which differs from discrete dynamical systems where the time evolution is defined using iterative maps [211].

In general, a continuous dynamical system θ of k components is mathematically defined using a first-order differential equation:

$$\theta = \frac{d\boldsymbol{x}}{dt} = f(\boldsymbol{x}) \tag{2.13}$$

where $\boldsymbol{x} \in \mathbb{R}^k$ and $f : \mathbb{R}^k \to \mathbb{R}^k$. In this compact representation, the vector $\mathbf{x}$ represents the dynamical variables, also referred to as state variables [108]. For example, in the Hodgkin-Huxley model described in 2.1.1.2, we have four state variables describing a membrane potential and three gating variables for currents generated by the motion of Na^+ and K^+ ions.

A dynamical system can be characterised by its phase plane, which is a plot of the system, *i.e.*, the time derivatives, versus the state variables [108, 211]. In particular, the dynamical system typically lives in a subset of the $(\frac{dx}{dt}, x)$ plane, which is called phase portrait. To better understand the importance of phase portraits in dynamical systems analysis, it is essential to introduce the concept of trajectory. A trajectory in the phase plane is the set of points representing the dynamical system for $t \to \infty$ [211]. Trajectories typically vary when changing the initial conditions *i.e.*, the point in the phase plane where the system is at $t = 0$. A *phase portrait* can be defined as the set of all the possible trajectories of a dynamical system in the phase plane. The phase portrait can only be created numerically for non-linear systems in cases where an analytical solution is impossible [211]. This is crucial to understand as many systems of interest in neuroscience are non-linear dynamical systems [108].

Of particular importance in dynamical systems are fixed points, *i.e.*, points where $\frac{d\mathbf{x}}{dt} = 0$. In such points, the system is in a steady state (equilibrium), which can be asymptotically stable, unstable or a saddle point (if $k \geq 2$) [211]. Following the definition of Poincaré, a fixed point θ^* is stable if it attracts all nearby trajectories:

$$\lim_{t \to \infty} \theta^* + \Delta\theta = \theta^* \tag{2.14}$$

for any small Δ. In the phase portrait, these points are where nearby trajectories fall into for t large enough. On the other hand, trajectories starting near unstable points will tend to diverge or move towards a stable fixed point [108, 211].

A dynamical system equilibrium is described not only by simple points. Limit cycles or fractal attractors, also called strange attractors, might also be present in higher dimensional dynamical systems [211].

The dynamical system may also be described by parameters [211]. The system's dynamic, such as the number and the nature of fixed points, can be influenced by those parameters. A valuable tool to study the influence of the parameters of a dynamical system is the bifurcation diagram [211]. A bifurcation happens when the number, the nature of the fixed point or both change due to the effect of changing a specific parameter. The dynamic of relatively simple systems like the logistic map may differ from having a single fixed point to a chaotic attractor based on relatively small changes in one parameter [7].

Now that we have an introductory knowledge about dynamical systems, we can apply this framework to our neural mass model of recurrent excitatory neurons described by Eq. (2.12). The fixed points of such a system can be found by solving Eq. (2.12) for $\frac{dr}{dt} = 0$. Such a solution is not guaranteed to exist, and if F is a non-linear function, numerical simulations might be needed to find it. With a sigmoidal F-I curve, like in Eq. (2.10), the system's fixed points depend on the value of the parameters. For example, Figure 2.4 presents the cases for $w = 4$ and $w = 4.2$. Interestingly, even a tiny change in the parameter value can cause the system to dramatically change behaviour, bringing the system from having three fixed points to only one.

2.1.2.2 The Wilson-Cowan model

In the previous section, I presented a simple neuronal network model with only one single excitatory population. However, it is also important to consider other kind of neuronal population to describe more complex cognitive functions. To extend the approach to excitatory and inhibitory neuronal populations, a simple yet powerful model is the Wilson-Cowan rate model [238].

Two coupled differential equations describe the Wilson-Cowan model, each describing the dynamics of the excitatory or inhibitory population:

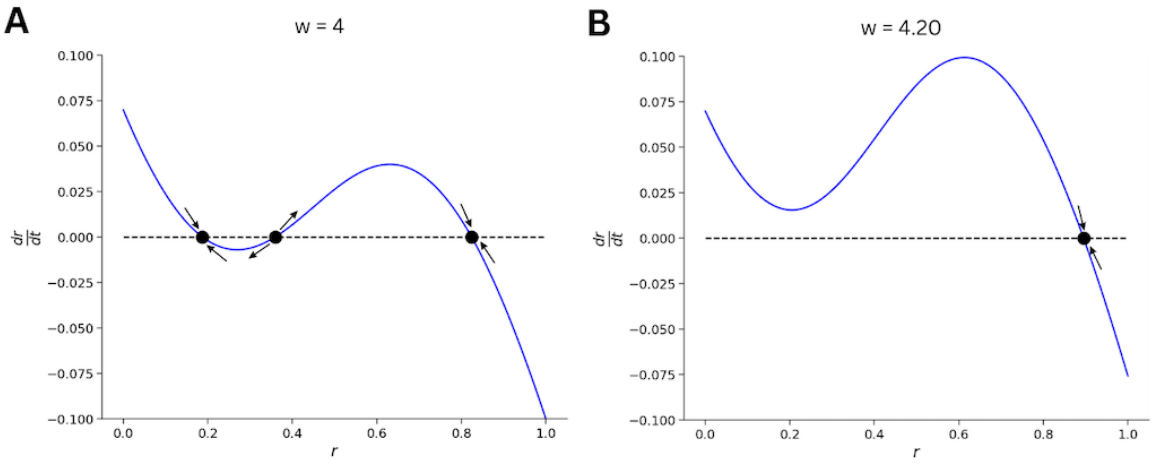

Figure 2.4: Phase portrait of a neural mass model of a single excitatory neuron population. (A) Phase portrait for the model with excitatory coupling parameter $w = 4$. Three fixed points are observed, with two being stable (arrows pointing inwards) and one unstable; (B) Phase portrait for $w = 4.2$. Only one stable fixed point is observed.

$$\begin{aligned} \tau_E \frac{dr_E}{dt} &= -r_E + F_E(w_{EE} r_E - w_{EI} r_I + I_E^{ext}; a_E, \theta_E) \\ \tau_I \frac{dr_I}{dt} &= -r_I + F_I(w_{IE} r_E - w_{II} r_I + I_I^{ext}; a_I, \theta_I) \end{aligned} \tag{2.15}$$

where $r_E(t)$ and $r_I(t)$ represent the average activation (firing rate) for the excitatory and inhibitory population, respectively, τ_E and τ_I are the characteristic times of the dynamics of each population, and w represents the connection strengths, e.g. w_{EI} is the connection strength from the excitatory population to the inhibitory one. The transfer functions (or F-I curves) $F_E(x; a_E, \theta_E)$ and $F_I(x; a_I, \theta_I)$ can be different for the excitatory and the inhibitory populations.

When studying a two-dimensional system, plotting the two variables in the phase plane is the standard procedure. In particular, of great interest are the nullclines, defined as the set of points in which the derivative of one of the two variables is zero. In the Wilson-Cowan model, we have the excitatory nullcline where $\frac{dr_e}{dt} = 0$ and the inhibitory nullcline where $\frac{dr_I}{dt} = 0$. The points of intersection of the nullclines, if they exist, have both derivatives equal zero, *i.e.*, they are fixed system points. Each nullcline divides the phase plane into two regions: one where the derivative is positive and the other where it is negative. This guides the behaviour of the system and can be visualised using a vector field, *i.e.*, the map of the tangent to the trajectory in each point (r_E, r_I) of the phase plane (Figure 2.5).

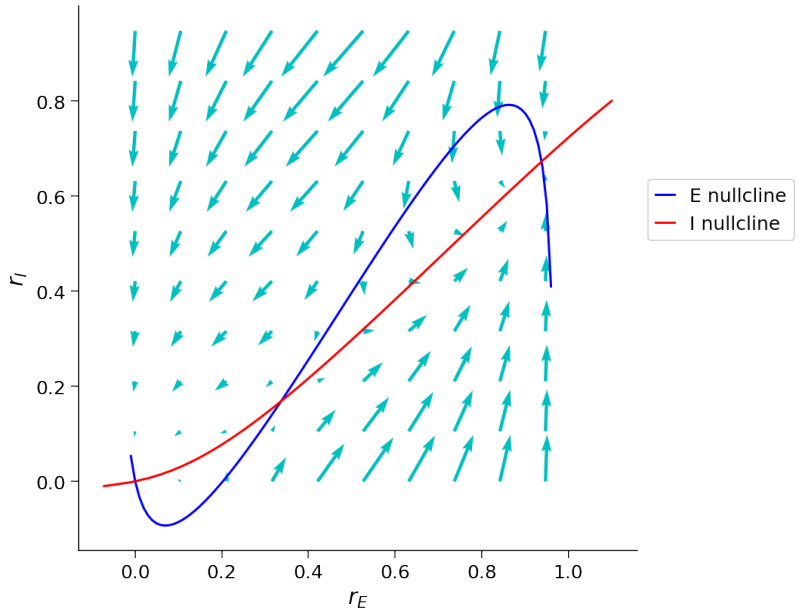

Figure 2.5: Phase plane and vector field of the Wilson-Cowan model. The blue line represents the excitatory nullcline, while the red line is the inhibitory nullcline. The intersection points are the fixed points of the system. The vector field (teal arrows) shows that the point in the middle is unstable while the others are stable.

Excitatory and inhibitory interacting pools help study one of the most critical processes that the brain performs: decision-making [73]. In particular, the next sections will focus on perceptual decision-making and explain how competing dynamics of excitatory and inhibitory neural populations provide a perfect framework to study the problem.

2.1.3 Modelling perceptual decision-making

Through perceptual decision-making, an organism is able to acquire and integrate sensory information from the external world to make choices or decisions [87]. The study of perceptual decision-making is an interdisciplinary field which combines behavioural experiments [187, 146, 18], neuroimaging [192, 90] and computational modelling [242] which is the focus of this thesis.

Behavioural experiments are generally the starting point to gain a qualitative and quantitative description of the phenomena we want to study. Typically, behavioural experiments are designed to measure the responses of subjects to stimuli that can be of different origins, like visual, auditory or any other which involves sensory perception. The responses are then used to infer the underlying decision-making mechanisms and can be used to create a model. In addition to the responses, experiments can focus

on other measures like the subjects' reaction time to the stimuli or physiological responses like pupil dilation [157, 224].

One of the most interesting paradigms is the visual decision-making. Here, stimuli can span from a range of simple choices like single dots or lines to more complex ones such as images [146], faces or scenes [90]. The two-alternative forced-choice task (2AFC) is a typical experimental setting to study visual decision-making. Here, subjects are presented with a noisy visual stimulus and are asked to make a binary decision (choose between 2 factors) about the stimulus. The subjects are not allowed to give different answers beyond the two possible choices (forced choices); hence, they have to answer even in a situation of uncertainty. This kind of tasks are performed to measure the subject's sensitivity to the stimuli as well as their decision criterion. In this way, it is possible to quantify the level of accumulated evidence that the subjects require to make a choice.

A notable example of a 2AFC paradigm is the random dot motion discrimination task [187]. Here, subjects watch a combination of a pattern of random dots moving in all possible directions, with a certain percentage of dots moving coherently in one single direction. Such visual stimulus is also referred to as random dot kinematogram (RDK). The possible directions are two, for example, left or right, and subjects are instructed to choose the perceived direction by pressing one of two buttons. The stimuli can be of a different degree of complexity depending, for example, on the total number of dots, and motion direction can be more or less ambiguous depending on the number of dots moving coherently [187]. Typically, behavioural experiments are performed on different coherence levels; a high coherence level indicates that the proportion of moving dots is high, and subjects usually make a higher fraction of correct choices in less time. On the other hand, lower coherence levels indicate a more difficult task with longer reaction times and less decision accuracy.

This task is interesting as it can be used to estimate the subject's decision criterion, *i.e.*, the level of evidence that the subjects need to accumulate before making a decision, together with the time required to make a decision (reaction time). The results can then be used to create a model of perceptual decision-making and can be supported by other experiments like neural activity measurements. In the next section, I will present in detail one of the most successful models of perceptual decision-making in a 2AFC task.

2.1.3.1 The Wang-Wong model

The Wang-Wong model is a two-variables neural mass model based on the Wilson-Cowan model, which implements the evidence accumulation involved in decision-making during a perceptual decision-making task [242]. The model consists of a mean-field approximation of a large-scale spiking neuronal network model [232] composed of 7200 leaky IF neurons of the lateral intra-parietal area (LIP). The activity of the LIP area has been observed to be correlated with both the response time and the choice of the subject performing the perceptual decision-making task [187]. In particular, the Wang-Wong model successfully reproduced the behavioural performances and the reaction times of a monkey performing a random dot motion discrimination task [242, 187]. Each variable accumulates evidence for one of the two directions of the dots (leftward or rightward) and competes with the other via mutual inhibitory connections. A self-excitatory mechanism is also present. The input is modelled as two different synaptic currents generated by the stimulus and represents the momentary evidence supporting the two alternatives. In addition to the stimulus currents, the model considers a common non-selective background input. Here, we consider an extension of the model [126] where the background current is modulated by a factor β to take into account the speed-accuracy trade-off observed in some studies [208, 91].

In the Wang-Wong model, the input is represented by a constant current for each population. The current responsible for the correct choice (left or right) has a higher intensity,

$$I_l = J_{ext}\mu_0(1+c) \tag{2.16}$$

$$I_r = J_{ext}\mu_0(1-c) \tag{2.17}$$

where $c \in [-1, 1]$ is the coherence with negative values referring to rightward movements, μ_0 is the stimulus strength, and J_{ext} is the average synaptic coupling with external synapses. When $c = 1$ or $c = -1$, all the dots are moving towards a single direction, and one of the two inputs would be 0 as no dots are moving in that direction. These currents contribute to the synaptic currents of the two populations,

$$I_{syn,l} = J_{ll}S_l - J_{lr}S_r + I_l + \beta I_0 + I_{noise,l} \tag{2.18}$$

$$I_{syn,r} = J_{rr}S_r - J_{rl}S_l + I_r + \beta I_0 + I_{noise,r} \tag{2.19}$$

where J_{ll} and J_{rr} are the self-excitatory coupling strengths within same populations, J_{lr} and J_{rl} the inhibitory coupling strengths across different populations, S_l and S_r

are the synaptic gating variables representing the activities of the two populations, β is the modulation of the background current I_0 and $I_{noise,l}, I_{noise,r}$ are Gaussian noisy currents for the left and right population, respectively. This simplified model consider a general case in which excitatory and inhibitory synaptic interactions are not symmetrical. In this work, we followed the approach of the authors in the original paper [242], setting $J_{ll} = J_{rr}$ and $J_{lr} = J_{rl}$ after following a consideration derived from [185]. In particular, neurons of the two different populations interact each other via an all-to-all connection pattern. Thus, when considering a mean-field approach, it is reasonable to assume that while excitatory and inhibitory strengths between two single neurons may vary depending on the population the neurons belong to (or even within the same population), it can be considered constant when referring to the average synaptic strength within and across populations, as the neurons belonging to the system can be considered biologically very similar. We can note that the excitatory coupling contributes positively to the population's current corresponding to the correct choice because it is multiplied by the corresponding synaptic gating variable. On the other hand, the inhibitory coupling has a negative sign and multiplies the synaptic variable, representing the other choice.

The temporal evolution of the synaptic variables is given by the following equations,

$$\frac{dS_l}{dt} = S_l - \frac{S_l}{\tau_s} + \gamma r_l(1 - S_l) \tag{2.20}$$

$$\frac{dS_r}{dt} = S_r - \frac{S_r}{\tau_s} + \gamma r_r(1 - S_r) \tag{2.21}$$

where $r_{l,r}$ are the firing rates of the neurons of the two populations, τ_s is the characteristic decaying time of the synaptic activity, and γ is a kinetic parameter to fit the synaptic activity [242].

The equations for the firing rates are,

$$r_l = \phi(I_{syn,l}) = \frac{aI_{syn,l} - b}{\exp(-d(aI_{syn,l} - b))} \tag{2.22}$$

$$r_r = \phi(I_{syn,r}) = \frac{aI_{syn,r} - b}{\exp(-d(aI_{syn,r} - b))} \tag{2.23}$$

where the parameters a, b and d define the input-output function that generates a firing rate from a current [1].

The noisy current in (2.18) and (2.19) evolves in time following,

$$\frac{dI_{noise,i}}{dt} = I_{noise,i} - \frac{1}{\tau_{noise}}(-I_{noise,i}(t) + \eta(t)\sqrt{\tau_{noise}\sigma^2_{noise}}), \tag{2.24}$$

where $i = l$ or r whether the current refers to the left or right population, respectively, σ_{noise} and τ_{noise} are the amplitude and the characteristic time of decay of the noisy current, respectively, and $\eta(t)$ is a Gaussian white noise with zero mean and unit variance.

The phase plane analysis of the Wang-Wong model provides a practical biological interpretation of the model equations. We repeat the same procedure as the one applied to the Wilson-Cowan model in section 2.1.2.2 by plotting the nullclines *i.e.*, the set of points where $\frac{dS_1}{dt} = 0$ and $\frac{dS_2}{dt} = 0$ on the (S_1, S_2) plane. In Figure 2.6A, we can observe the phase plane in the absence of a coherent stimulus ($c = 0$ in Eq. 2.16 and Eq. 2.17) *i.e.*, when all the dots are moving randomly without a preferred direction. In such a situation, the two nullclines are symmetrical and intersect 3 times in 2 stable fixed points separated by one saddle point, which attracts trajectories moving on the $S_1 = S_2$ direction and repel towards the orthogonal direction, driving the system towards one of the two stable fixed points. The two stable fixed points represent the model's choices, one for the leftward movement and the other for the rightward. The system is symmetrical without coherent motion, and the choice will depend on the initial condition. Each stable point has a basin of attraction, which can be defined as the set of points in the phase plane where that stable point attracts the trajectories. Without coherent movement, the model will choose based on which basin of attraction the initial condition is. Adding a coherent stimulus breaks the symmetry of the phase plane by increasing one of the two basins of attraction (Figure shows the case of $c = 0.2$). By increasing the coherence, the basin of attraction of the correspondent choice will grow more extensively than the one for the other choice. In the limit of a perfect coherent stimulus ($c = 1$), the two nullclines will intersect in one stable point, which becomes the system's trivial choice.

2.2 Graph representations of biological neural networks

So far, we have been focusing on tools and models where certain specific properties of single neurons are essential. Starting from single neuron models, we moved through neural mass models, where several neurons are grouped into neural populations to study local circuits at a microscopic level. Now, I will present one of many possible macroscopic approaches to studying the human brain. The term macroscopic refers to

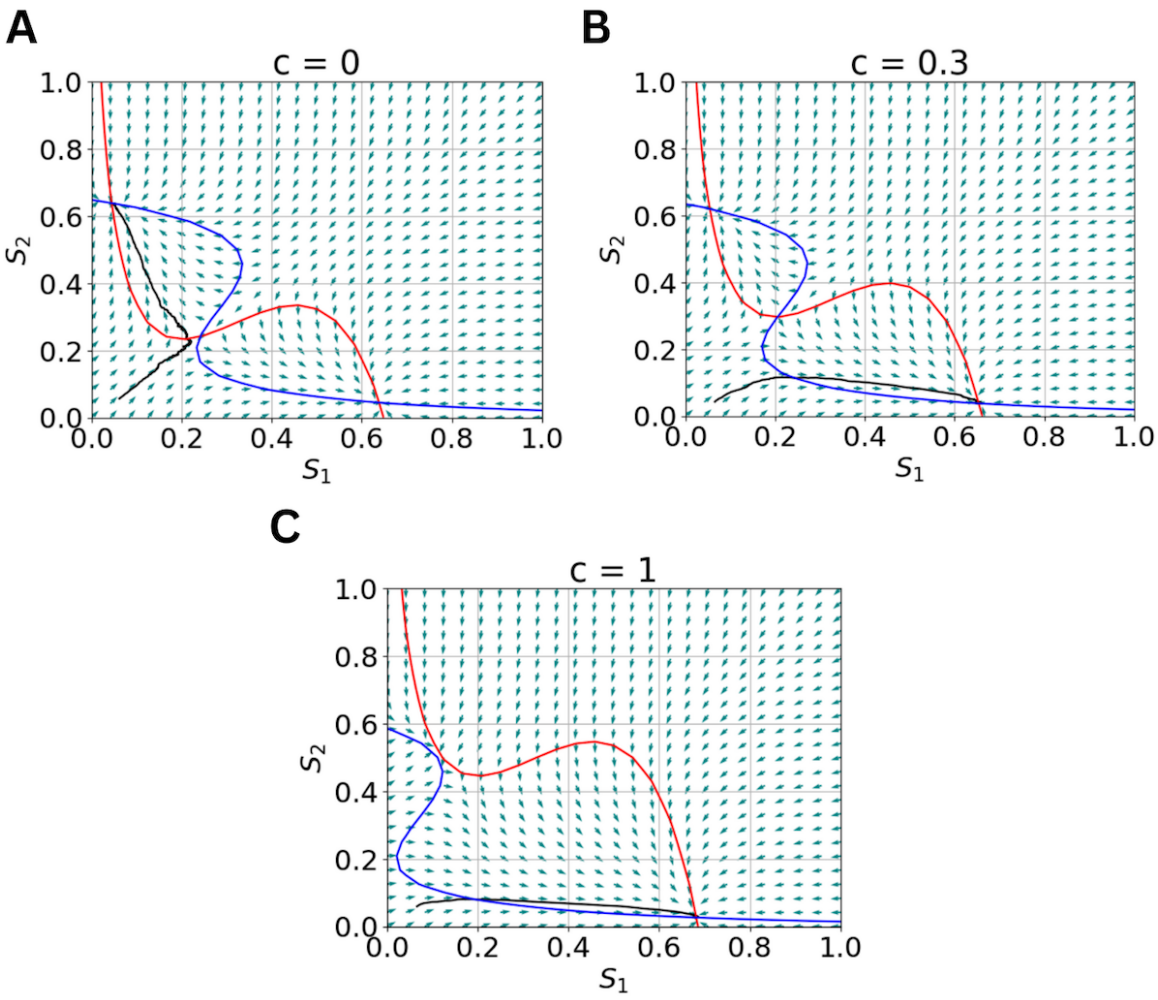

Figure 2.6: Phase plane and vector field of the Wang Wong model. The blue and red lines represent the nullclines for the two choices. The black line represents one possible trajectory of the system. Teal arrows represent the vector field. (a) Phase plane in the absence of a coherent stimulus ($c = 0$). The plot is symmetrical, with three fixed points, two stable, separated by one saddle point; (b) Phase plane for 30% coherence($c = 0.3$). The basin of attraction corresponding to the correct choice is larger than the one corresponding to the wrong choice (c) Phase plane for completely coherent stimulus ($c = 1$). There is only one stable fixed point corresponding to the trivial choice.

the fact that the models presented in the following sections can be applied to studying the whole brain (while not necessarily only limited to such scale [158]). Naturally, at such a scale, it is difficult to keep the finer details of single neurons, for example, those biologically plausible details, or describe neurons or neural population using dynamical equations. Instead, I will present a mathematical tool called Graph or Network to study the connectivity within the brain and how this is important to explain certain brain features.

The main difference between a dynamical system approach and a graph-theoretical analysis applied to neural network is that the first method focused more on temporal evolution of the system with a particular attention to behaviour and stability, while the latter focuses on concepts like information flow or resilience to damages. Clearly, the boundary between the scopes of these different approaches is quite nuanced and an effort should be put into integrative approaches to get the best of the different methods. For example, Graph Theory can also be applied to study dynamics [103].

In this thesis, in addition to the consideration about the difference in the scales of application, the dynamical system approach presented in 4 relates equilibrium states with perceptual decision making, while graph theory will be used to study how brain structure relates to its functioning (Chapter 3).

2.2.1 Graph Theory, Network Science and biological networks

Graphs or networks are concepts derived from Graph Theory, a branch of Mathematics [9, 83]. A *graph* is a system composed of elements called nodes connected with links[1]. A more formal mathematical definition of a graph is typically written in the form of

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}) \tag{2.25}$$

Here, a graph $\mathcal{G}$ is defined as a collection of vertices $\mathcal{V}$ and edges $\mathcal{E}$. An edge going from vertex $u \in \mathcal{V}$ to vertex $v \in \mathcal{V}$ is defined as $(u, v) \in \mathcal{E}$.

A Graph is usually represented by an adjacency matrix $\mathbf{A} \in \mathbb{R}^{|\mathcal{V}|\times|\mathcal{V}|}$ [9, 83]. In the case of a binary or unweighted graph, all the edges have the same strength and

[1]In literature, the term graph is often interchangeable with network [9, 83]. Similarly, the terms vertex and node usually represent the same thing. The same applies to the terms edge and link. The terms graph, vertex and edge are typically used when referring to the mathematical abstract object used to represent a specific instance in a real system (we most often use the term social network and not social graph). Unless otherwise specified, the terms graph and network, edges and vertices or vertex and nodes are interchangeable in the present work.

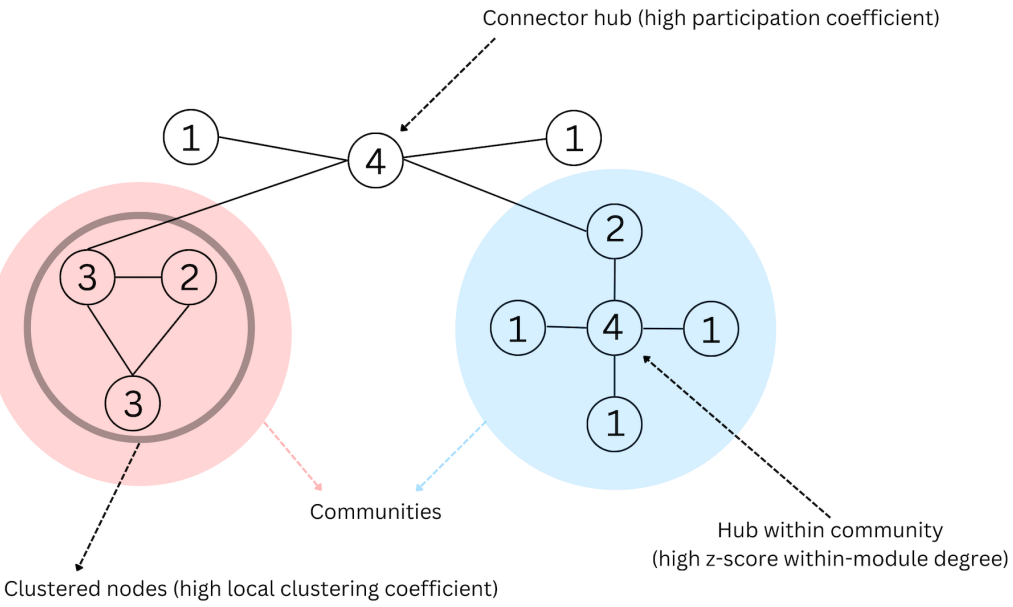

Figure 2.7: Example of a simple graph with some properties outlined. The numbers represent the node degree. Coloured circles identify different communities. The grey circle outline a set of 3 nodes fully connected to give an example of a highly clustered neighbourhood.

the elements of $\mathbf{A}$ can only assume the value $\mathbf{A}[u, v] = 1$ if there is an edge between the vertices u and v, 0 otherwise. If the graph is undirected, *i.e.*, $(u, v) \in \mathcal{E} \Leftrightarrow (v, u) \in \mathcal{E}$, $\mathbf{A}$ is symmetric. Social networks are examples of networks represented by undirected Graphs (if vertices represent friendships and Bob is Alice's friend, then Alice is automatically Bob's friend). At the same time, the World Wide Web is an example of an undirected graph since an edge represented by a URL is unidirectional.

The adjacency matrix is a convenient tool to represent a graph, but it has some limitations [9, 83]. First, the number of edges per vertex can vary dramatically, leading to sparse adjacency matrices. Sparse matrices are usually space-inefficient when we need computational analysis of many large graphs. Another caveat is that adjacency matrices are not permutation invariant. This requires an indexing of the vertices, which must be maintained across the analysis, and it could be more problematic in some state-of-the-art applications of machine learning on graphs. In such circumstances, using the edges list $\mathcal{V}$ or adjacency list is an elegant and memory-efficient way of representing sparse matrices.

As mentioned before, when using graphs as models for a system we want to investigate, we are usually interested in the connectivity pattern among nodes instead of the properties of the single node itself. Here, I will present some of the most funda-

mental measures of Graph Theory [9, 83]. Some of these properties can be computed node-wise, while others are global properties of the entire graph.

Node degree. Nodes can have different roles within a graph. It is convenient to use specific features or measures to define the properties of a node. A typical starting point is the node degree. Starting from the adjacency matrix $\mathbf{A}$, the node degree is defined as,

$$d_u = \sum_{u \in \mathcal{V}} \mathbf{A}_{u,v} \tag{2.26}$$

In an undirected, unweighted graph, node degree represents the number of links that a specific node forms with other node. In Figure 2.7, degrees are outlined inside each node. For a directed graph, the degree can be easily extended to an in-degree and out-degree, *i.e.*, the number of links starting from the node u and the number of links ending to the node u, respectively. For a weighted graph, the degree represents the sum of the strengths of the link that the node u forms with other nodes. If there is a link between u and v, or equivalently $(u, v) \in \mathcal{V}$, u and v are called neighbours. As we will see in more detail, these simple features can already discriminate important nodes in the network called hubs, *i.e.*, nodes connecting with many other nodes.

Average node degree. Single-node features can be extended to all the nodes in the graph, providing full-graph features or global features which are helpful to summarise important properties of the graph. In the example of node degree, we can define the average degree by simply averaging the degree of all the nodes in a graph,

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i = \frac{2L}{N} \tag{2.27}$$

where N is the number of nodes in the graph, and L is the number of edges. This is an important property but does not give important information about the graph's structure.

Degree Distribution. A more descriptive degree-based global feature is the degree distribution,

$$P(k) = \frac{n_k}{N} \tag{2.28}$$

where n_k is the number of nodes with degree k and N is the total number of nodes. The degree distribution represents the probability that picking a random node in a network, that node has degree k.

The degree distribution is one of the most crucial global features as it can identify essential classes of graphs. Consider, for example, the random graph or Erdős-Rényi graph model. This is a model to generate a graph by considering a set of nodes and assigning edges between nodes randomly with probability p. The average degree of random graphs is well approximated by a Gaussian distribution, meaning that most nodes have a typical number of connections while just a few have a very low or very high degree. This distribution, however, is not observed in real networks where usually the degree distribution follows a power law distribution defined as

$$P(k) \sim k^{-\gamma} \tag{2.29}$$

where γ is called degree exponents and characterises the distribution. Networks with a power law degree distribution are called scale-free networks because they lack a typical scale (for example, random networks have a scale represented by the average degree, which is the top value of the Gaussian degree distribution).

The scale-free property is important because it unveils how real systems organise into this kind of network. This has been studied by Barabási-Albert, who introduces the concept of preferential attachment. Given a graph $\mathcal{G}$ and a new node u trying to connect to $\mathcal{G}$, it will link to a node of $\mathcal{G}$ with probability,

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \tag{2.30}$$

meaning that the higher the degree of the node of $\mathcal{G}$, the higher the probability u will connect with it. In other words, real networks do not form new connections randomly, but hubs are more likely to get new connections.

Node centrality. Hubs in a graph are typically important nodes with central roles. However, more than a node degree is required to determine the importance of a node in a graph. A different measure of a property called node centrality can give information about the role of a node within a graph, like how influential that node is. Consider, for example, a social network where links represent the diffusion of news. A hub represents a famous newspaper with many followers, but it is not guaranteed to be the most influential node. Indeed, a lower degree node helps spread the news more quickly because it has a higher centrality.

Among the possible centrality measures, two rely on the definition of distance. In a graph, the distance or shortest path length between two nodes u and v is the minimum number of edges that must be crossed to reach v starting from u.

The betweenness centrality of a node v can be defined as,

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{2.31}$$

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v excluding those with v as destination. In other words, betweenness centrality measures how often a node lies on the shortest path between two other nodes. In our information passing example, the information flow often crosses a high central node and can diffuse information very effectively.

Another centrality measure is closeness centrality defined as,

$$C_B = \frac{1}{\sum_y d(y, x)} \tag{2.32}$$

where $d(y, x)$ is the distance between x and y. This is a measure of the average shortest path length between a node and all other nodes.

In the following subsection, I will introduce a novel graph approach to characterise nodes with different centrality roles.

Clustering coefficient. The clustering coefficient is a local graph measure which gives information about how densely connected a given node's neighbours are. It is defined as,

$$c_u = \frac{|(v_1, v_2) \in \mathcal{E} : v_1, v_2 \in \mathcal{N}(u)|}{\binom{d_u}{2}} \tag{2.33}$$

where $\mathcal{N}(u)$ are the neighbours of the node u. The denominator calculates how many pairs of nodes are in the neighbourhood of u. A clustering coefficient of 1 indicates that all the neighbours of a node are connected, forming a fully connected sub-graph.

Watts and Strogatz have observed that the clustering coefficient is higher in real networks than in random ones [233]. They also linked this feature to one of the most essential properties of real networks: small-worldness. A small-world is a network where the average distance between any two nodes is small relative to the number of nodes in the network. This means that in a small world network, it does not matter how many nodes there are; a node can most likely be reached with relative ease from any other node.

Communities. In real networks, nodes are organised into dense groups called communities. A community is a subset of nodes $\mathcal{S} \subset \mathcal{N}$ such that a node $s \in \mathcal{S}$ is much more likely to connect to another node $r \in \mathcal{S}$ in the same community rather than to any other node outside that community. Two different communities can be easily identified in Figure 2.7.

The problem of identifying communities in a graph is called community detection, and the most effective algorithms to solve it are based on hierarchical clustering. Hierarchical clustering is based on constructing a similarity matrix, whose elements x_{ij} represent the distance between node i and j. Then, two possible approaches can be used starting from the similarity matrix: agglomerative algorithms that merge highly similar nodes into the same community or divisive algorithms that remove links between nodes with low similarity. A hierarchical tree or dendrogram is generated in both cases to predict the possible community partitions. Two examples of agglomerative and divisive algorithms are the Ravasz and Girvan-Newman algorithms, respectively.

Identifying communities in a network can provide valuable insights into the structure and dynamics of the network. In particular, Communities can help identify critical nodes in a network, such as those that bridge different communities or have a high degree of centrality within a community. These nodes can be considered influential, essential or central in the network. Such node influence can be quantified using several measures, for example, the participation coefficient and the within-module z-degree.

Participation coefficient. The participation coefficient measures the diversity of connections of a node in a network [79]. It is a value between 0 and 1 and indicates the proportion of a node's connections absent in its most connected community. A node with a participation coefficient of 0 is fully embedded in its most connected community, while a node with a participation coefficient of 1 has an equal number of connections to all communities in the network. Community detection algorithms often use the participation coefficient to identify nodes that bridge different communities in a network.

It can be calculated using the following formula

$$P_i = 1 - \sum_{c=1}^{C} \left(\frac{k_i^c}{k_i} \right)^2 \tag{2.34}$$

where C is the number of communities in the network, k_i^c is the number of connections of node i within community c and k_i is the total degree of node i.

Within-module degree z-score. The within-module z-score measures the relative connectivity of a node in a network with respect to its community or module [79]. It quantifies how well a node connects to the other nodes in its community compared to how well it would connect to the other nodes in the network if its connections were randomly distributed. It can be calculated using the following formula:

$$z_i = \frac{k_i - \bar{k}_{s_i}}{\sigma_{k_{s_i}}} \tag{2.35}$$

where k_i is the degree of node i, s_i refers to the community to which node i belongs, $\bar{k}_{s_i}$ and $\sigma_{k_{s_i}}$ are the average degree and standard deviation of degree in the community s_i.

More advanced approaches, which will not be used in this thesis, include the spectral analysis of the Laplacian matrix of brain networks [42, 2], exponential random graph models [198, 199] and network motifs analysis [207].

2.2.2 Network Neuroscience

Given the interacting nature of neurons and areas in the brain, a network model might be a good choice. However, the problem of defining nodes and links is not trivial. We may be tempted to follow the first intuitive approach, modelling single neurons as nodes and synapses as edges. This approach has two significant issues:

1. As already discussed, the number of neurons and synapses is vast. Therefore, a whole human brain simulation would require a computational power far higher than the one we can access. To make a comparison, the only neural system analysed at single neuron and synapse level is the nervous system of the nematode worm *Caenorhabditis elegans*, which consists of 2,287 synapses connecting 279 neurons [221].

2. Even if it was possible to create a perfect full-scale replica of a human brain, the model would be just as complex as the original system, making it hard to extract knowledge from it.

However, the brain possesses a particular structure of groups of strongly connected neurons specialised in the same task [205, 218]. This allows researchers to focus only on some of these areas and study the whole brain at a larger scale by analysing their interactions.

So far, an optimal size of brain areas in terms of number of neurons has not been determined. Hence, brain networks have been studied at different scales, from whole brain networks consisting of 70 nodes [231] to 140,000 nodes [52]. This is advantageous because the specific network model can be constructed differently based on the requirements of each study [22]. However, it is essential to be careful as the number of nodes significantly affects the network properties [231].

Once the problem of node definition is solved, the next step is clearly defining what can be identified as a link. Based on the type of interaction represented, current approaches typically follow three different frameworks [63]:

1. **Structural networks**. This is probably the first intuitive approach as well as the oldest historically. In structural networks, brain areas are anatomical portions of the brain tissue which are physically connected by white matter tracks.

2. **Functional networks**. Here, links are identified in terms of some similarity measure between brain areas activities *i.e.*, temporal co-activation. In general, links cannot be interpreted as real physical connections between regions.

3. **Effective networks**. This is a similar approach to functional networks, with the difference that links represent effective interaction estimated using mechanistic models that rely on causality measures to estimate the strength of the connections.

Modelling the structural connectivity (SC) and the functional connectivity (FC) of the human brain as a network provides the possibility to apply the well-developed approaches of Graph Theory and Network Science [62]. Some remarkable results on SC and FC organisation include the discovery of nodes with a high number of connections, also called hubs [167, 223, 179], and their tendency to connect to each other creating dense clusters called rich clubs [13, 222]. Moreover, peripheral nodes have been often found directly linking to hubs creating the so-called core-periphery organisation [13]. On a higher scale, it was observed that the brain is organised in hierarchical communities, i.e. dense clusters of highly connected sub-networks further divisible in smaller communities [13, 19] and that it has the small-world property [19].

These encouraging results, as well as the growing amount of data and the evolution of computer science tools and methods, gave birth to a new discipline called Network Neuroscience [156]. Network Neuroscience provided useful insights in developmental science [114] and ageing [217], psychiatric conditions [148] and neurological diseases [20, 10], cognitive science [178, 154], metabolism [160] and many other fields.

2.2.3 Novel graph metric: Graphlets and Orbits

Recently, a new technique based on graphlets, defined as small connected induced subnetworks, has been proposed [181]. The main difference between graphlets and network motifs is that graphlets are connected and are used to identify local topological features, while motifs do not have to be induced or connected, and they need to be over-represented in the studied network compared to random networks [181]. If a graphlet is not an all-to-all connected graph (each node connected to all the others), different nodes may have different topological roles. These differences are addressed by introducing the concept of automorphism orbits; we say that nodes belong to different orbits if they have a different topological role (see [182] for a formal definition). For example, in the 3-node graphlet where node 1 is connected to node 2 and node 2 to node 3 (the second graphlet in Figure 2.8A), nodes 1 and 3 are topologically identical but different from node 2. Since graphlets are small, they represent more complex topological information about the neighbourhood of a node compared to classical graph theory measures like node degree or clustering coefficient. Graphlet-based measures have been used to compare real networks with theoretical network models [181, 182, 254], providing a comparison method which outperforms the older ones [254] and to understand how biological function is connected to topological structure in biological networks [149, 239].

In neuroscience, graphlet may potentially offer a powerful tool to characterise local topological roles of important nodes as well as offering a potential biomarker for the identification of specific conditions such as a disease. I will provide different example of potential applications in neuroscience of the tools described in the following sections.

2.2.3.1 Graphlet Degree Vector

We can associate to every single node in a graph a Graphlet Degree Vector (GDV) where the components are the degree of a given orbit, i.e. the times that node appears in the considered orbit [254]. For example, In Figure 2.8B, node V appears twice in orbit 5; hence, the correspondent component of the GDV of V will be 2. The number of times the given node appears in orbit 0, which is the first component of a GDV, is the degree of the node. In this context, the GDV can be interpreted as an extension of the degree measure of a node for the description of local topology [254]. This is useful as now we are able to discriminate between different types of important nodes

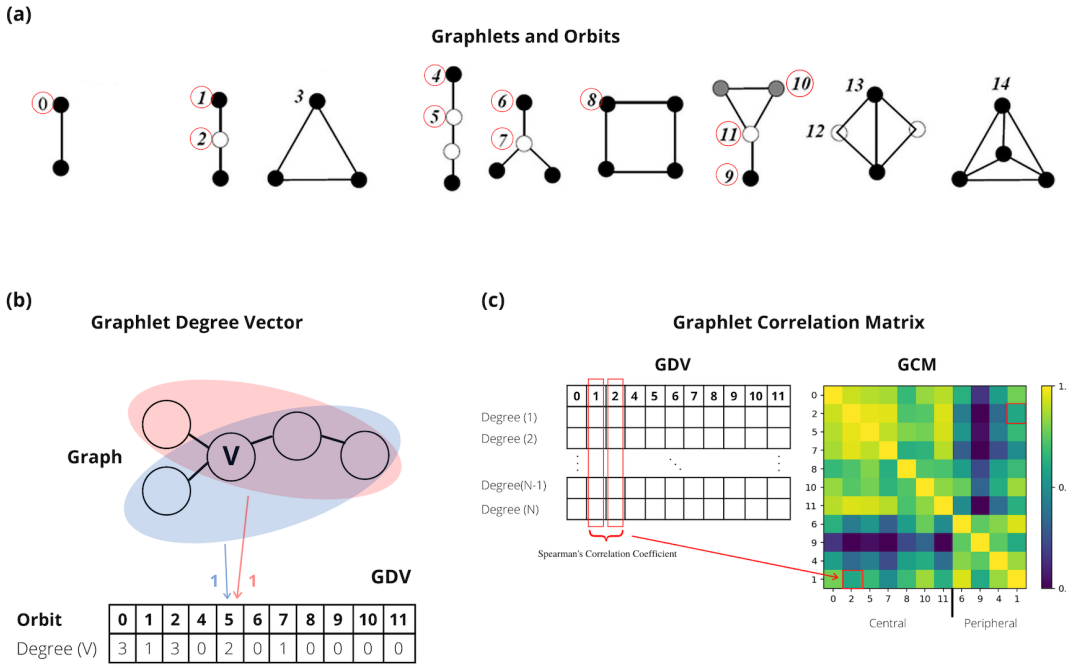

Figure 2.8: Illustration of graphlet-related methods. (a) All the 14 automorphism orbits of graphlets up to 4 nodes. The 11 linearly independent orbits are highlighted in red circles; (b) Construction of the Graphlet Degree Vector. For each node V of a graph, the elements of the vector are the number of times the node V appears in a given orbit. In the example, node V appears in the position of orbit 5 two times (outlined in red and blue). Hence, the GDV value of node V corresponding to orbit 5 is 2. Only the 11 independent orbits of graphlets up to 4 nodes have been considered. Repeating the procedures for each of the N nodes of a given Graph yields an N x 11 matrix where each row is the GDV of the correspondent node, and each column refers to a specific orbit; (c) Construction of a Graphlet Correlation Matrix. Each element of the GCM is the Spearman's correlation coefficient of two columns of the GDV. This value gives information about the correlation of two given orbits within a Graph. In the example, the Spearman correlation coefficient between columns of the GDV correspondent to orbits 1 and 2 has been calculated. The so-defined GCM is symmetric.

(hubs) in the brain while node degree only allowed to discriminate between hubs and low-connected nodes.

We will see in more details that graphlets and orbits can help discriminating between highly local nodes and nodes participating in topologically long-range connections. While for some kinds of socio-economic networks this is not the case, for brain networks nodes can have both roles, helping in identifying important brain areas involved in multi-modal activity. In addition, comparing information from "normal" subjects, graphlets can provide a way to assess how connectivity in brain structure or function associates with different states or diseases.

Considering all the graphlets up to 4 nodes, 15 different orbits can be identified. However, only 11 are linearly independent [254] (Figure 2.8A in red circles).

2.2.3.2 Graphlet Correlation Matrix

Stacking the GDVs of all the nodes in a graph, we construct an $N \times 11$ matrix where N is the dimension of the graph and 11 are the independent orbits of up to 4 node graphlets. We can now create the Graphlet Correlation Matrix (GCM) where the nm component is the Spearman's correlation coefficient of the n-th and the m-th columns of the matrix of stacked GDVs [254]. This procedure condenses the topology of a graph of any size into an 11×11 symmetric matrix where the values range from -1 to 1 [254]. Moreover, Graphlet Correlation Distance (GCD), defined as the Euclidean distance of the lower diagonal values of the GCMs of two networks, is a graph distance measure which outperforms all the other previously defined [254].

GCMs offer a powerful and elegant tools to condense topological information of a network of any dimension into a condensed 11×11 matrix, offering an efficient tool for comparison. This, could potentially provide a quick way to compare brains from different experimental settings or discriminate between different states or diseases (assuming that such conditions causes a change in the topology of the brain).

2.3 Potential contributions of Machine Learning to Neuroscience

In the previous sections I provided and introduction to some of the most successful computational tools starting from the smallest scale, the single neuron, going through neural mass models of neural population, up to graph theory which overlook the fine biological details and the use of differential equations to describe the neural system under investigation. Since graph theory is typically applied to study the whole brain,

I kind of completed the bottom-up approach anticipated, from single neuron to the whole brain. However, there is still a missing piece which is difficult to associate to a particular scale, but it is an essential computational tools which in the last decades has exploded in popularity given its staggering success in solving complex problem with ease. The scope of this section is to introduce such approach which is Artificial Intelligence and Machine Learning in particular, and providing two notable example of application of these tools to solve two active problems in neuroscience.

In the last decades, an increasing effort has been made to enable machines to perform tasks that intelligent entities perform efficiently [85, 127]. Artificial Intelligence (AI) has seen a real explosion, with many new publications each year, patents, and open-source code libraries. Providing an exhaustive overview of the broad field of AI is beyond the scope of this thesis. The focus instead will be on how AI, particularly Deep Learning, are beneficial to solving practical problems in Neuroscience. It is also convenient to note that since the best example of general intelligence we have so far comes from the human brain, AI research has intertwined with Neuroscience with considerable mutual benefits [85]. Many advances in AI were directly inspired by human capabilities at solving specific tasks and even the human brain's biological structure [85]. On the other hand, understanding how AI solves specific tasks may give insights into how information might be encoded and processed by human brain areas, helping advance neuroscience research [85].

2.3.1 Supervised vs. Unsupervised learning

In general, a machine learning model is a mapping function $f(x;\theta) : \mathrm{R}^I \rightarrow \mathrm{R}^O$ where I and O represent the dimensions of the input and output space, respectively. In particular, the mapping is defined between two statistical distributions, $P(I)$ and $P(O)$, which are the input and output data distributions, respectively. In practice, the true mapping function is unknown because we could usually access a relatively small subset of only $P_{train} \subset P(I,O)$ of the actual input and output distributions used to train such a model. This is the reason why machine learning models are considered approximation methods [74].

Machine learning algorithms mostly fall into two broad classes depending on the learning strategy and the task they have to solve: supervised and unsupervised learning [74].

In a typical supervised learning approach, the model is presented with data that has been previously labelled. The dataset used to perform supervised learning consists of a series of inputs and the respective labels or targets. The model is then trained

in an iterative way to map the input to the output in such a way as to minimise the prediction error. The learning is iterative because it works as follows: first, the model's parameters are randomly initialised, and a prediction on the input data is performed. Such prediction is compared with the ground truth (the labels), and a loss or cost function is used to determine the error committed by the model in making the prediction. The loss function can typically estimate the difference in the prediction and the ground truth, meaning it can estimate how far the model prediction is from the actual value. The gradient of the loss function is then computed. The gradient is a vector that indicates the parameter variation that causes the highest increase in the loss function. Since a machine learning approach aims to minimise the error between the prediction and the actual labels, a small portion of the gradient is subtracted from the respective parameters. In this way, the loss will be lower at the next iteration, and the predictions will be better. This procedure is called gradient descent, and each iteration is called a training epoch or simply epoch. Gradient descent is usually repeated until convergence, although we will see that choosing the optimal convergence point is not straightforward. Gradient descent can be expressed in mathematical terms as follows:

$$\bar{\theta}_n = \bar{\theta}_{n-1} - \alpha \nabla_{\bar{\theta}} \mathcal{L}(\hat{x}, \hat{y}; \theta) \tag{2.36}$$

where $\bar{\theta}$ are the model's parameters, n refers to the training epoch, $\mathcal{L}$ is the loss or cost function, $(\hat{x}, \hat{y}) \sim P(I, O)$ is a single data point including the input feed to the model and the generated output, and α is a parameter that controls the rate of the model's parameters update called learning rate. In practice, the gradient of the loss is not generally computed using a single data point. Instead, the input is divided into batches and the gradient is computed on the average loss of the data points within each batch. This approach is called mini-batch gradient descent.

Supervised learning tasks are typically divided into two main categories: classification and regression. In classification, the goal of the model is to assign a category or class to each input data point. In regression, the model learns a mapping from the input to a continuous value instead. A notable example of a supervised learning application is image classification, a problem to which researchers have dedicated a tremendous amount of resources in the past decades [125, 197, 89, 49].

In unsupervised learning instead, the input data does not come with labels. Hence, the aim of the algorithm is to discover hidden patterns, structures or relationships in the data. Some examples include clustering techniques [53], anomaly detection [164], or dimensionality reduction. In particular, dimensionality reduction techniques, like

Principal Component Analysis (PCA) [115] or t-SNE [140], are useful to reduce the dimension of the data while preserving their essential structure. In the next section, I will present PCA in more details and introduce one of the many potential application which will be used in Chapter 5.

It is also worth briefly mentioning reinforcement learning (RL), a different paradigm based on virtual agents that receive rewards or punishment when they perform a desired or undesired action, respectively [212]. Recently, reinforcement learning has seen a growing interest as a computational tool for neuroscience because of its promising applications to understand better how planning works in the human brain [48].

2.3.2 Principal Component Analysis (PCA)

In the previous section, I introduced the concept of dimensionality reduction techniques as a class of unsupervised learning algorithms. Here, I will present in more details one of the most popular dimensionality reduction algorithms: Principal Component Analysis (PCA) [174]. Similarly to other dimensionality reduction techniques, PCA's primary goal is to reduce the dimensions of a dataset, transforming it into a lower-dimensional representation. It is designed in such a way as to preserve the essential features in the data, with a particular focus on the sources of variation. PCA comprises various sequential steps and typically starts with data preprocessing, such as standardisation or normalisation. The first step consists of the computation of the covariance matrix of the input data. This matrix represents the variation of each variable with respect to all the others. In other words, the covariance matrix encodes how each data point varies with respect to the others. Then, PCA identifies the eigenvectors of the covariance matrix along with the respective eigenvalues through eigendecomposition or singular value decomposition (SVD). Eigenvectors will be the principal components, and their eigenvalues represent the fraction of variance in the data captured (explained) by each component. Principal components are orthogonal, *i.e.*, they are uncorrelated. Hence, they represent the relationships between the original data points, providing insights into their linear dependencies. Finally, principal components are sorted based on the fraction of explained variance, meaning that the first component is the one that captures the most variance, and so on. We can then select only a subset of components based on the desired level of explained variance. For example, a typical approach is to set a variance threshold (e.g. 90%) and estimate how many principal components are needed to explain such a level of variance. When all PCA procedures are completed, the original data can be projected onto the selected principal components in order to obtain a reduced-dimensional representation

of the original dataset. Such projected data can then be used for further processing or other kinds of analysis or predictive modelling. PCA has many applications in various fields, including image and signal processing, data compression, feature extraction, exploratory data analysis, and more [116]. Although PCA is a powerful dimensionality reduction method, it is essential to note that it makes a strong assumption, like the linear relation of the input data. Datasets in which data points have more complex relations than simply linear may not benefit from applying PCA. Later in this chapter, I will discuss how PCA can help identify latent components of neural activity. This will be beneficial in better understanding how the brain encodes information from the external world as well as providing a more efficient tool for neural decoding, *i.e.*, translating brain activity into a desired output. Chapter 5 will show a possible application for motor behaviour decoding.

2.3.3 Deep Learning

Deep Learning is a revolutionising sub-field of Machine Learning, the branch of computer science focused on enabling machines to 'learn', *i.e.*, improving performances in specific tasks without external human supervision [74]. The term *(*deep) refers to the use of artificial neural networks (ANNs), also called deep neural networks (DNNs). These networks generally comprise a sequence of basic processing units ensembles. The ensembles take the name of layers while the basic units are called neurons, because they are inspired by the functioning of real neurons in the human brain[2]. Each layer can be seen as a mathematical transformation of the input data. Each of these transformations projects the input data into a different space, transforming it into increasingly complex representations as we go deeper into the sequence of layers. The goal of the model is to learn an optimal representation, also called embedding, which helps solve a specific task.

The birth of Deep Learning is usually located in the 1950s and 1960s and attributed to Frank Rosenblatt, who developed the perceptron algorithm [189] as well as to Paul Werbos who introduced the backpropagation algorithm for training neural networks [235]. However, the field did not immediately get the interest he deserved.

[2]The fact that artificial neurons are inspired by real neurons should not be interpreted as the fact that they are necessarily a good model of real biological neurons. Similarly, even if DNNs reach comparable performances as humans in specific tasks, this does not guarantee that the computation carried by the DNNs is comparable to the ones performed by the human brain. I will briefly present a proof of concept of a DNN ensemble to mimic the behavioural results of human participants performing a 2AFC task in Chapter 4, discussing the advantages as well as the limitations of these kinds of approaches to study the human cognition.

One of the main reasons was that perceptrons were strongly criticised for being unable to solve a problem as simple as learning the XOR logic function [151]. The interest in Deep Learning rose again in the mid-1980s with a series of studies culminating in the formulation of the Universal Approximation Theorem (UAT), which states that a conceptually simple DNN with sufficient processing units can approximate any function at any level of accuracy [99].

When introducing the broad field of Deep Learning, a typical starting point is defining the feedforward neural network (FNN), also known as a feedforward network or an artificial feedforward network (AFN) [74]. This is the conceptually most straightforward Deep Learning algorithm as the information flows in a unique direction: from the input to the output layer, passing sequentially through the hidden layers without making a cycle. The number of hidden layers, as well as the number of hidden units or neurons[3] is variable and can be adapted based on the complexity of the task the algorithm needs to solve. These numbers, together with the learning rate and other parameters we will encounter in this section, are called hyperparameters, and they can be tuned to optimise the model's performance.

The data is fed into the input layer, and each layer applies a linear projection followed by a non-linear transformation to the data, known as activation functions, before passing it on to the next layer:

$$y = \sigma(xA^T + b) \tag{2.37}$$

where y represents the output of a unit, x is the input to the unit, σ a non linear activation function, A^T is the transpose parameters matrix and b a bias term. The most common choices for the activation function includes sigmoid, tanh or ReLU. Under this framework, a FNN, as well as more complex architectures, can be seen as a nested function in the form of

$$y = f_n(f_{n-1}(...(f_2(f_1(x))))). \tag{2.38}$$

The final model's output is usually the output of the final layer, also called the output layer. The number of neurons in the output layer determines the dimensionality of the output. The output can be a single value, discrete for classification or continuous for regression. However, it can assume a more complex form like vectors, for example, when the desired output is in the form of sequences, images, or reconstruction of the input.

[3]In the field of AI, the term neuron or artificial neuron are usually interchangeable as there is no ambiguity with biological neurons.

A specific type of FNN is the Multi-Layer Perceptron (MLP), a neural network composed of multiple layers of artificial neurons. An MLP is typically fully connected, meaning that the output of each neuron of the previous layer is fed into all of the neurons of the next layer. The number of artificial neurons in a single layer determines the dimensionality of the space where the linear projection is applied. For example, in the specific case of a layer with three artificial neurons and a single feature x, Eq. (2.37) becomes

$$y = \sigma(a_1 x + a_2 x + a_3 x + b). \tag{2.39}$$

Given their conceptual simplicity, relative ease of implementation and versatility, MLPs are the most used Deep Learning architectures with a variety of applications.

In section 2.3.1, machine learning was defined as an approximation method, and later, the Universal Approximation Theorem (UAT) was introduced. More specifically, the UAT states that a feedforward neural network with a single hidden layer composed of a finite number of neurons can approximate any continuous function, given enough hidden neurons [99]. The proof of the theorem will not be reported here in detail. However, it is based on the simple intuition that non-linear activation functions used in Deep Learning are smooth, non-polynomial functions that can be linearly combined to approximate any other smooth function. It is important to note that the UAT provides theoretical proof of the feasibility of a universal approximation with a DNN. In practice, a neural network with many hidden neurons may still not be able to approximate a function well because of overfitting or other practical issues.

More complex mathematical functions require more units per layer to be approximated, and in some cases, the dimension of such an algorithm would be unreasonably high. For this reason, part of the effort in deep learning research has been put into finding more efficient ways to perform tasks using different kinds of data. For example, images are defined as grids of $h \times w \times c$ where h and w are the height and width in pixels, respectively, and c is the number of channels (for example, $c = 1$ for grey-scale images or $c = 3$ for RGB images). In such cases, instead of passing each pixel to a single unit in an MLP, processing the input image with operations like convolutions and pooling to identify space-invariant features allows the algorithm to generalise to different tasks involving visual data. These operations are the essential components of Convolutional Neural Networks (CNN) first introduced in [131]. On the other hand, data that have a temporal component (like time series) or that are more generally ordered in something "before" and "after" (like stock index prices or words in sentences) are best processed by recurrent architectures, in which a hidden state is

recurrently passed to the following processing unit to keep track of the previous state of the sequence. Such layers are the basis of Recurrent Neural Networks (RNN), first introduced in [191]. Specifically, given an input sequence x_t, each layer computes and outputs y_t as well as a hidden state h_t, which keeps track of the information contained in previous elements of the sequence,

$$\begin{aligned} h_t &= \sigma_h(W_h x_t + U_h h_{t-1} + b_h) \\ y_t &= \sigma_y(W_h h_t + b_y) \end{aligned} \tag{2.40}$$

where σ represents a non-linear activation function, W and U are parameter matrices and b a bias term. When training RNNs, since the gradient is computed at each point in time (element of the sequence), the following products of gradients could become very small (vanishing gradient) or large (exploding gradient), making it difficult to continue the training procedure. This problem has been addressed by researchers with the introduction of variations of RNNs, such as long short-term memory (LSTM) networks [94] and gated recurrent units (GRUs) [27]. These kinds of architectures are able to retain information over a more extended period of time (more distant elements in the sequence), and they have been proven to help avoid vanishing or exploding gradients [247].

In the following two sections, I will present two of the many possible applications that Deep Learning has in Neuroscience.

2.3.4 Deep Learning for parameter recovery

Parameter recovery is the first problem in Neuroscience that can benefit from using Deep Learning. In general, parameter recovery is the problem of finding the optimal parameters of a model that fit the data collected from experimental studies [100, 236]. In psychology and Neuroscience, this framework is instrumental for many reasons. Here, I will present two main applications which will be used in 4.

The first task that can benefit from parameter recovery is assessing a model's performance when reproducing specific cognitive functions. By finding the parameters that optimise a model's performance, researchers can more easily spot pitfalls and implement improvements to the model [21].

The second task is identifying differences across cognitive functions or different groups of participants. For example, one potential application is recovering the optimal parameters of a cognitive model that fits the behavioural data of a group of healthy participants and repeat the procedure for a group of subjects affected by

neurological diseases or other conditions [184, 25]. In particular, the potential application to healthy subjects for early diagnosis of neurodegenerative disease is a field of growing interest for the potential improvement it can have on people's mental health span.

Classical methods for parameter recovery include Bayesian estimation [152], Markov chain Monte Carlo [250], and Swarm Intelligence algorithms [51]. Recently, Deep Learning has become a valuable ally for solving the problem of parameter recovery. In fact, computational models of cognition, especially those more biologically plausible, have become increasingly complex, meaning they include more parameters. As the number of parameters in a model increases, so does the difficulty in recovering them. Moreover, parameter recovery of models with parameters with a similar effect on the output can also be challenging because of trade-off effects [236]. Deep learning helps solve these issues thanks to its capabilities of learning features by itself without heavy data preprocessing, different from other statistical or machine learning applications where feature engineering is an essential step and might require a considerable amount of resources.

Deep Learning also has some drawbacks that must be carefully considered. One such drawback is that powerful Deep Learning algorithms' are usually highly complex, meaning they have many trainable parameters. A direct consequence is that a large amount of data is usually required for successful training to make sure the algorithm generalises well to unseen data (the problem of training an algorithm with few data samples, the so-called one-shot learning problem, has not yet a unique, satisfying solution and it is currently an active field of research in AI [124, 227]). On the other hand, while data availability is limited in many cases, data from cognitive computational models can be generated at will, with the only limitation being the computational resources available.

It is also worth mentioning that when the model's interpretability is a hard requirement, instead of just a nice-to-have, the use of Deep Learning may pose some additional challenges. Indeed, the high complexity of Deep Learning algorithms makes it highly challenging to identify what a learned feature represents. This is because learned features are generally abstract, non-linear combinations of the inputs that humans cannot interpret directly. These challenges are currently tackled by the field of interpretability or explainability of AI algorithms to make future AI systems trustworthy and aligned with humans' values and goals [193].

In general, Deep Learning is exceptionally suitable for parameter recovery. In Chapter 4, a simple MLP will be used for parameter recovery of a biologically plausible

model of perceptual decision-making [242] introduced in Section 2.1.3.1.

2.3.5 Decoding continuous behaviour using Deep Learning

The second problem treated in this thesis that benefits from using Deep Learning is making predictions about the outside world by interpreting activity recorded from the brain, usually referred to as neuronal decoding [69]. Neuronal decoding typically uses a computational model called a decoder that takes neuronal activity in various forms as input and generates an output to predict what information is encoded in the neuronal activity. The problem of how the brain encodes information is still an active field of research [188]. As anticipated in Chapter 1, the human brain is an exceptional encoder because it can take a large amount of data in the form of sensory inputs, as well as other internal neural activity, and combine them to create an efficient neural representation that can be used as a model of the external world. Similarly to the case of embeddings in an artificial neural network, a comprehensive understanding of what computation the brain makes to generate such representations is still missing. Neuronal decoding can help shed light onto this yet obscure mechanism, allowing researchers to understand better what information about the external world and internal states is represented by neural activity. It can be a valuable tool for researchers exploring the neural mechanisms underlying various cognitive and behavioural processes such as attention, memory, and decision-making [234]. Moreover, an efficient decoding model that can translate neural activity into behavioural output, especially if in real time, made it possible for the development of advanced neurotechnologies like Brain-Computer Interfaces (BCI) or neuro-prosthetic [240]. These technologies are already helping several people who suffered the loss of limb control either from a traumatic loss [186], spinal cord injuries [15], or because of other neurological conditions [121] which may also impair other critical day-to-day activities like speaking [139] or even seeing [98]. Moreover, BCIs have been shown promising applications for neurological and psychiatric diagnoses such as Alzheimer's disease, Parkinson's disease, and schizophrenia [200]. Of particular interest is the combined usage of BCIs and non-invasive neurostimulation approaches such as deep brain stimulation (DBS) or transcranial magnetic stimulation (TMS) to regulate abnormal neural activity and treat various neurological conditions [190].

2.3.5.1 Latent dynamics analysis of neuronal data

Usually, neuronal decoding is performed on the entire dataset coming from a neuronal recording session. However, the data may be high dimensional in some cases,

making it challenging to perform neuronal decoding in the whole dataset. Neuronal data usually present different correlation patterns that vary based on the specific brain region, experimental conditions or the type of neural activity being recorded [30]. In such cases, dimensionality reduction techniques such as PCA, discussed in section 2.3.2, can help identify lower-dimensional representations of the neural data. Such representations are usually referred to as latent variables or latent dynamics, and they constitute unobserved or hidden variables underlying the observed neural data [38]. The term 'latent' comes from the fact that these variables can not be directly measured but are typically inferred based on statistical models and analysis techniques. Latent variables can be thought of as primary sources of variability that give rise to the observed neuronal responses [38] and may represent different aspects of neural activity, such as firing rates, oscillatory patterns, or network dynamics.

Among the several applications, latent dynamics analysis can provide valuable insights into the neural basis of motor behaviour and enable accurate decoding of motor-related information from neuronal data. By leveraging the latent variables that capture the underlying motor dynamics, it becomes possible to decode specific aspects of motor behaviour, such as movement trajectories, kinematics, and even intended actions, from the patterns of neural activity [29, 168]. Using latent dynamics instead of the original high-dimensional data for neuronal decoding of motor behaviour is also advantageous to improve the signal-to-noise ratio by capturing the shared variability across neurons or time points related to the motor behaviour, effectively filtering out noise and uninformative fluctuations [249]. Moreover, latent dynamics analysis can provide a compact representation of the neural data that generalises well across unseen trials or different experimental conditions [3].

In chapter 5, I will use PCA to identify the latent variables able to account for 80% and 90% of the explained variable and use such variables to see how decoding performances vary with respect to the case of the decoding of the whole neural dataset.

Chapter 3

Graphlet analysis of structural and functional connectomes

3.1 Introduction

Integrating data from different imaging modalities, such as structural and functional MRI (fMRI), electroencephalography (EEG) or magnetoencephalography (MEG), is an active field of research given its promising applications to uncover the relationship between the structure and the functioning of the brain. In particular, the effective integration of fMRI and MEG-derived functional connectivity (FC) has received an enormous interest in the last decades because of the dual nature of temporal/spatial resolutions of the modalities [88]. Data from fMRI has a high spatial resolution but poor temporal resolution, while EEG/MEG possesses an excellent temporal resolution at the cost of a reduced spatial resolution. Integrating fMRI and MEG can combine the potential of each approach, minimising their drawbacks. Several integration approaches have been proposed. The most conceptually simple are the so-called fusion methods, which combine the features extracted from the data by averaging or concatenating them [201]. However, while the simplicity of such approaches is definitely an advantage, these methods suffer from the risk of minimising or even ignoring the specific strengths of each modality. A more rigorous approach is multimodal data analysis, a set of statistical methods, such as multivariate analysis, to model the relationship between the modalities. In this framework, we may find powerful methods such as Partial Least Squares (PLS), Canonical Correlation Analysis (CCA), or Independent Component Analysis (ICA) [252].

In this work, I will focus on another class of methods called *connectivity-based* approaches, specifically using graph-theoretic tools and measures [106]. Graph-theory comprises a large set of tools to analyse the local and global properties of brain

networks. These approaches have been extremely helpful in understanding the brain structure and functioning, combining different type of networks (structural, functional or effective) [155, 97, 128]. Moreover, graph-theoretical approaches can be adapted to integrate data from various imaging modalities, like MRI, fMRI or EEG/MEG among the others, to study the relationships between structure and function [203, 97]. Graph-theory-based integration of structural and functional connectomes has been proven effective in uncovering new properties of the brain and explaining how complex brain dynamics arise from its structure [169]. At the same time, fMRI and MEG have been used to identify critical parts of the human brain involved in complex tasks like expressive language [246] or to identify biomarkers of neurological diseases [84] for early diagnosis and treatments.

Recently, a novel Graph Theory approach has been proposed by making use of graphlets. Graphlets are small, induced, connected subgraphs. They are similar to network motifs, with the exception that motifs are defined as under or over-represented compared to a random graph null model. The frequency of different graphlets and their correlation is able to characterise the topology of the network under investigation and can provide markers to classify networks from different domains [181]. The frequency of graphlets can be evaluated for each node using the Graphlet Degree Vector (GDV), an extension of the concept of node degree, which is then used to estimate the Graphlet Correlation Matrix (GCM), a helpful tool whose dimension does not depend on the network size. These tools were presented in detail in Section 2.2.3. To the best of my knowledge, there has yet to be a comprehensive study on the graphlet statistics of human brain connectome across multiple modalities, and this offers an excellent opportunity to use advanced graph-theory tools for multimodal integration. The current study proposes graphlet-based analysis on whole-brain SC (DWI) and FC (fMRI and MEG) to address the main topological differences across these modalities.

This chapter's analysis will be guided by three main questions. First, how do graphlet-based measures vary across imaging modalities? In other words, can graphlet-based approaches be a suitable graph-theory-based multimodal integration method? Second, how does graphlet-based topology vary between groups of brain regions (brain clusters)? In particular, the large-scale brain organisation suggested by [70] will be used. Third, how do grahplet-based measures relate to conventional graph measures? This question is vital to ensure that graphlets can effectively uncover new topological information unavailable to conventional graph metrics and provide a quick and effective way to condense such information, which normally require several different

graph measures and extensive computational resources. In particular, I will present the results of experiments designed to understand how graphlet-based measures relate to conventional graph measures by correlating specific graphlet-based statistics with six classical graph measures: clustering coefficient, average degree neighbour, participation coefficient, within-module degree z-score, betweenness centrality and closeness centrality.

3.2 Methods

3.2.1 Neuroimaging

Data used in this chapter come from three different non-invasive neuroimaging techniques: Magnetic Resonance Imaging (MRI), functional MRI (fMRI) and Magnetoencephalogram (MEG). In this section, I will present a brief introduction to each method.

3.2.1.1 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is based on the principles of nuclear magnetic resonance (NMR), a phenomenon that occurs when nuclei of certain atoms are placed in a magnetic field and exposed to radiofrequency (RF) pulses. Hydrogen is the principal atom used in magnetic resonance imaging (MRI) because of its large abundance in biological tissues.

When a subject is positioned in the MRI scanner, under the influence of a static magnetic field B_0, the spins of hydrogen nuclei (protons) in tissues are aligned along the direction of B_0. These aligned proton spins are not static. Instead, a precession motion around the axis of the magnetic field is observed. Such precession happens at a characteristic frequency called the Larmor frequency, which is proportional to the strength of the external magnetic field. Next, a brief radiofrequency pulse (RF pulse) at the Larmor frequency is applied, momentarily perturbing the protons' spins which will not align with B_0 anymore. When the RF pulse is switched off, the protons spin relax to their initial equilibrium state, in alignment with B_0. This decay process is referred to as relaxation, and the signal takes the name of Free Induction Decay (FID) signal. Relaxation phenomena involve T1 relaxation (longitudinal relaxation) and T2 relaxation (transverse relaxation).

Through T1 relaxation, also known as spin-lattice relaxation, protons transfer energy to their neighbouring atoms (the lattice) and realign with the magnetic field B_0. The time constant $T1$ quantifies the time it takes for the longitudinal magnetisation

to recover approximately 63% of its original value. Different tissues have different T1 relaxation times, allowing for contrast between different types of tissues in T1-weighted images. For example, fat has a shorter T1 time compared to water, leading to a brighter signal in fat and a darker signal in water on T1-weighted images.

T2 relaxation, or spin-spin relaxation, involves the dephasing of proton spins in the transverse plane due to interactions between neighbouring spins. The time constant T2 is the time it takes for the transverse magnetisation to decay to about 37% of its initial value. T2-weighted images provide contrast based on the differences in T2 relaxation times of various tissues. Tissues with high water content, such as cerebrospinal fluid, have longer T2 relaxation time and appear brighter on T2-weighted images, while tissues like white matter, with shorter T2 relaxation times, appear darker.

In addition to T2 relaxation, there is a phenomenon known as T2* decay, which accounts for magnetic field inhomogeneities and susceptibility effects. T2* decay occurs faster than T2 decay due to these additional factors, leading to a more rapid signal loss. Hence, it is possible to generate specific MRI pulse sequences to create images with particular sensitivity to magnetic field inhomogeneities and T2* relaxation effects. This makes such sequences perfectly suited for another neuroimaging technique called functional MRI (fMRI) focused on detecting changes in blood oxygenation levels, which indirectly relate to neuronal activity.

The contrast in MRI images is primarily determined by the differences in T1 and T2 relaxation times of tissues, as well as the proton density (the number of hydrogen nuclei per unit volume). By adjusting the timing parameters of the RF pulses and the data acquisition, it is possible to create images that emphasise T1 contrast (T1-weighted images), T2 contrast (T2-weighted images), or proton density. Two parameters are of particular importance, Repetition Time (TR) and Time to Echo (TE). Normally, a series of RF pulses is applied instead of just one single pulse. TR identifies the amount of time between successive pulse sequences applied to the same portion of tissue. TE instead, is the difference in time that occurs from the delivery of the RF pulse and the receipt of the relaxation signal (echo). To obtain T1-weighted images, short TR (repetition time) and short TE (echo time) values are used to emphasise T1 differences, making fat appear bright and water appear dark. On the other hand, T2-weighted images can be obtained using long TR and long TE values, making water appear bright and fat appear darker. Finally, long TR (similar to T2-weighted images) and short TE values highlight differences in proton

density, providing a different type of tissue contrast which is the base of proton density-weighted images.

A commercial MR scanner is typically composed of four principal components. The primary magnet creates the high-intensity static magnetic field which causes the nuclei to polarise and spins to align with the magnetic field lines. The Radio-frequency coils (RF coils) generate the variable magnetic field. The gradient coils apply a smooth (linear) small variation to the static field at different locations. Such variation, slightly changes the spin precession frequency or phase, making it easy to encode the spatial information of an imaged object in the frequency associated with a position in space. The auxiliary coils, compensate for potential inhomogeneities in the principal magnetic fields.

In the next two subsections, we briefly introduce two MRI modalities used in the current chapter: Diffusion-Weighted Imaging (DWI) and functional MRI (fMRI).

3.2.1.2 Diffusion-Weighted Imaging

DWI is a non-invasive neuroimaging technique that measures the random Brownian motion of water molecules within a voxel of tissue in vivo and non-invasively [215] with MR. In tissues, diffusion is often restricted by cellular structures such as membranes, fibres, and macromolecules. DWI exploits these differences in water diffusion to generate contrast in MR images.

Specifically, the T2 signal attenuation is based on how easily water molecules diffuse in a specific region. The further the distance covered by water molecules (easier diffusion) the lower the amount of the initial T2 which persists. For example, water within cerebrospinal fluid (CSF) generally diffuses very easily, causing the areas to be black because no signal remains. Conversely, water within the grey matter cannot move as easily because of the numerous cell membranes acting as obstacles and therefore the initial T2 signal is only slightly attenuated. DWI uses strong paired gradients to sensitise the MRI signal to the movement of water molecules. These diffusion-sensitising magnetic gradients are applied in different directions, and the signal attenuation observed is related to the extent of water diffusion in those directions. The first gradient dephases the spins of water protons, and the second gradient, applied after a certain time interval, attempts to re-phase them. If water molecules have moved (diffused) between the application of these gradients, the rephasing is incomplete, leading to signal attenuation. The degree of signal attenuation is proportional to the amount of diffusion, allowing for the creation of diffusion-weighted images.

DWI first focuses on the extraction of diffusion information from the tissue, obtaining a T2* weighted image. Next, diffusion attenuation is applied using the b-value. The b-value is a crucial parameter that controls the sensitivity of the MRI signal to the diffusion of water molecules. It is defined as

$$b = \gamma^2 G^2 \delta^2 \left(\Delta - \frac{\delta}{3} \right) \tag{3.1}$$

where γ is the gyromagnetic ratio, a constant specific to the type of nucleus being imaged (for hydrogen, it is approximately 42.58 MHz/T), G is the strength of the magnetic gradient applied, δ is the duration of each gradient pulse in ms, Δ is the time between the onset of the two gradient pulses (ms). Higher b-values increase the sensitivity to diffusion but also result in greater signal attenuation. Typical b-values range from 0 (the first T2*-weighted image with no diffusion weighting) to 1000 s/mm^2 or higher, depending on the application. Larger b values can be obtained by applying stronger magnetic gradients as well as of longer duration, and by increasing the interval between paired gradient pulses. Water diffusion within tissues is quantified by the apparent diffusion coefficient (ADC). It can be calculated by acquiring images at different b-values and fitting the data to the following equation:

$$S(b) = S(0)e^{-b \cdot D} \tag{3.2}$$

where $S(b)$ is the signal intensity with diffusion weighting, $S(0)$ the signal without diffusion weighting $b = 0$, and D is the ADC. The ADC reflects the degree of water diffusion in tissues, with higher ADC values indicating more free diffusion and lower values indicating restricted diffusion.

3.2.1.3 Functional MRI

Functional MRI is another non-invasive MR-based technique, focused on the study of the time evolution of a particular signal to infer functional correlation across different brain areas [113]. It does not detect neuronal activity directly but instead measures changes in blood flow, blood volume, and blood oxygenation, which are coupled to neuronal activity. This phenomenon is known as neurovascular coupling. The most common fMRI technique relies on the Blood Oxygenation Level-Dependent (BOLD) signal, which reflects changes in the magnetic properties of blood due to varying levels of oxygenation. When neurons become active, they consume more oxygen. In response, local blood flow increases to supply the active region with oxygen, leading to a complex sequence of haemodynamic changes. Changes in blood oxygenation alter

the local magnetic field, which affects the MR signal. Oxygenated haemoglobin is diamagnetic (weakly repelled by a magnetic field), while deoxygenated haemoglobin is paramagnetic (attracted to a magnetic field). The relative concentration of these two forms of haemoglobin affects the T2* relaxation time, which is what the BOLD signal measures. Functional MRI offers good spatial resolution, typically in the range of 1-3 mm, allowing for detailed mapping of brain activity. However, its temporal resolution is limited by the haemodynamic response, which occurs over several seconds. This temporal delay means that fMRI cannot capture the rapid neural activity but instead provides an averaged response over a short period.

3.2.1.4 Magnetoencephalogram

Magnetoencephalography (MEG) is a non-invasive neuroimaging technique used to measure the magnetic fields produced by neuronal activity in the brain. Unlike fMRI, which relies on changes in blood flow or magnetic properties of blood, MEG directly captures the magnetic signals generated by the electrical currents flowing through neurons. This direct measurement offers unique advantages in temporal resolution although it does not have a good spatial resolution compared with other typical MR-based techniques.

Neuronal activity in the brain generates tiny magnetic fields, on the order of magnitude of 10 fT, primarily from the synchronous activity of pyramidal neurons in the cortex. According to Maxwell's laws, when these neurons fire, they produce electrical currents that create corresponding magnetic fields. MEG sensors, typically superconducting quantum interference devices (SQUIDs), detect these magnetic fields outside the head with high sensitivity and precision. Given the extremely small magnitude of the magnetic field generated by neuronal currents, MEG scanners are required to be located in magnetically shielded rooms (Faraday cages). During an MEG recording session, the subject typically sits or lies down with their head positioned within an array of MEG sensors. The sensors detect the magnetic fields generated by neuronal activity in real-time, providing millisecond-level temporal resolution. This high temporal resolution makes MEG particularly valuable for studying the dynamics of brain processes.

After MEG data acquisition, several pre-processing steps need to be performed. For example, artefacts like eye blinks or other minor events that introduce noise in the recordings need to be filtered. Following the preprocessing step, an essential part of MEG data processing is source reconstruction. Since MEG sensors record magnetic fields outside the head, it is necessary to reconstruct the sources of these signals

within the brain. This process, known as source localisation or source reconstruction, involves mathematical modelling to estimate the locations and strengths of the neuronal sources generating the observed magnetic fields. In section 3.2.2.11, the algorithm used for this study will be conceptually explained and compared to other common source reconstruction methods.

3.2.2 Imaging data acquisition and processing

In this section I will thoroughly present the procedures carried for data acquisition and pre-processing. All the procedures presented before Section 3.2.2.13 were carried by different member of the team involved in the project.

3.2.2.1 Participants

The participant cohort for this study consists of a group of 29 individuals who are in good health and were selected from the participant panel of Cardiff University School of Psychology. Ages range from 18 to 35 years and a mean age of 21.1 $\pm$2.94 years. None of the participants have a documented history of neurological or psychiatric disorders. The study was approved by the Cardiff University School of Psychology Research Ethics Committee, and all participants provided their informed consent in written form.

3.2.2.2 MRI data acquisition

The diffusion-weighted images (DWI) were obtained for the whole brain, multi-shell, of all participants from a Siemens 3T Connectom MRI scanner (Siemens Medical Systems) at the Cardiff University Brain Research Imaging Centre (CUBRIC). A HARDI protocol was utilized for spin-echo echoplanar imaging (EPI) pulse sequence (echo time 59 ms, repetition time 3000 ms, voxel size 2×2×2 mm). The application of diffusion sensitizing gradients followed in 20 isotropic directions at b-values of 200 and 500 s/mm^2, in 30 isotropic directions at a b-value of 1200 s/mm^2 and in 61 isotropic directions at b-values of 2400, 4000, 6000 s/mm^2. Thirteen volumes without diffusion weighting (b = 0 s/mm^2) were obtained with an interleaved order across the sequence. A correction for susceptibility-induced distortions was applied by acquiring three images at b = 0 s/mm^2 and 30 diffusion directions at b=1200 s/mm^2 with the opposite phase encoding direction. Participants also went through high-resolution T1-weighted magnetization prepared rapid gradient echo scanning (MP-RAGE: echo

time 3.06 ms; repetition time 2250 ms sequence, flip angle 9°, field of view=256×256 mm, acquisition matrix 256×256, voxel size 1×1×1 mm).

A multi-band protocol (echo time 35 ms, repetition time 1500 ms, voxel size 2 × 2 × 2 mm, multi-band factor 3, flip angle 70°, AC-PC alignment with a posterior-down tilt) was used for the acquisition of the final BOLD-sensitive T2*-weighted EPI images for the whole brain. A total of 315 volumes were obtained from an interleaved sequence. The instructions given to participants during the scan were to rest with eyes open and focus on a red dot projected on a grey background. Analogously to MRI data acquisition, a red dot was projected to a grey background, and participants were instructed to fixate their eyes on it. The duration of each recording session was about 8 minutes.

3.2.2.3 MEG data acquisition

MEG recordings for the whole brain were collected with a 275-channel CTF radial gradiometer system (CTF Systems, Canada) in a magnetically shielded chamber with a sampling rate of 1200 Hz. During the analysis of the primary sensors, it was discovered that one sensor exhibited an excessive level of noise, rendering its data unreliable for further examination. To cancel out any residual noise effectively, 29 additional reference channels were recorded while analysing the primary sensors as synthetic third-order gradiometers [229]. Blinks and eye movements were tracked using Continuous horizontal and vertical bipolar electro-oculograms (EOG). To minimise the potential effects of head movement on the recorded data, a chin rest was utilised while participants were seated in the MEG chair. This arrangement aimed to restrict excessive head movements during the experimental sessions by providing stable and comfortable support for the chin. This precautionary measure aimed to ensure the acquisition of high-quality and reliable data by minimising artefacts that could arise from unintentional head movements. A Polhemus FASTRAK (Colchester, Vermont) was used to digitise the head shape with the position of coils for MEG/MRI co-registration.

3.2.2.4 Cortical reconstruction

The T1-weighted MP-RAGE images were processed using Freesurfer (version 5.3.0, `http://surfer.nmr.mgh.harvard.edu`). The pre-processing consisted of motion correction, intensity normalisation, skull-stripping, white-matter segmentation, tessellation, smoothing, inflating and spherical registration [58]. Following pre-processing, the surface of grey matter/white matter with inner skull, scalp and pial images were

created. The mean non-diffusion image (b = 0 s/mm^2) was used as the base for the registration of conformed and intensity normalised T1-weighted image with a boundary-based rigid body registration with six degrees of freedom [76]. For each participant, the forward and inverse transformation matrices between the native DWI space and T1 space were computed for the following co-registration and tractography analyses.

The brain surface was parcellated into 360 regions using the Multimodal Human Connectome Project-based (HCP-MMP) atlas [70]. After careful examination and analysis, it was determined that errors were present in the hippocampus regions of several subjects. As a result, these regions had to be discarded from further consideration in the study. Despite this setback, the remaining brain regions were deemed reliable and suitable for analysis. Consequently, 358 region surfaces were included in the subsequent investigations, allowing for a comprehensive examination of the neural activity and connectivity within the remaining brain regions. The range of dimensions of the brain regions was wide (122 mm^2 smallest region to 3198 mm^2 largest region) in the original HCP-MMP atlas. More extensive regions in the HCP-MMP atlas were subdivided into a determined number of sub-regions using the "mris_divide_parcellation" command from Freesurfer to obtain more homogeneous size across ROIs (average largest region size was 454 mm^2). Smaller and larger subdivisions were checked manually and corrected. The final structural connectome contained 664 regions of interest (ROI).

Annotation files were registered in the *fsaverage* space to the subject space, and the surface was sampled into volume to create subject-level parcellation of subdivided HCP-MMP atlas. For each subject, cortical parcellation volume was co-registered to native DWI space with no resampling and closest white-matter voxels were assigned to grey matter regions. The assignation was performed by searching the grey-matter voxels less distant than 1.73mm in a 5x5x5mm cube centred at the white-matter voxel.

Native space-registered Freesurfer segmentation images, including white matter and dilated grey matter/white matter boundary used in tractography, were used to create the white matter mask.

3.2.2.5 DWI data preprocessing

DWI data format was converted from DICOM to NIfTI format using dcm2nii. Thermal noise was removed from the images with the MP-PCA method, and skull-stripping was performed using FSL BET [220]. Drift correction was applied [228], as

well as subsequent susceptibility-induced distortions, eddy currents and head motion corrections using FSL eddy and "topup" functions (FSL 6.0.1). Following gradient nonlinearity, [71] and Gibbs ringing artefacts [120] corrections, a mean non-diffusion image (b = 0 s/mm^2) was generated.

3.2.2.6 Tractography

Following the computation of input response functions for the cerebrospinal fluid (CSF), grey matter (GM) and single fibre white matter (WM) from multishell DWI data [47], the distributions of fibre orientation were estimated using multi-tissue constrained spherical deconvolution [111] for each tissue compartment in all voxels. The "mtnormalise" tool from MRTrix was used to normalise multi-tissue fibre orientation functions for multisubject comparison.

Next, seed masks for tractography were generated with the following steps: creating multi-tissue fibre orientation functions by resampling each subject's left and right hemisphere surfaces into volume and registering to native DWI space with the transformation matrix from the previous step; creating a labelled WM/GM boundary image by multiplying binarised WM/GM boundary image with subject-based dilated HCP-MMP atlas image; adding the sub-cortical segmentation created in the recon-all procedure to the labelled image. A total of 375 seed masks (358 cortical, 17 sub-cortical) were generated.

Region-to-region tractography was computed using a probabilistic algorithm based on the second-order integration over fibre orientation [219]. To avoid undesired partial volume effects, seed masks were restricted to the HCP-MMP atlas parcellations on the extended GM/WM boundary. Tractography was only performed on white matter. Two hundred streamlines were generated for each seed-voxel [196]. The minimum streamline length was set to 10mm. For tractography computational power optimisation, larger seed masks were subdivided into smaller masks while preserving the target masks [196], and the tractography was executed in parallel.

Streamlines ending outside the target mask or transverse WM were discarded. An outlier removal based on clustering was also used [34]. To establish connections between various brain regions, a sampling procedure was employed along the length of the streamlines. These streamlines were then grouped into clusters based on their proximity, as determined by the Euclidean distance between them. Whenever the distance between two streamlines exceeded a predefined threshold, a new cluster was created. Through iterative testing, the threshold was ultimately set to a value of 5 after evaluating multiple alternatives. It was found that smaller clusters could

have provided more substantial insights and were therefore disregarded in the subsequent analysis. This methodology allowed for identifying and examining meaningful connections between different brain regions while ensuring a consistent and reliable clustering approach.

3.2.2.7 Structural Connectome

To derive the structural connectome, a total of 358 cortical ROIs from the HCP-MMP atlas were utilised. The extraction process involved quantifying the number of valid streamlines connecting each pair of regions. This approach enabled the establishment of connectivity patterns between different brain areas. To determine the density of connections, the number of streamlines was divided by the sum of those pairs of regions that were linked. In order to focus on more robust connections, any connections with fewer than 50 streamlines were excluded from further analysis. ROIs were grouped into 22 clusters according to their spatial location for a more comprehensive understanding of the interrelationships and functional subdivisions within the brain [70].

3.2.2.8 fMRI preprocessing

The preprocessing of resting state fMRI (3D rsfMRI) images was performed using the Connectome Computation System (CCS) [244], which includes Freesurfer [57], FSL [110] and AFNI [35]. The aim was to project 3D rsfMRI images onto 2D cortical surfaces to increase the test-retest reliability of rsfMRI analyses [253]. Images underwent the following preprocessing steps with CSS: (1) removal of the first 10 seconds of each scan for signal equilibration, (2) slice timing correction, (3) 3D motion correction, (4) 4D global mean-based intensity normalization, (5) regressing out the white matter, cerebrospinal fluid and motion parameters, (6) band-pass temporal filtering (0.01–0.1 Hz), (7) linear and quadratic trends removal, (8) Application of a rigid boundary-based transformation (BBR) algorithm for co-registrating between individual functional and anatomical images [76], and (10) projection of functional images onto the fsaverage5 cortical surfaces in the standard MNI space (10,242 vertices per hemisphere and 4 mm inter-vertex gap on average) [245].

3.2.2.9 fMRI derived functional connectome

From the subparcellated HCP-MMP atlas, a signal-average time series was calculated for each of the 358 ROIs, and the connectivity was determined as the Pearson

correlation coefficient between each pair of ROI time series. Connections with a non-significant correlation were discarded.

3.2.2.10 MEG preprocessing

MEG data preprocessing was based on previous pipelines [214]. Continuous raw MEG data underwent the following preprocessing steps: (1) import to Fieldtrip [165], (2) 256 Hz downsampling, (3) application of 1-100 Hz 4th order two-pass Butterworth bandpass filter, (4) line noise removal with 50 and 100 Hz notch filters, (5) Removal of visual and cardiac artefacts using ICA decomposition with the fastica algorithm [105] with the help of simultaneous EOG recordings. Removed components were in the range of two to five for each subject.

3.2.2.11 MEG source reconstruction

Brainstorm software [213] was used to import Freesurfer and MRI scan-generated surfaces of the inner skull, scalp, pial, and grey matter/white matter boundary. These data were then aligned to the MNI coordinate system with an automated procedure. The iso2mesh software Brainstorm implementation [55] was used to generate a cortical surface of dipole locations by identifying the midpoint between the pial surface and grey matter/white matter boundary and downsampling it to 10,000 homogeneously spaced vertices. Analogously, the inner skull surface was downsampled to 500 vertices. Matlab was subsequently used to align these structural surfaces with the MEG digitizers. A realistic, subject-specific, single shell forward model was then created from the aligned MEG gradiometers, inner skull surface, and cortical surface [161].

The source activity was reconstructed using exact low-resolution electromagnetic tomography (eLORETA) [170, 171], a linear, regularized, weighted minimum-norm inverse solution with exact, zero error localization [170, 214]. The eLoreta method has been shown to have high performances on (parcellated) resting-state data [136, 214] and whole brain synchronization [56, 172] compared to other methods. In addition, eLORETA caused low source leakage, a situation where the presence of spurious correlation between voxels' activity causes a non-precise estimation of functional connectivity, resulting in lower-quality functional networks [32]. Other possible choices for source reconstruction algorithms are the linearly constrained minimum variance (LCMV) beamformer [225], weighted LCMV which is a LCMV with depth normalisation, [93] and other methods that use least-squares minimum norms under different prior assumptions of source covariance like eLORETA. Specifically, these algorithms are the minimum norm estimate (MNE) [82], weighted MNE [65], variance-normalized

and sLORETA [173]. The orientations of the dipole were set normal to the cortical surface [39, 92] under the hypothesis that M/EEG signals are mainly generated by the postsynaptic currents in the dendrites of large vertically oriented pyramidal neurons in layers III, V, and VI of the cortex [163].

The cortical surface was registered to the MEG-optimized reduction [214] of the HCP-MMP atlas (115 cortical ROIs per hemisphere) in Freesurfer. Each ROI time series was computed as the time course of the first principal component of all voxels inside the ROI.

3.2.2.12 MEG derived functional connectome

Functional connectome was generated using phase and amplitude measures within four frequency bands (theta 4-8Hz; alpha 8-13 Hz; beta 13-30 Hz; gamma 30-100Hz). Amplitude envelope correlation (AEC) was calculated after multivariate orthogonalization [32] between ROI pairs. This particular metric was chosen following previous proof of its high reliability [33].

3.2.2.13 Network binarization

The connectomes derived from DWI, fMRI and MEG data are represented as unweighted adjacency matrices. In order to apply the graphlet approach, a standard 10% relative threshold has been used to binarise the matrices, meaning that only 10% of the most robust connections (matrix elements) were kept and transformed to 1 while all other elements were set to 0. The Python library bctpy, a Python implementation of the MATLAB Brain Connectivity Toolbox (BCT), has been used for this procedure. This threshold value was set after testing for different threshold levels and observing that 10% was an optimal level to preserve a power law degree distribution, a fundamental property of real-world biological network as discussed in Chapter 2. Setting higher threshold levels resulted in the disappearance of the power-law degree distribution in favour of Gaussian degree distribution, typical of random networks, and this could be a sign that the network was too dense. On the other hand, lower threshold levels led to irregular degree distributions and multiple isolated nodes.

3.2.3 Graphlets and Orbits

3.2.4 Brain plots on central orbits proportions

The starting point of the graphlet-based analysis of human SC and FC is the computation of the Graphlet Correlation Matrix (GCM), presented in detail in Chapter 2.2.3.2. GCMs help identify groups of correlated orbits. For example, in Barabasi-Albert scale-free networks, orbits can be grouped into 3 categories based on the topology of the nodes belonging to them: central orbits (0,2,5,7), near-central clustering orbits (10,11) and peripheral orbits (1,4,6,9) [254]. However, in most real networks, the topological role of orbits 10 and 11 does not differ considerably from the central orbits [254]. Hence, we can consider 2 categories only: central orbits (0,2,5,7,10,11), associated mainly with nodes connected to many other nodes (hubs) and peripheral orbits (1,4,6,9), primarily associated with lower-degree peripheral nodes [254]. This binary separation is helpful because we can now characterise a brain region using the percentage of its central orbits over the total. We introduce a new orbit-based centrality measure called Orbit Centrality Percentage (OCP), defined as the z-transformed values of the percentages of the central orbits over the total.

3.2.5 Associating between modality-specific and theoretical GCMs

To infer which theoretical model is closer to brain orbit topology, we construct two theoretical GCMs. The Barabasi-Albert network is a useful model for generating networks through preferential attachment [9]. As such, it is a model widely used to describe some properties of real-world networks, like the emergence of the power-law in the degree distribution (scale-free property). However, it remains a theoretical model built on simplistic assumptions, and as such it is unable to reproduce specific properties. For example, real-world networks possess a more clustered structure, resulting in a higher clustering coefficient. It can be shown that in Barabasi-Albert scale-free GCM, 3 groups of strongly correlated orbits (0,2,5,7), (10,11) and (6,9,4,1) can be observed [254]. Following this observation, the first theoretical GCMs have been constructed as 3 groups of perfectly correlated orbits (0,2,5,7), (10,11) and (6,9,4,1) while the correlation across groups was set to 0. Although this GCM does not precisely represent a Barabasi-Albert scale-free network, it is attractive as a model due to the lack of correlation between central orbits (0,2,5,7) and near-central high-clustered orbits (10,11). The ideal 2 orbit groups GCM will have maximum correlation across central orbits (0,2,5,7,10,11) or peripheral orbits (6,9,4,1) while 0 correlation

between orbits of different groups. This resembles more a situation where networks possess the scale-free like Barabasi-Albert ones, with a stronger clustering in addition. This is more often observed in real networks. The next step is computing the GCD between the lower diagonal of the brain GCMs and the two model GCMs. We repeat this analysis at a whole-brain level and on another case for each of the 21 clusters. The GCM of a cluster is constructed considering the full GDVs of the nodes within that cluster, including orbits that involve links outside the cluster. Considering, for example, the node in Figure 3.1B, the count of orbits 5 will still be 2 even if the 5 nodes of the network belong to 2 or more different clusters. Clusters 1 and 2 covering the occipital lobe were combined due to the small number of regions in the two clusters.

3.2.6 Associating between Graphlet and classical graph-based metrics

Central and peripheral orbits were associated with hubs and degree-1 nodes, respectively [254]. However, this definition needs to give a more unambiguous interpretation of nodes in a network with a high correlation between central and peripheral orbits (for example, biological networks [254]). To better investigate the independent topological properties of orbits and apply the results to brain networks, we calculate the distribution of partial correlation coefficients of central orbits count π with four classical graph measures defined for each node i, partialling out peripheral orbits count χ and vice versa the distribution of partial correlation coefficient of peripheral orbits count partialling out central orbits count. This method is also helpful in investigating to which degree graphlets can provide additional new information about brain network topology compared to analysis based on 4 classical graph measures: local clustering coefficient [159], average neighbour degree [159], within-module degree z-score [79] and participation coefficient [79].

3.3 Results

3.3.1 Graphlet correlation matrices from multi-modal whole-brain connectome

For each modality, we computed the GDV of each node (ROI) and then estimated the whole-brain GCM as Spearman's correlation over the number of 11 possible orbits. Figure 3.1 shows the average GCMs for SC, fMRI FC and the 4 frequency bands of MEG FC. The mean CGD across each pair of modalities is shown in Figure 3.1G.

The mean GCMs show 2 separate groups of central and peripheral orbits for each modality. However, the correlation between the two groups is high (higher than 0.5). The mean GCDs between modalities showed that the mean SC is more similar to fMRI FC than MEG FC ($F(4, 26) = 29.384, p < 0.001, \eta_p^2 = 0.531, BF_{inc} = 5.629e^{13}$). For MEG, α and β FC are less distant to pairwise comparisons of other frequency bands ($F(3, 26) = 12.438, p < 0.001, \eta_p^2 = 0.27, BF_{inc} = 1.301e^{15}$).

3.3.2 Spatial variance in graphlet orbit distribution

From the stacked GDVs of all ROI within a network, we calculated the percentage of central orbits over the total for each ROI. We repeated this for each network and z-transformed the results. We named this new measure Orbit Centrality Percentage (OCP). The percentages are averaged first across subjects and then over ROIs within a cluster. Figure 3.2 shows the mean cortical surface plots of the z-transformed percentages of central orbits per cluster. Figure 3.2G shows the similarity of the plots calculated as the Spearman's correlation coefficient between plots.

3.3.3 Comparisons with theoretical-derived GCMs

We calculated the distance, estimated using the GCD, between all the networks across modalities and the two theoretical GCMs models 3 orbit groups and 2 orbit groups. Such theoretical GCMs represent an ideal Barabási-Albert network, and an ideal clustered network, respectively. In Figure 3.3 the distributions of the distances with theoretical models are shown. Figure 3.3A represents the theoretical GCM of an ideal 3-orbit groups network, while Figure 3.3B is the GCM of an ideal network where central and peripheral orbits are perfectly correlated within them and not correlated to each other. Overall, all the modalities have a lower distance from the two groups model indicating a similarity with the clustered ideal network which is closer to real networks.

In Figure 3.4, the same similarity with theoretical models is applied to each of the 21 clusters of the brain. MEG FC is not shown because most clusters were disconnected due to the small number of nodes and this led to several NaN values in the GCMs. The results indicate that both structural and functional brain clusters are better characterised by the two orbits groups model, indicating a high correlation between near-central clustered orbits 10 and 11 and other central orbits, similarly to the case of the whole brain and other real-world networks [254].

3.3.4 Correlation between graphlet and classical graph-based metrics

We calculated the partial correlation coefficient of each node's central and peripheral orbits count across networks with six node-wise classical graph measures: Clustering coefficient C, average neighbour degree δ, within-module degree z-score z, participation coefficient P, beetweenness centrality β and closeness centrality χ. When the correlation with central orbits is calculated, peripheral orbits count is partialled out and vice versa. The bar plots with a 95% confidence interval of the distributions of partial correlation coefficients are shown in Figure 3.5. The results generally show a clear difference in the role of central and peripheral orbits for structural connectomes, while the difference is of a lower degree for functional connectivity, although some notable exceptions are present like the correlation with the within-module z-score for MEG γ.

3.4 Discussion

In this study, we performed graphlet-based analyses on whole-brain SC (DWI) and FC (MEG and fMRI). We introduced a new metric based on the proportion of central and peripheral orbit counts that shows spatial heterogeneity. We compared CGMs of whole-brain SC and FC with two theoretical GCM models, one inspired by the Barabási-Albert scale-free network model and one derived from empirical observations of real networks. Furthermore, we investigated the link between graphlet-based measures and conventional graph metrics.

The GCMs across modalities show 2 agglomerates of strongly correlated orbits and a lower correlation between the orbits of different types. Similarly, networks of the whole brain and brain clusters appear to be more similar to the 2 orbit groups theoretical model while preserving a certain degree of similarity with the 3 orbit groups model inspired by the Barabási-Albert scale-free model. The partial correlation coefficients show a different behaviour, indicating that nodes with similar graphlet degrees may have different topological roles across modalities.

To start with, connectivity matrices show two coarse modules in DWI (corresponding to the two hemispheres) and four in fMRI, with structured functional connections between the two hemispheres. Conversely, MEG presents a higher number of smaller modules, indicating another kind of topological structure which is similar across frequency bands. GCD takes into account such topological difference in a very efficient way, while traditional methods (like degree distribution distance) only look at coarser

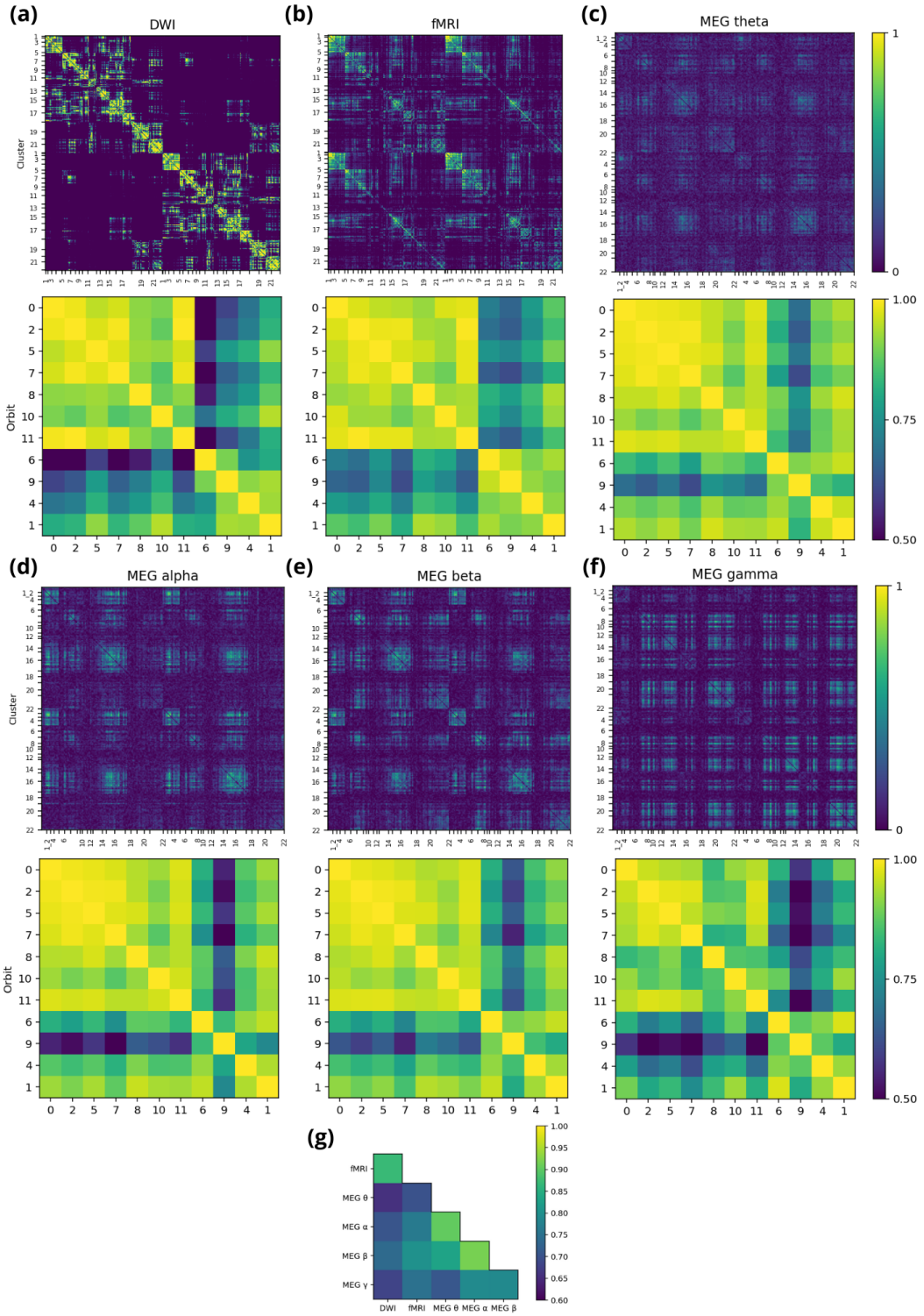

Figure 3.1: Mean unweighted connectivity matrix at 10% relative threshold and correspondent mean Graphlet Correlation Matrix across 28 subjects. (a) DWI; (b) fMRI; (c) MEG theta band; (d) MEG alpha band; (e) MEG beta band; (f) MEG gamma band; and (g) Mean GCD across modalities.

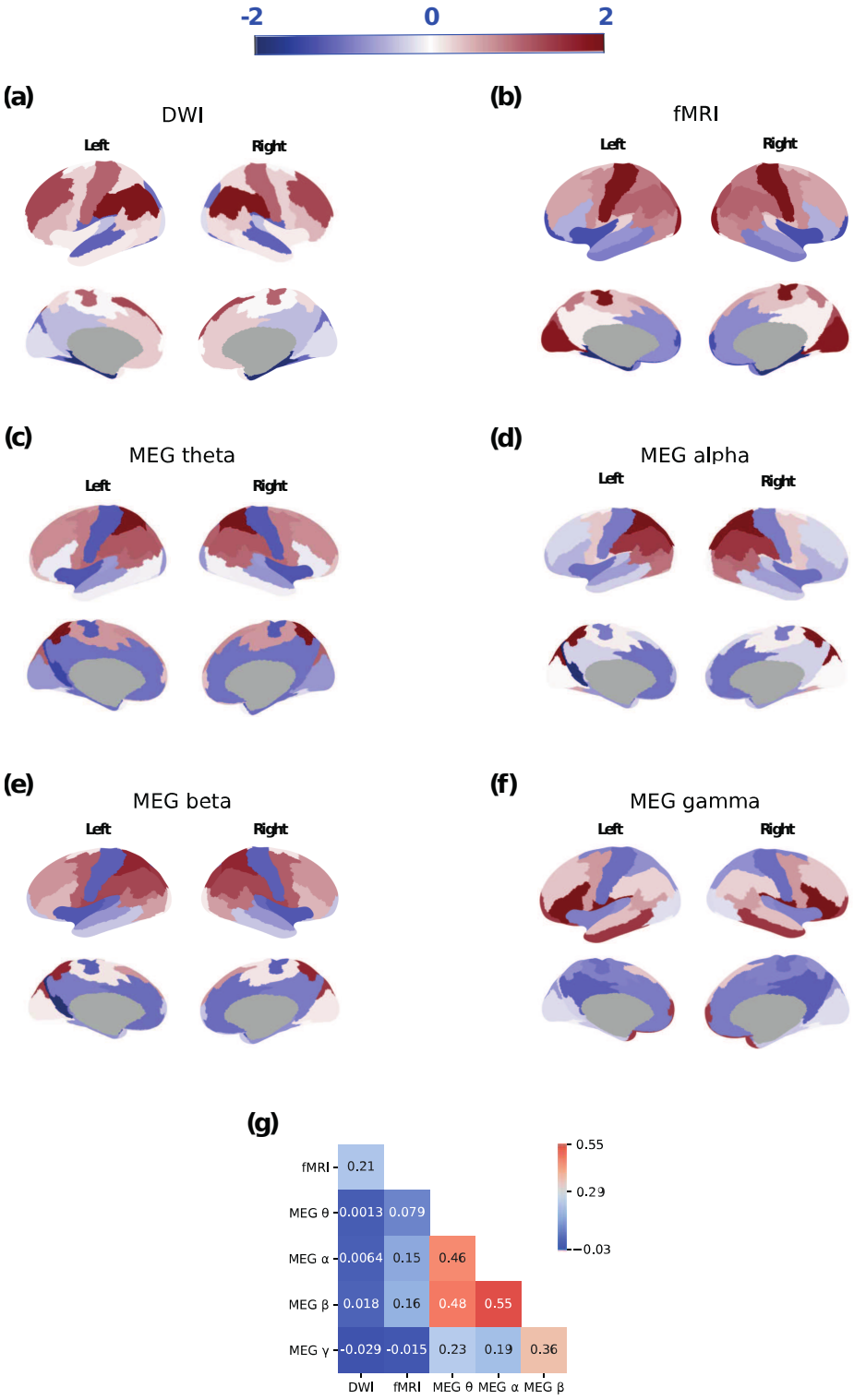

Figure 3.2: Cortical plots of the Orbit Centrality Percentage (OCP) *i.e.*, the z-transformed percentage of central orbits over total orbits (central + peripheral) (a) DWI; (b) fMRI; (c) MEG theta band; (d) MEG alpha band; (e) MEG beta band; (f) MEG gamma band: and (g) similarity across plots calculated as the mean Spearman's correlation coefficient.

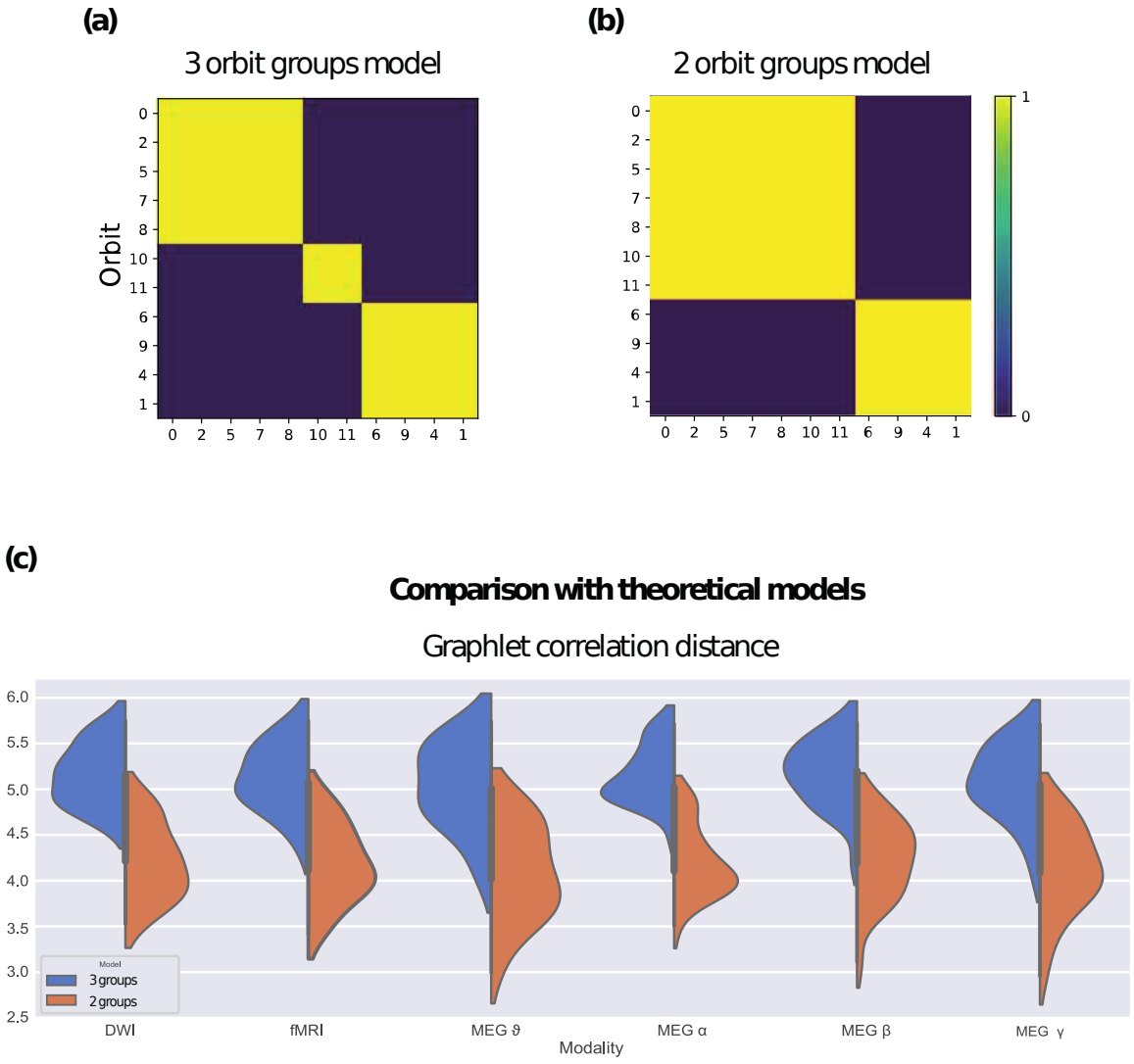

Figure 3.3: GCM model-based analysis of the whole network (a) GCM of an ideal 3 orbit groups network; (b) GCM of an ideal 2 orbit groups network; (c) Distribution of the GCD between the network and 2 models.

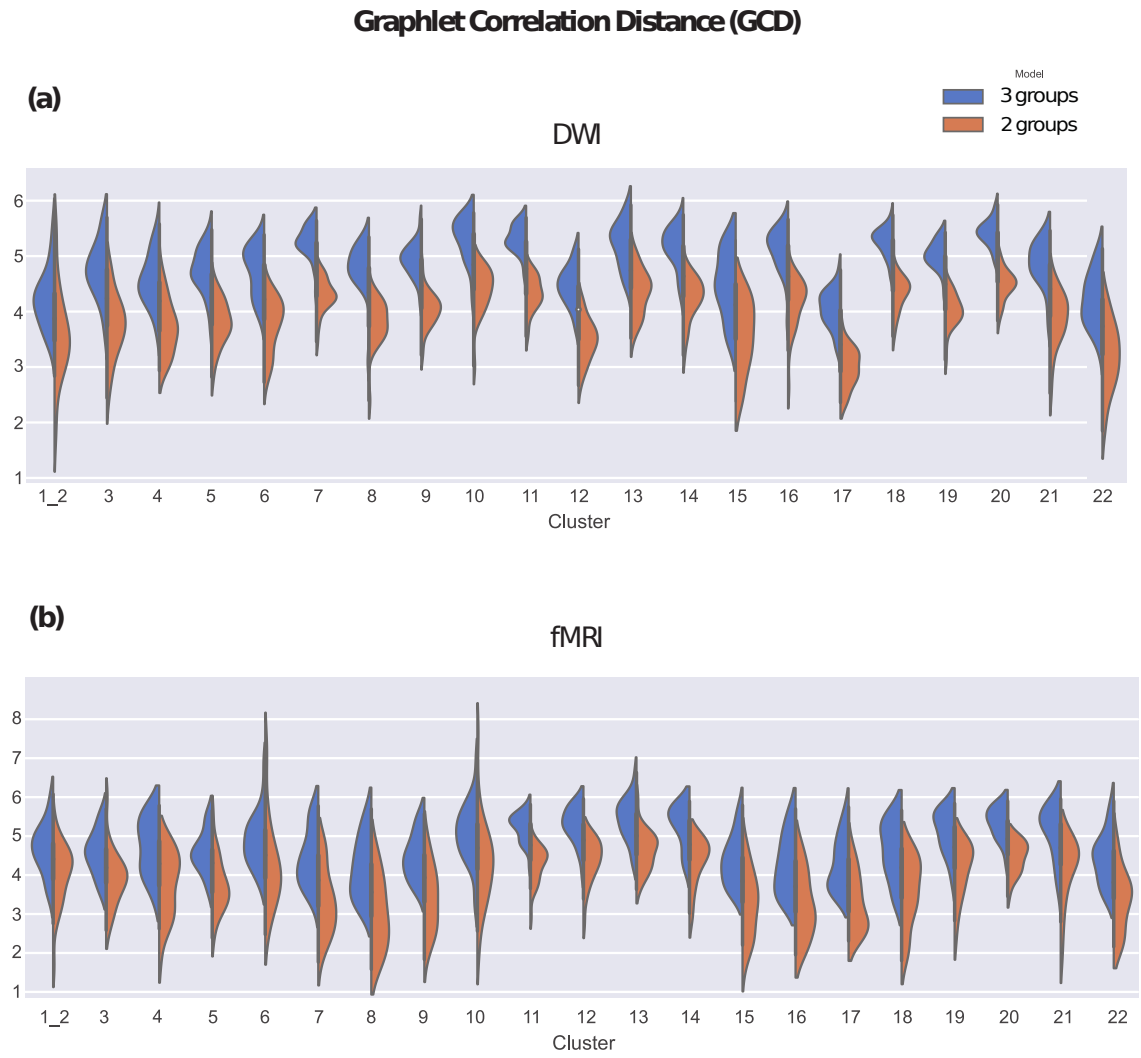

Figure 3.4: Distribution of the GCD between the 21 clusters and 2 models. (a) DWI; (b) fMRI.

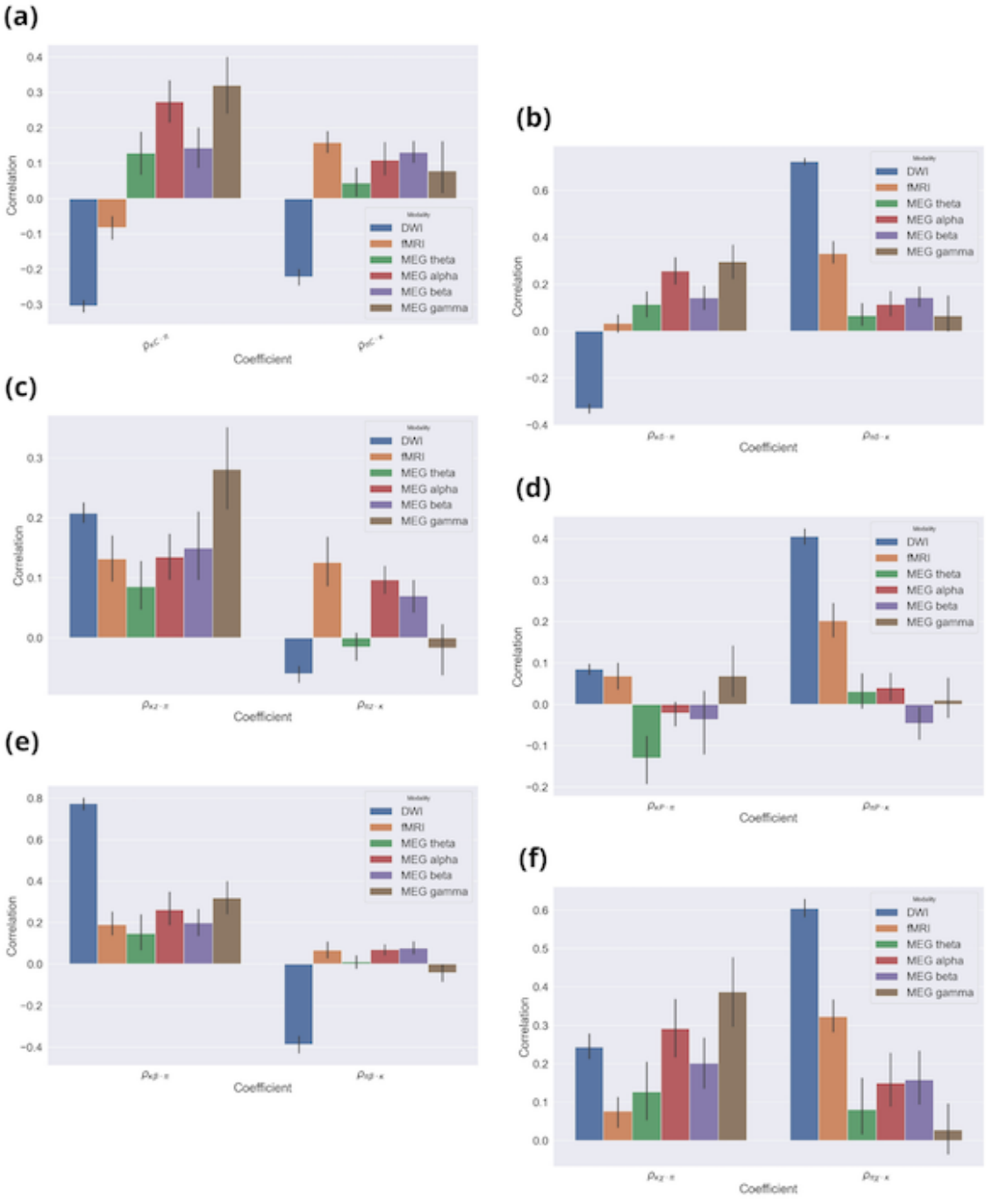

Figure 3.5: Partial correlation coefficient of central κ and peripheral π orbits count with 6 "classical" graph measures. When the central orbits count correlation is calculated, peripheral orbits count is partialled out and vice versa. (a) Clustering coefficient C; (b) average neighbour degree δ; (c) within module z-score z; (d) participation coefficient P; (e) Betweenness Centrality β; (f) Closeness Centrality χ.

details and may not be suitable. Conversely, other methods like the Laplacian distance [254] are too sensitive to finest differences and may be more suitable to study finer problems like inter-subject variability. This result is even more clear when looking at the partial correlation coefficients, showing that GCD can effectively discriminate between certain categories of network without the need for complex analysis based on multiple traditional network measures. A general low GCD is a sign of similarity in term of topology and this translates into similar ways to propagate information in the network, as well as similarity in underlying properties which are universal to resting-state brain networks. As such, given the qualitative similarity between DWI and fMRI or between the various frequency bands of MEG, it is not surprising to have corresponding low GCD for these pairs of networks.

Unlike the ideal Barabasi-Albert scale-free network, GCMs of brain networks show a robust correlation among orbits (0,2,5,7) and (10,11). This result suggests that central nodes are brokers, *i.e.*, they act as connectors between different clusters/communities. In addition, they simultaneously posses a clustered neighbourhood, meaning that such broker nodes also have an essential role in keeping together the nodes within a community. Moreover, the correlation between central and peripheral orbits is lower than between orbits of the same type, but it is still higher than 0.5, indicating that nodes are likely to be clustered/broker and peripheral nodes simultaneously. This suggests that the broker nodes are likely to participate in long-range connections or, in other words, links between broker nodes put topologically distant communities together, a result compatible with the description of the rich-clubs organisation of the brain [75, 122]. This is not surprising, as the human brain needs to be organised in several specialised modules, topologically identify by communities, to accomplish complex goals. Rich-clubs topology describe this situation where highly connected nodes tends to connect to other highly connected nodes, and since such nodes also act as hubs for their communities, these connections are indeed connections between communities. Central nodes not correlated to peripheral nodes can be interpreted as within community or provincial hubs, while peripheral nodes not correlated to central nodes can be interpreted as low-connected peripheral brain regions [179]. These two types of nodes are less common, indicating that only few areas of the brain are low-connected to other distant areas. These results support the evidence of the human brain organisation in communities while providing a powerful yet simple tool to outline the role of the nodes in the brain organisation. It has been observed that the organisation of the human brain in communities allows for complex cognitive tasks and mechanisms [36].

The spatial distribution of orbits' centrality percentage shows a high central region in the SC fronto-parietal area and occipital lobe. Cortical plots for fMRI FC show a slightly different pattern, with higher central orbits in the motor cortex and lower in the other regions. This result is in accordance with previous studies investigating the role of hubs in the human brain resting state fMRI FC [134]; short-range hubs were identified in the primary cortex, which in our case corresponds to a high OCP, while long-range hubs were found in the multimodal association cortex, which in our results correspond to slightly lower OCPs, indicating regions with a slightly higher presence of peripheral orbits. Primary cortex is involved in motor activity. High OCP in such area indicates that that connections are mainly formed with proximal areas, like the premotor cortex or sensory areas. On the other hand, multimodal association cortex is involved in higher functions which requires connections with several different areas in the brain, even topologically distant. Is is important to note that short-range and long-range hubs refer to the concept of topological distance, which in FC-derived connectomes might not indicate proximity in anatomical terms. Short-range hubs can be associated with ROIs with a high OCP, indicating a within-community or provincial hub keeping together many lower-degree nodes inside a community. In contrast, intermediate OCP values indicate an ROI which is both central and peripheral and can be identified as a connector hub [79].

The motor cortex is dominated by peripheral orbits in all the frequency bands of MEG FC. Moreover, OCP of theta, alpha and beta FC show high similarity, while OCP of gamma FC appears different from the other bands apart from the beta band. As central and peripheral orbits are positively correlated, we can associate high OCP values to provincial hubs while low OCP derives from low-degree peripheral nodes. Intermediate values of OCP refer to nodes with similar numbers of central and peripheral orbits, and they can be identified as connector hubs. SC shows a coarse OCP plot. Extreme values are only possessed by the inferior parietal region (high OCP) and medial temporal (low OCP). Other regions show intermediate OCP values, indicating a balanced percentage of central and peripheral orbits.

The OCP plot of fMRI FC follows a gradient similar to the gradient of myelin content, meaning that regions with higher OCP are also regions with high myelin content, which are correlated to uni-modal regions [104]. Moreover, it can be seen from MEG FC that the OCP plot is similar to temporal activation of sensory areas, where higher OCP areas correspond to activation of more extended events in time [104].

The partial correlation coefficients show interesting features at two different levels of observation. First, central and peripheral orbit counts relate to graph metrics differently, indicating that central and peripheral orbits have a distinct topological role in the brain. In particular, participation coefficient and within-module degree z-score show substantial differences across types of orbits. Second, different types of orbits correlate differently across modalities, especially between DWI-derived SC and FC. In particular, the results suggest that SC peripheral nodes connect different communities with topologically long-range links. Indeed, the correlation between peripheral orbits counts and participation coefficient is positive and significant (around 0.4), while the correlation with the within-module degree z-score is closer to 0. On the other hand, central nodes are negatively correlated with average neighbour degree and positively correlated with the within-module degree z-score, indicating that they tend to connect to lower degree nodes within a community, creating the core-periphery structure already identified in the human brain, *i.e.* they can be identified as provincial hubs. The role of central nodes as provincial hubs is preserved across modalities, while other results are less significant, especially for MEG FC. In general, FC shows more clustered modules than SC (this can also be observed qualitatively from the adjacency matrices) and the higher inter-subjects variability makes it difficult to draw a clear conclusion from some partial correlations. Overall, FC seems to be organised in several different modules on both hemispheres highly connected, but the role of peripheral hubs could be more evident in SC.

In summary, graphlet-based measures are a simple yet effective tool for topological analysis of the human brain connectome, although not especially suited for research of novel properties. Most of the results about topological properties of SC and FC of the brain obtained with graphlet-based analysis were already present in the literature using conventional graph-theory approaches. GCM and OCP are tools which combine effectiveness in describing complex topological features of the whole brain with a surprising simplicity of visualisation of the results. Although at this level, graphlet-based analysis is not well suited for uncovering new properties of the human brain topology, it has promising potential developments towards the applications to directed networks [195], weighted networks [204], dynamical networks [102], or the combination with spectral theory [239].

Chapter 4

Deep Learning for parameter recovery of a perceptual decision making model

4.1 Introduction

4.1.1 Decision making and sequential sampling models

Thanks to the advancement in psychology and neuroscience and the improvements in neuroimaging techniques, detecting and studying neural activity within the brain has become more accessible. For instance, it has been observed that the lateral intraparietal (LIP) region of the brains of non-human primates shows a strong correlation between neural activity and the time it takes for the individual to reach a decision during decision-making tasks such as random dot motion discrimination [187, 101]. Consequently, a growing interest in developing biologically more realistic models is being developed.

When studying decision-making, the so-called Sequential Sampling Models (SSM) represent a starting point in terms of simplicity. These decision-making models are based on the hypothesis that individuals accumulate noisy evidence towards a specific decision until a threshold is reached. When the accumulated evidence reaches the threshold, the decision is made. One notable example is the Drift Diffusion Model (DDM), which represents a decision maker that accumulates competing evidence towards an upper or lower stopping boundary. DDM is widely used, and it helped to advance our understanding of cognitive processes [112] as well as characterising how ageing impacts cognitive performances [209] or how conditions such as depression can alter decision-making processes [130]. However, DDMs are not biologically realistic,

and strong efforts have been put into the improvement of this kind of model by the inclusion of more biological details.

Developing more biologically detailed models of cognition usually means adding more parameters to describe specific processes at neuronal level. In addition, differential equations that describe the evolution of such systems are usually more complex. This allows researchers to fit behavioural data better and reproduce essential aspects of neural dynamics that are difficult to study with simpler models. On the other hand, essential procedures such as parameter recovery are usually more challenging to perform on more complex models. In this chapter, I will tackle the problem of parameter recovery of a biologically plausible perceptual decision-making model [242] described in detail in section 2.1.3.1, which is more complex than a typical DDM.

4.1.2 Parameter recovery

Parameter recovery refers to the task of identifying the optimal parameters to provide the best fit possible between the model's predictions and empirical data. In psychology and neuroscience, it is vital to analyse and create models that reproduce the results of behavioural experiments or represent patterns of neural activity. For complicated models with no analytical solutions, numerical optimisation algorithms are often necessary. Such algorithms usually operate on several iterations until convergence, *i.e.*, when the model predictions approximate the real data. Optimisation methods have the issue of being sensitive to local minima, meaning that it is not uncommon for the algorithm to converge without finding the best possible solution. To solve this problem, these methods use global optimisation techniques to find the global minimum, and they may be further improved by introducing an element of randomness into the search process to circumvent local minima. However, even in the case of stochastic global optimisation strategies, the solution associated with the global minimum is not guaranteed to be found. Moreover, they can be computationally intensive algorithms, and concerns have been raised regarding the need for more computational resources for effective parameter recovery [5].

Recently, parameter recovery has been applied in the context of behavioural decision-making models, such as the diffusion model for conflict tasks [236] or the leaky competing accumulator model [150]. These approaches are typically based on a severe restriction of the parameter values to a range that is only able to reproduce real-world behaviour in normal conditions. However, it is essential to be able to recover the parameter of models reproducing more extreme behavioural data associated

with, for example, the presence of neurological conditions or particular experimental settings.

To the best of my knowledge, parameter recovery has not yet been extensively explored in the case of the biologically plausible Wang-Wong decision-making model presented in Section 2.1.3.1. Examining the model's equations (Eq. 2.16 to 2.24), it is clear that it is a complex model, making parameter recovery more difficult than on simpler models like DDM. Consequently, fitting the Wang-Wong model to behavioural data using conventional parameter recovery techniques has been challenging. However, as I will present in more detail in the next section, recent advancements in Deep Learning have allowed us to address the challenge of parameter recovery in the context of more intricate decision-making models.

4.1.2.1 Deep Learning for Parameter Recovery

In the present study, we applied parameter recovery to a neural mass model focusing on a subset of the parameters to investigate the performance as well as the limitations of this approach. We used a deep learning approach as it has many advantages compared to global optimisation procedures, and it is computationally efficient.

The number of publications about deep learning has exploded in the last few years, and it is still constantly increasing. Thus, it is very well known how to make this approach as efficient as possible as well as to tackle the most common issues that can emerge when using a deep learning algorithm. Creating specific libraries for deep learning, data analysis, feature engineering and related problems, and a vast community constantly giving free insights when needed made deep learning an easy-to-deploy solution to many problems. It is often easier to design a deep learning algorithm than applying statistical methods or a traditional, more conceptually simple machine learning approach where thorough feature engineering work may be required in order to get valuable results and satisfactory performances.

One of the main advantages of deep learning is that algorithms are able to find the best representation of the raw data without external supervision, often making it easy to apply the same core method to very different research areas. Of course, this may have some drawbacks. For example, the features identified by the complex algorithm may be too abstract to find a straightforward real-world interpretation [132]. Furthermore, we need to pay attention to avoid over-fitting. In this situation, the algorithm memorises the data we used to train it, losing its ability to generalise to unseen data. Another limitation might be the need for more data. Training supervised deep learning algorithms usually requires a large amount of data, which

is not always easy to acquire. However, this does not represent an issue in our case, as samples of data can be generated at will, making deep learning particularly suitable for parameter recovery. Moreover, training on a large amount of data offers some significant advantages, like the fact that it is easier to find the global minimum without worrying about local minima.

4.1.2.2 Deep Learning as a model of cognition

Parameter recovery is one of the many problems that can benefit from using deep learning algorithms. Originally, artificial neural networks themselves were further considered to be a model of human cognition, although extremely simplified without the presence of biological details [189]. Later developments have focused on improving the capabilities of deep learning to enable it to solve complex tasks with similar or even higher performances than the human brain. Recently, the interest in neuro-inspired AI has risen again to provide deep learning algorithms with human capabilities like intuition, memory or planning [85]. However, although the connection between deep learning research and cognitive science is becoming stronger than ever [176], the use of AI as a model of cognition presents several issues mainly concerning a strong bias towards the desire to publish positive results leading to a credibility crisis [16]. In the second half of this chapter, I will present a proof-of-concept of the use of deep learning as a model of cognition. Specifically, I will use an elementary CNN and RNN ensemble with the only scope being generating a psychometric response similar to primates while performing a random-dot motion discrimination task.

4.2 Methods

4.2.1 Neural mass models

We simulated the data using the two-variable neural mass model introduced in Section 2.1.3.1. The Wang-Wong model reproduces the evidence accumulation involved in decision-making during a random dot motion discrimination task [242]. Each of the variables in the model accumulates evidence in favour of either the leftward or rightward direction of the dots while engaging in competitive interactions through inhibitory connections. Additionally, the model includes a self-excitatory mechanism. The input is represented by two distinct synaptic currents, corresponding to the evidence supporting each of the two alternatives. Alongside these stimulus currents, the model includes a shared non-selective background input. In this context, we have extended the model to include modulation of the background current by a factor β.

Parameter	Minimum value	Original value [242]	Maximum value
$J_{ll} = J_{rr}$	0.25 nA	0.2609 nA	0.30 nA
$J_{lr} = J_{rl}$	0 nA	0.0497 nA	0.25 nA
J_{ext}	-	5.2×10^{-4} nA·Hz^{-1}	-
I_0	-	0.3255 nA	-
μ_0	-	15Hz	-
β	0.95	1	1.05
a	-	270 (VnC)$^{-1}$	-
b	-	108 Hz	-
d	-	0.154 s	-
γ	-	6.41×10^{-4}	-
τ_{nd}	100 ms	-	500 ms
σ_{noise}	-	$I_0/16.275$	-
τ_{noise}	-	2 ms	-

Table 4.1: Original parameters of the Wang-Wong model plus minimum and maximum values of the four parameters used for parameter recovery: J_{ll}, J_{lr}, β, τ_{nd}.

This extension makes the model able to reproduce the observed trade-off between speed and accuracy in certain studies.

We started from the original values used in [242]; The stimulus strength μ_0 has been fixed to 15Hz (for an analysis of the effects of varying μ_0 see [242]), the average synaptic coupling with external synapses $J_{ext} = 5.2 \times 10^{-4}$nA· Hz^{-1}, the self-excitatory coupling strengths within same populations $J_{ll} = J_{rr} = 0.2609$nA, the inhibitory coupling strengths across different populations $J_{lr} = J_{rl} = 0.0497$nA, the background current $I_0 = 0.3255$nA, the input-output function parameters $a = 270$(VnC)$^{-1}$, $b = 108$Hz and $d = 0.154$s [1], the kinetic parameter $\gamma = 6.41 \times 10^{-4}$ [242] and the noisy current $I_{noise,l,r}$ has an amplitude of $\sigma_{noise} = I_0/16.275$ [208] and characteristic time of decay $\tau_{noise} = 2$ms.

4.2.2 Data sampling and processing

We proceeded by focusing our attention on recovering four parameters: the excitatory and inhibitory coupling strengths J_{ll} and J_{lr}, the background current modulation β, and we defined a new parameter called non-decision time τ_{nd} as a simple constant of time added to the reaction time of each decision. The parameter values have been drawn from the following uniform distributions:

- $J_{ll} = J_{rr} \in [0.25, 0.30]$nA

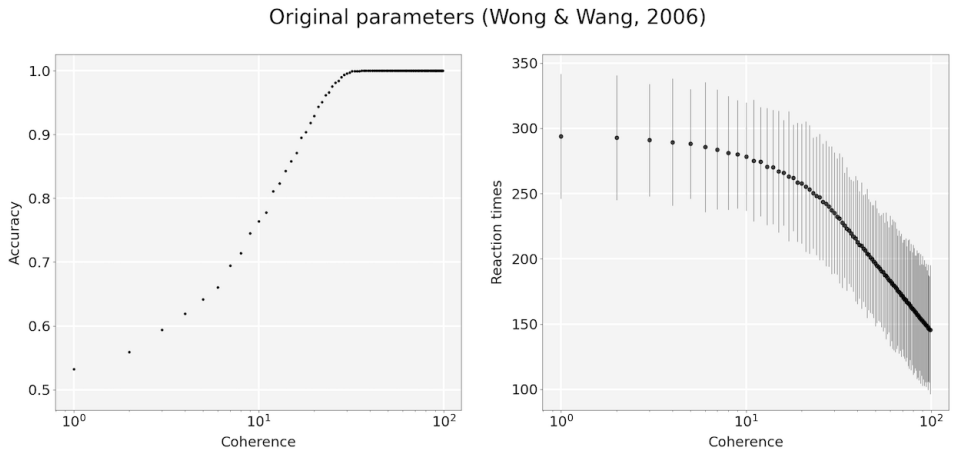

Figure 4.1: Accuracy and reaction time versus coherence plots for the original parameters of the model by Wang and Wong.

- $J_{lr} = J_{rl} \in [0, 0.25]$nA
- $\beta \in [0.95, 1.05]$
- $\tau_{nd} \in [100, 500]$ms

The intervals were determined after a careful plausibility analysis of the results. We noted that parameters outside these intervals may lead to infinitely long reaction times, which is implausible in a 2AFC task setting. We computed the accuracy and reaction time for each parameter at different coherence levels. We were interested in the range of parameter values in which the model's results were reasonably comparable to a real subject. For example, we expected the accuracy to be close to 0.5 at 0 coherence and increasing following a sigmoid-like function to saturate at 1 for high coherence levels (Figure 4.1)[242, 187]. Similarly, reaction times were expected not to be extremely low without a loss in accuracy (*i.e.*, speed-accuracy trade-off). In Figure 3.2, we show some of the accuracy and reaction time versus coherence plots for $J_{ll} = 0.3$, $J_{lr} = 0.01$ and $\beta = 1.05$. In the original paper, $J_{ll} = 0.2609$, $J_{lr} = 0.7049$ and the background current was not modulated (analogous to the case of $\beta = 1$).

We generated 5,000 combinations of the four parameters, each drawn randomly from the abovementioned intervals. For each set of parameters, we computed the equations for a total of 10,000 trials using a coherence of 0.15 (15%). Each trial had a maximum duration of 2.5s for 2,500 iterations of 1 ms each. A decision is made if the activity $S_{l,r}$ of one of the two accumulators reaches a threshold set at 15Hz as in the original paper [242]. For simplicity, we assumed a positive value for the coherence,

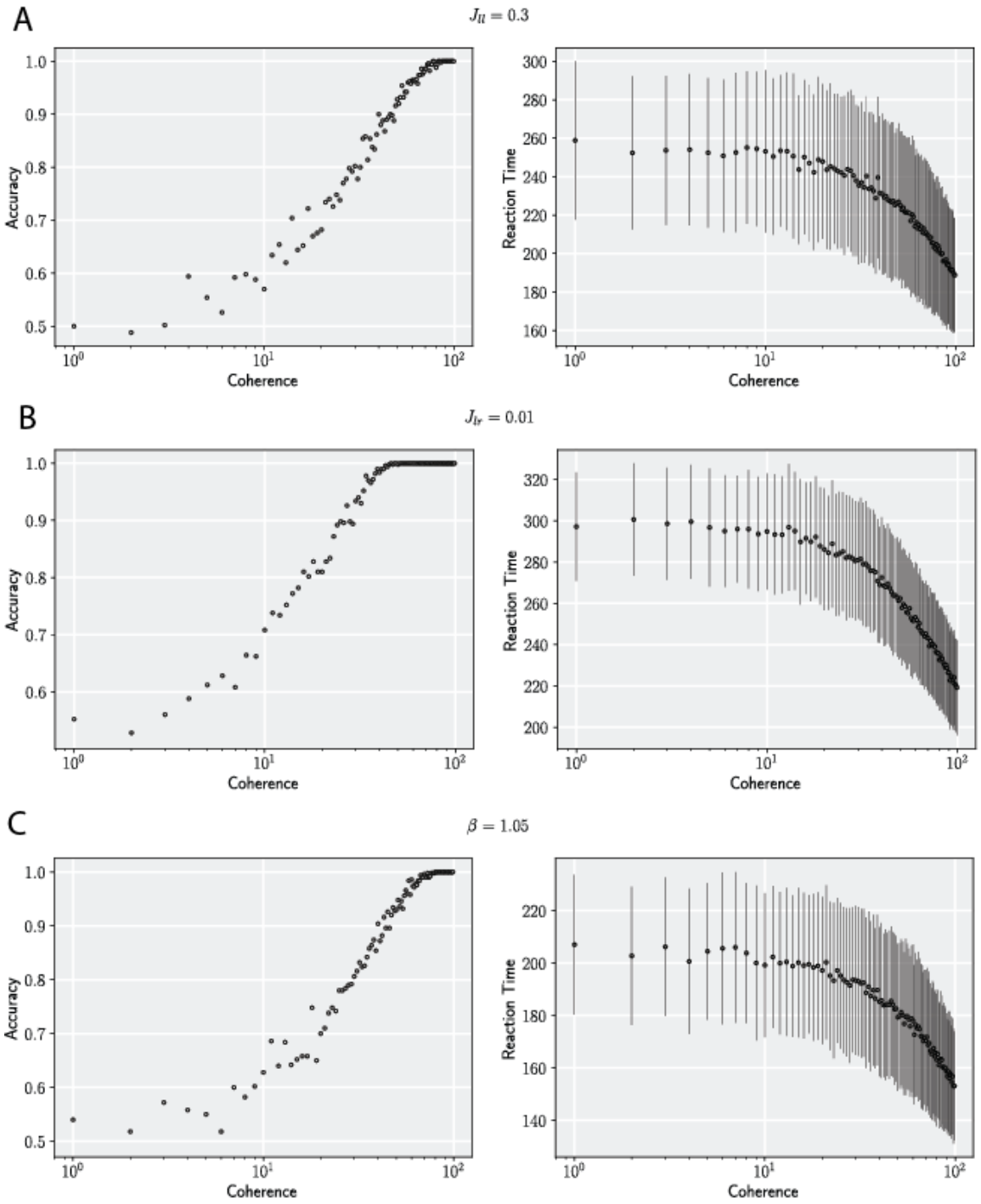

Figure 4.2: Accuracy and reaction time versus coherence plots. **a**, results for $J_{ll} = 0.3$ instead of 0.2609 of the original paper. Faster reaction times and lower accuracy indicate a speed-accuracy trade off when increasing the self-excitatory coupling strength. **b**, results for $J_{lr} = 0.01$ instead of 0.0497 of the original paper. Faster reactions without speed-accuracy trade off is observed. **c**, results for $\beta = 1.05$ instead of no modulation (correspondent to $\beta = 1$) of the original paper. Here we can see similar speed-accuracy trade off as in the case of higher J_{ll}.

meaning that the correct direction of the points was always leftward. Of course, this procedure is legitimate only for artificial model training as in an actual experiment, it may cause a bias in the decision-making process after the subject learns that the dots always move leftward. The output of each trial was 0 for rightward/wrong decisions or 1 for leftward/correct decisions. If the trial ended without either accumulator activity reaching the threshold, an output of -1 indicating a non-decision has been generated. The number of iterations to reach a decision has been stored to identify the reaction time plus the non-decision time τ_{nd}. The results were used to compute a set of features describing the model's performance for each set of parameters. Even if deep learning algorithms can automatically identify useful features, as discussed in 2.3.3, it is more efficient to compute simple statistics describing the data instead of passing the large dimensional raw trial results as input. We initially identified 14 features:

- **Accuracy** The fraction of correct choices over the total choices.
- **Indecision** The fraction of trials ended without a decision over the total trials.
- **Mean Reaction Time** The average reaction times for correct and incorrect choices.
- **Percentiles of Reaction Time Distribution** The 10th, 30th, 50th, 70th and 90th percentiles of the reaction time distributions for wrong and correct choices.

We decided not to use the indecision rate as input for the deep learning model because, in a real experiment, the subject is instructed to make a decision even if they are not entirely confident. Consequently, a framework that uses the non-decision rate cannot be extended to the case of experiments involving real subjects. For the same reason, the accuracy is calculated over the total choices, not considering the iterations ended without a decision. However, the indecision rate was an efficient way to identify certain combinations of parameters that generate extreme behaviours of the model. In fact, in the plausibility analysis of the results, we did not consider the complex problem of how the parameters interact between them. For example, we observed that the model could not reach a single decision in 10,000 trials for some set of parameters. We decided to discard the results, which had an indecision rate lower than 0.2. Similarly, we excluded the results with an accuracy higher than 0.999 because the reaction time distribution for these trials was concentrated on shallow values, indicating extreme behaviour where the model reached the decisions

immediately for all the trials. In both cases, it is reasonable to assume that this kind of behaviour cannot be observed in experiments with real subjects. Consequently, we did not lose any useful information. Even excluding these values, however, some parameters led to decisions with low reaction times compared to subjects in actual experiments, although the accuracy distribution was reasonable. We decided to keep these values to avoid constraining the parameter intervals too much and affecting the generalisability of the algorithm. Moreover, reducing the boundaries of the parameter distributions could prevent the model from predicting more extreme values of the parameters that might be associated with particular conditions like specific pathology [119]. We discarded 1183 of the 5000 initial combinations of parameters, such that a total of 3817 combinations remained as our clean dataset. The final distributions of the parameters are shown in Figure 4.5.

The script for running the models' equations and generating the outputs has been written in Cython [11]. This Python library compiles the code in C, dramatically increasing efficiency and improving the running time. All other data cleaning and feature extraction scripts have been written in Python.

4.2.3 Deep Learning for parameter recovery

The 13 features described previously were used as the input of our deep learning algorithm. We trained the algorithm to predict all four parameters in a multiple-regression problem. We used a simple deep neural network composed of 3 fully connected layers, each followed by a ReLU non-linear activation (see Figure 4.3 for a schematisation of the entire procedure and Figure 4.4 for a simplified flow diagram of the parameter recovery approach). The network has been trained using a Mean Squared Error (MSE) loss and Adam optimiser with momentum. We used hyperparameter tuning to find the best architecture in terms of performance. We used Python's library PyTorch to define the deep learning model as well as the training and hyperparameter optimisation procedures. The models were trained using a single GPU of a 10 GPUs Nvidia DGX-1 module owned by the Cardiff University Brain Research Imaging Centre (CUBRIC). In contrast, for hyper-parameter tuning, we used two GPUs to train two models at a time in parallel.

4.2.4 Neural ensemble

The neural ensemble for reproducing the psychometric curve consists of a simple CNN module followed by an RNN layer (Figure 4.6).

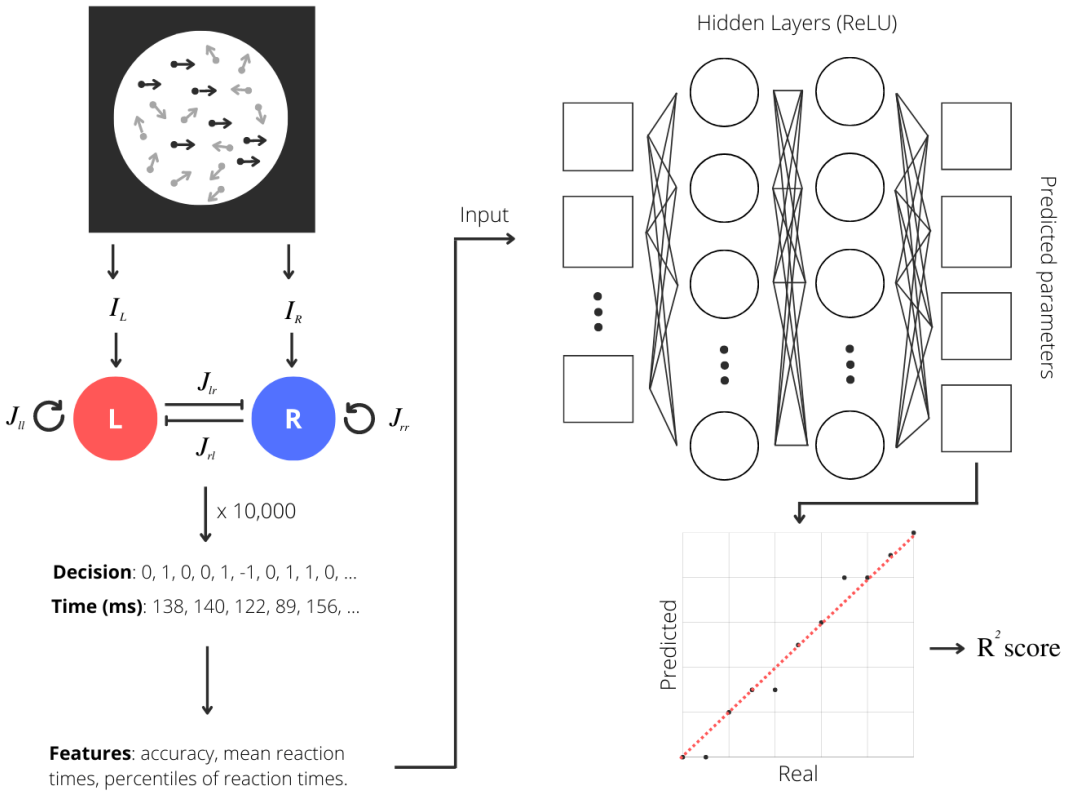

Figure 4.3: Schematisation of the entire procedures of data sampling and parameter recovery. First, a simulated input from a random dot motion discrimination task is passed to the two-variable neural mass model in the form of an input current. The model producing each output is defined by a specific set of parameter. For each set of parameters, the behavioural output of 10,000 single trials are used to compute statistics to be used as input features for the deep learning algorithm. The algorithm is a DNN with fully connected layers and nonlinear ReLU activation. The predicted parameters are compared with the real ones using R^2 scores.

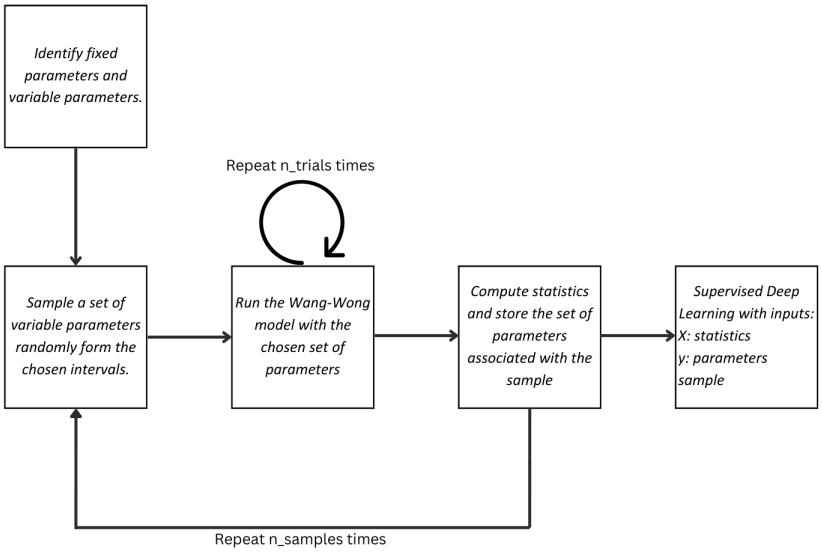

Figure 4.4: Simplified low diagram of the parameter recovery procedure. First, the variable parameters are chosen. Second, a set of variable parameters is drawn randomly from the specified distribution (sample). Third, using the parameters drawn, the model is run several times (trials). Fourth, the statistics are computed and the parameters specific to that sample stored. The sampling procedure is repeated until enough data for the Deep Learning is generated.

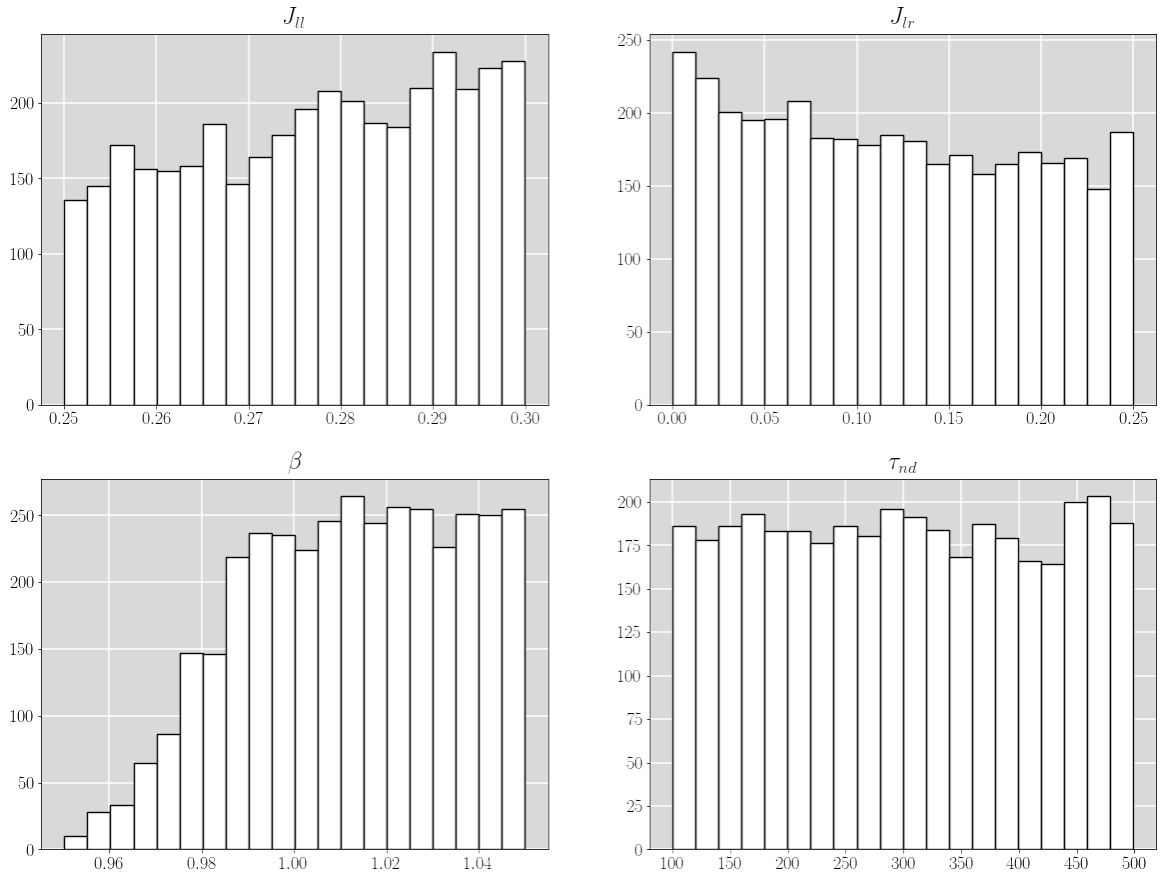

Figure 4.5: Distributions of the four parameters after removing samples generated by an extreme behaviour of the model (for example 100% accuracy and 0ms reaction time). It can be seen that values of β lower than 1 are more likely to generate an extreme behaviour and consequently they have been discarded. The values of J_{ll} and J_{lr} are slightly shifted towards higher and lower values respectively. The non-decision time τ_{nd} doesn't show a significant shift from the starting uniform distribution.

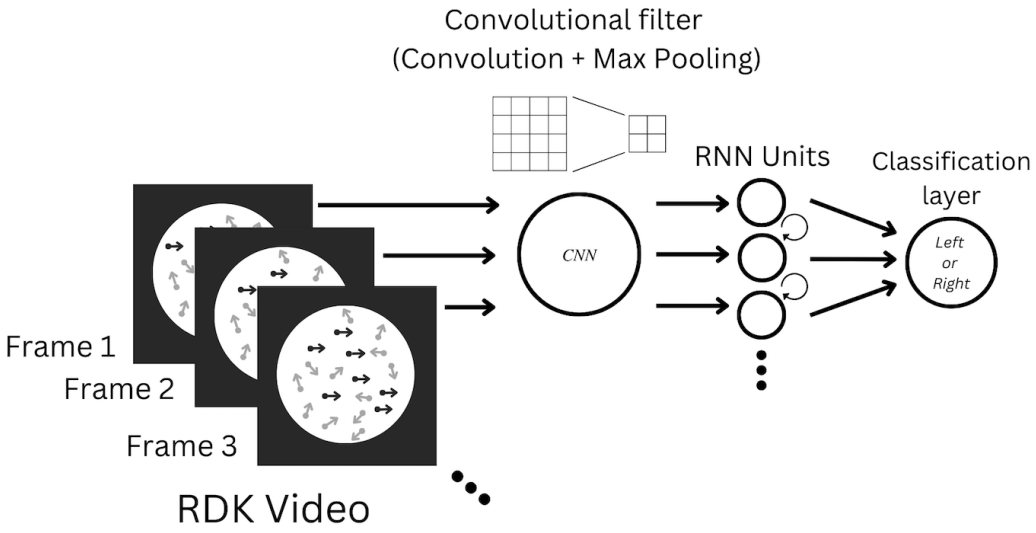

Figure 4.6: Schematisation of the CNN + RNN ensemble for RDK discrimination task. Each frame of the RDK video is passed to a single CNN filter which also applies a pooling. The output is a sequence of the same temporal dimension as the input and it is passed to a single RNN layer. The output is transformed into a single value representing the probability that the model estimates of the dots movement in the RDK being leftward or rigthward.

The CNN applies a single convolutional filter of dimension 5×5 to a sequence of 60 frames of a random dot kinematogram (RDK). Each sequence is generated on a screen at 60 Hz, meaning that each trial lasts for one second. The frames are transformed into single-channel images (greyscale) and resized to 128×128 pixels. The convolution was followed by a two-dimensional max pooling with a stride of 2 and a kernel size of 2. A zero padding of type "same" was used to avoid dropping pixels from the images. The CNN output of each frame was 64×64 matrices, and they were stored into a list forming a sequence of processed frames. The sequence of CNN outputs was passed to a single RNN layer with ten hidden units. The RNN output, a sequence of 60 matrices 64×64, was linearly projected into a single value followed by a sigmoid activation function to obtain a probability.

The neural ensemble has been trained on examples of random dot motion of 100% coherence in both directions, left and right, which were labelled as 0 and 1, respectively. With 50 trials for each direction in the training set, the model reached 100% accuracy in identifying the correct direction of movement in 10 epochs on average. The accuracy has been evaluated on a test set of 10 trials for each direction. The model trained on non-noisy fully coherent RDK has been evaluated on noisy trials of different coherence. In particular, a total of 31 trials at different levels of coherence has been generated: from 1% to 20% with an increasing step of 1%, from 25 to 50 with an increase of 5% and from 60% to 100% with an increase of 10%. The model has been trained on the fully coherence data and evaluated ten times on each coherence level to avoid large fluctuations in the accuracy level. The average accuracy for each coherence level was plotted and compared to the accuracy of the Wang-Wong model evaluated on 10,000 trials for each of the 31 coherence levels.

4.3 Results

The predictions of the model on the four parameters are shown in Figure 4.7. We can see that the model fails to predict J_{ll} values ($R^2 = 0.1097$) and perform slightly better for the other parameters with R^2 scores of 0.86, 0.69 and 0.88 for J_{lr}, β and τ_{nd} respectively. We hypothesised that the difficulty of prediction might be due to the concurrency of J_{ll} and β with a similar influence on the outputs of the model. To test this hypothesis, we generated two new batches of samples; in the first, we kept $J_{ll} = 0.2609$ constant as in the original paper. In the second we set $\beta = 1$ meaning no modulation of the background current I_0. We repeated the hyper-parameter tuning and the training procedures. In Table 4.2 we show the hyper-parameters leading to

Parameters	Batch Size	LR	Mom	l1 size	l2 size	l3 size
4 parameters	50	0.0001	0.52	32	512	32
J_{ll} constant	75	0.0003	0.84	512	512	32
β constant	50	0.0005	0.80	128	128	256

Table 4.2: Hyperparameter values for optimal performances of the algorithms trained for the 3 cases: all four parameters prediction, predictions excluding J_{ll} and prediction excluding β, respectively. The hyperparameters to be optimised via fine-tuning are batch size, learning rate (LR), momentum (Mom), and the number of units (neurons) in each of the 3 layers separately (ln size). The optimal combinations of hyperparameter values have been obtained by looking for the maximum performances of a model after repeating the training process with a random selection of hyperparameter values.

the best performances in the three cases. The predictions with the absence of J_{ll} are shown in Figure 4.8A, while the results with the absence of modulation β are shown in Figure 4.8B. We can see an increasing in the performance of the algorithm especially when β is kept constant. When J_{ll} is constant performances calculated as R^2 scores are 0.94, 0.90 and 0.92 for J_{lr}, β and τ_{nd} respectively. When β is constant performances are 0.95, 0.99 and 0.99 for J_{ll}, J_{lr} and τ_{nd} respectively.

Figure 4.9 shows the parameter recovery performances on J_{lr}, β and τ_{ndt} obtained with input features computed on different number of trials. The performance tendentiously increases up to 10,000 trials (7500 for τ_{ndt}) and then shows an undulatory behaviour that would appear noisy in a finer binning. Figure 4.10 shows the results of the average accuracies of the neural ensemble compared to the Wang-Wong model at different coherence levels. It can be seen that the accuracy of the neural ensemble approximates well the Wang-Wong one, which in turn approximates the performance of real subjects at every accuracy level [187].

4.4 Discussion

For each set of parameters drawn randomly from a uniform distribution with specific boundaries, we generated several samples of 10,000 trials each using a neural mass model simulating a subject performing the random dot motion discrimination task. Each trial generated a decision with an associated reaction time. Trials with high indecision rate or extremely low reaction times had been discarded. For each sample, we computed the accuracy and 12 statistics to describe the reaction time distributions for correct and incorrect decisions of the 10,000 trials. We trained a simple 3-layers

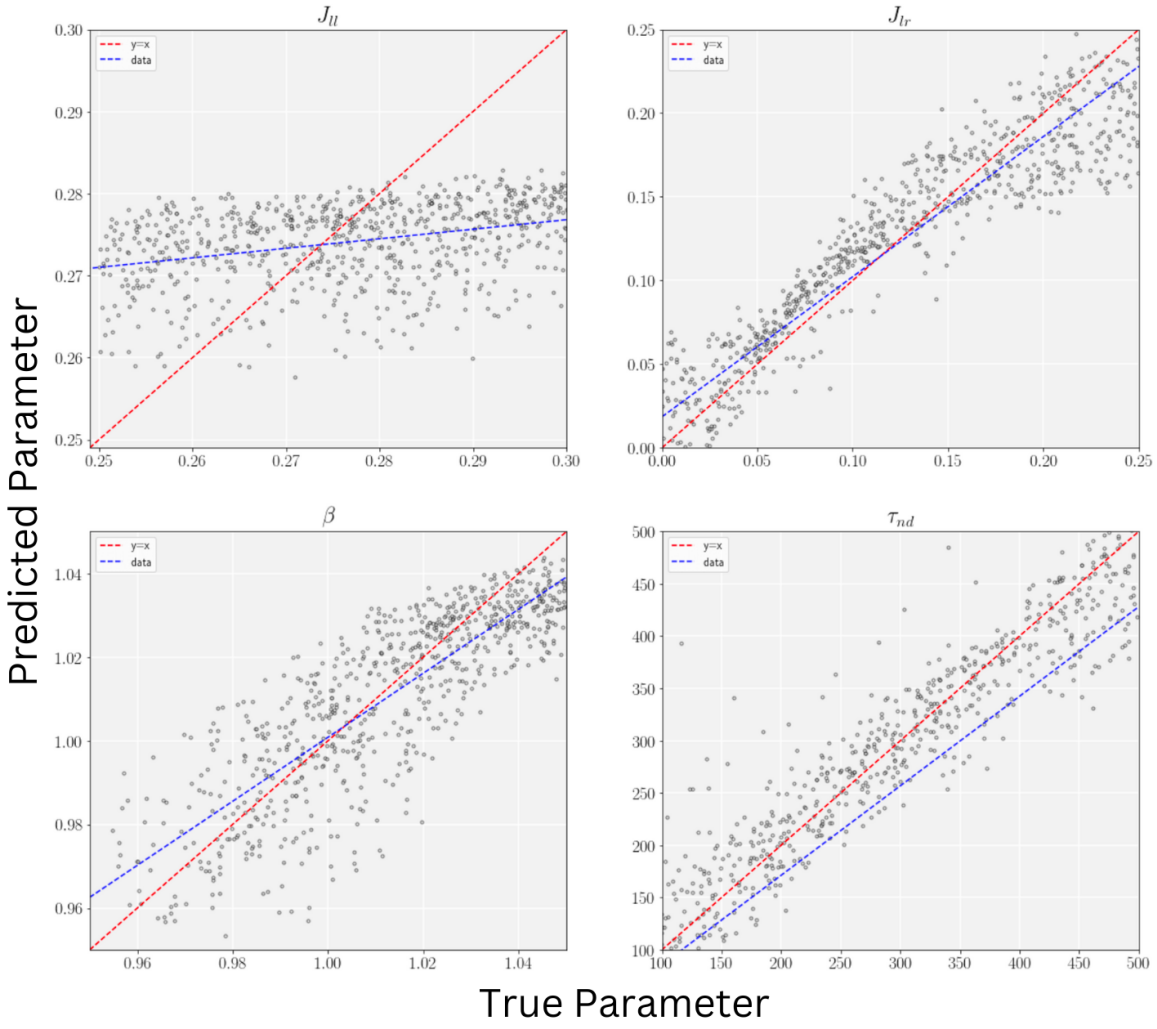

Figure 4.7: Model's predictions (y-axis) compared to the expected values (x-axis) for multiple regression of all the 4 parameters. The red dashed line represent the $x = y$ straight line, *i.e.* the best result possible. The blue dashed line is a linear fitting of the data. It can be seen that the model fails to predict J_{ll} ($R^2 = 0.11$) and doesn't show a high accuracy for the others parameters, especially for β (($R^2 = 0.69$)).

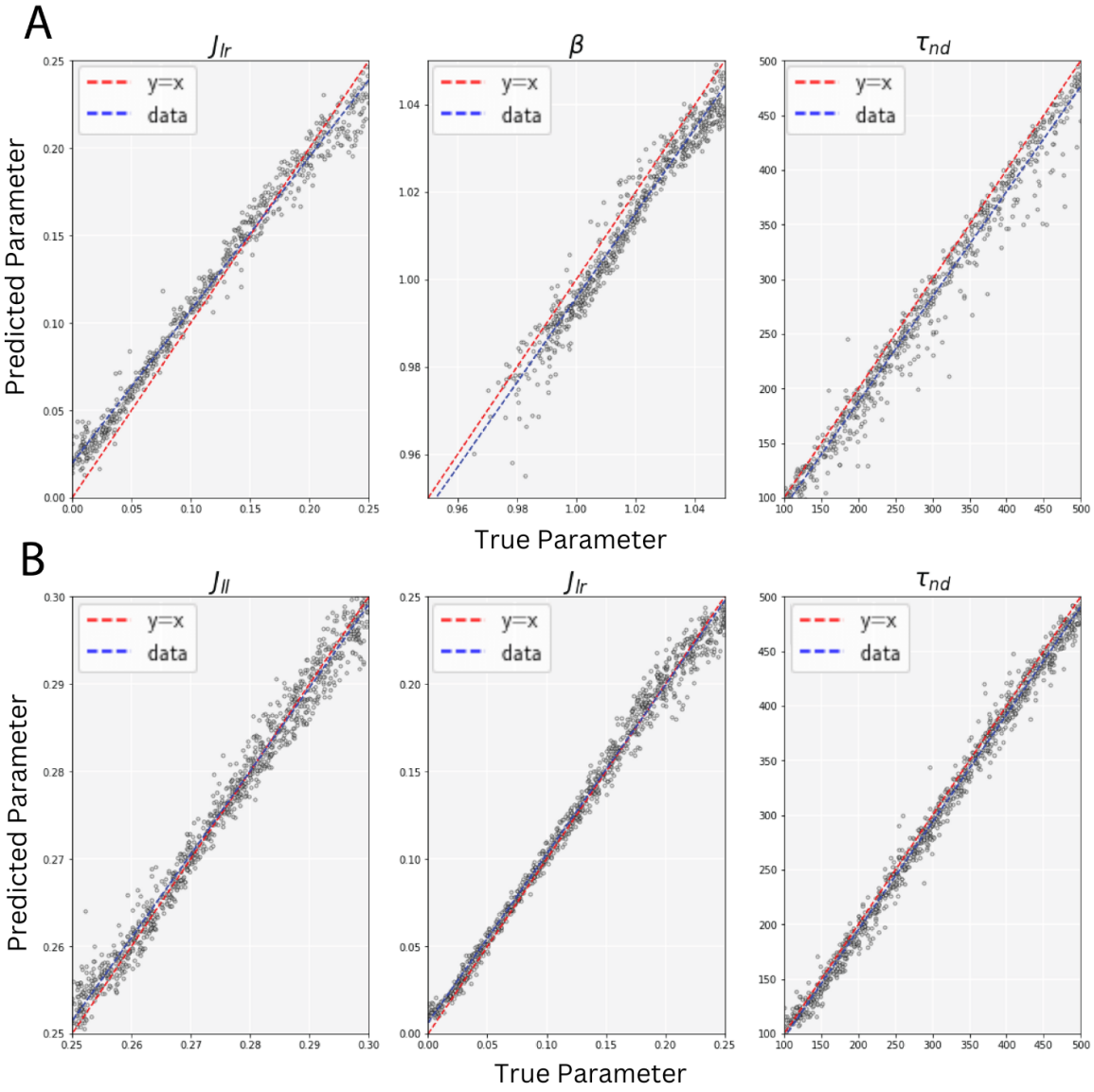

Figure 4.8: Model's predictions (y-axis) compared to the expected values (x-axis) for multiple regression of 3 parameters. **a**, predictions for J_{lr}, β and τ_{nd} (J_{ll} excluded). The prediction accuracy is better than the 4 parameters case although significant fluctuations can still be observed. R^2 scores are 0.94, 0.90 and 0.92 for J_{lr}, β and τ_{nd} respectively. **b**, predictions for J_{ll}, J_{lr} and τ_{nd} (β excluded). Here the predictions are satisfactory with smaller fluctuations. R^2 scores are 0.95, 0.99 and 0.99 for J_{ll}, J_{lr} and τ_{nd} respectively.

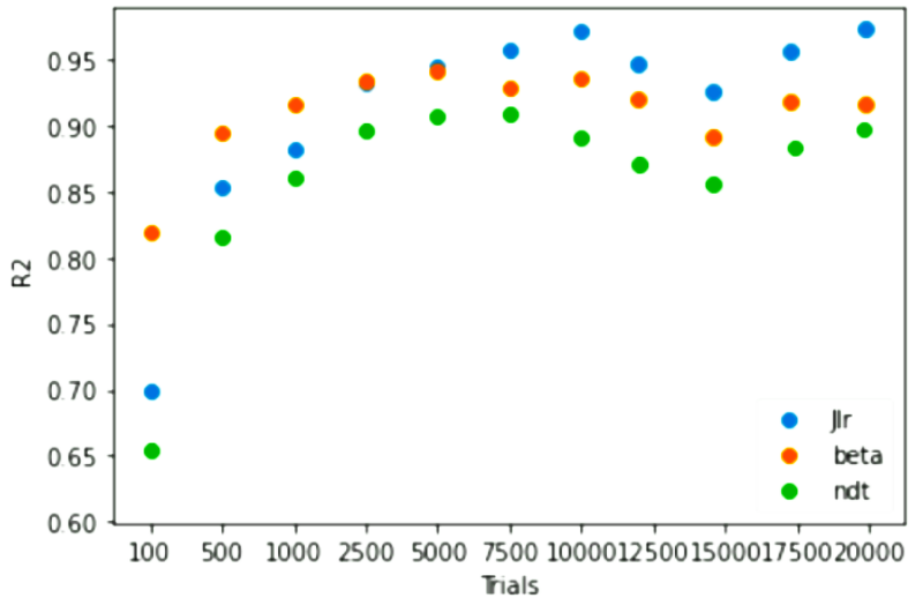

Figure 4.9: Performances of parameter recovery of the three parameters J_{lr}, β and τ_{nd} with input data features evaluated on different numbers of trials.

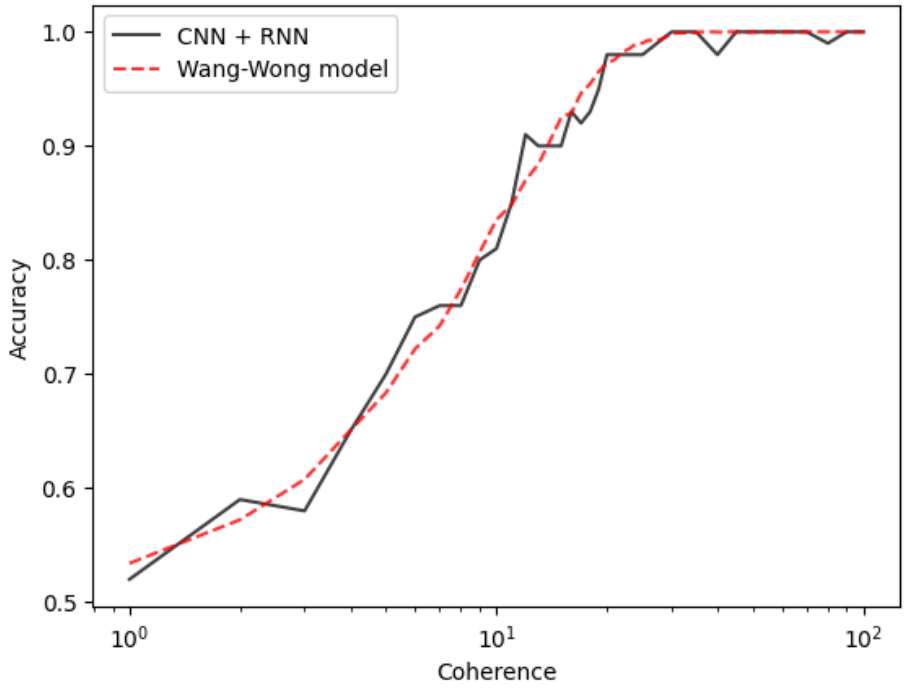

Figure 4.10: Accuracy versus coherence plot of a simple CNN + RNN neural ensemble (black line) compared to plot of the Wang-Wong model (red dashed line). The results of the neural ensemble are averaged over 10 retraining while the Wang-Wong ones are obtain using 10,000 trials for each coherence level.

deep neural network to take the 13 features as input and predict the corresponding parameter values that generated that specific features. We found that the model performs well when either J_{ll} and β were kept constant (three parameters recovery) but performed poorly for all of the four parameters recovery. Finally, I implemented a simple CNN + RNN ensemble to act directly on visual frames of an RDK thus simulating real subjects performing the random dot motion discrimination task. The ensemble, although extremely simple, approximates well the Wang-Wong model and the psychometric data.

A particular attention has been posed to selecting the intervals of the parameters such that we had the best possible biological plausibility without sacrificing generalisability. For this reason, we decided to keep combinations of parameters that gave lower reaction times than expected in a real subject. This is one of the most debatable procedures as there may exist several factors that might influence the selection of the parameter values to predict. Moreover, further studies need to investigate whether the model we used fits well experimental data from healthy subjects, and is also able to describe pathological subjects or results from a different task. Indeed, a model can often generate results even when set with extreme parameter values but it is likely that those results don't fit behavioural data collected in different circumstances than the ones which the model was originally designed to describe. This fact should be taken into account when applying our deep learning approach for parameter recovery of such different task or subject conditions in a real-world scenario as a poor recovery performance might be caused by the inadequacy of the model to generalise to such conditions instead of the recovery algorithm itself.

Our results suggest that, because of the inherent interactions of some parameters, they may have a similar effect on the outputs hence making recovery extremely difficult, not only for deep learning. For this reason, it is important to carefully design an experiment and combine the parameter selection with previous knowledge about the effects of those parameters and their variability across subjects or conditions. For example, suppose we wanted to use our algorithm to recover the self-excitatory synaptic strength in healthy and pathological subjects to prove that it is lower in the presence of pathology. We might design an experiment where the subjects perform a random dot discrimination task, pass their performances to the algorithm and hypothetically observe that indeed J_{ll} is significantly higher in healthy subjects. However, to conclude that J_{ll} is lower in patients with a certain pathology, we first need to address the possibility that such difference in the behavioural output is caused for example

by the modulation β of the background current I_0 as we've observed them having a similar effects on the outputs in their range of values (speed-accuracy trade-off).

While Deep Learning provides an extraordinary versatile tool for solving the problem of parameter recovery, it also has some limitations. In this chapter, we fixed four parameters for simplicity making the problem relatively easy. However, in a scenario with more complex models (or simpler models with a higher number of free parameters), the complexity of the DL algorithm used should also increase to maintain a good predictive power. If the interpretability of the model (and of course, the computational overhead) is not a concern then Deep Learning remains an outstanding tool combining practicality and powerfulness. However, if the number of parameters to be recovered is high, classical optimisation algorithms might become easier to apply. Current research is trying to create more effective DL algorithms by exploiting specific geometry of the parameter search space [23].

A word of caution is needed about the results of the neural ensemble. Indeed, this approach has to be intended as a simple proof-of-concept and has many limitations. For example, no information about reaction times or other kinds of output is available. Recent work has tackled the problem of creating neural architectures with an intrinsic reaction time [183]. A similar result on accuracy may enable future research to compare human neural representations with the embedding learned by the artificial neural architecture. Embeddings are vector representations of an object (or more generally, an input) which is of lower dimension compared to the real world representation (*i.e.*, we may describe an apple using a 3-dimensional embedding made of colour, dimension, sweetness). In some way, human brain has to use an analogous approach using the neural code, as it would be impossible to represent with high-fidelity the entire world using our limited, although enormous, number of neurons and synapses. However, we should be careful in interpreting high correlations of the representations, as this does not necessarily mean that the computations performed by the neural ensemble are comparable to the ones in the brain. On the other hand, these kinds of results may help advance the research in neuro-inspired AI to create algorithms that perform well in advanced complex tasks that human performs efficiently [85] but not necessarily advance our knowledge of the computation the human brain carries.

Chapter 5

Deep Learning for Neural Decoding of Motor Behaviour

5.1 Introduction

The problem of deciphering the neural code, *i.e.*, how the brain encodes the information from the external world into neural representations, has received a high level of attention in the past decades. It is essential to understand the representations the brain uses to picture the external world and how they are elaborated and translated into actionable plans to achieve complex goals. Not only does this have an inherent scientific significance, but it also has profound implications for several practical applications. In recent decades, the interest in brain-computer interfaces (BCIs) and neurotechnologies has exploded [109, 4]. Such devices have been studied for promising clinical applications like neuro-prosthetics [78], rehabilitation [141], therapeutics [147], and advanced diagnostics [202] for neurological diseases and conditions. In addition, BCIs find increasing applications in other fields like gaming [145] or enhancing human capabilities [14].

In particular, BCIs are based on neuronal decoding of the activity in the brain, *i.e.*, the translation of neural representations into action and behaviour. This active interdisciplinary approach is found in an intersection of neuroscience, computer science and engineering. Neural representations carry latent information related to cognitive states, sensory perceptions or intentions that precede motor actions. Given the complexity of the brain, finding such latent representations is an arduous task which can find enormous benefits from the use of computational tools and machine learning in particular.

Several tools have been developed for neural decoding. One of the most typical approaches is the well-established linear regression, with or without regularisation

[96]. Other options include, for example, the Kalman filter [243] or Naive Bayes decoder [123]. Such algorithms have many advantages, being computationally simple, easy to interpret and less susceptible to overfitting, a common problem when using machine learning tools. On the other hand, such simplicity comes at the cost of a strong assumption. For instance, linear regression assumes that input and output variables are linearly correlated, and noise follows a Gaussian distribution. These assumptions may be instrumental in specific contests, but generally, they may not hold, thus making the results completely unreliable.

In contrast, deep learning methods are much more flexible thanks to their approximation capabilities. As discussed in Chapter 2, deep learning's neural architectures are universal approximators, meaning that complex enough neural networks can approximate any functions, even the most intricate. Also, flexibility comes at a significant cost, namely, the increased computational overhead and susceptibility to overfitting, a situation where the model fails to generalise to unseen data.

The brain possesses a highly intricate and nonlinear structure and functioning patterns. As such, linear relationships between input and output cannot generally be assumed. Therefore, deep learning is a practical choice for neural decoding despite its high computational cost, provided that careful attention is given to all of the standard procedures to avoid overfitting. In this thesis's chapter, I will use a recurrent neural architecture to decode continuous actions from neural recordings, because such model structure suits data with a temporal component. Recurrent neural networks can keep information about a sequence's previous element by propagating a hidden state. This is very important as brain activity casually related to motor behaviour may start several hundred milliseconds before the actual movement and even present some correlated activity after. I will also study the latent dynamics represented as the most explaining variables identified through PCA. An important question to answer will be whether latent dynamics can carry enough information to achieve satisfactory decoding performances and even if they are able to eliminate noise sources, thus improving performance.

5.2 Methods

5.2.1 Data collection

This chapter's analysis is performed on a publicly available dataset [177]. A subset of the data has been analysed in another study [129].

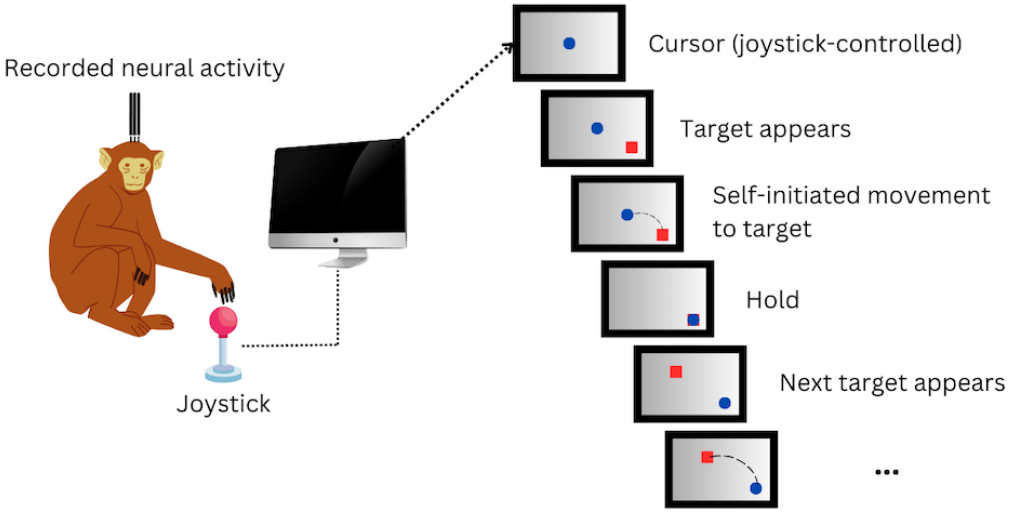

Figure 5.1: Schematisation of the reaching task performed by a monkey during the behavioural experiment. The monkey controls a cursor using a joystick while its neural activity is recorded. The monkey learns to reach a target and hold for 100ms before the next target is triggered.

The recordings come from two different monkeys of the species Rhesus Macaque, M and T (they are referred to as MM and MT), performing a reaching task in which a computer cursor is controlled via arm movements (Figure 5.1).

During the task, a primate chair sat the monkey while it controlled a two-link planar manipulator. Such manipulator consists of two rigid links connected in a serial chain by two rotational joints (often referred to as revolute joints), and operates in a two-dimensional (planar) workspace. Hand movements were confined in a horizontal plane with a square workspace of 20 cm x 20 cm. A 2 cm square acted as an on-screen visual cue indicating the target location for each reach. When the cue appears on the screen, the monkey has to reach it. After each reach, the previous visual cue disappears and a new one will appear in a random location of the screen. After up to four sequential successful reaches to the targets, a liquid reward was offered to the monkey. This series of successful reaches identifies a trial. The successive target location was chosen following a pseudo-random procedure and located with an annulus (radius = 5-15 cm, angle = 360 degrees) centred on the current target. The pseudo-random procedure was used as follows: first, a distance between 5 cm and 15 cm and a relative angle between 0° and 360° with respect to the current target are randomly chosen for the new target. Then, if the new target is located outside the

workspace, a 90° angle is added, and the distance is set to 5 cm. Step 2 is repeated until the target falls within the workspace.

For each trial, the first reach (made from rest) had a different timing for target presentation compared to the following three reaches. For the first reach, after the target was presented, the monkey was allowed to move without an instructed delay period. For the other reaches, the monkey initiated the next target, keeping the cursor within a 2 cm x 2 cm box centred on the current target. Consequently, the successive target is triggered but not displayed, with a 100ms delay. Due to the specificity of the software, the next target is displayed after roughly 100ms, so, in total, the successive target was displayed at 200ms on average after the monkey successfully reached the current target. In addition, a 100ms holding period was imposed simultaneously when the next target was triggered. Thus, the time interval between the reach of the current target by the monkey and when the moment was allowed again was 200ms. The holding period has been used to force the monkeys to decelerate when approaching the target and generate relatively smooth arm movements of variable distance.

All surgical and experimental procedures were performed following the standard for the care and use of laboratory animals. The institutional animal care and use committee of Northwestern University approved the procedures. The data collection was performed using 100-electrode arrays (Blackrock Microsystems, Salt Lake City, UT) implanted in the PMd region (1mm electrode shaft length) and in M1 (1.5mm length). Spike sorting with standard procedures has been performed manually, identifying spikes from raw waveforms (raw signals).

Monkey MM performed a total of 140 trials of 4 reaches each in a single trial for a total of 496 valid reaches. From the primary motor cortex (M1) region, the activity from 67 neurons has been recorded, while from the dorsal premotor cortex (PMd), the number of neurons was 94. Monkey MT performed three sessions while only activity from PMd was recorded. A total of 419 valid reaches were completed in the first session, and 49 neurons were recorded. The second session consisted of 646 reaches and 46 neurons recorded. The third session consisted of 652 and 57 neurons. Invalid reaches or neuron recordings were present due to errors in the data collection, and such data were discarded. In particular, for monkey MM 5 invalid reaches in the single session were excluded, while monkey MT performed 6, 8 and 7 invalid reaches in the three sessions, respectively.

The generated dataset contained two sets of data: raw and heavily processed. The raw data has been used in the analysis presented in this chapter to have maximum

flexibility. Raw data consisted of recordings of neural activity and four different behavioural outputs: position, speed and acceleration of the cursor moved by the monkeys plus the force applied to the controller only for monkey MT. Only the first three outputs were analysed in this chapter. All of the recordings were collected in bins of 1ms each. At such time resolution, each time bin contained one spike at maximum.

5.2.2 Data preparation

The data was processed using an existing processing and decoding toolkit [69]. Input and output data were binned in 50ms time bins. Input bins contained the number of spikes in each 50ms time window, while output bins contained an average output value for each bin. A functional connectivity matrix between neurons has been estimated using simple cross-correlation to perform graphlet-based analysis, similar to Chapter 3.

Following a standard procedure [69], decoding the behavioural output at a specific time was performed using a total of 13 time bins: six before, six after and the one corresponding to the same time bin of the output. In other words, a total of 650ms of 300ms before, 300ms after, and 50ms concurrent to the binned output have been used to decode 50ms of average output within that time window. To decode an output $y_{t(0)}$, the input consisted of $X_{t(-6)}, X_{t(-5)} \dots X_{t(0)} \dots X_{t(6)}$. An analysis of how decoding performances vary with the number of time bins in the input has been performed for velocity decoding in the PMd region of monkey MM.

The data has been further processed by computing the z-score and 0-centring values to help convergence of the DL decoding algorithm.

5.2.3 Neural decoding

Given the continuous nature of the behavioural patterns in the dataset, decoders with a recurrent mechanism are the most suited for this kind of analysis. In particular, a simple Recurrent Neural Network (RNN) has been chosen among its more complex derivations, such as GRU or LSTM. The reason is that GRU and LSTM are more suited to "remember" information far in the past for long sequences, while in our case, the sequences are composed of 13 elements close in time. The RNN takes a sequence of 13 components (time bins) and, for each component, generates an embedding of a dimension equal to the number of hidden units used. It applies Eqs. (2.40) to the input x_t with $t = 6$. A number of 400 hidden units have been used for regular

analysis, meaning that h_t in Eqs. (2.40) has a dimension of 400 and hence U_h a dimension of 400×400, W_h a dimension of 400×13 and b_h a dimension of 400. In the case of the analysis of decoding performances using input data of varying numbers of bins, 300 hidden units have been used. After the projection in the higher dimensional embedding space, a non-linear function (tanh) has been applied. Finally, a fully connected layer following Eq. (2.37), is used to project the RNN's output of dimension $400 \times 13 = 5200$ or $300 \times 13 = 3900$ to a two-dimensional behavioural output for the X and Y components, respectively. The training has been performed on the 70% of the data and tested on another 15% chosen such that there is no overlap between any of the neural data time windows. A batch size of 32 was used, and the training lasted for 25 epochs (15 for the different number of time bins analysis). The MSE Loss has been selected for minimisation to improve the decoder performance, and the RMSprop optimiser has been used with a learning rate of 0.0005. The trained decoder has been evaluated using the remaining 15% as validation data, and its performance was quantified using the R^2 score.

5.3 Results

The neural recording data consist of timed arrays of 0 and 1, where 1 indicates that a spike happened. The raw dataset has been put into bins of 50ms each by summing all the spikes that happened in a specific time bin. The behavioural outputs have been binned as well by averaging the data within each window of 50ms. The output data to decode has been constructed by picking 13 time bins of spiking data for each output bin (concurrent bin, six before and six after).

To start with, the connectivity matrices of the PMd area for the three sessions performed by the monkey MT plus PMd and M1 areas of monkey MM have been estimated. Figure 5.3 shows the connectivity matrices across sessions and regions and the respective GCMs obtained using the same approach as in Chapter 3. The GCMs show 2 clusters similar to the case of SC and FC-derived graphs, but with a lower correlation between orbits of different types (central and peripheral). These results indicate similar graphlet substructures in connectivity matrices between single-unit recordings from non-human primates and whole-brain structural-functional data from humans as shown in Chapter 3 (3.1).

The following figures show a comparison between the predicted behavioural output for the X and Y components separately and the actual output: Figure 5.4 for position decoding, Figure 5.5 for velocity, and Figure 5.6 for acceleration. The performances

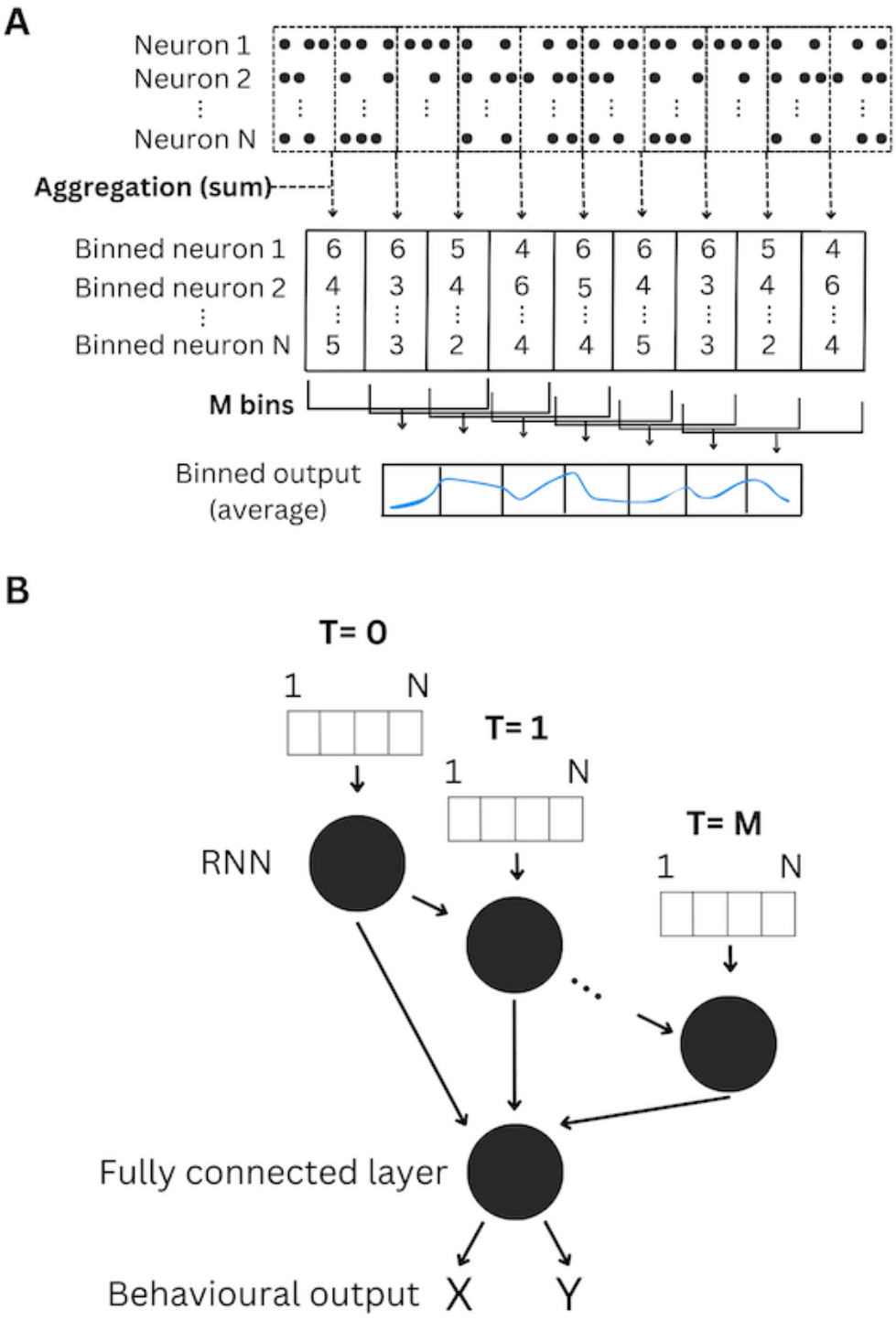

Figure 5.2: Schematisation of data preparation and neural decoding procedure. (a) Depiction of data preparation method including binning input and output and associating a series of M input bins, before and after the concurrent output time; (b) Simple diagram of the neural architecture (RNN) used for neural decoding.

of the decoder have been evaluated using the R^2 scores. Overall, the best decoding performances are obtained with velocity decoding, with R^2 scores comprised between 0.60 and 0.77. Following, position decoding with R^2 scores between 0.30 and 0.64. Finally, position decoding with R^2 scores between 0.22 and 0.53. The M1 region of monkey MM performing the single session offered the highest performances most of the time, while the decoding of the PMd region of MT did not perform better than the PMd region in MT, independently from the fact that more neurons were recorded.

Figure 5.7 presents the fraction of variance explained by the components identified using PCA. The red dashed line highlights the number of components necessary to obtain 80% of explained variance, while the green dashed line the 90% of explained variance. Such values are relatively high compared to previous studies focused on different areas [26, 117]. That is, we observed a large variety of neural activity patterns in the current data, which led to a high number of principal components needed to explain a sufficient amount of variance.

Figure 5.8, 5.9 and 5.10 shows again the predicted behavioural outputs with the difference that the input data is composed by the PCA components of the spiking data that explained 90% of the variance. Generally, decoding performances on PCA-reduced data are lower, but in the case of position decoding, using PCA components instead of the unaltered data leads to a significant increase in performances for position decoding from the PMd area of monkey MT.

Figure 5.11 shows the R^2 scores obtained using inputs of different dimensionality, covering all possible combinations of the number of bins before and after ranging from 0 to 10. Considering monkey MM and PMd region as example, the highest performances are reached when the bins before are increased while the bins after are decreased. This result indicates that considering the correct number of bins before and after the concurrent output bin, it is possible to reach higher decoding performances while maintaining a similar computational overhead.

5.4 Discussion

In this study, we performed neural decoding of continuous motor behaviour from neural spiking data from two monkeys performing a 2D reaching task. We utilised a recurrent neural network and used the binned data, in the whole form and PCA-reduced, as input to decode three different behavioural outputs: position, velocity and acceleration of the cursor controlled by the monkey's arm.

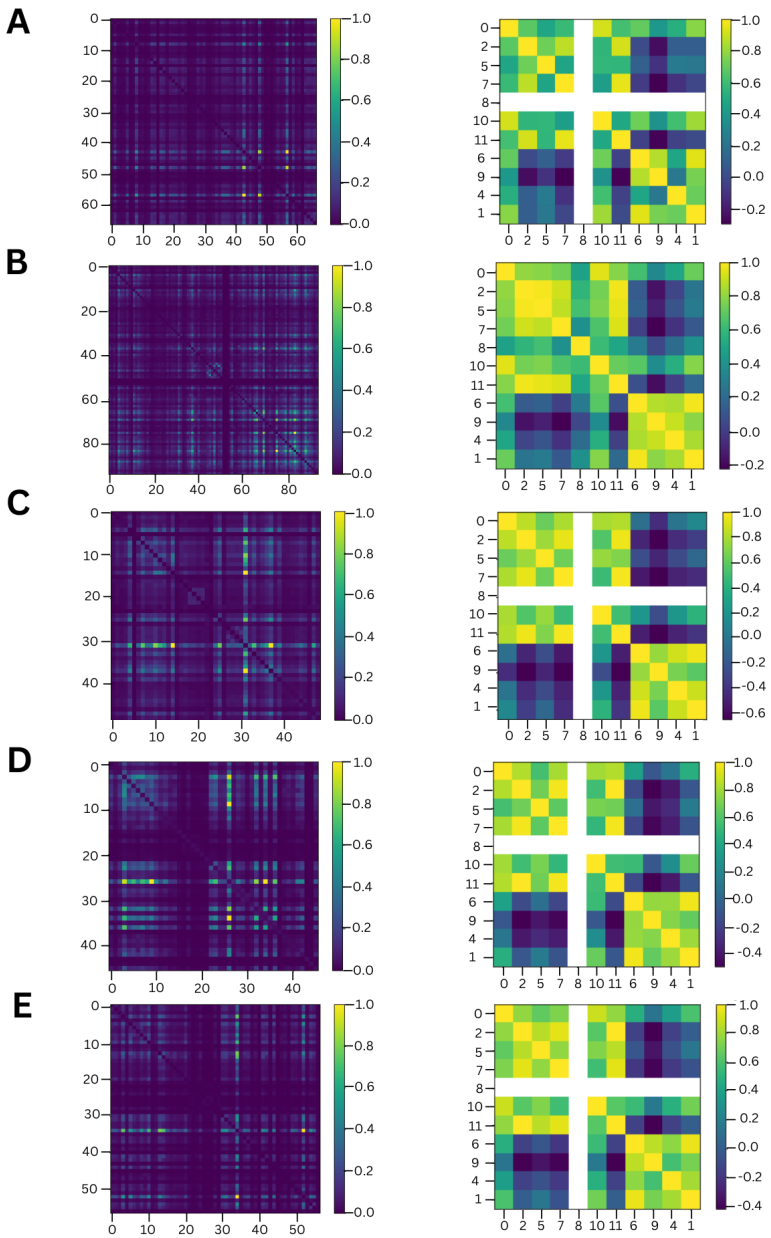

Figure 5.3: Connectivity matrices of neurons estimated using cross-correlation between spike trains (left) and respective GCM (right). (a) Monkey MM, M1 area; (b) Monkey MM, PMd area; (c) Monkey MT, PMd area, session 1; (d) Monkey MT, PMd area, session 2; (e) Monkey MT, PMd area, session 3.

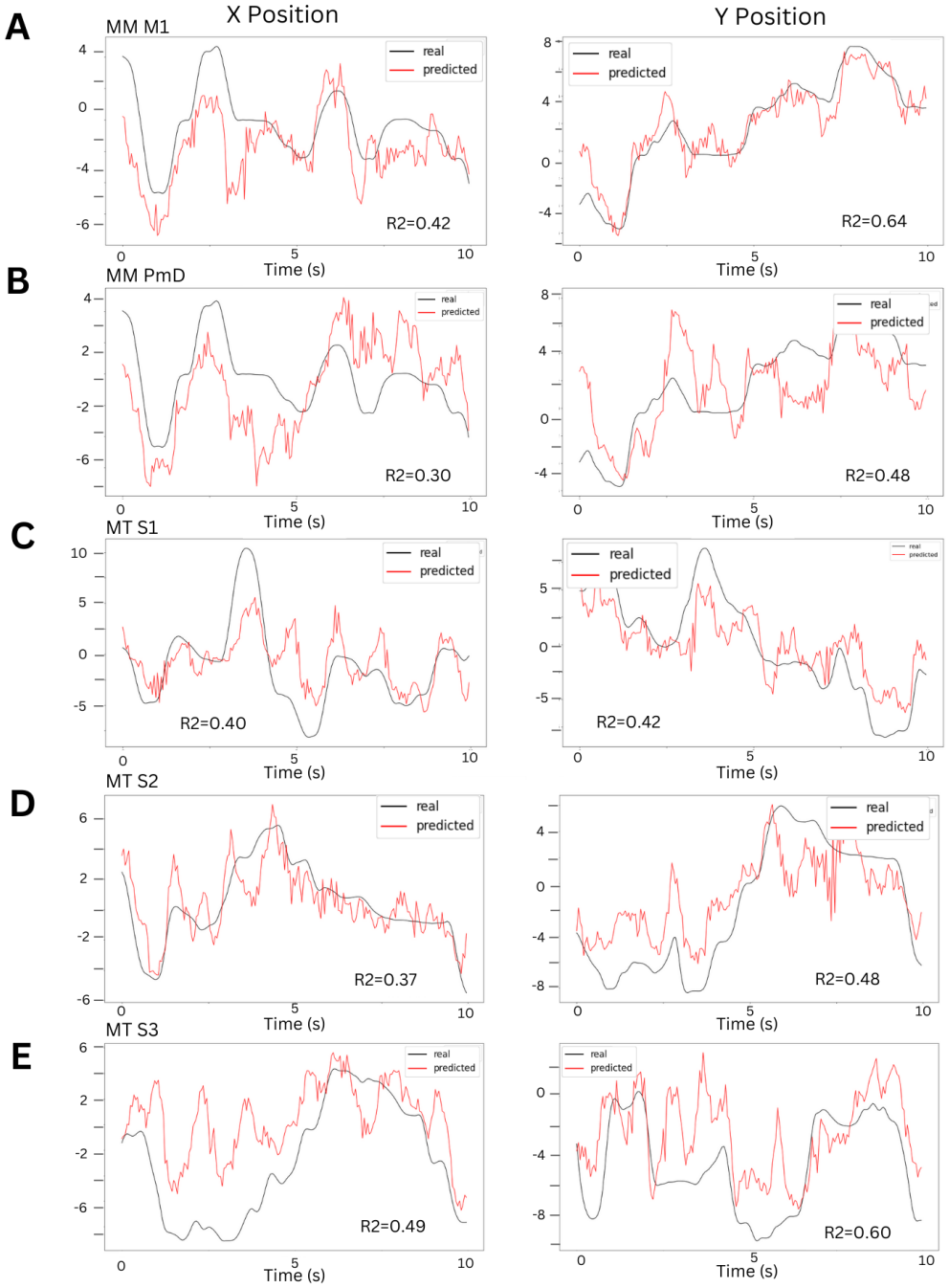

Figure 5.4: X component (left) and Y component (right) of real and predicted position of the cursor controlled by the monkey arm movement. (a) Monkey MM, M1 area; (b) Monkey MM, PMd area; (c) Monkey MT, PMd area, session 1; (d) Monkey MT, PMd area, session 2; (e) Monkey MT, PMd area, session 3.

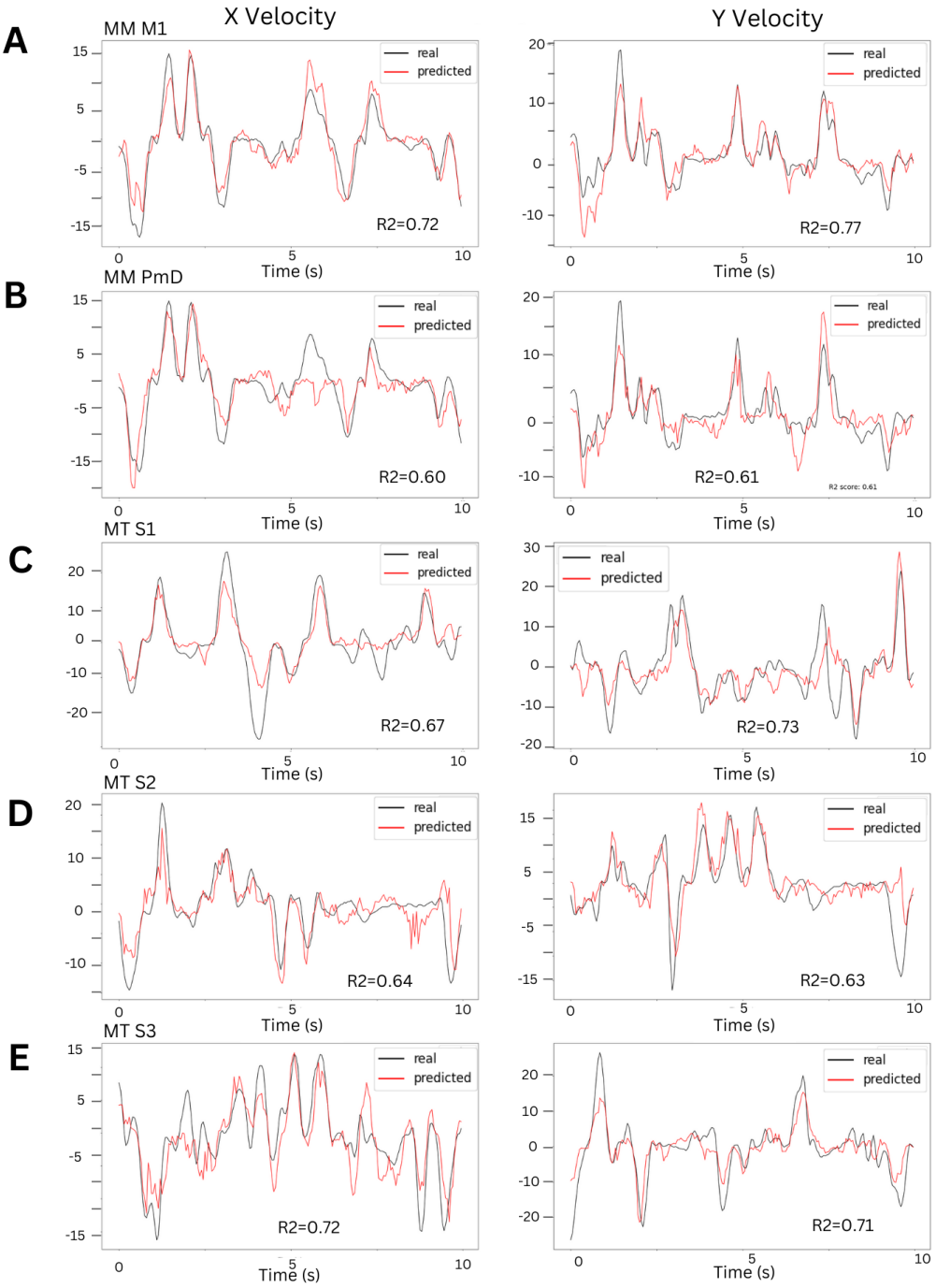

Figure 5.5: X component (left) and Y component (right) of real and predicted velocity of the cursor controlled by the monkey arm movement. (a) Monkey MM, M1 area; (b) Monkey MM, PMd area; (c) Monkey MT, PMd area, session 1; (d) Monkey MT, PMd area, session 2; (e) Monkey MT, PMd area, session 3.

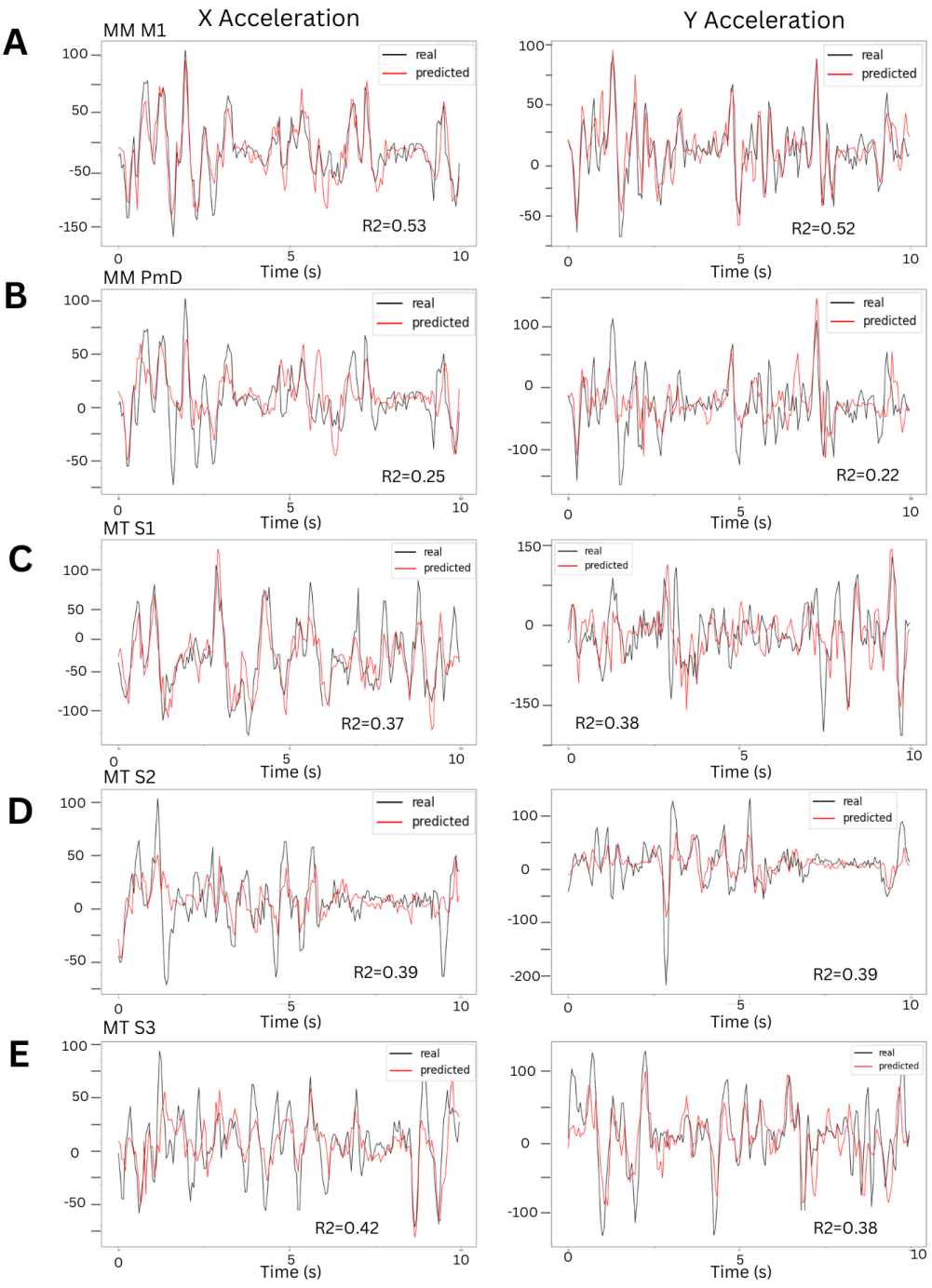

Figure 5.6: X component (left) and Y component (right) of real and predicted acceleration of the cursor controlled by the monkey arm movement. (a) Monkey MM, M1 area; (b) Monkey MM, PMd area; (c) Monkey MT, PMd area, session 1; (d) Monkey MT, PMd area, session 2; (e) Monkey MT, PMd area, session 3.

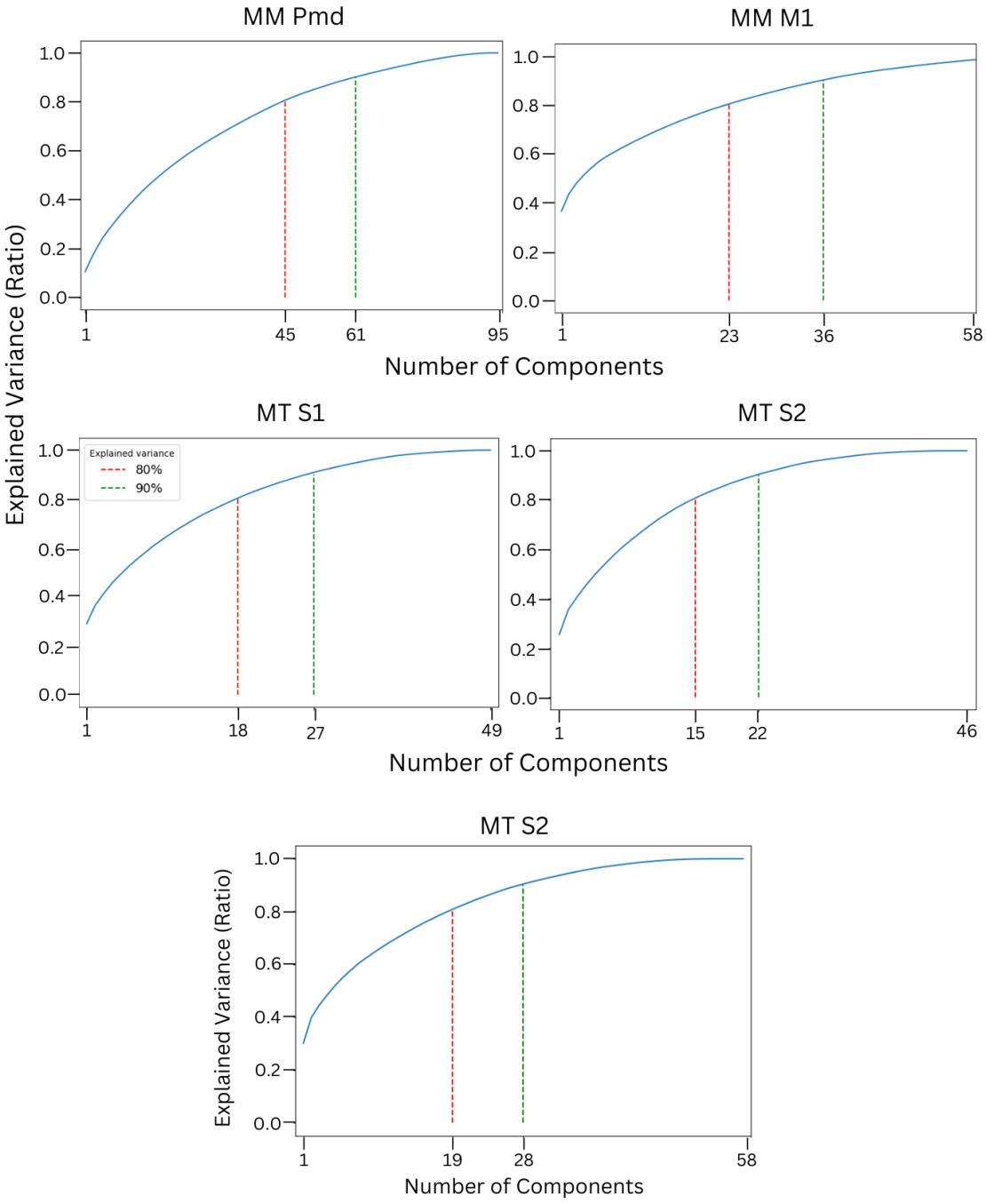

Figure 5.7: Explained variance plots of principal components identified with PCA across all the regions and sessions. The blue line indicates the cumulative explained variance. The red and green dashed lines indicate the number of components to reach 80% and 90 % of explained variance, respectively.

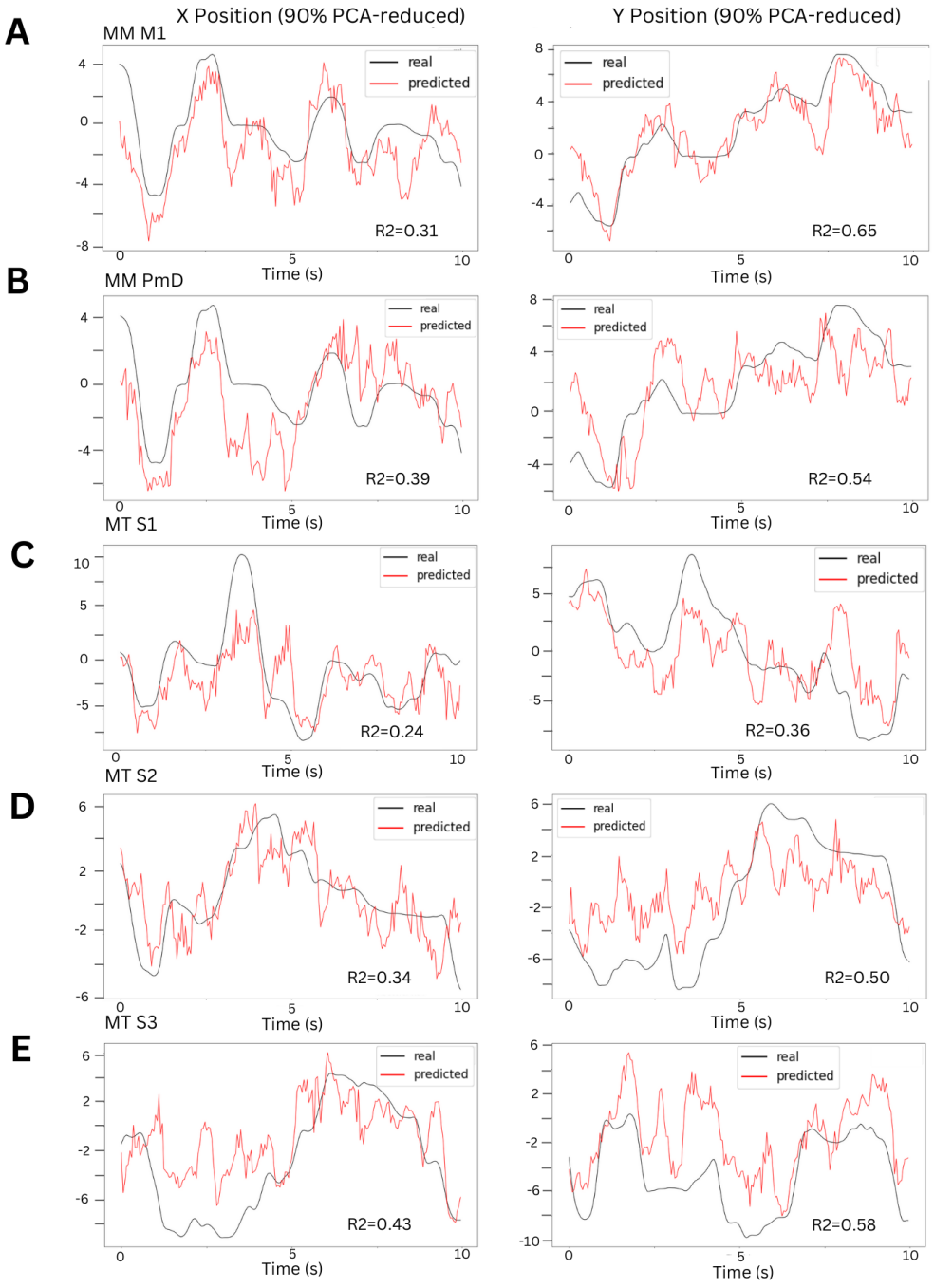

Figure 5.8: X component (left) and Y component (right) of real and predicted position of the cursor from neural data PCA reduced to explain 90% of the variability. (a) Monkey MM, M1 area; (b) Monkey MM, PMd area; (c) Monkey MT, PMd area, session 1; (d) Monkey MT, PMd area, session 2; (e) Monkey MT, PMd area, session 3.

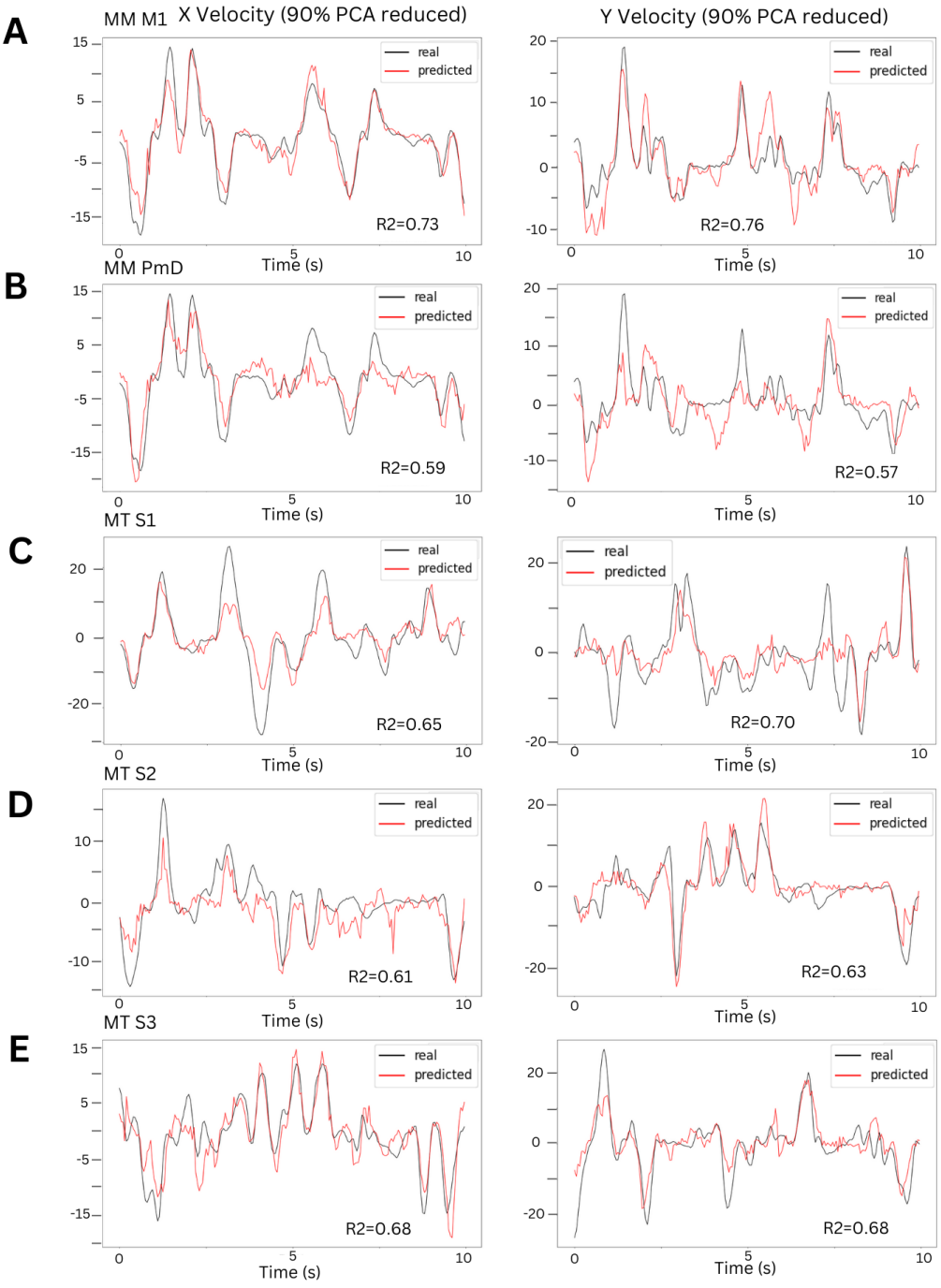

Figure 5.9: X component (left) and Y component (right) of real and predicted velocity of the cursor from neural data PCA reduced to explain 90% of the variability. (a) Monkey MM, M1 area; (b) Monkey MM, PMd area; (c) Monkey MT, PMd area, session 1; (d) Monkey MT, PMd area, session 2; (e) Monkey MT, PMd area, session 3.

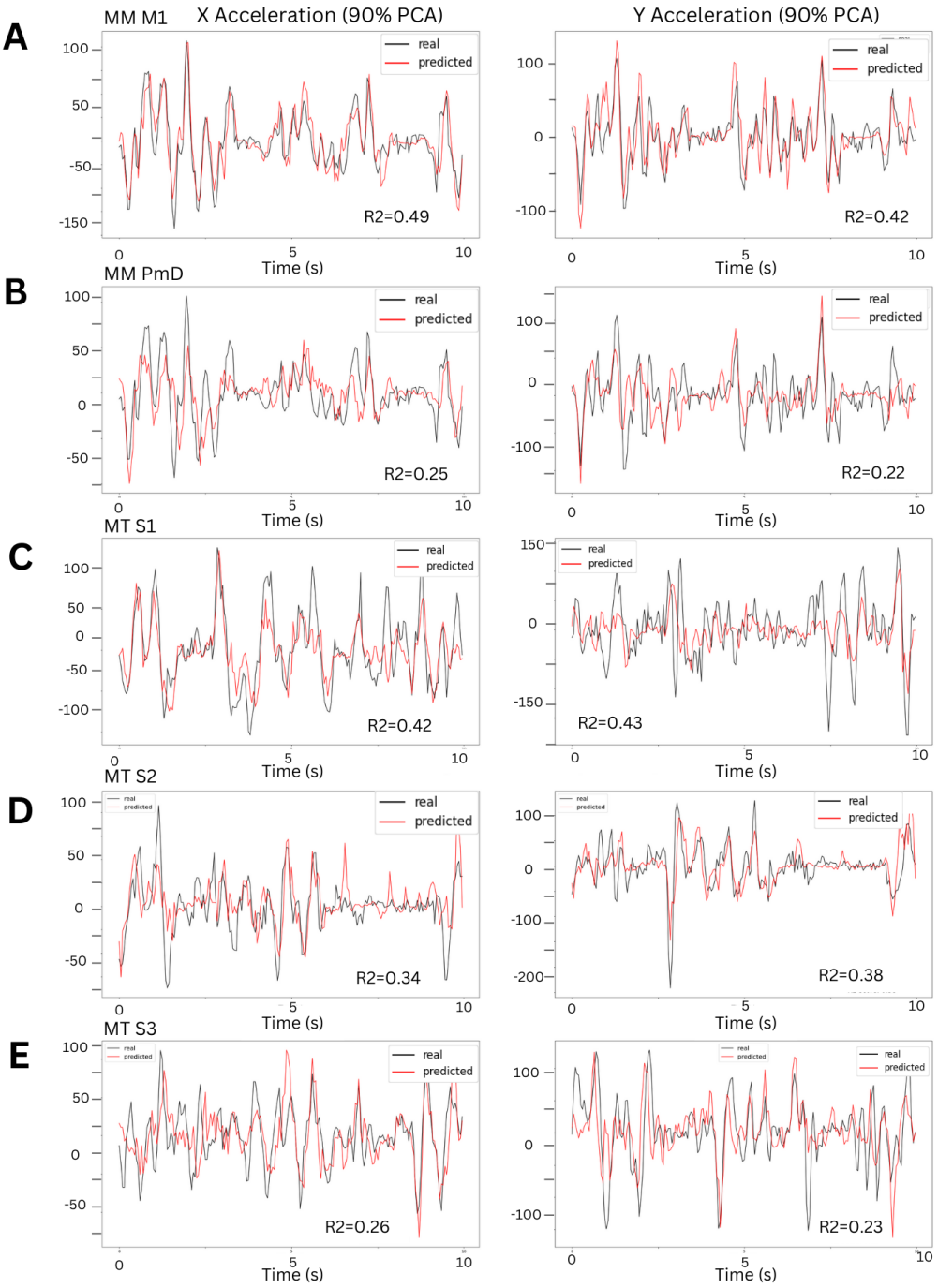

Figure 5.10: X component (left) and Y component (right) of real and predicted acceleration of the cursor from neural data PCA reduced to explain 90% of the variability. (a) Monkey MM, M1 area; (b) Monkey MM, PMd area; (c) Monkey MT, PMd area, session 1; (d) Monkey MT, PMd area, session 2; (e) Monkey MT, PMd area, session 3.

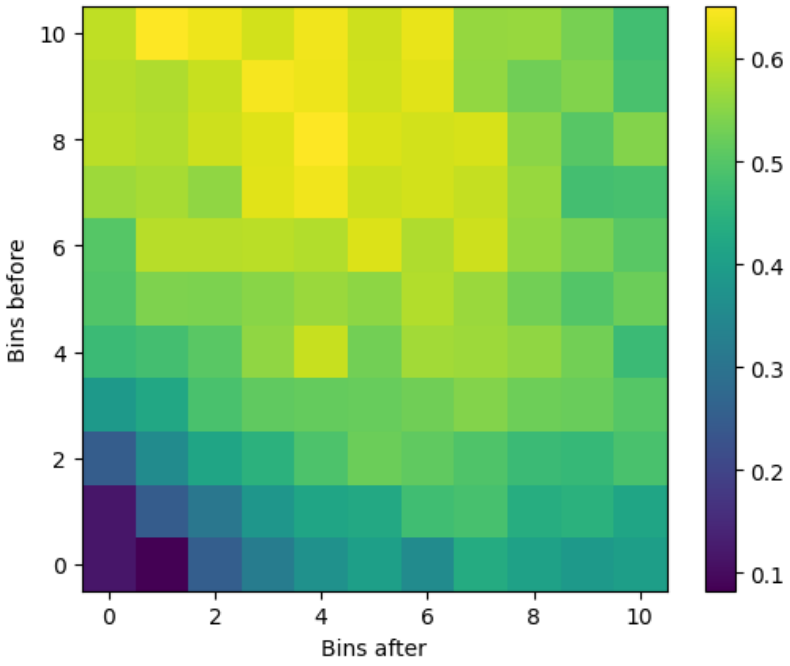

Figure 5.11: Decoding performance as R^2 scores with input data consisting of all the possible combinations from 0 up to 10 bins after and before. The decoding is performed on velocity from neurons of MM monkey's PMd region.

A critical premise is that Machine Learning, particularly Deep Learning, is generally easy to apply but more challenging to optimise. The choice of the optimal algorithm is already a daunting task, given our vast pool of choices in that sense. Here, RNNs were chosen following other studies [69] or simple considerations about the tasks that can be performed using this specific architecture. However, this does not guarantee that RNN is the best choice for decoding performance. Moreover, Deep Learning algorithms typically have several hyperparameters to optimise, which usually requires fine-tuning operations. This is important because the same decoder may need to be tuned differently on each dataset to reach optimal performances. In addition, the same algorithm can even perform differently based on the hardware is running on (the so-called "no free lunch theorem" [241]). Given the high number of model training procedures on different datasets, fine-tuning has not been performed in this chapter to avoid an exaggerated computational overhead. However, it is reasonable to assume that fine-tuning would only slightly increase some performances. Below, I will focus on discussing the most significant results.

Good decoding performances do not necessarily imply that a brain area encodes that specific behavioural output, nor that the decoder performed analogous compu-

tation as the nervous system. Specifically, the top performances reached by velocity decoding do not necessarily imply that M1 and PMd encode more information about velocity than position or acceleration. Instead, our results suggested that the neural activity in M1 and PMd and velocity output have a more straightforward correlation type, a less noisy relation, or both with respect to other motor outputs. In addition, it is generally not possible to conclude that motor behaviour is a direct cause of the specific neural activity used for decoding, even if one precedes the other in time. We can accurately decode a motor output from a specific pool of neurons involved in a different process caused by another pool, which is the main responsible for movement. One notable case which has been widely studied in the previous decades is motor learning. It has been observed that PMd has a vital role in motor learning, a process which temporally follows motor activity [80, 81]. Although motor learning comes after the motor output, we may still be able to decode it from the neural ensembles involved. To assess this effects, a cause-effect analysis should be carried on which is a complicated problem I overlooked in this case. In general, assessing causality is a challenging task, and different experiments from various fields may be necessary for a comprehensive understanding of how neural activity causally leads to motor behaviour [66, 81, 12].

Another exciting aspect of these results is how M1 performed better than PMd in MM. Again, this may support the fact that M1 is more specialised in directly translating neural activity into motor behaviour. At the same time, PMd is also involved in more complex backward activities, *i.e.*, those neural activities which propagates in the opposite direction of standard forward motor activities. An example is motor learning, with neural activity propagating backward with respect to standard motor tasks like limb controls. This, could add noise to the decoding task of limbs movement. However, it is interesting to note that in the third session of MT monkey experiments, performances were similar to the MM case, except for acceleration decoding. In addition, in all three sessions, PMd decoding in MT performed better than PMd decoding in MM. Given the similar experimental conditions and subject, these results strongly suggest that the presence of activity in PMd is not directly correlated to motor behaviour. In fact, PMd neurons recorded in MM were significantly more than in MT, indicating that more noise, *i.e.*, neural activity related to different motor tasks, may have also been added.

If neurons have strong cross-correlation or merely introduce noise unrelated to behaviour, it is useful to apply dimensionality reduction to capture the strongest sources of variance in the data. However, in the current study, PCA-reduced data

did not increase decoding performances, with the only notable exception of position decoding in PMd of monkey MM. This may be because velocity is well represented by neural activities in M1 and PMd, meaning that reducing the input data dimensions does not provide any advantage in denoising. Position decoding might be different because PMd has been observed to better encode the prediction of future occurrence of a target compared to M1 in a 2D reaching task [86]. Potential future research may focus on the study of position decoding using a different number of time bins before or after the concurrent output bin to evaluate better the role of PMd in position encoding.

It is also interesting to note that the number of components to explain 80% and 90% of the variance is relatively high compared to other brain regions or neural data. This is particularly evident for monkey MM PMd, while the effect is weaker for M1 decoding, where a lower number of components is needed to explain a higher level of variance. Again, this could be because different sources of variance are present in the recorded data, with a more substantial presence of competing representation in PMd than in M1. An exciting research direction is the analysis of the first components, as previous studies indicate that different components may encode different information [216].

Another vital aspect to remember when discussing the potential origin of noisy information from the input data is the number of bins used for neural decoding, especially the ones after the concurrent output bin. Although other studies have used up to six time bins after (300ms) [69], the results of this chapter show that adding an excess of time bins after the concurrent time causes a drop in performances. Given that the computational overhead is similar, including more time bins before the concurrent time and leaving a maximum of one or two (50-100ms) bins after the concurrent time is better. Using an excessive number of time bins after the concurrent time may introduce neural activity unrelated to the behaviour we are trying to decode, thus making input data more noisy and hard to decode. This may not be true for different brain areas, and potential future improvements of this framework may focus on integrating different sources of neural activity [162].

Chapter 6

Discussion

6.1 Summary of the results

In this thesis, I presented a small but powerful set of results from the application of computational models at different scales for neuroscience.

Proceeding with a top-bottom approach, in Chapter 3 I presented first an analysis of the whole brain connectome using an approach from graph theory and network science, two disciplines that overlook fine-grained biological details and describe in mathematical terms the human brain as a collection of nodes and links. More specifically, I applied a novel approach based on graphlets, which are defined as small, induced, connected subgraphs. The results of the application of the graphlet approach to human structural (DWI) and functional (fMRI and MEG) connectomes show a concise, efficient and elegant way to summarise many topological results based on network neuroscience. After identifying the first linearly independent orbits from graphlet of up to 5 nodes, I computed the Spearman's correlation coefficient between each pair of orbits to obtain the Graphlet Correlation Matrix (GCM), a useful tool that can condensate many topological information of a graph of any dimension in an 11×11 matrix, offering an efficient and universal comparison tool [254]. We found that differently from socio-economical networks, in brain networks, all the orbits are strongly correlated similarly to other biological networks. However, the correlation is stronger between central orbits and between peripheral orbits, indicating the presence of provincial hubs as well as connector hubs in the human brain. I computed the Orbit Centrality Percentage (OCP), *i.e.*, the z-transformed percentage of central orbits over the total, for each brain region of the HCP human brain atlas. This analysis offered an easy way to identify and characterise important regions in the brain. Indeed, network neuroscience intensively investigated the role of hyper-connected nodes, namely the hubs of the network. Hubs can have different roles, and characterising such nodes

using only their degree risks to overlook other essential features. For example, hubs can be brokers/connectors if they act as a bridge between several nodes of different communities, or they can be provincial hubs if they connect to many lower-degree nodes within the same community [79].

In Chapter 4, I utilised a simple MLP for parameter recovery of the Wang-Wong model of perceptual decision-making (Eq. 2.16 to 2.24). I drew several combinations of 4 different parameters: excitatory synaptic strength J_{ll}, inhibitory strength J_{lr}, background current modulation β and non-decision time τ_{nd}. With the help of accuracy and reaction time vs coherence plots, the parameter distributions have been adjusted, discarding those parameter values leading to an unreasonable model's behaviour. The input layer of the MLP took a set of representative statistics computed from the model's raw results (decisions and reaction times). The output layer provided the parameter values that most likely generated those specific statistics. In the case of 3 parameters decoding, the model performed very well with R^2 scores around 0.9. In contrast, the model failed to reach acceptable performances for the four-parameter decoding because of the concurrent effect of J_{ll} and β on the output. The decoding performances have been evaluated on statistics computed from different model runs (trials), outlining an optimal performance of around 10,000 trials. Finally, a proof-of-concept neural ensemble has been presented to act directly on a sequence of frames of a random dot kinematogram, showing that even a simple CNN and RNN combination can provide a psychometric curve similar to the one of human subjects performing the same task.

In Chapter 5, I performed a neuronal decoding task on spiking neuron data of two monkeys performing a 2D reaching task. The monitored brain areas were M1 and PMd with 58 and 95 neurons, respectively, in one session for one monkey and PMd in three different sessions for the other with 49, 46 and 58 neurons, respectively. Using RNN, I found that velocity decoding performed better than position and acceleration, with R^2 scores around the value of 0.7 on average. In addition, M1 decoding performed slightly better in general for all the behavioural outputs. The performance has also been tested with respect to input data of different bins, proving that it is better to limit the number of bins after the concurrent output time and focus on adding more bins before. In the second part of the experiment, I performed PCA on the neural data, proving that the numbers of components needed to explain 80% and 90% of the variance were relatively high. Then, I performed the same decoding task, this time using the PCA-reduced data as input for the decoder. The decoding performance

generally dropped significantly, except for the position decoding from PMd of the first monkey, where PCA reduction of input data was beneficial.

6.2 Discussion

The human brain is arguably the most complex object in the known universe. As such, neuroscience can enormously benefit from the use of computational tools. The zoo of computational tools for neuroscience includes many approaches. In this thesis, I decided to catalogue them based on the application scale. This is just one of the many possible choices to classify such tools, and it helped address the specific research questions I posed at the beginning of the thesis.

Intuitively, larger-scale approaches necessitate several approximations, particularly reducing biological details to some essential features. We have seen graph theory to fit particularly well in this framework as it models a network as a mathematical object which can be represented as a square matrix. At this level of analysis, we want to answer research questions related to the higher organisation of the brain or to the broader activity patterns that may explain whole-brain states or describe biomarkers used to discriminate a specific group of individuals from others. Moving to a smaller scale is far from being a straightforward procedure, as the research questions may change dramatically. The Wang-Wong model introduced in Chapter 4 was created as a model for the LIP area in the brain that has been proven to accumulate noisy evidence before reaching a decision threshold in a 2AFC. Reducing the scale of analysis allowed us to consider biological parameters such as synaptic strengths and currents. This comes with the necessity to overlook details like the effect of other brain regions. The input is simulated as an input current, but no information about the origin of this current is provided. To bridge the gap between large-scale and mid-scale computational tools, I presented a proof-of-concept neural ensemble to simulate a system able to solve the random dot motion discrimination task from a series of frames (video). This model has many limitations compared to the Wang-Wong model, the biggest being the non-biological plausibility. However, it provides a proof-of-concept for potential future development in integrating tools at different scales (see Section 6.3).

In general, the choice of the model is guided by many factors. However, two specific aspects are essential: interpretability and complexity. Although not necessarily, those two aspects are usually inversely correlated: the higher the complexity of the model, the more challenging its interpretability. Hence, we should expect a loss in

interpretability when considering more powerful approaches. AI provides a typical example. The MLP used in Chapter 4 for parameter recovery and the RNN used in Chapter 5 for neural decoding are not natively interpretable models. That means that they are not able to answer research questions such as "Which specific part of the inputs guided the model to make a specific prediction?" or "Are the computations carried by AI similar to the ones performed by the human brain?". This lack of native interpretability is caused by the complex nonlinear computation that DNNs can perform. Conversely, the meaning of correlation coefficients between orbits in Chapter 3 were easier to understand and related to real-world scenarios, helping to characterise different real-world networks. However, the predictive power of the graphlet approach is far from the one of AI. It is essential to find a good balance between interpretability and complexity [133]. We may prefer more complex models because we have the necessary computational resources. However, a model as complex as the system we want to study is not helpful. For example, the original proposal of the Human Brain Project (HBP) [143] was criticised because it aimed at recreating a perfect computational replica of a whole brain using neuromorphic computing [61]. In such a scenario, researchers complained about the lack of well-defined research questions and the non-clear usefulness of such large-scale simulation [61].

In this discussion, it is also worth mentioning alternative models to accomplish the goal set at the beginning of the thesis. For example, Dynamic Causal Modelling (DCM) is a framework that specifies, fit and compare models [64]. It is widely used to estimate the influence that an area of the brain has on another using multivariate time series like fMRI or MEG. Such approach first focuses on the fitting procedure of different models to the time-varying data, applying a subsequent Bayesian approach to select the best model in term of some type of performances of interest. As described, this approach fits perfectly in the framework of the problem I addressed in Chapter 3, *i.e.*, the multi-modal integration problem. While offering potential improvements over a conceptually simple approach like Graph Theory, DCMs are usually more computationally expensive and there are high risks of over-fitting, especially when dealing with complex models and a lack of data [137]. In Chapter 4 and 5, Deep Learning was the main tool used to answer the relative research questions. As discussed, supervised DL algorithms estimate the joint probability $p(\hat{X}, \hat{y})$, with $\hat{X}$ and $\hat{y}$ referring to a subset of input and output data, respectively. This is in turn an approximation of $p(X, y)$ where X and y refer to the real complete data which for practical reasons are unavailable. A similar approach is followed by Bayesian methods. In particular, to solve the problem of information encoding in the brain, Bayesian approaches look

for the joint distribution $p(D, I)$ (where D refers to the information encoded by a specific area, and I the sensory input) [226]. Bayesian models have been extensively used to study visual processing, sensory integration, sensorimotor integration, and collective decision making [175]. However, compared to DL approaches, Bayesian methods suffer from similar disadvantages like complexity and lack of interpretability. In addition, specifying appropriate prior distributions and likelihoods requires domain expertise and can be subjective, potentially influencing results. On the other hand, Deep Learning does not inherently provide measures of uncertainty in predictions, which can be a limitation in critical applications. A novel field of research is working towards the unification of the two frameworks in the so-called Bayesian Deep Learning [230].

Another critical aspect of computational tools is that, by definition, they are used with computers, and sometimes, the optimisation of computational resources can be as important as any other field-specific research question. In all of the chapters of this thesis, I paid attention to the problem of computational overhead and how to improve it. I have shown in Chapter 3 that graphlet analysis can efficiently represent the topology of large brain networks in just an 11×11 matrix. In Chapter 4, the main computational overhead was given by the several runs of the Wang-Wong model to generate task responses. I made sure to use the most efficient library in Python, Cython, which compiles the code in C, and verified that the number of trials chosen was optimal to avoid running unnecessary simulations. Finally, in Chapter 5, I presented the performance of a decoder on PCA-reduced data, showing that in most cases, the decrease in the R^2 scores can be compensated by the reduced computational overhead and, in one case that it can even be more beneficial to use PCA-reduced data as it is less contaminated with noise due to unrelated neural activity.

Computational tools in neuroscience are instrumental to better understanding how the brain works and have potential applications in improving people's lives. The approaches presented in this thesis, as well as many others, can be used for the diagnosis of neurological conditions or even rehabilitation, regardless of the scale of the model. In Chapter 3, the discussion focused on the multimodal integration of SC and FC or FC from different sources. However, analogous methods can be applied to classify different groups of subjects. There is increasing evidence of the modification of the structure or functional patterns of the brain in neurological disease, or even decades before the onset of the specific condition [60]. In such a framework, a powerful and efficient tool to study the topology of brain networks, like the graphlet

approach, provides hope in developing tools for early diagnosis of neurological conditions, enabling early treatments that could improve people's healthspan and save lives. Effective multimodal integration provides a more powerful basis for potential future diagnostic tools. Future research should be focused on applying existing tools, especially elegant and efficient ones like the graphlet approach, to clinical studies, instead of focusing on just trying to find more powerful computational approaches. This is also important especially for Deep Learning, as the vast majority of publications in the field are increasingly focused on improving predictive performances by often just a small percentage. A similar conclusion can be drawn from the results in Chapter 4. It has been observed that a neurological condition, even if not diagnosed yet, may cause people to perform differently in a task than healthy participants [28]. However, associating between different performances and the early onset of a neurological disease is not straightforward. Parameter recovery may help by identifying parameter changes referred to a biological property. The variation of a biological property is usually a more robust signal of the potential future onset of a neurological condition, thus providing another valuable tool for potential early diagnoses [180]. Future research should be focused on clinical study where parameter recovery can effectively estimates parameters of healthy brain and pathological ones. The number of potential applications is vast given the wide range of pathological conditions. As such, integrating between different fields of research like pathological neuroscience or neurosurgery is imperative.

We should remember that a successful computational tool is not only one that has evident practical application, even if far in the future. Theoretical implications are also significant as they may verify or confute models and other theoretical paradigms, which can help advance the research in the field [138]. For example, the graphlet approach challenged the vision of a complex graph theory-based analysis of networks, which requires several metrics to be computed and interpreted, substituting it with a simple representation in the form of a matrix. This is also relevant to one of the most theoretical and practical issues, which will be discussed in the next session: the unification of different computational tools [72].

6.3 Limitations and future directions

Computational tools have revolutionised every research field. However, modern approaches still require many improvements in different aspects. In this section, I will

discuss the limitations of the tools used in the thesis, presenting potential improvements to guide future research.

In this thesis, I presented the results on a top-bottom scale, presenting whole-brain analysis first down to single-neuron activity recording. While reducing the scale of analysis, I also focused on more task-oriented problems, from the analysis of resting state connectomes with graph theory, passing through a simple 2AFC where the responses are made in fractions of seconds, to more complex 2D reaching task which lasted several minutes. Given the broad range of scales of approaches treated in this thesis, I focused on these specific problems because it is generally more challenging to apply more powerful tools to large-scale problems. However, larger-scale computational approaches can be extended to task-oriented analysis and more complex brain dynamics. For example, in Chapter 3, I cited different works based on the extension of the graphlet approach to more complex problems in network neuroscience, like dynamic networks, which are helpful for task-based analysis. Another promising research field is the application of neural architectures specifically designed to operate on graphs. Graph Neural Networks (GNNs) are increasing their popularity and find several applications in whole-brain analysis. For example, GNNs can be used to simulate brain states and generate data for subsequent analysis [118]. On a more practical aspect, GNNs can be used to characterise and analyse brains affected by neurological diseases [37, 251].

On the other hand, finer-scale approaches like NMM can also be extended to larger-scale problems [45]. However, particular attention should be put to avoid an excessive increase in complexity that could make the tool unable to answer the research questions it was designed for.

Again, for simplicity requirements, given the broad range of scales considered, I focused more on answering research questions with the most straightforward possible tool instead of asking which approach could give the best increment in performance. This is particularly the case of AI approaches. The simple MLP in Chapter 4 or the basic RNN in Chapter 5 are certainly not the optimal models. However, they acted as a good balance between performance and computational efficiency. The field of AI is in constant evolution, with models of surprising performances coming out every month. One notable example is transformers, robust architectures based on attention scores on the input data, which outperform RNNs and their derivation in basically every task performed on sequence data [135].

As I discussed in the previous section, even if ideal in some way, a computational tool cannot be simultaneously powerful and interpretable. The emphasis to be put

on one of the two aspects depends on the research question we want to answer or the potential practical aspect. For example, neuronal decoding is one of the approaches that should be as much interpretable as possible for many reasons. In particular, to understand how the input is processed and transformed in a behavioural output, *i.e.*, which computations are performed by the brain to transform sensory data and previous internal representations, we need a model simple enough to be interpretable but powerful enough to approximate specific aspect of the highly intricate nonlinear nature of the brain. The RNN used in Chapter 5, and deep learning in general, has solid predictive capabilities but lacks interpretability. This is also particularly important for future applications, like BCI, to ensure ethical standards and trustworthiness [31]. While developing more powerful interpretable models is difficult, a new research field focused on developing external tools to interpret black box models like deep learning is quickly developing [193]. Currently, the need for more powerful AI tools is pushing researchers to focus more on the capabilities of an AI system instead of its interpretability. Few large companies have whole teams working on AI safety and explainability. However, AI interpretability should not be separated from any other AI research to ensure the creation of more powerful computational tools natively interpretable.

A broader limitation of current computational tools is that a unification of different approaches needs to be included. Future research efforts should be put into multimodal integration of experimental procedures and tools for implementing specific tools. For example, in Chapter 3 I performed all the analysis with a clear research question in mind: How do topology and other characteristics change across different imaging techniques? A similar approach is essential for neuronal decoding as well to improve decoding performances using complementary techniques like EEG and fNIRS or to enable less invasive but still powerful BCIs. A different kind of unification regards the specific tools to implement computational models and approaches [72]. Recently, new open-source libraries for research in computational neuroscience were released [210, 194]. However, many researchers still prefer to create ad-hoc implementations for their studies. Future effort should be put into making the code of a specific computational tool open and available to the community not only to reproduce the results of the specific study but also to build new knowledge quickly.

While I put a strong emphasis on clinical applications, it is imperative to clearly state that most approaches are still very preliminary to have a strong impact on people lives. However, first results are encouraging and the potential of these approaches to positively influence people's lives is high, making it worth the effort and

resources invested in advancing this field of study. A notable exception is given by brain-computer interfaces. BCI technology is constantly evolving and extraordinary advancements have been made, as discussed extensively in Chapter 5. Two main factors are essential for succesful BCI applications, namely the accuracy of the data acquisition method, and the performances of the decoding algorithms. While improvements should be made on both sides, more effort should be put in the theoretical understanding of the brain encoding mechanisms. This is to ensure future exciting development of BCIs, not only limiting to the important problem of neuroprosthetic, but providing effective rehabilitation therapies or even enhancing human cognitive capabilities.

Bibliography

[1] Larry F Abbott and Frances S Chance. Drivers and modulators from push-pull and balanced synaptic input. *Progress in brain research*, 149:147–155, 2005.

[2] Farras Abdelnour, Michael Dayan, Orrin Devinsky, Thomas Thesen, and Ashish Raj. Functional brain connectivity is predictable from anatomic network's laplacian eigen-structure. *NeuroImage*, 172:728–739, 2018.

[3] Misha B Ahrens, Jennifer M Li, Michael B Orger, Drew N Robson, Alexander F Schier, Florian Engert, and Ruben Portugues. Brain-wide neuronal dynamics during motor adaptation in zebrafish. *Nature*, 485(7399):471–477, 2012.

[4] Hadeel Alharbi et al. Identifying thematics in a brain-computer interface research. *Computational Intelligence and Neuroscience*, 2023, 2023.

[5] Solène Ambrosi, Mathieu Servant, Agnès Blaye, and Boris Burle. Conflict processing in kindergarten children: New evidence from distribution analyses reveals the dynamics of incorrect response activation and suppression. *Journal of experimental child psychology*, 177:36–52, 2019.

[6] Iulia Andras, Elio Mazzone, Fijs WB van Leeuwen, Geert De Naeyer, Matthias N van Oosterom, Sergi Beato, Tessa Buckle, Shane O'Sullivan, Pim J van Leeuwen, Alexander Beulens, et al. Artificial intelligence and robotics: a combination that is changing the operating room. *World journal of urology*, 38:2359–2366, 2020.

[7] Marcel Ausloos. *The logistic map and the route to chaos: From the beginnings to modern applications*. Springer Science & Business Media, 2006.

[8] Laurent Badel, Sandrine Lefort, Thomas K Berger, Carl CH Petersen, Wulfram Gerstner, and Magnus JE Richardson. Extracting non-linear integrate-and-fire models from experimental data using dynamic i–v curves. *Biological cybernetics*, 99:361–370, 2008.

[9] Albert-László Barabási. Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987):20120375, 2013.

[10] Danielle S Bassett and Edward T Bullmore. Human brain networks in health and disease. *Current opinion in neurology*, 22(4):340, 2009.

[11] Stefan Behnel, Robert Bradshaw, Craig Citro, Lisandro Dalcin, Dag Sverre Seljebotn, and Kurt Smith. Cython: The best of both worlds. *Computing in Science & Engineering*, 13(2):31–39, 2011.

[12] Sven Bestmann, Orlando Swayne, Felix Blankenburg, Christian C Ruff, Patrick Haggard, Nikolaus Weiskopf, Oliver Josephs, Jon Driver, John C Rothwell, and Nick S Ward. Dorsal premotor cortex exerts state-dependent causal influences on activity in contralateral primary motor and dorsal premotor cortex. *Cerebral cortex*, 18(6):1281–1291, 2008.

[13] Richard F. Betzel and Danielle S. Bassett. Multi-scale brain networks. *NeuroImage*, 160:73 – 83, 2017. Functional Architecture of the Brain.

[14] Marc Jonathan Blitz and Woodrow Barfield. Memory enhancement and brain–computer interface devices: Technological possibilities and constitutional challenges. In *Policy, Identity, and Neurotechnology: The Neuroethics of Brain-Computer Interfaces*, pages 207–231. Springer, 2023.

[15] Chad E. Bouton, Ammar Shaikhouni, Nicholas V. Annetta, Marcie A. Bockbrader, David A. Friedenberg, Dylan M. Nielson, Gaurav Sharma, Per B. Sederberg, Bradley C. Glenn, W. Jerry Mysiw, et al. Restoring cortical control of functional movement in a human with quadriplegia. *Nature*, 533(7602):247–250, 2016.

[16] Jeffrey S Bowers, Gaurav Malhotra, Federico Adolfi, Marin Dujmović, Milton L Montero, Valerio Biscione, Guillermo Puebla, John H Hummel, and Rachel F Heaton. On the importance of severely testing deep learning models of cognition. *Cognitive Systems Research*, 82:101158, 2023.

[17] Michael Breakspear and Cornelis J Stam. Dynamics of a neural system with a multiscale architecture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457):1051–1074, 2005.

[18] Kenneth H Britten, Michael N Shadlen, William T Newsome, and J Anthony Movshon. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *Journal of Neuroscience*, 12(12):4745–4765, 1992.

[19] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience*, 10(3):186–198, 2009.

[20] Ed Bullmore and Olaf Sporns. The economy of brain network organization. *Nature Reviews Neuroscience*, 13(5):336–349, 2012.

[21] Jerome R Busemeyer and Zheng Wang. Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology*, 44(1):171–189, 2000.

[22] Carter T Butts. Revisiting the foundations of network analysis. *science*, 325(5939):414–416, 2009.

[23] Wenming Cao, Zhiyue Yan, Zhiquan He, and Zhihai He. A comprehensive survey on geometric deep learning. *IEEE Access*, 8:35929–35949, 2020.

[24] Jessica A Cardin, Larry A Palmer, and Diego Contreras. Cellular mechanisms underlying stimulus-dependent gain modulation in primary visual cortex neurons in vivo. *Neuron*, 59(1):150–160, 2008.

[25] James F Cavanagh, Thomas V Wiecki, Michael X Cohen, Carlos M Figueroa, Johan Samanta, Svetlana J Sherman, and Michael J Frank. Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. *Nature Neuroscience*, 14(11):1462–1467, 2011.

[26] John K Chapin and Miguel AL Nicolelis. Principal component analysis of neuronal ensemble activity reveals multidimensional somatosensory representations. *Journal of neuroscience methods*, 94(1):121–140, 1999.

[27] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[28] Marie-Josée Chouinard and Isabelle Rouleau. The 48-pictures test: A two-alternative forced-choice recognition test for the detection of malingering. *Journal of the International Neuropsychological Society*, 3(6):545–552, 1997.

[29] Mark MM Churchland, John P Cunningham, Matthew T Kaufman, Stephen I Ryu, and Krishna V Shenoy. Neural population dynamics during reaching. *Nature*, 487(7405):51–56, 2012.

[30] Marlene R Cohen and Adam Kohn. Measuring and interpreting neuronal correlations. *Nature Neuroscience*, 14(7):811–819, 2011.

[31] Allen Coin, Megan Mulder, and Veljko Dubljević. Ethical aspects of bci technology: what is the state of the art? *Philosophies*, 5(4):31, 2020.

[32] Giles L Colclough, Matthew J Brookes, Stephen M Smith, and Mark W Woolrich. A symmetric multivariate leakage correction for meg connectomes. *Neuroimage*, 117:439–448, 2015.

[33] Giles L Colclough, Mark W Woolrich, PK Tewarie, Matthew J Brookes, Andrew J Quinn, and Stephen M Smith. How reliable are meg resting-state connectivity metrics? *Neuroimage*, 138:284–293, 2016.

[34] Marc-Alexandre Côté, Eleftherios Garyfallidis, Hugo Larochelle, and Maxime Descoteaux. Cleaning up the mess: tractography outlier removal using hierarchical quickbundles clustering. In *23rd ISMRM annual meeting. Toronto, Canada*, 2015.

[35] Robert W Cox. Afni: what a long strange trip it's been. *Neuroimage*, 62(2):743–747, 2012.

[36] Nicolas A Crossley, Andrea Mechelli, Petra E Vértes, Toby T Winton-Brown, Ameera X Patel, Cedric E Ginestet, Philip McGuire, and Edward T Bullmore. Cognitive relevance of the community structure of the human brain functional coactivation network. *Proceedings of the National Academy of Sciences*, 110(28):11583–11588, 2013.

[37] Hejie Cui, Wei Dai, Yanqiao Zhu, Xiaoxiao Li, Lifang He, and Carl Yang. Brainnnexplainer: An interpretable graph neural network framework for brain network based disease analysis. *arXiv preprint arXiv:2107.05097*, 2021.

[38] John P Cunningham and Byron M Yu. Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, 17(11):1500–1509, 2014.

[39] Anders M Dale, Arthur K Liu, Bruce R Fischl, Randy L Buckner, John W Belliveau, Jeffrey D Lewine, and Eric Halgren. Dynamic statistical parametric mapping: combining fmri and meg for high-resolution imaging of cortical activity. *Neuron*, 26(1):55–67, 2000.

[40] Olivier David and Karl J Friston. A neural mass model for meg/eeg:: coupling and neuronal dynamics. *NeuroImage*, 20(3):1743–1755, 2003.

[41] Peter Dayan, Laurence F Abbott, et al. Theoretical neuroscience: computational and mathematical modeling of neural systems. *Journal of Cognitive Neuroscience*, 15(1):154–155, 2003.

[42] Siemon de Lange, Marcel de Reus, and Martijn Van Den Heuvel. The laplacian spectrum of neural networks. *Frontiers in computational neuroscience*, 7:189, 2014.

[43] Gustavo Deco, Viktor Jirsa, Anthony R McIntosh, Olaf Sporns, and Rolf Kötter. Key role of coupling, delay, and noise in resting brain fluctuations. *Proceedings of the National Academy of Sciences*, 106(25):10302–10307, 2009.

[44] Gustavo Deco, Viktor K Jirsa, Peter A Robinson, Michael Breakspear, and Karl Friston. The dynamic brain: from spiking neurons to neural masses and cortical fields. *PLoS computational biology*, 4(8):e1000092, 2008.

[45] Gustavo Deco, Adrián Ponce-Alvarez, Patric Hagmann, Gian Luca Romani, Dante Mantini, and Maurizio Corbetta. How local excitation–inhibition ratio impacts the whole brain dynamics. *Journal of Neuroscience*, 34(23):7886–7898, 2014.

[46] Bo Deng. Alternative models to hodgkin–huxley equations. *Bulletin of Mathematical Biology*, 79:1390–1411, 2017.

[47] Thijs Dhollander, David Raffelt, and Alan Connelly. Unsupervised 3-tissue response function estimation from single-shell or multi-shell diffusion mr data without a co-registered t1 image. In *ISMRM Workshop on Breaking the Barriers of Diffusion MRI*, volume 5, page 5, 2016.

[48] Bradley B Doll, Dylan A Simon, and Nathaniel D Daw. The ubiquity of model-based reinforcement learning. *Current opinion in neurobiology*, 22(6):1075–1081, 2012.

[49] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[50] David A Drachman. Do we have brain to spare? *Neurology*, 64(12):2004–2005, 2005.

[51] Russell C Eberhart, Yuhui Shi, and James Kennedy. *Swarm intelligence*. Elsevier, 2001.

[52] Victor M Eguiluz, Dante R Chialvo, Guillermo A Cecchi, Marwan Baliki, and A Vania Apkarian. Scale-free brain functional networks. *Physical review letters*, 94(1):018102, 2005.

[53] Absalom E Ezugwu, Abiodun M Ikotun, Olaide O Oyelade, Laith Abualigah, Jeffery O Agushaka, Christopher I Eke, and Andronicus A Akinyelu. A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. *Engineering Applications of Artificial Intelligence*, 110:104743, 2022.

[54] Emily B Falk and Danielle S Bassett. Brain and social networks: fundamental building blocks of human experience. *Trends in cognitive sciences*, 21(9):674–690, 2017.

[55] Qianqian Fang and David A Boas. Tetrahedral mesh generation from volumetric binary and grayscale images. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1142–1145. Ieee, 2009.

[56] Holger Finger, Marlene Bönstrup, Bastian Cheng, Arnaud Messé, Claus Hilgetag, Götz Thomalla, Christian Gerloff, and Peter König. Modeling of large-scale functional brain networks based on structural connectivity from dti: comparison with eeg derived phase coupling networks and evaluation of alternative methods along the modeling path. *PLoS computational biology*, 12(8):e1005025, 2016.

[57] Bruce Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.

[58] Bruce Fischl, Martin I Sereno, and Anders M Dale. Cortical surface-based analysis: Ii: inflation, flattening, and a surface-based coordinate system. *Neuroimage*, 9(2):195–207, 1999.

[59] Richard FitzHugh. Impulses and physiological states in theoretical models of nerve membrane. *Biophysical journal*, 1(6):445–466, 1961.

[60] Alex Fornito and Edward T Bullmore. Connectomics: a new paradigm for understanding brain disease. *European Neuropsychopharmacology*, 25(5):733–748, 2015.

[61] Yves Frégnac and Gilles Laurent. Neuroscience: Where is the brain in the human brain project? *Nature*, 513(7516):27–29, 2014.

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

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

[64] Karl J Friston, Lee Harrison, and Will Penny. Dynamic causal modelling. *Neuroimage*, 19(4):1273–1302, 2003.

[65] Manfred Fuchs, Michael Wagner, Thomas Köhler, and Hans-Aloys Wischmann. Linear and nonlinear current density reconstructions. *Journal of clinical Neurophysiology*, 16(3):267–295, 1999.

[66] Qing Gao, Huafu Chen, and Qiyong Gong. Evaluation of the effective connectivity of the dominant primary motor cortex during bimanual movement using granger causality. *Neuroscience letters*, 443(1):1–6, 2008.

[67] Wulfram Gerstner, Werner M Kistler, Richard Naud, and Liam Paninski. *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press, 2014.

[68] Wulfram Gerstner, Henning Sprekeler, and Gustavo Deco. Theory and simulation in neuroscience. *science*, 338(6103):60–65, 2012.

[69] Joshua I Glaser, Ari S Benjamin, Raeed H Chowdhury, Matthew G Perich, Lee E Miller, and Konrad P Kording. Machine learning for neural decoding. *Eneuro*, 7(4), 2020.

[70] Matthew F Glasser, Timothy S Coalson, Emma C Robinson, Carl D Hacker, John Harwell, Essa Yacoub, Kamil Ugurbil, Jesper Andersson, Christian F Beckmann, Mark Jenkinson, et al. A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171–178, 2016.

[71] Matthew F Glasser, Stamatios N Sotiropoulos, J Anthony Wilson, Timothy S Coalson, Bruce Fischl, Jesper L Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R Polimeni, et al. The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, 80:105–124, 2013.

[72] Nigel H Goddard, Michael Hucka, Fred Howell, Hugo Cornelis, Kavita Shankar, and David Beeman. Towards neuroml: model description methods for collaborative modelling in neuroscience. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 356(1412):1209–1228, 2001.

[73] Joshua I Gold and Michael N Shadlen. The neural basis of decision making. *Annu. Rev. Neurosci.*, 30:535–574, 2007.

[74] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[75] David S Grayson, Siddharth Ray, Samuel Carpenter, Swathi Iyer, Taciana G Costa Dias, Corinne Stevens, Joel T Nigg, and Damien A Fair. Structural and functional rich club organization of the brain in children and adults. *PloS one*, 9(2):e88297, 2014.

[76] Douglas N Greve and Bruce Fischl. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, 48(1):63–72, 2009.

[77] John S Griffith. On the stability of brain-like structures. *Biophysical journal*, 3(4):299–308, 1963.

[78] Christoph Guger, Werner Harkam, Carin Hertnaes, and Gert Pfurtscheller. Prosthetic control by an eeg-based brain-computer interface (bci). In *Proc. aaate 5th european conference for the advancement of assistive technology*, pages 3–6. Citeseer, 1999.

[79] Roger Guimera and Luís A Nunes Amaral. Functional cartography of complex metabolic networks. *nature*, 433(7028):895–900, 2005.

[80] U Halsband and H-J Freund. Premotor cortex and conditional motor learning in man. *Brain*, 113(1):207–222, 1990.

[81] Ulrike Halsband and Regine K Lange. Motor learning in man: a review of functional and clinical studies. *Journal of Physiology-Paris*, 99(4-6):414–424, 2006.

[82] Matti S Hämäläinen and Risto J Ilmoniemi. Interpreting magnetic fields of the brain: minimum norm estimates. *Medical & biological engineering & computing*, 32:35–42, 1994.

[83] William L Hamilton. Graph representation learning. *Synthesis Lectures on Artifical Intelligence and Machine Learning*, 14(3):1–159, 2020.

[84] Zulfi Haneef and Sharon Chiang. Clinical correlates of graph theory findings in temporal lobe epilepsy. *Seizure*, 23(10):809–818, 2014.

[85] Demis Hassabis, Dharshan Kumaran, Christopher Summerfield, and Matthew Botvinick. Neuroscience-inspired artificial intelligence. *Neuron*, 95(2):245–258, 2017.

[86] Nicholas Hatsopoulos, Jignesh Joshi, and John G O'Leary. Decoding continuous and discrete motor behaviors using motor and premotor cortical ensembles. *Journal of neurophysiology*, 92(2):1165–1174, 2004.

[87] Christopher K. Hauser and Emilio Salinas. *Perceptual Decision Making*, pages 1–21. Springer New York, New York, NY, 2013.

[88] Bin He and Zhongming Liu. Multimodal functional neuroimaging: integrating functional mri and eeg/meg. *IEEE reviews in biomedical engineering*, 1:23–40, 2008.

[89] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[90] Hauke R Heekeren, Sean Marrett, Peter A Bandettini, and Leslie G Ungerleider. A general mechanism for perceptual decision-making in the human brain. *Nature*, 431(7010):859–862, 2004.

[91] Richard P Heitz and Jeffrey D Schall. Neural mechanisms of speed-accuracy tradeoff. *Neuron*, 76(3):616–628, 2012.

[92] Arjan Hillebrand and Gareth R Barnes. The use of anatomical constraints with meg beamformers. *Neuroimage*, 20(4):2302–2313, 2003.

[93] Arjan Hillebrand, Gareth R Barnes, Johannes L Bosboom, Henk W Berendse, and Cornelis J Stam. Frequency-dependent functional connectivity within resting-state networks: an atlas-based meg beamformer solution. *Neuroimage*, 59(4):3909–3921, 2012.

[94] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[95] Alan L Hodgkin and Andrew F Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500, 1952.

[96] Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[97] Christopher J. Honey and et al. Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences*, 106(6):2035–2040, 2009.

[98] Ralf Hornig, Marcus Dapper, Eric Le Joliff, Robert Hill, Khalid Ishaque, Christoph Posch, Ryad Benosman, Yannick LeMer, José-Alain Sahel, and Serge Picaud. Pixium vision: first clinical results and innovative developments. *Artificial Vision: A Clinical Guide*, pages 99–113, 2017.

[99] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.

[100] Ronald Hübner and Thomas Pelzer. Improving parameter recovery for conflict drift-diffusion models. *Behavior Research Methods*, 52:1848–1866, 2020.

[101] Alexander C Huk and Michael N Shadlen. Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *Journal of Neuroscience*, 25(45):10420–10436, 2005.

[102] Yuriy Hulovatyy, Huili Chen, and Tijana Milenković. Exploring the structure and function of temporal networks with dynamic graphlets. *Bioinformatics*, 31(12):i171–i180, 2015.

[103] Mark D Humphries. Dynamical networks: Finding, measuring, and tracking neural population activity using network science. *Network Neuroscience*, 1(4):324–338, 2017.

[104] Julia M Huntenburg, Pierre-Louis Bazin, and Daniel S Margulies. Large-scale gradients in human cortical organization. *Trends in cognitive sciences*, 22(1):21–31, 2018.

[105] A Hyvarinen and E Oja. Independent component analysis: A tutorial. notes for international joint conference on neural networks (ijcnn'99), washington dc, 1999.

[106] Md Rafiqul Islam, Xiaoxia Yin, Anwaar Ulhaq, Yanchun Zhang, Hua Wang, Noreen Anjum, and Thomas Kron. A survey of graph based complex brain network analysis using functional and diffusional mri. *American Journal of Applied Sciences*, 14(12):1186–1208, 2018.

[107] Eugene M Izhikevich. Simple model of spiking neurons. *IEEE Transactions on neural networks*, 14(6):1569–1572, 2003.

[108] Eugene M Izhikevich. *Dynamical systems in neuroscience*. MIT press, 2007.

[109] Ravichander Janapati, Vishwas Dalal, and Rakesh Sengupta. Advances in modern eeg-bci signal processing: A review. *Materials Today: Proceedings*, 80:2563–2566, 2023.

[110] Mark Jenkinson, Christian F Beckmann, Timothy EJ Behrens, Mark W Woolrich, and Stephen M Smith. Fsl. *Neuroimage*, 62(2):782–790, 2012.

[111] Ben Jeurissen, Jacques-Donald Tournier, Thijs Dhollander, Alan Connelly, and Jan Sijbers. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion mri data. *NeuroImage*, 103:411–426, 2014.

[112] David J Johnson, Christopher J Hopwood, Joseph Cesario, and Timothy J Pleskac. Advancing research on cognitive processes in social and personality psychology: A hierarchical drift diffusion model primer. *Social Psychological and Personality Science*, 8(4):413–423, 2017.

[113] Marcia K Johnson, Carol L Raye, Karen J Mitchell, Erich J Greene, William A Cunningham, and Charles A Sanislow. Using fmri to investigate. *Cognitive, Affective, & Behavioral Neuroscience*, 5(3):339–361, 2005.

[114] Mark H Johnson, Richard Griffin, Hanife Halit Gergely Csibra, Teresa Farroni, Michelle De Haan, Leslie A Tucker, Simon Baron-Cohen, and John Richards. The emergence of the social brain network: Evidence from typical and atypical development. *Development and psychopathology*, 17(3):599, 2005.

[115] Ian T. Jolliffe. Principal component analysis and factor analysis. *Principal Component Analysis*, 2(2):115–128, 1986.

[116] Ian T. Jolliffe. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 359(1787):1–19, 2002.

[117] Awais M. Kamboh and Andrew J. Mason. Computationally efficient neural feature extraction for spike sorting in implantable high-density recording systems. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 21(1):1–9, 2013.

[118] Xuan Kan, Hejie Cui, Joshua Lukemire, Ying Guo, and Carl Yang. Fbnetgen: Task-aware gnn-based fmri analysis via functional brain network generation. In *International Conference on Medical Imaging with Deep Learning*, pages 618–637. PMLR, 2022.

[119] G Kashyap, D Bapat, D Das, R Gowaikar, RE Amritkar, G Rangarajan, V Ravindranath, and G Ambika. Synapse loss and progress of alzheimer's disease-a network model. *Scientific Reports*, 9(1):1–9, 2019.

[120] Elias Kellner, Bibek Dhital, Valerij G Kiselev, and Marco Reisert. Gibbs-ringing artifact removal based on local subvoxel-shifts. *Magnetic resonance in medicine*, 76(5):1574–1581, 2016.

[121] Muhammad Ahmed Khan, Rig Das, Helle K Iversen, and Sadasivan Puthusserypady. Review on motor imagery based bci systems for upper limb post-stroke neurorehabilitation: From designing to application. *Computers in biology and medicine*, 123:103843, 2020.

[122] Dae-Jin Kim and Byoung-Kyong Min. Rich-club in the brain's macrostructure: Insights from graph theoretical analysis. *Computational and Structural Biotechnology Journal*, 18:1761–1773, 2020.

[123] Fabian Kloosterman, Stuart P Layton, Zhe Chen, and Matthew A Wilson. Bayesian decoding using unsorted spikes in the rat hippocampus. *Journal of neurophysiology*, 111(1):217–227, 2014.

[124] Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. Siamese neural networks for one-shot image recognition. *ICML Deep Learning Workshop*, 2, 2015.

[125] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[126] Dominik Krzemiński and Jiaxiang Zhang. Imperfect integration: Congruency between multiple sensory sources modulates decision-making processes. *Attention, Perception, & Psychophysics*, pages 1–17, 2022.

[127] Brenden M Lake, Tomer D Ullman, Joshua B Tenenbaum, and Samuel J Gershman. Building machines that learn and think like people. *Behavioral and brain sciences*, 40:e253, 2017.

[128] Helmut Laufs and et al. Electroencephalographic signatures of attentional and cognitive default modes in spontaneous brain activity fluctuations at rest. *Proceedings of the National Academy of Sciences*, 100(19):11053–11058, 2003.

[129] Patrick N Lawlor, Matthew G Perich, Lee E Miller, and Konrad P Kording. Linear-nonlinear-time-warp-poisson models of neural activity. *Journal of computational neuroscience*, 45:173–191, 2018.

[130] Victoria M Lawlor, Christian A Webb, Thomas V Wiecki, Michael J Frank, Madhukar Trivedi, Diego A Pizzagalli, and Daniel G Dillon. Dissecting the impact of depression on decision-making. *Psychological medicine*, 50(10):1613–1622, 2020.

[131] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[132] Dian Lei, Xiaoxiao Chen, and Jianfei Zhao. Opening the black box of deep learning. *arXiv preprint arXiv:1805.08355*, 2018.

[133] Daniel Levenstein, Veronica A Alvarez, Asohan Amarasingham, Habiba Azab, Zhe S Chen, Richard C Gerkin, Andrea Hasenstaub, Ramakrishnan Iyer, Renaud B Jolivet, Sarah Marzen, et al. On the role of theory and modeling in neuroscience. *Journal of Neuroscience*, 43(7):1074–1088, 2023.

[134] Xu-Hong Liao, Ming-Rui Xia, Ting Xu, Zheng-Jia Dai, Xiao-Yan Cao, Hai-Jing Niu, Xi-Nian Zuo, Yu-Feng Zang, and Yong He. Functional brain hubs and their test–retest reliability: a multiband resting-state functional mri study. *Neuroimage*, 83:969–982, 2013.

[135] Tianyang Lin, Yuxin Wang, Xiangyang Liu, and Xipeng Qiu. A survey of transformers. *AI Open*, 2022.

[136] Quanying Liu, Marco Ganzetti, Nicole Wenderoth, and Dante Mantini. Detecting large-scale brain networks using eeg: impact of electrode density, head modeling and source localization. *Frontiers in neuroinformatics*, 12:4, 2018.

[137] Gabriele Lohmann, Kerstin Erfurth, Karsten Müller, and Robert Turner. Critical comments on dynamic causal modelling. *Neuroimage*, 59(3):2322–2329, 2012.

[138] Bradley C Love. Levels of biological plausibility. *Philosophical Transactions of the Royal Society B*, 376(1815):20190632, 2021.

[139] Shiyu Luo, Qinwan Rabbani, and Nathan E Crone. Brain-computer interface: applications to speech decoding and synthesis to augment communication. *Neurotherapeutics*, 19(1):263–273, 2022.

[140] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.

[141] Ravikiran Mane, Tushar Chouhan, and Cuntai Guan. Bci for stroke rehabilitation: motor and beyond. *Journal of neural engineering*, 17(4):041001, 2020.

[142] Henry Markram, Joachim Lübke, Michael Frotscher, and Bert Sakmann. Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. *Science*, 275(5297):213–215, 1997.

[143] Henry Markram, Karlheinz Meier, Thomas Lippert, Sten Grillner, Richard Frackowiak, Stanislas Dehaene, Alois Knoll, Haim Sompolinsky, Kris Verstreken, Javier DeFelipe, et al. Introducing the human brain project. *Procedia Computer Science*, 7:39–42, 2011.

[144] David Marr. *Vision: A computational investigation into the human representation and processing of visual information.* W.H. Freeman, New York, NY, 1983.

[145] David Marshall, Damien Coyle, Shane Wilson, and Michael Callaghan. Games, gameplay, and bci: the state of the art. *IEEE Transactions on Computational Intelligence and AI in Games*, 5(2):82–99, 2013.

[146] Carrie J McAdams and John HR Maunsell. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area v4. *Journal of Neuroscience*, 19(1):431–441, 1999.

[147] Dennis J McFarland, Janis Daly, Chadwick Boulay, and Muhammad A Parvaz. Therapeutic applications of bci technologies. *Brain-Computer Interfaces*, 4(1-2):37–52, 2017.

[148] Vinod Menon. Large-scale brain networks and psychopathology: a unifying triple network model. *Trends in cognitive sciences*, 15(10):483–506, 2011.

[149] Tijana Milenković and Nataša Pržulj. Uncovering biological network function via graphlet degree signatures. *Cancer informatics*, 6:CIN–S680, 2008.

[150] Steven Miletić, Brandon M Turner, Birte U Forstmann, and Leendert van Maanen. Parameter recovery for the leaky competing accumulator model. *Journal of Mathematical Psychology*, 76:25–50, 2017.

[151] Marvin L Minsky and Seymour A Papert. Perceptrons: expanded edition, 1988.

[152] Jonas Močkus. On bayesian methods for seeking the extremum. In *Optimization Techniques IFIP Technical Conference: Novosibirsk, July 1–7, 1974*, pages 400–404. Springer, 1975.

[153] Catherine Morris and Harold Lecar. Voltage oscillations in the barnacle giant muscle fiber. *Biophysical journal*, 35(1):193–213, 1981.

[154] Malaak Nasser Moussa, Crystal D Vechlekar, Jonathan H Burdette, Matt R Steen, Christina E Hugenschmidt, and Paul J Laurienti. Changes in cognitive state alter human functional brain networks. *Frontiers in human neuroscience*, 5:83, 2011.

[155] Sarah F. Muldoon and et al. Stimulation-based control of dynamic brain networks. *Nature Communications*, 7:12408, 2016.

[156] Sarah Feldt Muldoon and Danielle S. Bassett. Network and multilayer network approaches to understanding human brain dynamics. *Philosophy of Science*, 83(5):710–720, 2016.

[157] Peter R Murphy, Joachim Vandekerckhove, and Sander Nieuwenhuis. Pupil-linked arousal determines variability in perceptual decision making. *PLoS computational biology*, 10(9):e1003854, 2014.

[158] Carl J Nelson and Stephen Bonner. Neuronal graphs: A graph theory primer for microscopic, functional networks of neurons recorded by calcium imaging. *Frontiers in Neural Circuits*, 15:662882, 2021.

[159] Mark Newman. *Networks: An Introduction*. Oxford university press, 2009.

[160] Martin Niethammer, Chris C Tang, Andrew Feigin, Patricia J Allen, Lisette Heinen, Sabine Hellwig, Florian Amtage, Era Hanspal, Jean Paul Vonsattel, Kathleen L Poston, et al. A disease-specific metabolic brain network associated with corticobasal degeneration. *Brain*, 137(11):3036–3046, 2014.

[161] Guido Nolte. The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Physics in Medicine & Biology*, 48(22):3637, 2003.

[162] Catherine L Ojakangas, Ammar Shaikhouni, Gerhard M Friehs, Abraham H Caplan, Mijail D Serruya, Maryam Saleh, Daniel S Morris, and John P Donoghue. Decoding movement intent from human premotor cortex neurons for neural prosthetic applications. *Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society*, 23(6):577, 2006.

[163] Piotr Olejniczak. Neurophysiologic basis of eeg. *Journal of clinical neurophysiology*, 23(3):186–189, 2006.

[164] Salima Omar, Asri Ngadi, and Hamid H Jebur. Machine learning techniques for anomaly detection: an overview. *International Journal of Computer Applications*, 79(2), 2013.

[165] Robert Oostenveld, Pascal Fries, Eric Maris, and Jan-Mathijs Schoffelen. Fieldtrip: open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Computational intelligence and neuroscience*, 2011, 2011.

[166] TB OpenAI. Chatgpt: Optimizing language models for dialogue. *OpenAI*, 2022.

[167] Satu Palva and J. Matias Palva. Discovering oscillatory interaction networks with m/eeg: challenges and breakthroughs. *Trends in Cognitive Sciences*, 16(4):219 – 230, 2012.

[168] Chethan Pandarinath, Daniel J O'Shea, Jenna Collins, Rafal Jozefowicz, Sergey D Stavisky, Jonathan C Kao, Stephen I Ryu, Leigh R Hochberg, Jaimie M Henderson, Krishna V Shenoy, et al. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature Methods*, 15(10):805–815, 2018.

[169] Hae-Jeong Park and Karl Friston. Structural and functional brain networks: from connections to cognition. *Science*, 342(6158):1238411, 2013.

[170] Roberto D Pascual-Marqui. Discrete, 3d distributed, linear imaging methods of electric neuronal activity. part 1: exact, zero error localization. *arXiv preprint arXiv:0710.3341*, 2007.

[171] Roberto D Pascual-Marqui. Theory of the eeg inverse problem. *Quantitative EEG analysis: methods and clinical applications*, pages 121–140, 2009.

[172] Roberto D Pascual-Marqui, Dietrich Lehmann, Martha Koukkou, Kieko Kochi, Peter Anderer, Bernd Saletu, Hideaki Tanaka, Koichi Hirata, E Roy John, Leslie Prichep, et al. Assessing interactions in the brain with exact low-resolution electromagnetic tomography. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369(1952):3768–3784, 2011.

[173] Roberto Domingo Pascual-Marqui et al. Standardized low-resolution brain electromagnetic tomography (sloreta): technical details. *Methods Find Exp Clin Pharmacol*, 24(Suppl D):5–12, 2002.

[174] Karl Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(11):559–572, 1901.

[175] William Penny. Bayesian models of brain and behaviour. *International Scholarly Research Notices*, 2012, 2012.

[176] Pietro Perconti and Alessio Plebe. Deep learning and cognitive science. *Cognition*, 203:104365, 2020.

[177] Matthew G Perich, Patrick N Lawlor, Konrad P Kording, and Lee E Miller. Extracellular neural recordings from macaque primary and dorsal premotor motor cortex during a sequential reaching task. *CRCNS. org*, 2018.

[178] Michael I Posner, Mary K Rothbart, and Pascale Voelker. Developing brain networks of attention. *Current opinion in pediatrics*, 28(6):720, 2016.

[179] Jonathan D Power, Bradley L Schlaggar, Christina N Lessov-Schlaggar, and Steven E Petersen. Evidence for hubs in human functional brain networks. *Neuron*, 79(4):798–813, 2013.

[180] J Pozueta, R Lefort, and ML Shelanski. Synaptic changes in alzheimer's disease and its models. *Neuroscience*, 251:51–65, 2013.

[181] Natasa Pržulj, Derek G Corneil, and Igor Jurisica. Modeling interactome: scale-free or geometric? *Bioinformatics*, 20(18):3508–3515, 2004.

[182] Nataša Pržulj. Biological network comparison using graphlet degree distribution. *Bioinformatics*, 23:e177–e183, 2007.

[183] Farshad Rafiei and Dobromir Rahnev. Rtnet: A neural network that exhibits the signatures of human perceptual decision making. *bioRxiv*, pages 2022–08, 2022.

[184] Roger Ratcliff, Anjali Thapar, and Gail McKoon. Aging, practice, and perceptual tasks: A diffusion model analysis. *Psychology and Aging*, 26(3):678–689, 2011.

[185] A Renart, N Brunel, and XJ Wang. Mean-field theory of recurrent cortical networks: from irregularly spiking neurons to working memory. *Computational neuroscience: A comprehensive approach*, pages 431–490, 2003.

[186] Neethu Robinson, Ravikiran Mane, Tushar Chouhan, and Cuntai Guan. Emerging trends in bci-robotics for motor control and rehabilitation. *Current Opinion in Biomedical Engineering*, 20:100354, 2021.

[187] Jamie D Roitman and Michael N Shadlen. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of neuroscience*, 22(21):9475–9489, 2002.

[188] Edmund T Rolls and Alessandro Treves. The neuronal encoding of information in the brain. *Progress in neurobiology*, 95(3):448–490, 2011.

[189] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

[190] Benjamin Rosin, Milos Slovik, Raphy Mitelman, Michal Rivlin-Etzion, Simon N Haber, Zvi Israel, Eilon Vaadia, and Hagai Bergman. Closed-loop deep brain stimulation is superior in ameliorating parkinsonism. *Neuron*, 72(2):370–384, 2011.

[191] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

[192] Paul Sajda, Marios G Philiastides, and Lucas C Parra. Single-trial analysis of neuroimaging data: inferring neural networks underlying perceptual decision-making in the human brain. *IEEE reviews in biomedical engineering*, 2:97–109, 2009.

[193] Wojciech Samek, Thomas Wiegand, and Klaus-Robert Müller. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *IEEE Signal Processing Magazine*, 35(6):20–30, 2017.

[194] Paula Sanz Leon, Stuart A Knock, M Marmaduke Woodman, Lia Domide, Jochen Mersmann, Anthony R McIntosh, and Viktor Jirsa. The virtual brain: a simulator of primate brain network dynamics. *Frontiers in neuroinformatics*, 7:10, 2013.

[195] Anida Sarajlić, Noël Malod-Dognin, Ömer Nebil Yaveroğlu, and Nataša Pržulj. Graphlet-based characterization of directed networks. *Scientific reports*, 6(1):1–14, 2016.

[196] Michael Schirner, Simon Rothmeier, Viktor K Jirsa, Anthony Randal McIntosh, and Petra Ritter. An automated pipeline for constructing personalized virtual brains from multimodal neuroimaging data. *NeuroImage*, 117:343–357, 2015.

[197] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[198] Sean L Simpson, Satoru Hayasaka, and Paul J Laurienti. Exponential random graph modeling for complex brain networks. *PloS one*, 6(5):e20039, 2011.

[199] Sean L Simpson, Malaak N Moussa, and Paul J Laurienti. An exponential random graph modeling approach to creating group-based representative whole-brain connectivity networks. *Neuroimage*, 60(2):1117–1126, 2012.

[200] Ranganatha Sitaram, Tomas Ros, Luke Stoeckel, Sven Haller, Frank Scharnowski, Jarrod Lewis-Peacock, Nikolaus Weiskopf, Maria Laura Blefari, Mohit Rana, Ethan Oblak, et al. Closed-loop brain training: the science of neurofeedback. *Nature Reviews Neuroscience*, 18(2):86–100, 2017.

[201] Juan Song, Jian Zheng, Ping Li, Xiaoyuan Lu, Guangming Zhu, and Peiyi Shen. An effective multimodal image fusion method using mri and pet for alzheimer's disease diagnosis. *Frontiers in digital health*, 3:637386, 2021.

[202] Zuoting Song, Tao Fang, Jing Ma, Yuan Zhang, Song Le, Gege Zhan, Xueze Zhang, Shouyan Wang, Hui Li, Yifang Lin, et al. Evaluation and diagnosis of brain diseases based on non-invasive bci. In *2021 9th International Winter Conference on Brain-Computer Interface (BCI)*, pages 1–6. IEEE, 2021.

[203] Tananun Songdechakraiwut, Li Shen, and Moo Chung. Topological learning and its application to multimodal brain network integration. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24*, pages 166–176. Springer, 2021.

[204] Hossein Azari Soufiani and Edo Airoldi. Graphlet decomposition of a weighted network. In *Artificial Intelligence and Statistics*, pages 54–63. PMLR, 2012.

[205] Olaf Sporns. Network attributes for segregation and integration in the human brain. *Current opinion in neurobiology*, 23(2):162–171, 2013.

[206] Olaf Sporns, Joseph A Gally, GEORGE N Reeke Jr, and Gerald M Edelman. Reentrant signaling among simulated neuronal groups leads to coherency in their oscillatory activity. *Proceedings of the National Academy of Sciences*, 86(18):7265–7269, 1989.

[207] Olaf Sporns and Rolf Kötter. Motifs in brain networks. *PLoS Biol*, 2(11):e369, 2004.

[208] Dominic Standage, Da-Hui Wang, and Gunnar Blohm. Neural dynamics implement a flexible decision bound with a fixed firing rate for choice: a model-based hypothesis. *Frontiers in neuroscience*, 8:318, 2014.

[209] Jeffrey J Starns and Roger Ratcliff. The effects of aging on the speed–accuracy compromise: Boundary optimality in the diffusion model. *Psychology and aging*, 25(2):377, 2010.

[210] Marcel Stimberg, Romain Brette, and Dan FM Goodman. Brian 2, an intuitive and efficient neural simulator. *elife*, 8:e47314, 2019.

[211] Steven H Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. CRC press, 2018.

[212] Richard S. Sutton and Andrew G. Barto. Introduction to reinforcement learning. *Adaptive computation and machine learning*, 135, 1998.

[213] François Tadel, Sylvain Baillet, John C Mosher, Dimitrios Pantazis, and Richard M Leahy. Brainstorm: a user-friendly application for meg/eeg analysis. *Computational intelligence and neuroscience*, 2011, 2011.

[214] Luke Tait, Ayşegül Özkan, Maciej J Szul, and Jiaxiang Zhang. Cortical source imaging of resting-state meg with a high resolution atlas: An evaluation of methods. *BioRxiv*, 2020.

[215] DG Taylor and MC Bushell. The spatial mapping of translational diffusion coefficients by the nmr imaging technique. *Physics in medicine & biology*, 30(4):345, 1985.

[216] David Thura, Jean-François Cabana, Albert Feghaly, and Paul Cisek. Unified neural dynamics of decisions and actions in the cerebral cortex and basal ganglia. *BioRxiv*, pages 2020–10, 2020.

[217] Dardo Tomasi and Nora D Volkow. Aging and functional brain networks. *Molecular psychiatry*, 17(5):549–558, 2012.

[218] Giulio Tononi, Olaf Sporns, and Gerald M Edelman. A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences*, 91(11):5033–5037, 1994.

[219] J Donald Tournier, Fernando Calamante, and Alan Connelly. Improved probabilistic streamlines tractography by 2nd order integration over fibre orientation distributions. In *Proceedings of the international society for magnetic resonance in medicine*, volume 1670. Ismrm, 2010.

[220] J-Donald Tournier, Robert Smith, David Raffelt, Rami Tabbara, Thijs Dhollander, Maximilian Pietsch, Daan Christiaens, Ben Jeurissen, Chun-Hung Yeh, and Alan Connelly. Mrtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage*, 202:116137, 2019.

[221] Emma K Towlson, Petra E Vértes, Sebastian E Ahnert, William R Schafer, and Edward T Bullmore. The rich club of the c. elegans neuronal connectome. *Journal of Neuroscience*, 33(15):6380–6387, 2013.

[222] Martijn P Van Den Heuvel and Olaf Sporns. Rich-club organization of the human connectome. *Journal of Neuroscience*, 31(44):15775–15786, 2011.

[223] Martijn P van den Heuvel and Olaf Sporns. Network hubs in the human brain. *Trends in cognitive sciences*, 17(12):683–696, 2013.

[224] Jochem van Kempen, Gerard M Loughnane, Daniel P Newman, Simon P Kelly, Alexander Thiele, Redmond G O'Connell, and Mark A Bellgrove. Behavioural and neural signatures of perceptual decision-making are modulated by pupil-linked arousal. *Elife*, 8:e42541, 2019.

[225] Barry D Van Veen, Wim Van Drongelen, Moshe Yuchtman, and Akifumi Suzuki. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on biomedical engineering*, 44(9):867–880, 1997.

[226] Iris Vilares and Konrad Kording. Bayesian models: the structure of the world, uncertainty, behavior, and the brain. *Annals of the New York Academy of Sciences*, 1224(1):22–39, 2011.

[227] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, and et al. Matching networks for one shot learning. In *Advances in Neural Information Processing Systems (NIPS)*, 2016.

[228] Sjoerd B Vos, Chantal MW Tax, Peter R Luijten, Sebastien Ourselin, Alexander Leemans, and Martijn Froeling. The importance of correcting for signal drift in diffusion mri. *Magnetic resonance in medicine*, 77(1):285–299, 2017.

[229] Jiri Vrba and Stephen E Robinson. Signal processing in magnetoencephalography. *Methods*, 25(2):249–271, 2001.

[230] Hao Wang and Dit-Yan Yeung. A survey on bayesian deep learning. *ACM computing surveys (csur)*, 53(5):1–37, 2020.

[231] Jinhui Wang, Liang Wang, Yufeng Zang, Hong Yang, Hehan Tang, Qiyong Gong, Zhang Chen, Chaozhe Zhu, and Yong He. Parcellation-dependent small-world brain functional networks: A resting-state fmri study. *Human brain mapping*, 30(5):1511–1523, 2009.

[232] Xiao-Jing Wang. Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, 36(5):955–968, 2002.

[233] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.

[234] Haiguang Wen, Junxing Shi, Yizhen Zhang, Kun-Han Lu, Jiayue Cao, and Zhongming Liu. Neural encoding and decoding with deep learning for dynamic natural vision. *Cerebral cortex*, 28(12):4136–4160, 2018.

[235] Paul Werbos. Beyond regression: New tools for prediction and analysis in the behavioral sciences. *PhD thesis, Committee on Applied Mathematics, Harvard University, Cambridge, MA*, 1974.

[236] Corey N White, Mathieu Servant, and Gordon D Logan. Testing the validity of conflict drift-diffusion models for use in estimating cognitive processes: A parameter-recovery study. *Psychonomic bulletin & review*, 25:286–301, 2018.

[237] Robert W Williams and Karl Herrup. The control of neuron number. *Annual review of neuroscience*, 11(1):423–453, 1988.

[238] Hugh R Wilson and Jack D Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal*, 12(1):1–24, 1972.

[239] Sam F L Windels, Noël Malod-Dognin, and Nataša Pržulj. Graphlet Laplacians for topology-function and topology-disease relationships. *Bioinformatics*, 35(24):5226–5234, 06 2019.

[240] Jonathan R. Wolpaw and Elizabeth Winter Wolpaw. *Brain-Computer Interfaces: Principles and Practice*. Oxford University Press, 2012.

[241] David H. Wolpert and William G. Macready. No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.*, 1:67–82, 1997.

[242] Kong-Fatt Wong and Xiao-Jing Wang. A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*, 26(4):1314–1328, 2006.

[243] Wei Wu, M Black, Yun Gao, M Serruya, A Shaikhouni, J Donoghue, and Elie Bienenstock. Neural decoding of cursor motion using a kalman filter. *Advances in neural information processing systems*, 15, 2002.

[244] Ting Xu, Zhi Yang, Lili Jiang, Xiu-Xia Xing, and Xi-Nian Zuo. A connectome computation system for discovery science of brain. *Science Bulletin*, 60(1):86–95, 2015.

[245] BT Thomas Yeo, Fenna M Krienen, Jorge Sepulcre, Mert R Sabuncu, Danial Lashkari, Marisa Hollinshead, Joshua L Roffman, Jordan W Smoller, Lilla Zöllei, Jonathan R Polimeni, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*, 2011.

[246] Vahab Youssofzadeh and Abbas Babajani-Feremi. Mapping critical hubs of receptive and expressive language using meg: a comparison against fmri. *NeuroImage*, 201:116029, 2019.

[247] Yong Yu, Xiaosheng Si, Changhua Hu, and Jianxun Zhang. A review of recurrent neural networks: Lstm cells and network architectures. *Neural computation*, 31(7):1235–1270, 2019.

[248] Ekim Yurtsever, Jacob Lambert, Alexander Carballo, and Kazuya Takeda. A survey of autonomous driving: Common practices and emerging technologies. *IEEE access*, 8:58443–58469, 2020.

[249] Philip M Zeman, Bernie C Till, Nigel J Livingston, James W Tanaka, and Peter F Driessen. Independent component analysis and clustering improve signal-to-noise ratio for statistical analysis of event-related potentials. *Clinical Neurophysiology*, 118(12):2591–2604, 2007.

[250] Lixin Zhan, Bart Piwowar, Wing-Ki Liu, PJ Hsu, SK Lai, and Jeff ZY Chen. Multicanonical basin hopping: A new global optimization method for complex systems. *The Journal of chemical physics*, 120(12):5536–5542, 2004.

[251] Shuoyan Zhang, Jiacheng Yang, Ying Zhang, Jiayi Zhong, Wenjing Hu, Chenyang Li, and Jiehui Jiang. The combination of a graph neural network technique and brain imaging to diagnose neurological disorders: A review and outlook. *Brain Sciences*, 13(10):1462, 2023.

[252] Yu-Dong Zhang, Zhengchao Dong, Shui-Hua Wang, Xiang Yu, Xujing Yao, Qinghua Zhou, Hua Hu, Min Li, Carmen Jiménez-Mesa, Javier Ramirez, et al. Advances in multimodal data fusion in neuroimaging: Overview, challenges, and novel orientation. *Information Fusion*, 64:149–187, 2020.

[253] Xi-Nian Zuo and Xiu-Xia Xing. Test-retest reliabilities of resting-state fmri measurements in human brain functional connectomics: a systems neuroscience perspective. *Neuroscience & Biobehavioral Reviews*, 45:100–118, 2014.

[254] Ömer Yaveroğlu, Noël Malod-Dognin, Darren Davis, Zoran Levnajic, Vuk Janjic, Rasa Karapandza, Aleksandar Stojmirovic, and Nataša Pržulj. Revealing the hidden language of complex networks. *Scientific Reports*, 4(4547), 2015.