Data-Driven Modelling of the Brain Using EEG Data with Exogenous Input

A Dynamic Network Identification Approach to Determine Brain Connectivity

S. L. Bakels

Delft University of Technology

TU Delft

Data-Driven Modelling of the Brain Using EEG Data with Exogenous Input

A Dynamic Network Identification Approach to Determine Brain Connectivity

Master of Science Thesis

For the degree of Master of Science in Systems and Control and Mechanical Engineering at the Delft University of Technology

November 6, 2023

by

S. L. Bakels

Student number: 4480279

Project Duration: November 2022 - November 2023

Thesis committee: Dr. Ir. M. Jafarian (Supervisor)
Dr. Ir. M. L. van de Ruit (Supervisor)
Dr. Ir. D. Boskos (Graduation Committee)
Dr. Ir. W. Mugge (Graduation Committee)

TU Delft

Delft Center for Systems and Control

Abstract

The human brain, with its intricate web of billions of neurons and trillions of synaptic connections, is a remarkable organ responsible for performing complex cognitive processes. While brain imaging techniques like functional magnetic resonance imaging (fMRI) and electroencephalogram (EEG) provide insights into neural activity, there is no broadly accepted mathematical framework for the collaborative activity of neuronal populations and their communication. In the interdisciplinary fields of neuroscience and computational modelling, the pursuit of simplification is paramount for unravelling the enigma of the human brain's network complexity.

A data-driven mathematical framework is introduced representing the brain's cortical network obtained from EEG data while incorporating an exogenous input. This input signifies an external stimulus that impacts the brain's activity. The mathematical framework employs subspace identification to capture brain dynamics, resulting in the formulation of a state-space model. Additional manipulation of the state-space representation enables the evaluation of the statistical interrelationships among brain sources, commonly referred to as functional connectivity. The main objective of this mathematical framework is to model the brain dynamics and explore functional connectivity in individuals participating in a passive task that stimulates the brain, utilizing EEG data. Integrating an exogenous input into the dynamic model to represent stimuli within state-space models, pivotal for modelling brain networks during cognitive processes, fills a gap in the scientific literature.

The developed framework consists of four consecutive steps. Firstly, EEG data undergoes preprocessing to enhance signal quality and ensure data reliability. In the second step, the state-space model capturing brain dynamics is identified, employing two distinct subspace identification methods: N4SID and PO-MOESP. These methods are chosen to evaluate their effectiveness in representing intricate brain dynamics. The third step facilitates the transition from simulated states in the state space model to brain sources, enhancing outcome interpretability. The fourth stage involves the application of Granger causality for state-space models, a statistical hypothesis test used to evaluate the predictive value of one time series in relation to another. Employing Granger causality helps ascertain the functional connectivity between different brain sources.

Employed on a dataset that involves continuous perturbations of participants' wrists mimicking a sensorimotor task, the framework yields valuable insights. These findings validate the framework's capability to capture brain dynamics through a subspace identification method within a linear state-space model. Furthermore, simulation results indicate that the PO-MOESP method outperforms the N4SID method. Additionally, the framework is competent to assess functional connectivity between brain sources, resulting in a network connectivity diagram that offers valuable insights into the statistical relationships between distinct brain regions. Altogether, it is concluded that the designed algorithm can determine the intricate interactions within the brain during a passively performed task stimulating the brain. Consequently, this achievement forms a robust foundation for advancing Brain-Computer Interfaces (BCIs) and enhancing the diagnosis of neurological disorders.

Acknowledgements

During my research, I have had lots of support from various persons. I would like to begin by extending my gratitude to my supervisors, Matin and Mark. It is through your support, the meetings, and your constructive feedback that I was able to conduct this research successfully. Pursuing a thesis for two masters can infuse a project with excitement and innovation, driven by the wealth of diverse perspectives and insights it affords. However, it also requires a high degree of adaptability, especially when exploring topics that may not exactly fit within the domains of your discipline or research. Additionally, I extend my gratitude to Prof. Dr. Ir. Michel Verhaegen for affording me the opportunity to engage in two insightful meetings. His expertise in subspace identification proved invaluable in the design and validation of the algorithm while also providing crucial insights to elucidate certain findings.

I also extend my thanks to my housemates, Joosje and Sophie. Throughout this year, which has seen its fair share of challenging moments, you were a constant source of support. Whether as attentive listeners, shoulders to lean on, or providers of comfort in the form of chocolate, you were always there for me. We have shared the burden of doing a thesis, and I could not have wished for other comrades during this period.

Lastly, I would like to express my heartfelt gratitude to my parents. Despite all the detours I took to get here, which ultimately prolonged my studies, it is thanks to your unwavering backing and interest that I have felt empowered. Dad, your continuous sharing of articles related to my studies reflects your deep interest in both my academic pursuits and me. Mum, your passion for healthcare has played a significant role in influencing my decision to add something medical to my engineering studies. Your dedication and hard work serve as an inspiring example of a strong and diligent woman. It is the unwavering support from both of you that has consistently empowered me to pursue growth, both in my personal life and in academia.

Delft, University of Technology — S. L. Bakels
November 6, 2023

Contents

List of Figures

List of Tables

1

Introduction

The human brain is an incredibly complex organ characterised by billions of neurons and trillions of synaptic connections, which serve as junctions for neuronal intercommunication. This intricate neural network is the foundation for the brain's ability to process information, enabling humans to engage in cognitive functioning and motor activities.

Brain imaging techniques such as functional magnetic resonance imaging fMRI and electroencephalography EEG reflect the collective activity of thousands of neurons. In the realm of neuroscience and computational modelling, simplification is the key to gaining insights into the complexity and diversity of the network that constitutes the human brain. Such simplification can be achieved by constructing dynamic models, which serve as simplified representations of real-world entities in equations or computer code. These models are intended to mimic how essential features of the studied system change over time while leaving out inessentials [1]. However, there is no broadly accepted mathematical theory for the collaborative activity of neuronal populations and their communication [2].

1.1. Motivation

Developing a dynamic model of large-scale brain activity holds significance for various objectives. Foremost, it facilitates an in-depth comprehension of cerebral functioning and the intricate communication dynamics among distinct neural sources. Additionally, it is helpful in multiple applications, including the development of real-time brain-computer interfaces (BCI) and neurofeedback systems, which rely on immediate interaction with brain dynamics. Moreover, dynamic models could play a pivotal role in clinical domains, facilitating the diagnosis of neurological conditions such as Alzheimer's disease and epilepsy, offering insights into how brain activity evolves across diverse conditions and aiding in therapy design.

The intricacies of movement, cognition, and perception arise from the coordinated activity of neurons within cortical circuits and across the expansive networks of the brain. To identify areas that subserve specific tasks, most neuroimaging studies have traditionally focused on investigating task-dependent changes in brain activity [3]. However, to gain a comprehensive understanding of brain functioning and the execution of cognitive processes, departing from the conventional practice of examining isolated brain regions separately is necessary, as brain functioning hinges on the pattern of interactions between regions [3]. Functional connectivity analysis, which involves assessing the statistical dependencies among cortical or subcortical brain regions, is a means of unveiling these interactions [4]. By integrating the ability to extract functional connectivity patterns from neuroimaging data, researchers are equipped with a powerful approach to conduct in-depth analyses of complex brain networks.

To model the dynamic behaviour of the brain and determine the functional connectivity, it is necessary to record the brain's activity. Three commonly used non-invasive techniques for measuring brain activity are functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), and electroencephalography (EEG). Notably, EEG stands out as a versatile and cost-effective solution, advantageous in dynamic environments where participants may not be constrained to stationary positions [5]. Furthermore, EEG has a high temporal resolution, making it an attractive option for researchers seeking to model the temporal dynamics of brain networks [6]. Hence, employing and processing EEG to capture electrical signals from the scalp enables the evaluation of cortical brain activity. This approach, coupled with participants' engagement in simple tasks stimulating the brain during experimental sessions, is a favourable strategy for initiating the model's construction.

The conventional mathematical framework for modelling brain dynamics using EEG data relies on the utilisation of linear vector autoregressive (VAR) functions [7, 8]. These models have demonstrated their ability to effectively capture a substantial portion of brain activity [9, 10]. However, time series data originating from diverse applications often exhibit a moving-average component, which may not be efficiently captured by a finite-order VAR model [11]. This moving-average characteristic assumes significance when dealing with data subjected to filtration, downsampling, and noise, as with neurophysiological data. Consequently, the dynamics should be explained by vector autoregressive moving-average (VARMA) models rather than VAR models [11, 12].

As an alternative, studies have proposed the utilisation of state-space models to capture the brain's dynamics from EEG data [7, 13, 14] or fMRI data [15]. This choice is driven by the established equivalence between the VARMA models and state-space models [11]. State-space equations provide a robust and linear mathematical framework for modelling and analysing dynamic systems over time, accommodating the incorporation of moving-average components. Additionally, state-space models equip researchers to easily leverage tools from control theory and system identification to gain deeper insights into and manipulate dynamics [16]. Furthermore, the capacity of state-space models to accommodate missing data is particularly advantageous in neuroscience, where data quality and availability can be variable and inconsistent [16]. To date, research has focused on applying state-space models to evaluate brain connectivity in input-free datasets devoid of specified external stimuli, primarily during resting state experiments.

Data-driven identification techniques can be used to determine the parameters of the state-space representation using EEG data. Within the realm of data-driven system identification, subspace identification emerges as a promising approach [17, 18]. This method unveils the underlying system's hidden characteristics directly from input-output measurements without requiring an exhaustive understanding of the system's physical equations [18]. Unlike several other identification algorithms, such as the prediction error method, subspace identification eliminates the need for model parameterisation [16]. Additionally, the system model is obtained noniteratively, enhancing computational efficiency and reducing model complexity [16]. Applying subspace identification for state-space model identification opens the path to extracting functional connectivity from the model.

Granger causality analysis is a widely adopted statistical approach for detecting functional connectivity topology from data obtained through EEG. This method uses statistical techniques to infer the relationship between time series based on the observed data [19]. Traditionally, Granger causality is applied within the context of VAR models. However, Barnett and Seth [11] have formulated an approach for using Granger causality to state-space models. It is worth noting that this method is designed for state-space representations that lack an input, similar to the brain at rest without external stimuli.

1.2. Research objectives

To facilitate an in-depth comprehension of brain dynamics and functional connectivity in individuals engaged in passively performed tasks stimulating the brain, a mathematical framework is developed to simplify the complex dynamics. This framework holds promising potential for various applications, including BCI, and aiding in diagnosing neurological disorders. Integrating an exogenous input into the dynamic model to represent external stimuli within state-space models, pivotal for modelling brain networks during cognitive functioning and motor activities, fills a gap in the scientific literature. Hence, the research objective is:

> **Develop a data-driven dynamic mathematical model representing the brain cortical network from EEG data in the presence of an exogenous input.**

The research objective is achieved by developing a multi-step mathematical framework. This framework involves preprocessing the acquired EEG data, applying subspace identification, enhancing the interpretability of the identified state-space model, and using a modified Granger causality method designed for state-space systems with exogenous input.

Several significant contributions are introduced in this research. Firstly, it presents an innovative framework that shifts from conventional VAR models to state-space models. Within state-space models, it integrates an exogenous input to represent external stimuli. Hence, this study applies and modifies the Granger causality method for state-space representations that include an input. This decision is driven by the recognition that the brain operates in an interactive environment, constantly engaging with and responding to external stimuli. The selection of an input stimulus is deliberate, aiming to elicit a steady-state response within the brain. This choice is motivated by the fact that steady-state responses are more amenable to modelling and analysis than transient responses.

Furthermore, this study seeks to implement two different subspace identification techniques. Prior research has effectively investigated the estimation of functional brain connectivity using Granger causality within the context of state-space models that lack an exogenous input [13, 14]. However, it is worth noting that existing literature often omits detailed descriptions of the specific subspace identification methods utilised. This research aims to introduce distinct methodologies while offering comprehensive insights into their application, thereby improving the understandability of the resulting state-space model.

The efficacy of the developed framework is confirmed through its application to a dataset from Vlaar et al. [20]. The results obtained from applying this framework fulfil several crucial objectives. Firstly, they indicate the framework's viability and whether it is capable of effectively capturing the intricate dynamics of the brain. This evaluation depends on whether the brain dynamics can, in this case, be accurately characterised as linear. Second, the results provide insights into the performance of the two subspace identification techniques. Comparing these techniques aids in identifying the most appropriate and accurate method for modelling the brain's dynamic behaviour under specified conditions. Furthermore, this research culminate in creating a network diagram that visually represents the brain connectivity of the participants involved in the experiment.

1.3. Background

The research is built upon a broader context. Delving into the fundamental information underpinning this research is essential to provide a strong foundation for the upcoming chapters. The following section provides this foundational information, focusing on the employed brain imaging tool, EEG, and the methodology utilised for state-space model identification, specifically, subspace identification.

Understanding the theoretical components is crucial for comprehending the research at hand, such as the mathematical prerequisites, the applied methodologies of subspace identification, and the utilisation of Granger causality to assess functional connectivity. Hence, a more in-depth examination of these concepts will be conducted within the theoretical framework presented in Chapter 2.

1.3.1. Electroencephalography (EEG)

Electroencephalography (EEG) is a non-invasive neuroimaging technique measuring the brain's electrical activity, captured using numerous electrodes placed on the scalp. These electrodes record extracellular currents that arise from postsynaptic potentials, mostly from pyramidal cells, in the outermost section of the cerebral cortex. Postsynaptic potentials are changes in the membrane potential of the postsynaptic terminal of a chemical synapse. The potentials are derived from the release of excitatory and inhibitory neurotransmitters in the endpoints of the axon, the dendrites. An excitatory postsynaptic potential at an apical dendrite locally results in an intracellular current source, the region with the higher potential, and an extracellular current sink, the region with the lower potential. Conversely, there is an intracellular current sink and an extracellular current source at the cell body. These source-sink configurations are also known as current dipoles, forming the primary source of potentials measured by EEG [21]. The source of brain electrical activity detected by EEG is displayed in Figure 1.1.

The electrical potential from a single neuron is too weak for scalp electrodes to detect. Thus, measurable electric potentials require the simultaneous activity of neuronal populations with similar spatial orientation. This process enables the integration of electrical currents, which disseminate homogeneously, regardless of their frequency spectra, across the brain volume and subsequently through the pia mater, subarachnoid space, arachnoid, pura mater, skull, muscles, fat, and skin to the EEG recording electrodes [21].

It is crucial to emphasize that the electrical signals generated by neurons in the brain are weak, on the order of a millionth of a volt, and these signals have to pass through various tissues before reaching the electrodes on the scalp. Hence, EEG only captures a fraction of the brain's electrical activity, and this signal is obtained in the presence of other types of physiological activity creating noise, such as cardiac, ocular, and muscular activity, and environmental noise [23].

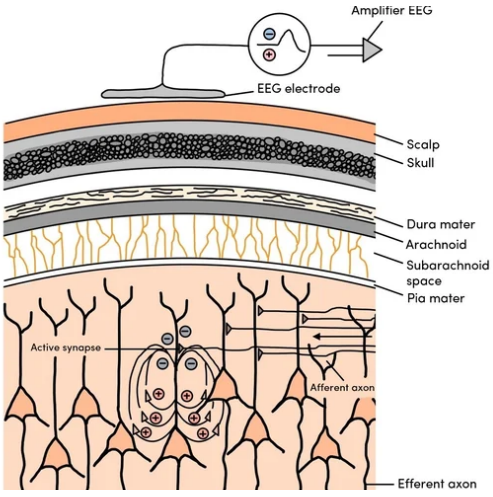


Figure 1.1: The source of the electrical activity measured by EEG in the brain, where electrical fields are generated by the aligned pyramidal cells. Adapted from [22].

1.3.2. Subspace identification

System identification is a specialized discipline focused on modelling and comprehending the behaviour of dynamic systems using input-output data to establish mathematical models that capture the underlying dynamics of a system, enable prediction, control, and deeper insight into its functioning [16]. Within the system identification domain, subspace identification is a prominent data-driven approach. The method excels in extracting the hidden representation of the system directly from input-output measurements without requiring a thorough understanding of the physical equations [18]. The term "subspace" refers to a lower-dimensional space that captures the essential dynamics of the system, making it particularly valuable for modelling complex dynamic systems.

Subspace identification encompasses a range of methods, each tailored to specific scenarios. Within this spectrum, there are open-loop and closed-loop approaches, as well as making subspace identification adaptable to Single Input Single Output (SISO) or Multiple Input Multiple Output (MIMO) systems [24].

Open-loop subspace identification applies to systems without feedback or control actions, as depicted in the block diagram shown in Figure 1.2.a. Commonly employed open-loop subspace identification techniques encompass the Numerical algorithms for Subspace State Space System Identification (N4SID) [25], Canonical Variate Analysis (CVA) [26], and the Multivariable Output Error State sPace (MOESP) methods, along with its variations Past Input Multivariable Output Error State sPace (PI-MOESP) and Past Output Multivariable Output Error State sPace (PO-MOESP) [17, 27].

However, the presence of feedback loops and the potential for system control using controllers may necessitate the application of closed-loop or partially closed-loop conditions [24]. A block diagram representing the basics of a closed-loop system is displayed in Figure 1.2.b. Closed-loop subspace identification models dynamic systems that operate in feedback-controlled environments. While open-loop subspace identification focuses on systems that lack feedback or control actions, closed-loop subspace identification thrives precisely in scenarios where feedback mechanisms play a central role. The traditional subspace identification techniques are biased under these conditions. Hence, Verhaegen [28] proposed a closed-loop subspace identification method via the identification of an overall open-loop state-space model followed by a reduction process to derive the state space representations for both the system and the controller [28]. After this pioneering work, numerous methodologies emerged to address this particular challenge.

The dynamic model in this study relies on output variables, namely the brain's electrical activity gauged through EEG, coupled with an exogenous input component. It is important to emphasize that the dynamic model is constructed using open-loop system identification. This distinction arises from the study's emphasis on experiments that elicit a response from the brain without active participation from the participants, referred to in this research as a simple passive task. The algorithm's validation is based on a dataset gathered during a

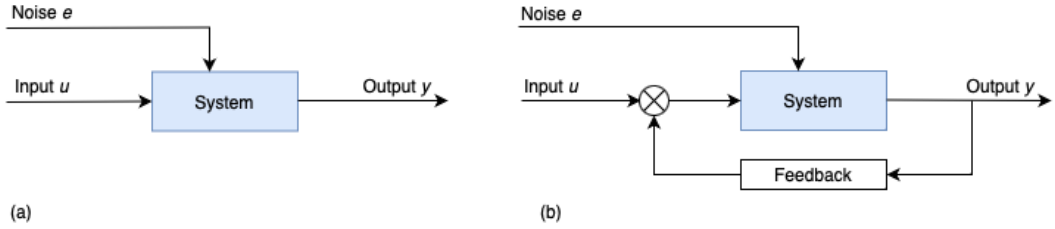


Figure 1.2: Comparison between the block diagram of open-loop and closed-loop systems. (a) Open-loop control system. (b) Closed-loop control system open-loop.

passively executed task in which participants experience wrist perturbations but do not actively partake in responsive behaviours. This differentiates it from the closed-loop scenarios where participants interact or respond intentionally.

1.4. Related work

Linear models have found extensive application in the analysis of brain dynamics, emerging as a practical approach that has garnered notable attention in recent decades, as demonstrated by various studies [8, 29]. These models serve as a valuable framework for understanding and interpreting the intricate dynamics of neural activity. Moreover, linear models have conventionally been employed to assess the association between the brain's motor cortex and related body muscles, often referred to as corticomuscular coherence [30, 31]. In these studies, linear modelling describes the brain's response to external stimuli, especially in tasks involving perturbations of the participant's muscles. Nonetheless, other research emphasizes the necessity of using nonlinear models for capturing brain dynamics resulting from the interaction between cortical oscillations and muscle activity [32, 33].

To validate the developed framework, the algorithm will be applied to a dataset from Vlaar et al. [20], which comprised EEG data. Studies have employed the same dataset [20, 32, 34] or using a similar approach for testing the cortical response to wrist manipulation [35, 36, 33]. In experimental settings, continuous wrist perturbation is anticipated to stimulate dynamic information flow among different brain sources involved in sensorimotor processing. This interaction encompasses sensory processes, including proprioceptive and tactile sensations, as well as motor processes. Research indicates that constant joint manipulation triggers the transmission of proprioceptive and tactile sensory information to the contralateral primary somatosensory cortex via the thalamus [32]. Other cortical areas that are likely to be activated during sensorimotor stimulation comprise the posterior parietal cortex [37], the insular cortex [38], and the secondary somatosensory cortices [39].

1.5. Thesis outline

This research is outlined as follows. First, the theoretical foundation for constructing the model is presented in Chapter 2. This chapter encompasses the foundational principles essential for developing a dynamic model, as well as an in-depth explanation of the two employed subspace identification techniques, namely, N4SID and PO-MOESP. Chapter 3 provides the algorithmic design of the framework, consisting of four steps to achieve the research objective. This chapter also delineates the experimental setup and conditions concerning the dataset employed for model validation. Subsequently, in Chapter 4, the simulation outcomes obtained by applying the developed framework to the dataset are presented. These results are determined for each framework step, culminating in the network connectivity diagrams. Finally, Chapter 5 serves as a summary of the research findings. This chapter responds to the research objective and assesses its success. In addition, this chapter provides recommendations for future research that can build upon the findings of this study.

2

Theoretical Framework

In neuroscientific research, the brain's cortical activity can be captured through EEG, yielding a substantial volume of data. The obtained dataset can serve as a valuable resource, facilitating the extraction of connectivity patterns within the brain to provide insights into the interactions and information processed between distinct brain regions. To accomplish this, the brain can be conceptualised as a dynamic system characterised by the changing activity of its constituent brain sources over time. This dynamic behaviour lends itself to mathematical modelling, enabling the description and simulation of the brain's intricate dynamics. Within this system, the complex interplay and connections among its components can be represented as a dynamic network. A dynamic network represents a specialised form of a dynamic system that focuses on the interactions and relationships between entities within a network or graph structure. The dynamic network captures how information, signals, or influences flow between different nodes. This perspective offers a framework for understanding the dynamic nature of the brain and the interconnections between its brain sources [40].

The construction of a dynamic model is achieved by applying system identification techniques. In the context of this research, particular emphasis is placed on the data-driven approach known as subspace identification. Once the dynamic systems are modelled, the topology of the dynamic network can be determined. Within this investigation lies the importance of revealing the statistical dependencies between individual cerebral brain regions, which can be accomplished by applying the Granger causality analysis.

This chapter introduces an in-depth examination of preliminaries intended to serve as a review of mathematical concepts and definitions to be used throughout the rest of the chapter and report. These principles and definitions, forming the framework of this research, are crucial for deciphering the complexities inherent in dynamic systems. Subsequently, subspace identification is explained by relying on these preliminaries, focusing on the algorithms of the two techniques N4SID and PO-MOESP and its corresponding assumptions. This chapter ends with an explanation of Granger causality analysis for state-space models to determine the statistical dependencies between time series to determine the topology of a dynamic network. Overall, this chapter explores the foundational concepts and methodologies underpinning the formation of dynamic models and connectivity networks.

2.1. Preliminaries

To establish a robust comprehension of the theoretical framework used for the formation of a dynamic model and identifying the dynamic network, this section expounds upon fundamental mathematical concepts and definitions integral to subspace identification and

Granger causality, in which these ideas will be frequently employed. The geometric tools are presented from a perspective rooted in linear algebra, independently of the specific context of subspace identification and Granger causality. In general, the materials of this section are borrowed from Verhaegen and Verdult [16].

The theoretical foundation of this report also relies upon a set of basic concepts and terms, which serve as the basis for formulating assumptions regarding dynamic systems. The key definitions are summarized in Section 2.2.1.

2.1.1. State-space model

A state-space model is a mathematical framework of equations describing how the system's internal states evolve over time in response to inputs and how these states generate outputs. The essential components of a state-space model are the state vector x, input vector u and output vector y. They are used to describe and analyze linear time-invariant (LTI) systems. An LTI system can effectively be characterised using a discrete-time state-space model, referred to as the process form:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + w(k), \\ y(k) &= Cx(k) + Du(k) + v(k), \end{aligned} \tag{2.1}$$

where the integer $k \geq 0$ represents the discrete time index, $x(k) \in \mathbb{R}^n$ is the state vector, $u(k) \in \mathbb{R}^m$ is the input vector, $w(k) \in \mathbb{R}^n$ is the state noise vector, $y(k) \in \mathbb{R}^l$ is the output vector and $v(k) \in \mathbb{R}^l$ is the output measurement noise. The system matrices A, B, C and D are of appropriate dimensions.

A state-space representation of a system is not unique. There are infinitely many representations. An arbitrary non-singular (or invertible) matrix $T \in \mathbb{R}^{n\times n}$, also called the transformation matrix, defines a state-coordinate transformation $z = Tx$ [41]. In the context of a state-space system:

$$\dot{x} = Ax + Bu, \quad y = Cx + Du,$$

the coordinate change $z = Tx$ leads to:

$$\dot{z} = \tilde{A}x + \tilde{B}u, \quad y = \tilde{C}x + \tilde{D}u,$$

with:

$$\begin{bmatrix} \tilde{A} & \tilde{B} \\ \tilde{C} & \tilde{D} \end{bmatrix} = \begin{bmatrix} TAT^{-1} & TB \\ CT^{-1} & D \end{bmatrix}. \tag{2.2}$$

2.1.2. Matrix properties

A state-space model comprises several matrices, referred to as the system matrices. A matrix, for explanation denoted as $A \in \mathbb{R}^{m\times n}$, is termed an orthogonal matrix when it satisfies the condition $A^TA = I$. The rank of the matrix A, denoted by rank(A), is defined as the number of linearly independent columns or linearly independent rows of A. The matrix A is considered to have full rank when rank$(A) = \min(m, n)$. A crucial property concerning the rank of the product of two matrices is stated in the subsequent lemma:

Lemma 2.1 (Sylvester's rank inequality)**.** Consider the matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$, then [42]:

$$\text{rank}(A) + \text{rank}(B) - n \leq \text{rank}(AB). \tag{2.3}$$

This inequality is frequently employed for assessing the rank of the product of the matrices AB in situations where both matrices A and B possess a full rank of n, with n being less than or equal to both m and p. Then, rank(AB) also equals n.

Linear subspaces
The matrix $A \in \mathbb{R}^{m\times n}$ defines a linear transformation from the vector space $\mathbb{R}^n$ to the vector space $\mathbb{R}^m$. A subspace is a set of vectors entirely contained within another vector space. Four fundamental subspaces related to the matrix A can be defined within these vector spaces.

The vector space spanned by the columns of the matrix A is a subspace of $\mathbb{R}^m$ and is denoted by range(A). Similarly, the rows of A span a vector space, being a subspace of $\mathbb{R}^n$, denoted by range(A^T). Another important subspace associated with matrix A is the kernel, also called the null space and the left null space. The null space, denoted by ker(A), consists of all vectors $x \in \mathbb{R}^n$ that satisfy $Ax = 0$. The left null space, denoted by ker(A^T), consists of all vectors $y \in \mathbb{R}^m$ that satisfy $A^T y = 0$. To summarize, the four fundamental subspaces of a matrix $A \in \mathbb{R}^{m\times n}$ are:

$$\begin{aligned} \text{range}(A) &= \{y \in \mathbb{R}^m : y = Ax \text{ for some } x \in \mathbb{R}^n\}, \\ \text{range}\left(A^{\mathrm{T}}\right) &= \left\{x \in \mathbb{R}^n : x = A^{\mathrm{T}} y \text{ for some } y \in \mathbb{R}^m\right\}, \\ \ker(A) &= \{x \in \mathbb{R}^n : Ax = 0\}, \\ \ker\left(A^{\mathrm{T}}\right) &= \left\{y \in \mathbb{R}^m : A^{\mathrm{T}} y = 0\right\}. \end{aligned} \tag{2.4}$$

2.1.3. Matrix decomposition
Working with smaller components is more manageable than dealing with large entities when performing more complex matrix operations [43]. Analogously, matrix decomposition is a technique that facilitates partitioning a matrix into several smaller matrix components. This process is referred to as matrix decomposition, as it enables the decomposition of a matrix into its constituent vectors and matrices. Two practical matrix decomposition methods used further in this chapter are explained.

Singular-Value Decomposition
Any matrix $A \in \mathbb{R}^{p\times j}$, that is not necessarily a square matrix, can be decomposed as [44]:

$$A = U\Sigma V^T, \tag{2.5}$$

where $U \in \mathbb{R}^{p\times p}$ and $V \in \mathbb{R}^{j\times j}$ are orthogonal matrices and $\Sigma \in \mathbb{R}^{p\times j}$ has its only nonzero elements along the diagonal. The elements on this diagonal σ_i, are called the singular values of A and are ordered such that:

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_k = 0,$$

for which $r = rank(A)$ and $k = min(p, j)$. The singular value decomposition (SVD) of the matrix A with rank r, such that $r < p$ and $r < j$, can be partitioned as follows:

$$A = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix},$$

where $U_1 \in \mathbb{R}^{p\times r}, U_2 \in \mathbb{R}^{p\times(p-r)}, \Sigma_1 \in \mathbb{R}^{r\times r}, V_1 \in \mathbb{R}^{j\times r}$, and $V_2 \in \mathbb{R}^{j\times(j-r)}$. From this relation, it can be concluded that the columns of the matrices U_1, U_2, V_1 and V_2 provide orthogonal bases for the four fundamental subspaces of the matrix A, which is:

$$\begin{aligned} \text{range}(A) &= \text{range}\,(U_1), \\ \ker\left(A^T\right) &= \text{range}\,(U_2), \\ \text{range}\left(A^T\right) &= \text{range}\,(V_1), \\ \ker(A) &= \text{range}\,(V_2). \end{aligned} \tag{2.6}$$

The SVD stands out as a numerically robust factorization method. Determining singular values remains insensitive to computational inaccuracies, including rounding errors [16].

QR decomposition
Any matrix $A \in \mathbb{R}^{p\times j}$ can also be decomposed using the QR factorization [44]:

$$A = QR, \tag{2.7}$$

where $Q \in \mathbb{R}^{p\times p}$ is an orthogonal matrix and $R \in \mathbb{R}^{p\times j}$ is upper triangular. For cases where $j > p$, the matrix R is augmented with columns on the right. Conversely, for cases where $p > j$, R is augmented with zero rows at the bottom. When the matrix A has rank r, such that $r < p$ and $r < j$, the QR factorization can be partitioned as follows:

$$A = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 & R_2 \\ 0 & 0 \end{bmatrix},$$

for $R_1 \in \mathbb{R}^{p\times r}, R_2 \in \mathbb{R}^{p\times(p-r)}, R_1 \in \mathbb{R}^{r\times r}$ and $R_2 \in \mathbb{R}^{r\times(j-r)}$. This relation shows that [16]:

$$\begin{aligned} \text{range}(A) &= \text{range}\left(Q_1\right), \\ \ker\left(A^T\right) &= \text{range}\left(Q_2\right), \\ \text{range}\left(A^T\right) &= \text{range}\left(R_1^T\right). \end{aligned} \tag{2.8}$$

The QR factorization of A is related to the RQ factorization of A^T.

The RQ factorization can be carried out with numerical reliability. Nevertheless, it is not a dependable method for ascertaining the rank of a matrix. To obtain a reliable rank determination, the SVD should be employed [16].

2.1.4. Projection matrices

Projection matrices are fundamental mathematical tools used in linear algebra and geometry. At their core, projection matrices allow to project points or vectors from a higher-dimensional space onto a lower-dimensional subspace, preserving specific properties of interest. Two primary types of projections are often encountered: orthogonal and oblique projections. This theory is based on Steward [45] and Hachicha et al. [46].

To better understand the geometric aspects involved, it is assumed that the system matrices $A \in \mathbb{R}^{p\times j}$, $B \in \mathbb{R}^{q\times j}$ and $C \in \mathbb{R}^{r\times j}$ are known. Additionally, the following notations are employed:

- $A^\dagger$ denotes the Moore-Penrose pseudoinverse of matrix A.
- $B^\perp$ represents the orthogonal complement of a subspace B. The orthogonal complement consists of all vectors in the vector space that are orthogonal to every vector in B. In other words, for every vector b in B and every vector x in $B^\perp$, the inner product of b and x is zero: $b \cdot x = 0$.

Orthogonal projections
An orthogonal projection is a fundamental concept that describes the process of projecting one vector onto another vector in a way that preserves the orthogonality (perpendicularity) between the vectors. The orthogonal projection operator is defined as Π_B, which projects the row space of a matrix onto the row space of the matrix B. The row space of a matrix refers to the subspace spanned by the rows (or row vectors) of that matrix. The orthogonal projection can be calculated as:

$$\Pi_B = B^T(BB^T)^\dagger B. \tag{2.9}$$

The projection of the row space of the matrix A onto the row space of the matrix B is defined by:

$$\frac{A}{B} = A\Pi_B = AB^T(BB^T)^\dagger B. \tag{2.10}$$

The $\Pi_{B^\perp}$ defines the geometric operator that projects the row space of a matrix onto the orthogonal complement of the row space of the matrix B:

$$\frac{A}{B^\perp} = A\Pi_{B^\perp}, \tag{2.11}$$

where:

$$\Pi_{B^\perp} = I_j - \Pi_B, \tag{2.12}$$

The combination of the projections Π_B and $\Pi_{B^\perp}$ decomposes a matrix A into two matrices of which the row spaces are orthogonal:

$$A = A\Pi_B + A\Pi_{B^\perp}. \tag{2.13}$$

Alternatively, Equation 2.13 can be written as:

$$A = L_B B + L_{B^\perp} B^\perp, \tag{2.14}$$

with:

$$\begin{aligned} L_B B &= \frac{A}{B}, \\ L_B B^\perp &= \frac{A}{B^\perp}. \end{aligned} \tag{2.15}$$

A geometric interpretation of the decomposition of matrix A using orthogonal projections is illustrated in Figure 2.1a. The property of an orthogonal projection is:

$$\Pi_B / \Pi_{B^\perp} = 0. \tag{2.16}$$

Oblique projections

An oblique projection, in the context of linear algebra, is a projection of a vector onto a subspace that is not orthogonal to the vector itself. This means that the projected vector can have a component in the direction of the subspace and another component orthogonal to it. The oblique projection of matrix A along the row space of matrix B onto the row space of matrix C, $A/_B C$, can be calculated as:

$$A/_B C = \begin{bmatrix} \frac{A}{B^\perp} \end{bmatrix} \begin{bmatrix} \frac{C}{B^\perp} \end{bmatrix}^\dagger C = A\Pi_{B^\perp} + (C\Pi_{B^\perp})^\dagger C. \tag{2.17}$$

Instead of decomposing A as a linear combination of two orthogonal matrices (B and $B^\perp$), the matrix can also be decomposed as a linear combination of two non-orthogonal matrices B and C [18]. This can be written as:

$$A = L_B B + L_C C + L_{B^\perp, C^\perp} \begin{bmatrix} B \\ C \end{bmatrix}^\perp, \tag{2.18}$$

with the matrix $L_C C$ as the oblique projection of the row space of A along the row space of B on the row space of C:

$$L_C C = A/_B C. \tag{2.19}$$

Figure 2.1b illustrates the oblique decomposition of matrix A. The properties of the oblique projections are:

$$\begin{aligned} B/_B C &= 0, \\ C/_B C &= C. \end{aligned} \tag{2.20}$$

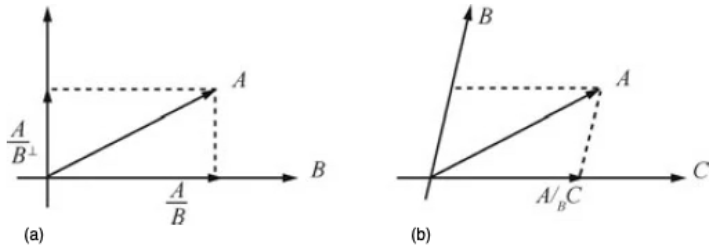


Figure 2.1: Projection matrices in the two-dimensional space. (a) The orthogonal projection. (b) The oblique projection. Retrieved from [46].

2.1.5. Kalman filering

For statistics and control theory, the Kalman filter is an algorithm that processes a series of measurements observed over time, accounting for statistical noise and inherent inaccuracies. Its primary objective is to generate estimations of the system states. These estimations are notably more precise than those derived from individual measurements, as the Kalman filter enables the optimal statistical reconstruction of the system's state within a given state-space model, characterized by the minimization of state-reconstruction errors conditioned on the available measurements. This section draws upon the insights and findings documented in Verhaegen and Verdult's work [16].

Discrete-time algebraic Riccati equation

The discrete-time algebraic Riccatti equation (DARE) is a nonlinear equation used to find the optimal gain matrix for a state-space control system, particularly for discrete-time linear systems [47]. The DARE can be represented as:

$$P = APA^T - (APC^T + S)(CPC^T + R)^{-1} \times (CPA^T + S^T) + Q, \tag{2.21}$$

where A and C are the system matrices, and S, R and Q are the covariance matrices. The goal of the DARE is to solve P in this equation. Once P is found, it can be used to design the Kalman gain K.

The Kalman filter for LTI systems

Consider the LTI model of Equation 2.1, with the process noise $w(k)$ and the measurement noise $v(k)$ assumed to be zero-mean white noise sequences. The estimate of the state, $\widehat{x}(k)$, can be improved by introducing a correction based on the difference between the measured output $y(k)$ and the estimated output, $\widehat{y}(k) = C\widehat{x}(k) + Du(k)$, as follows:

$$\widehat{x}(k+1) = A\widehat{x}(k) + Bu(k) + K(y(k) - C\widehat{x}(k) - Du(k)), \tag{2.22}$$

where the K is the steady state Kalman gain that must be chosen appropriately. The innovation Kalman filter can be denoted for further simplification as:

$$e(k) = y(k) - C\widehat{x}(k) - Du(k). \tag{2.23}$$

The difference $x_e(k)$ between the estimated state $\widehat{x}(k)$ and the real state $x(k)$ satisfies:

$$x_e(k+1) = (A - KC)x_e(k). \tag{2.24}$$

If the K is chosen such that $A - KC$ is asymptotically stable, the difference between the real state $x(k)$ and the estimated state $\widehat{x}(k)$ goes to zero for $k \to \infty$.

The K that satisfies this condition can be determined by first calculating the joint covariance matrix of the process and measurement noise:

$$E\left[\begin{bmatrix} v(k) \\ w(k) \end{bmatrix}\begin{bmatrix} v(j)^{\mathrm{T}} & w(j)^{\mathrm{T}} \end{bmatrix}\right] = \begin{bmatrix} Q(k) & S(k)^{\mathrm{T}} \\ S(k) & R(k) \end{bmatrix}\Delta(k-j) \geq 0, \tag{2.25}$$

where $\delta(k)$ is the unit pulse, a discrete-time signal that satisfies:

$$\Delta(k) = \begin{cases} 1, & \text{for } k = 0. \\ 0, & \text{for } k \neq 0. \end{cases} \tag{2.26}$$

Therefore, the joint covariance matrix is equal to:

$$\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \geq 0, \quad \text{with} R > 0. \tag{2.27}$$

If the pair (A, C) is observable and the pair $(A, Q^{\frac{1}{2}})$ is reachable, then the estimate of $x(k)$ at time instant $k-1$ is represented as $\widehat{x}(k|k-1)$ with properties:

$$\begin{aligned} E[x(k)] &= E[\widehat{x}(k)(k|k-1)], \\ E[(x(k) - \widehat{x}(k|k-1))(x(k) - \widehat{x}(k)|k-1)^T] &= P(k|k-1) \geq 0, \end{aligned} \tag{2.28}$$

The $P(k|k-1)$ represents the covariance matrix of the state estimation error $\widehat{x}(k|k-1)$, often assumed to be known as it relies on prior knowledge.

When $\widehat{x}(k|k-1) = E[x(k)]$ satisfies:

$$\lim_{k\to\infty} P(k|k-1) = P > 0, \tag{2.29}$$

for any symmetric initial conditions $P(0|-1) > 0$, where P satisfies the DARE, the P is unique. Then, the Kalman gain K can be formulated as:

$$K = (S + APC^T)(CPC^T + R)^{-1}. \tag{2.30}$$

This Kalman gain K ensures that $A - KC$ is asymptotically stable and, therefore, the difference between the actual state $x(k)$ and the estimated state $\widehat{x}(k)$ goes to zero for $k \to \infty$.

2.2. Methods of Subspace Identification: N4SID and PO-MOESP

As motivated in Section 1.3.2, this research focuses on open-loop subspace identification. In the domain of open-loop subspace identification methods, there are two primary categories. These categories aim to tackle the challenge of determining the system matrices of the state-space model described in Equation 2.1 while accommodating a set of similarity transformations. This research focuses on two distinct open-loop methods renowned for their ability to determine the desired state-space model: N4SID [25] and PO-MOESP [17, 27]. The N4SID method belongs to the first category, whereas PO-MOESP falls into the second category of methodologies [48]. The critical distinction between these two algorithms lies in the types of projection matrices used [46].

The selection of these algorithms is deliberate, driven by their classification as open-loop subspace identification techniques, their application to the dataset and their esteemed status in subspace identification. The rationale for choosing these methods is grounded in their well-established effectiveness, especially within the context of multivariable systems [49]. This alignment with the diverse nature of the study's scope makes them suitable candidates for the investigation. Furthermore, these methods benefit from applying

robust numerical tools such as QR decomposition and SVD [49]. These tools enhance the resilience when identifying state-space models from data often tainted by noise, a common occurrence when working with EEG datasets. This inherent robustness renders N4SID and PO-MOESP particularly well-suited for the specific objectives of this study, where accurate and reliable system identification is paramount to capture the intricacies of the dynamic brain processes. Additionally, both methods are supported by programming toolboxes, and their computational requirements are not excessive, making them operationally feasible within the constraints of available computational resources.

Notably, the objectives and a significant portion of the methodology employed by both N4SID and PO-MOESP are in concordance. The open-loop subspace identification problem starts from the perspective that if a system can be modelled as a LTI system, it can be represented by the state-space model of Equation 2.1. If the pair (A, C) is observable, one can design a Kalman filter for this system to estimate the state variables $\widehat{x}(k+1)$ of Equation 2.22, with the Kalman gain K and the innovation Kalman filter $e(k)$ of Equation 2.23. Ignoring the "^", indicating an estimate, the state-space model can be written in the innovation form:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + Ke(k), \\ y(k) &= Cx(k) + Du(k) + e(k). \end{aligned} \tag{2.31}$$

The goal of subspace identification is to estimate system matrices A, B, C, D, the Kalman gain K and the initial state $x(0)$ up to a similarity transformation directly from the input- and output measurements.

2.2.1. Mathematical tools

Subspace identification methods are based on the fact that storing the input- and output data into structured matrices makes it possible to retrieve the state sequence up to a similarity transformation [50]. This involves structuring input- and output measurement data and state matrices in a specific format or arrangement that optimally facilitates the application of these methods. This subsection sheds light on the essential aspects of data organisation. It presents examples playing a crucial role in the notations required for the utilised subspace identification methods N4SID and PO-MOESP. The materials of this section are based on Verhaegen and Verdult [16] and De Cock and De Moor [51].

Extended vectors and matrices

An extended vector is constructed by stacking multiple vectors into a single vector. The purpose of creating an extended vector is to represent a collection of relevant data or observations in a compact and structured format, which is easier to manipulate mathematically. Within the framework of subspace identification, the extended state vector is:

$$X_{i,N} = \begin{bmatrix} x(i) & x(i+1) & \dots & x(i+N-1) \end{bmatrix}^T. \tag{2.32}$$

The first entry of the subscript of $X_{i,N}$ refers to the upper time index, and the second subscript refers to the number of block rows.

On a related note, an extended matrix is formed by concatenating or combining multiple matrices. An example of this is the extended observability matrix O_N:

$$O_N = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{N-1} \end{bmatrix}, \tag{2.33}$$

where the subscript N denotes the number of block rows.

Hankel matrix
Hankel matrices are frequently used in subspace identification. A Hankel matrix is a symmetric matrix with constant elements across the anti-diagonals. The Hankel matrix of the output $y(k)$ can be constructed as follows:

$$Y_{i,s,N} = \begin{bmatrix} y(i) & y(i+1) & \cdots & y(i+N-1) \\ y(i+1) & y(i+2) & \cdots & y(i+N) \\ \vdots & \vdots & \ddots & \vdots \\ y(i+s-1) & y(i+s) & \cdots & y(i+N+s-2) \end{bmatrix}. \tag{2.34}$$

Hankel matrices of the same format can be formed for input and noise data. The first entry of the subscript of $Y_{i,s,N}$ refers to the time index of its top left entry, the second to the number of block rows, and the third to the number of columns. In general, $n < s \ll N$ [16].

Toeplitz matrix
A Toeplitz matrix is a special matrix type where each descending diagonal from left to right is constant. In other words, the elements along any given diagonal are equal such that the matrix is symmetric about its main diagonal. The lower block triangular Toeplitz matrix for the quadruple (A, B, C, D) is denoted as T_s:

$$T_s = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{s-2}B & CA^{s-3}B & \cdots & D \end{bmatrix}. \tag{2.35}$$

The lower block tiangular Toeplitz matrix for the triple (A, C, K) is defined by S_s:

$$S_s = \begin{bmatrix} I & 0 & \cdots & 0 \\ CK & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{s-2}K & CA^{s-3}K & \cdots & I \end{bmatrix}. \tag{2.36}$$

2.2.2. Foundational principles in N4SID and PO-MOESP algorithms

To establish the foundation of open-loop subspace identification methods, the following assumptions are imposed throughout Section 2.2:

- The system can be modelled as linear time-invariant.
- The eigenvalues of $A - KC$ are strictly inside the unit circle such that the system is asymptotically stable.
- The system is observable and controllable in the sense that (A, C) is observable and $(A, [BK])$ is controllable.
- The noise term e_k is a stationary, zero-mean, white noise process with a second-order moment given by $E(e_i e_j^T) = R_e \delta_{ij}$. Where δ_{ij} is the Kronecker delta.
- The input signal u_k is persistently exciting, which entails that it should vary over time, covering a wide range of values, to identify or estimate the parameters of a dynamic system effectively.
- The input u_k an d the noise sequence e_j are uncorrelated.

2.2.3. Algorithmic Comparison of N4SID vs. PO-MOESP

The N4SID and PO-MOESP methodologies entail a structured sequence of steps. The initiation of the algorithm involves the acquisition and embedding of data, where raw observations are transformed into a format suitable for subsequent analysis. In the second stage, the determination of the model order and state sequence is achieved, succeeded by the application of a least squares regression technique to estimate the state matrices [51].

Each step is discussed in a different paragraph using the preliminaries and definitions explained in Section 2.1 and the mathematical tools of Section 2.2.1. The methodology employed by both N4SID and PO-MOESP largely corresponds. However, for subspace decomposition in the second step, N4SID and PO-MOESP adopt divergent approaches. Therefore, two subparagraphs will provide a comprehensive and detailed exploration of the contrasting methodologies for this step.

Data collection and embedding

To initiate the subspace identification process, the input- and output data over some time must be collected from the system of interest. Subsequently, in the embedding step, the matrix that captures the dependencies and relationships between the input- and output variables is created. Both N4SID and PO-MOESP hinge on the state-space model in the innovation form, as represented in Equation 2.31. When filling in the state-space model at time instant k to an initial state $x(0)$ and an input signal from time 0 to k, the response can be found from the state equation by recursion [16]:

$$\begin{aligned}
x(1) &= Ax(0) + Bu(0) + Ke(0),\\
x(2) &= A^2x(0) + ABu(0) + Bu(1) + AKe(0) + Ke(1),\\
x(3) &= A^3x(0) + A^2Bu(0) + ABu(1) + Bu(2) + A^2Ke(0) + AKe(1) + Ke(2),\\
&\vdots\\
x(k) &= A^kx(0) + A^{k-1}Bu(0) + A^{k-2}Bu(1) + \ldots + ABu(k-2) + Bu(k-1)\\
&\quad + A^{k-1}Ke(0) + A^{k-2}Ke(1) + \ldots + AKe(k-2) + Ke(k-1).
\end{aligned} \tag{2.37}$$

In a concise representation, the response from time instant 0 to time instant k can be expressed as:

$$x(k) = A^kx(0) + + \sum_{i=0}^{k-1} A^{k-i-1}Bu(i) + \sum_{i=0}^{k-1} A^{k-i-1}Ke(i). \tag{2.38}$$

By invoking Equation 2.38 into the output of the state-space model of Equation 2.31, the following response from time instant 0 to time instant k can be observed:

$$y(k) = CA^kx(0) + C\sum_{i=0}^{k-1} A^{k-i-1}Bu(i) + C\sum_{i=0}^{k-1} A^{k-i-1}Ke(i) + Du(k-1) + e(k-1). \tag{2.39}$$

Through the concatenation of the output into an extended output vector, the following relationship between the input data batch $\{u(k)\}_{k=0}^{s-1}$ and the output batch $\{y(k)\}_{k=0}^{s-1}$ can be specified, making use of the extended matrix and Toeplitz matrices [16]:

$$\begin{bmatrix} y(0) \\ y(1) \\ y(2) \\ \vdots \\ y(s-1) \end{bmatrix} = \underbrace{\begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{s-1} \end{bmatrix}}_{O_s} x(0) + \underbrace{\begin{bmatrix} D & 0 & 0 & \cdots & 0 \\ CB & D & 0 & \cdots & 0 \\ CAB & CB & D & & 0 \\ \vdots & & \ddots & \ddots & \\ CA^{s-2}B & CA^{s-3}B & \cdots & CB & D \end{bmatrix}}_{\mathcal{T}_s} \begin{bmatrix} u(0) \\ u(1) \\ u(2) \\ \vdots \\ u(s-1) \end{bmatrix}$$
$$+ \underbrace{\begin{bmatrix} I & 0 & 0 & \cdots & 0 \\ CK & I & 0 & \cdots & 0 \\ CAK & CK & I & & 0 \\ \vdots & & \ddots & \ddots & \\ CA^{s-2}K & CA^{s-3}K & \cdots & CK & I \end{bmatrix}}_{\mathcal{S}_s} \begin{bmatrix} e(0) \\ e(1) \\ e(2) \\ \vdots \\ e(s-1) \end{bmatrix}, \tag{2.40}$$

where s is some arbitrary positive integer. This equation establishes a relationship between vectors originating from the input- and output data sequences, incorporating the unknown initial state $x(0)$ and the extended observability matrix O_s, as well as the Toeplitz matrices T_s and S_s.

Given the time-invariant nature of the system, the time-shifted input, state, and output vectors in Equation 2.40 correspond to the same matrices O_s, T_s, and S_s. As an illustration, consider a shift of k samples in Equation 2.40:

$$\begin{bmatrix} y(k) \\ y(k+1) \\ \vdots \\ y(k+s-1) \end{bmatrix} = O_s x(k) + T_s \begin{bmatrix} u(k) \\ u(k+1) \\ \vdots \\ u(k+s-1) \end{bmatrix} + S_S \begin{bmatrix} e(k) \\ e(k+1) \\ \vdots \\ e(k+s-1) \end{bmatrix}. \tag{2.41}$$

As stated, the matrices O_s, T_s, and S_s exhibit consistency. The relationships of Equation 2.40 and Equation 2.41 can be merged and concatenated for different time shifts, as permitted by the availability of input- and output samples, resulting in the following expression:

$$\begin{bmatrix} y(0) & y(1) & \dots & y(N-1) \\ y(1) & y(2) & \dots & y(N) \\ \vdots & \vdots & \ddots & \vdots \\ y(s-1) & y(s) & \dots & y(N+s-2) \end{bmatrix} = O_s \begin{bmatrix} x(0) \\ x(1) \\ \dots \\ x(N) \end{bmatrix} + T_s \begin{bmatrix} u(0) & u(1) & \dots & u(N-1) \\ u(1) & u(2) & \dots & u(N) \\ \vdots & \vdots & \ddots & \vdots \\ u(s-1) & u(s) & \dots & u(N+s-2) \end{bmatrix}$$
$$+ S_S \begin{bmatrix} e(0) & e(1) & \dots & e(N-1) \\ e(1) & e(2) & \dots & e(N) \\ \vdots & \vdots & \ddots & \vdots \\ e(s-1) & e(s) & \dots & e(N+s-2) \end{bmatrix}. \tag{2.42}$$

The above equation is referred to as the "data equation". The matrices formed using the input-, output- and noise data exhibit entries that remain consistent along the diagonals, indicating their nature as Hankel matrices. Therefore, using the definitions of Section 2.2.1, starting from time point i and accounting for data length s for N different time steps, the data equation can be expressed in a more compact and general form. The parameter s, representing the number of block rows, is a user-defined variable that demands careful selection. It should adhere to the condition where the model order n is smaller than s and s

is substantially smaller than the total number of data points denoted as N [16]. The model order n can be ascertained through the subspace decomposition step in the process. The following equation can be obtained:

$$Y_{i,s,N} = O_s X_{i,N} + T_s U_{i,s,N} + S_s E_{i,s,N}. \tag{2.43}$$

This equation establishes a crucial connection between matrices derived from the input-output data and matrices derived from the system's dynamics and disturbances. It serves as the foundational basis for reconstructing the system dynamics from the available input-output data, playing a pivotal role in the subspace identification process. Additionally, it constitutes the outcome for the embedding step.

Subspace decomposition
In the second step, factorization techniques are used to decompose the input and output data Hankel matrices into lower-dimensional subspaces, an essential process for extracting critical insights into the underlying system dynamics and determining the model's order. This step is the primary distinguishing feature between the two methods, N4SID and PO-MOESP. The subspace decomposition can be achieved through various methods, but this report employs the SVD and projection matrices.

Both N4SID and PO-MOESP use an instrumental variable Z_N to enhance the reliability and accuracy of the subspace identification process in the presence of noise and uncertainties in the input-output data [16]. This instrumental-variable matrix is carefully designed to possess specific properties that enhance its utility in the system identification context. These properties include:

$$\begin{aligned} &\lim_{N\to\infty} \frac{1}{N} E_{i,s,N} \Pi_{U^{\perp}_{i,s,N}} Z_N^T = 0, \\ &\operatorname{rank}\left(\lim_{N\to\infty} \frac{1}{N} X_{i,N} \Pi_{U^{\perp}_{i,s,N}} Z_N^T \right) = n. \end{aligned} \tag{2.44}$$

In words, this means that the instrumental-variable matrix is structured in such a way that, as the number of data samples approaches infinity, the expected value of a specific operation involving the matrix converges to zero. This property ensures that the instrumental variable remains unbiased and does not introduce systematic errors in the identification process. Furthermore, the rank of a specific matrix expression involving the instrumental-variable matrix and other relevant matrices remains constant even as the number of data samples increases towards infinity. This rank preservation property is essential for ensuring that the instrumental variable retains information about the system's dynamics.

The instrumental variable that satisfies these conditions, for both N4SID and PO-MOESP, is [16]:

$$Z_N = \begin{bmatrix} U_{0,s,N} \\ Y_{0,s,N} \end{bmatrix}. \tag{2.45}$$

A detailed proof for the choice of the instrumental variable can be found in Lemma 9.4-9.7 of the book "Filtering and System Identification: A Least Squares Approach" of Verhaegen and Verdult [16].

N4SID The N4SID method is designed to estimate the state sequence of a system by performing an oblique projection on subspaces generated from the input and output block Hankel matrices [50]. Specifically, this oblique projection is carried out on the extended state-space model represented in Equation 2.43. The projection occurs along the row space

of $U_{i,s,N}$, directed onto the row space of the instrumental variable Z_N. This operation is mathematically expressed as:

$$Y_{s,s,N}/_{U_{s,s,N}} Z_N = O_s X_{s,N}/_{U_{s,s,N}} Z_N + T_s U_{s,s,N}/_{U_{s,s,N}} Z_N + S_s E_{s,s,N}/_{U_{s,s,N}} Z_N. \tag{2.46}$$

In the context of this mathematical expression, it is evident that the last two components become void or insignificant. This phenomenon arises because the term $U_{s,s,N}/_{U_{s,s,N}} Z_N$ is constrained to be zero, aligning with the characteristic behaviour specified by the oblique projection property detailed in Equation 2.20. Additionally, the term $E_{s,s,N}/_{U_{s,s,N}} Z_N$ is nullified by the assumption that the noise is uncorrelated with past input and output data, which is the case in open-loop identification. Consequently, Equation 2.46 can be simplified to:

$$Y_{s,s,N}/_{U_{s,s,N}} Z_N = O_s X_{s,N}/_{U_{s,s,N}} Z_N. \tag{2.47}$$

This equation can be written as:

$$Y_{s,s,N}/_{U_{s,s,N}} Z_N = O_s \tilde{X}_s, \tag{2.48}$$

with $\tilde{X}_s = X_{s,N}/_{U_{s,s,N}} Z_N$.

The N4SID algorithm determines the order of the system and the observability matrix O_s directly from the SVD of the oblique projection [18]:

$$Y_{s,s,N}/_{U_{s,s,N}} Z_N = USV^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \simeq U_n S_n V_n^T. \tag{2.49}$$

The left-hand side of the equation consists solely of matrices containing known values derived from available input and output data. Consequently, the SVD can be solved.

In the presence of noise, the model order of the system can be estimated from the singular values of S_n and depends heavily on the gap between the n^{th} and the $(n+1)^{th}$ singular value. Equation 2.49 indicates that extended observability matrix can be determined as $U_n = O_s T$ and the estimated state sequence as $\tilde{X}_s = S_n V_n^T$.

PO-MOESP In contrast to N4SID, PO-MOESP aims at recovering the column space of the extended observability matrix using the orthogonal projection of the future input Hankel matrix $U_{s,s,N}$ defined as [16]:

$$\Pi_{U^{\perp}_{s,s,N}} = I_N - U^T_{s,s,N}(U_{s,s,N}U^T_{s,s,N})^{-1}U_{s,s,N}. \tag{2.50}$$

Multiplying the data equation of Equation 2.43 with the orthogonal projection gives:

$$Y_{s,s,N}\Pi_{U^{\perp}_{s,s,N}} = O_s X_{s,N}\Pi_{U^{\perp}_{s,s,N}} + T_s U_{s,s,N}\Pi_{U^{\perp}_{s,s,N}} + S_s E_{s,s,N}\Pi_{U^{\perp}_{s,s,N}}. \tag{2.51}$$

Utilizing the characteristics of the orthogonal projection as defined in Equation 2.16, the component $T_s U_{s,s,N}\Pi_{U^{\perp}_{s,s,N}}$ can be nullified. The instrumental-variable Z_N of Equation 2.45 is employed to remove the term $S_s E_{s,s,N}\Pi_{U^{\perp}_{s,s,N}}$. By applying this instrumental variable, Equation 2.51 can be deconstructed into:

$$Y_{s,s,N}\Pi_{U^{\perp}_{s,s,N}} Z_N^T = O_s X_{s,N}\Pi_{U^{\perp}_{s,s,N}} Z_N^T + S_s E_{s,s,N}\Pi_{U^{\perp}_{s,s,N}} Z_N^T. \tag{2.52}$$

Satisfying the first conditions set in Equation 2.44, it can be assumed that there exists no correlation between the noise term and the past input and output data. This leads to the conclusion that $E_{s,s,N}\Pi_{U^{\perp}_{s,s,N}} Z_N = 0$. Consequently, what persists is:

$$Y_{s,s,N}\Pi_{U^{\perp}_{s,s,N}} Z_N^T = O_s X_{s,N}\Pi_{U^{\perp}_{s,s,N}} Z_N^T. \tag{2.53}$$

Considering that $s \geq n$, implying that the number of block rows exceeds the system order, it can be asserted that:

$$\lim_{N\to\infty} \frac{1}{N} Y_{s,s,N} \Pi_{U^{\perp}_{s,s,N}} Z_N^T = \lim_{N\to\infty} \frac{1}{N} O_s X_{s,N} \Pi_{U^{\perp}_{s,s,N}} Z_N^T. \tag{2.54}$$

Due to the second condition of Equation 2.44, it can be assumed that:

$$\text{rank}(\lim_{N\to\infty} \frac{1}{N} X_{s,N} \Pi_{U^{\perp}_{s,s,N}} Z_N^T) = n. \tag{2.55}$$

Application of the Sylvester's inequality of Equation 2.3 to Equation 2.54 shows that:

$$\text{rank}(\lim_{N\to\infty} \frac{1}{N} Y_{s,s,N} \Pi_{U^{\perp}_{s,s,N}} Z_N^T) = n. \tag{2.56}$$

The properties of the instrumental variable Z_N outlined in Equation 2.44 guarantee that the multiplication by Z_N preserves the rank of Equation 2.54. Consequently the following relationship holds [16]:

$$\text{range}(\lim_{N\to\infty} \frac{1}{N} Y_{s,s,N} \Pi_{U^{\perp}_{s,s,N}} Z_N^T) = \text{range}(O_s). \tag{2.57}$$

From this relation, it becomes evident that an asymptotically unbiased estimate of the column space of O_s can be obtained from the SVD of the matrix $Y_{s,s,N} \Pi_{U^{\perp}_{s,s,N}} Z_N^T$.

$$Y_{s,s,N} \Pi_{U^{\perp}_{s,s,N}} Z_N^T = U_n \Sigma_n V_n^T = O_s T. \tag{2.58}$$

It is numerically more efficient to compute the SVD of the RQ factorization. The proposed RQ factorization is given as [16]:

$$\begin{bmatrix} U_{s,s,N} \\ \begin{bmatrix} U_{0,s,N} \\ Y_{0,s,N} \end{bmatrix} \\ Y_{s,s,N} \end{bmatrix} = \begin{bmatrix} R_{11} & 0 & 0 & 0 \\ R_{21} & R_{22} & 0 & 0 \\ R_{31} & R_{32} & R_{33} & 0 \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \end{bmatrix}, \tag{2.59}$$

such that it results in:

$$Y_{s,s,N} \Pi_{U^{\perp}_{s,s,N}} Z_N^T = R_{32} R_{22}^T, \tag{2.60}$$

$$\text{range}\left(\lim_{N\to\infty} \frac{1}{\sqrt{N}} R_{32}\right) = \text{range}\left(O_s\right). \tag{2.61}$$

The proof for Equation 2.60 comes, among other things, from the properties of the orthogonal Q matrices in the RQ factorization [16]:

- $Q_i Q_j^{\mathrm{T}} = 0$, for $i \neq j$.
- $Q_i Q_i^{\mathrm{T}} = I$.
- The sum of the individual orthogonal matrices' transpositions results in the identity matrix: $Q_1^T Q_1 + Q_2^T Q_2 + Q_3^T Q_3 + Q_4^T Q_4 = I_n$.

Based on Equation 2.59, the Hankel matrix of the future input data is equal to:

$$U_{s,s,N} = R_{11} Q_1. \tag{2.62}$$

Therefore, because of the definition for the future Hankel matrix of the input data and the attributes of the orthogonal Q matrices, the projection of the future input becomes:

$$\begin{aligned}\Pi_{U^{\perp}_{s,s,N}} &= I_N - Q_1^T R_{11}^T (R_{11} Q_1 Q_1^T R_{11}^T)^{-1} R_{11} Q_1, \\ \Pi_{U^{\perp}_{s,s,N}} &= I_N - Q_1^T Q_1, \\ \Pi_{U^{\perp}_{s,s,N}} &= Q_2^T Q_2 + Q_3^T Q_3 + Q_4^T Q_4.\end{aligned} \tag{2.63}$$

Derived from Equation 2.59, the Hankel matrix of the past output data is equal to:

$$Y_{s,s,N} = \left(R_{31} Q_1 + R_{32} Q_2 + R_{33} Q_3\right). \tag{2.64}$$

When the expression for the orthogonal projection of the future input Hankel matrix of Equation 2.63, the equation for the Hankel matrix of the past data of Equation 2.64 and the instrumental variable of Equation 2.45 are substituted into the expression for $Y_{s,s,N}\Pi^{\perp}_{U_{s,s,N}} Z_N^T$, it gives:

$$\begin{aligned}Y_{s,s,N}\Pi^{\perp}_{U_{s,s,N}} Z_N^T =& \left(R_{31} Q_1 + R_{32} Q_2 + R_{33} Q_3\right) \\ & \times \left(Q_2^T Q_2 + Q_3^T Q_3 + Q_4^T Q_4\right) Z_N^T, \\ =& \left(R_{32} Q_2 + R_{33} Q_3\right) \left(Q_1^T R_{21}^T + Q_2^T R_{22}^T\right), \\ =& R_{32} R_{22}^T.\end{aligned} \tag{2.65}$$

This concludes the proof for Equation 2.60.

To demonstrate the validity of Equation 2.61, it is necessary to have an input signal u that remains persistently exciting. A persistently exciting input signifies that the input should vary over time, encompassing a broad spectrum of values, thereby providing valuable insights into diverse facets of the system's behaviour. When this condition is met, it ensures the invertibility of the following matrix:

$$\lim_{N\to\infty} \frac{1}{\sqrt{N}} R_{22}.$$

Utilizing Sylvester's inequality as per Equation 2.3 demonstrates the following:

$$\begin{aligned}\operatorname{rank}\left(\lim_{N\to\infty} \frac{1}{\sqrt{N}} \mathcal{O}_s X_{s,N} \Pi^{\perp}_{U_{s,s,N}} \begin{bmatrix} U_{0,s,N} \\ Y_{0,s,N} \end{bmatrix}^{\mathrm{T}}\right) &= \operatorname{rank}\left(\lim_{N\to\infty} \frac{1}{\sqrt{N}} R_{32}\right), \\ &= n.\end{aligned} \tag{2.66}$$

This augmentation, together with Equation 2.54 and Equation 2.57, yields the desired result for proving Equation 2.61. The asymptotically unbiased estimate of the column space of O_s can be obtained from the SVD of the matrix R_{32}:

$$R_{32} = USV^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \simeq U_n S_n V_n^T. \tag{2.67}$$

The relationship between the column spaces can be expressed as $U_n = O_s T$. The singular values in the S matrix are inspected to estimate n, the order of the model, depending on the gap between the n^{th} and the $(n+1)^{th}$ singular value [16].

Estimate system matrices

In both N4SID and PO-MOESP methodologies, the observability matrix is obtained using subspace decomposition, leading to the expression $U_n = O_s T$, with $T \in \mathbb{R}^{n\times n}$ being a non-singular similarity transformation matrix. From the observability matrix, the system matrices can be obtained. The procedure starts by estimating A, C from:

$$U_n = O_s T = \begin{bmatrix} CT \\ CT(T^{-1}AT) \\ \vdots \\ CT(T^{-1}AT)^{s-1} \end{bmatrix} = \begin{bmatrix} C_T \\ C_T A_T \\ \vdots \\ C_T A_T^{s-1} \end{bmatrix}. \tag{2.68}$$

The matrix $\widehat{C}_T$ equals the first l rows of U_n, that is, $\widehat{C}_T = U_n(1 : l, :)$. The matrix $\widehat{A}_T$ is computed by solving the following equation, which, due to the condition $s < n$, has a unique solution:

$$U_n(1 : (s-1)l, :)A_T = U_n(l+1 : sl, :). \tag{2.69}$$

Given that the system matrices $\widehat{A}_T$ and $\widehat{C}_T$ are known, The matrices $\widehat{B}_T, \widehat{D}_T$ and the initial state $x_T(0)$ can be computed by solving the following least-square problem:

$$\begin{aligned} y(k) =& C_T A_T^k x_T(0) + \left(\sum_{\tau=0}^{k-1} u(\tau)^T \otimes C_T A_T^{k-\tau-1}\right) \text{vec}\,(B_T) + \left(u(k)^T \otimes I_\ell\right) \text{vec}\,(D_T) \\ &+ (\sum_{\tau=0}^{k-1} C_T A_T^{k-\tau-1} K_T e(\tau)) + e(k). \end{aligned} \tag{2.70}$$

This equation can be simplified by defining:

$$\begin{aligned} \phi(k)^T &= \left[\widehat{C}_T \widehat{A}_T^k \quad (\textstyle\sum_{\tau=0}^{k-1} u(\tau)^T \otimes C_T A_T^{k-\tau-1}) \quad (u(k)^T \otimes I_l)\right], \\ \theta &= \begin{bmatrix} x_T(0) \\ \text{vec}(B_T) \\ \text{vec}(D_T) \end{bmatrix}, \end{aligned} \tag{2.71}$$

such that Equation 2.70 becomes:

$$y(k) = \phi(k)^T \theta + (\sum_{\tau=0}^{k-1} C_T A_T^{k-\tau-1} K_T e(\tau)) + e(k). \tag{2.72}$$

In this context, the error term $e(k)$ is not correlated with $\phi(k)$. It can be omitted from the least-square solution due to a fundamental assumption inherent in linear regression and least-squares estimation [16]. This assumption relates to specific characteristics of the error term, including its independence, zero mean, normal distribution, and constant variance. Consequently, it becomes reasonable to exclude the error term from the least-squares estimation process. The primary objective of the least-squares method is to estimate the parameters of the underlying model while accommodating random variability caused by the error. In essence, the least-squares estimation seeks to identify parameter values that minimize the sum of squared errors, implicitly assuming that the errors conform to the mentioned assumptions.

Consequently, by employing the characterization in conjunction with the available output data, it becomes feasible to solve θ within a least-squares framework:

$$\min_\theta \frac{1}{N} \sum_{k=0}^{N-1} \left\| y(k) - \phi(k)^{\mathrm{T}} \theta \right\|_2^2. \tag{2.73}$$

By solving the optimization problem, the system matrices $\widehat{C}_T$ and $\widehat{D}_T$ along with the initial state $x_T(0)$ can be extracted. It is essential to highlight that, according to the principles of SVD, an alternate set of system matrices with a different similarity transformation could be derived if $U_n S_n^{\frac{1}{2}}$ is taken equal to the extended observability matrix O_s [16].

The remaining component of the state-space model in the innovation form described in Equation 2.31 that requires determination is the Kalman gain denoted as K. Using the estimated state matrices, assessing the state sequence $\widehat{X}_{i,N}$ while disregarding both the noise term and the Kalman gain is possible. The residuals, corresponding to the error term, can be calculated by solving a least-squares problem with the estimated system matrices:

$$\begin{bmatrix} \widehat{W}_{s,1,N-1} \\ \widehat{V}_{s,1,N-1} \end{bmatrix} = \begin{bmatrix} \widehat{X}_{s+1,N} \\ Y_{s,1,N-1} \end{bmatrix} - \begin{bmatrix} \widehat{A}_T & \widehat{B}_T \\ \widehat{C}_T & \widehat{D}_T \end{bmatrix} \begin{bmatrix} \widehat{X}_{s,N-1} \\ U_{s,1,N-1} \end{bmatrix}. \tag{2.74}$$

These residuals can be used to estimate the covariance matrices as follows:

$$\begin{bmatrix} \widehat{Q} & \widehat{S} \\ \widehat{S}^{\mathrm{T}} & \widehat{R} \end{bmatrix} = \lim_{N\to\infty} \frac{1}{N} \begin{bmatrix} \widehat{W}_{s,1,N} \\ \widehat{V}_{s,1,N} \end{bmatrix} \begin{bmatrix} \widehat{W}_{s,1,N}^{\mathrm{T}} & \widehat{V}_{s,1,N}^{\mathrm{T}} \end{bmatrix}. \tag{2.75}$$

Using the estimated covariance matrices, the DARE can be derived by solving:

$$\widehat{P} = \widehat{A}_T \widehat{P} \widehat{A}_T + \widehat{Q} - (\widehat{S} + \widehat{A}_T \widehat{P} \widehat{C}_T^T)(\widehat{C}_T \widehat{P} \widehat{C}_T^T + \widehat{R})^{-1}(\widehat{S} + \widehat{A}_T \widehat{P} \widehat{C}_T^T)T. \tag{2.76}$$

The estimate of the Riccati solution $\widehat{P}$ can subsequently be used to obtain an estimate of the Kalman gain $\widehat{K}$ for the system:

$$\widehat{K}_T = (\widehat{S} + \widehat{A}_T \widehat{P} \widehat{C}_T^T)(\widehat{R} + \widehat{C}_T \widehat{P} \widehat{C}_T^T)^{-1}. \tag{2.77}$$

As a result, the estimates of the system matrices $\widehat{A}_T, \widehat{B}_T, \widehat{C}_T$ and $\widehat{D}_T$ along with the initial state $\widehat{x}_T(0)$ and the Kalman gain $\widehat{K}$ are determined in this step. From here, the output $\widehat{Y}_{i,s,N}$ can be simulated for model verification and validation. This signifies the conclusion of the algorithms for both N4SID and PO-MOESP.

2.3. The Granger causality analysis

The Granger causality analysis is a statistical method employed to establish whether one time series can predict another time series, forming the statistical relationship between time series [19]. The fundamental premise of this technique is that if past values of a time series Y possess the ability to predict another time series X beyond what would be expected by the past values of X alone, then Y is said to 'Granger-cause' X. This method was initially implemented in econometrics and finance to determine causal relationships between economic variables. However, it is also possible to apply Granger causality analysis to infer directed functional connectivity between brain regions by utilising the statistical technique to examine whether the activity over time of one brain region exerts an influence on the activity of another brain region [11, 52, 53].

The Granger causality method traces its origins to initially debatable principles that serve as the foundation for the definition of causality. To turn these truths into testable conditions, two essential notations were introduced: Ω_n, representing the entirety of knowledge available in the universe up to time n, and $\Omega_n - Y_n$, signifying the subset of this knowledge that excludes the values of the time series Y at or before time n. Subsequently, the following hypotheses are presumed to hold within the Granger causality framework:

- The past and present may cause the future, but the future cannot cause the past.

- Ω_n contains no redundant information.

As a starting point for defining causality, this suggests:

Definition 2.1. Y_n is said to Granger-cause X_{n+1} concerning the entirety of knowledge available in the universe Ω_n up to time n if Prob $(X_{n+1} \in A|\Omega_n) \neq$ Prob $(X_{n+1} \in A|\Omega_n - Y_n)$ for a set A [54].

Thus, for causation to occur, in mathematical terms, Y_n causes X_{n+1}, the variable Y_n needs to have unique information about what value X_{n+1} will take in the near future [55].

However, Definition 2.1 cannot be used with actual data since all knowledge in the universe up to time n is not available. Therefore, a set of restrictions is proposed to make the definition workable for real-time data by defining J_n. J_n represents the information set available at time n that does not include any components of Y_n, which denotes the variable containing specific information gathered up to time n. The conditional probability distribution of X_{n+1} given J_n is denoted by $F(X_{n+1}|J_n)$. Now, the following conditions can be obtained:

Definition 2.2. Y_n is said to be the prima facie cause of X_{n+1} with respect to the information set $J_n + Y_n$ if $F(X_{n+1}|J_n) \neq F(X_{n+1}|J_n + Y_n)$ [54].

2.3.1. Assumptions for Granger causality

Several requirements must be met to perform the Granger causality analysis on state-space models to ensure the results are valid. These include:

- The relationship between the time series being analysed should be linear. Even if a system comprises both linear and nonlinear connections, Granger causality measures can still detect the linear parts of the system. However, their ability to identify weak nonlinear connections may be limited.
- The eigenvalues of the system are strictly inside the unit circle such that the system is asymptotically stable.
- The time series under consideration must, in principle, be stationary, which means that their statistical characteristics, such as the mean and variance, do not fluctuate over time [56]. In certain situations, the stationarity requirement may be relaxed, and it can partly be mitigated by using a windowing approach to take a dataset and partition it into subsections [11].
- The analysis is based on statistical tests, and the data size should be large enough to ensure that the statistical dependencies are significant. A large dataset size is necessary to determine the order with the best possible performance, such that the model order will not substantially affect the model's performance [53]. However, it is essential to note that selecting a model order that is too large can lead to overfitting, where the model not only fits the underlying relationship between variables in the system but also fits the noise unique to each observed sample [57].
- The error terms of the model, denoted as e_k, should exhibit a normal distribution. If the errors are not normally distributed, alternative methods such as non-parametric Granger causality can be used [58].

2.3.2. Application of Granger causality in state-space models

The Granger causality is predominantly applied within the framework of linear autoregressive modelling. Nevertheless, this approach has limitations as outlined in Section 1.1. Consequently, it is possible to apply the Granger causality analysis to state-space models. The materials of this section are borrowed from Barnett and Seth [11].

The starting point for Granger causality for state-space models is the discrete-time LTI system of Equation 2.1. Furthermore, the method makes use of the noise covariance matrix:

$$\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} = \mathbb{E}\left[\begin{bmatrix} w(k) \\ v(k) \end{bmatrix} \begin{bmatrix} w(k)^T & v(k)^T \end{bmatrix} \right], \tag{2.78}$$

where $Q = \mathbb{E}[ww^T]$, $R = \mathbb{E}[vv^T]$ and $S = \mathbb{E}[wv^T]$.

The state-space model and its associated covariance matrix can be transformed into the innovation form represented by Equation 2.31. In this transformation, the model prediction error covariance $\Sigma = \mathbb{E}[e(k)e(k)^T]$ is introduced. The innovation form represents a particular form of Equation 2.1, with specific relationships defined as $w(k) = Ke(k)$ and $v(k) = e(k)$ such that the noise covariance matrices are $Q = K\Sigma K^T$, $S = K\Sigma$ and $R = \Sigma$ [11].

It can be tested if state x_j is a Granger cause to state x_i for all $i \neq j$ by using the conditional Granger causality measure F_{ij}:

$$F_{ij} \equiv F_{x_j \to x_i | \text{ all other } x} = \log\left(\frac{\det \Sigma_{ii}^R}{\det \Sigma_{ii}}\right). \tag{2.79}$$

If $F_{ij} \neq 0$, then it can be concluded that state x_i is Granger caused by state x_j, in other words, this means that state x_i statistically depends on state x_j. In Equation 2.79, the Σ_{ii} and Σ_{ii}^R denote the terms on the diagonal (i,i) of the full model prediction error covariance Σ and its reduced model counterpart Σ^R. The reduced model is obtained by eliminating the influence of x_j from the full model. More simply stated, this is achieved by removing the j^{th} row of C in Equation 2.31.

As evident, the task at hand involves the determination of prediction error covariances, denoted as Σ and Σ^R. Given the system parameters in the state-space model, namely A, C, Q, R, S, these covariances can be derived by solving the DARE of Equation 2.21 for prediction error covariance matrices [59]. Then, the prediction error covariance and the Kalman gain are given by [42]:

$$\begin{aligned} \Sigma &= CPC^T + R, \\ K &= (APC^T + S)\Sigma^{-1}. \end{aligned} \tag{2.80}$$

After repeating Equation 2.79 for $j = 1, 2, \ldots, m$, where m corresponds to the total number of states, the statistical dependency between states can be evaluated. This is achieved by estimating conditional Granger causality measures F_{ij} for all pairs of states (i, j) and forming a matrix that encapsulates these values, whose diagonals are not in consideration. Notably, an essential characteristic of the Granger causality matrix is its invariance when subjected to a similarity transformation of the system [14].

3

Experimental and algorithmic framework

In the field of cognitive neuroscience, where the intricate dynamics of human behaviour remain not completely clear and insightful, experimental precision and algorithmic innovation are of paramount importance in unravelling the intricacies of processes like perception, memory and decision-making. These processes rely on a wide array of brain regions, highlighting the interdependence of the brain's complex network. To model the brain's dynamics and gain insights into the brain's cortical network, a crucial initial step involves conducting a neuroscience experiment. Such experiments may entail the generation and manipulation of large amounts of both raw and processed data. There is wide variability in the data types collected, from the form and behaviour of individual neurons to measures of brain functioning [60]. Within the scope of this research, a foundational model will be developed using data from an experiment in which participants passively participate in a task involving brain stimulation, allowing for the incorporation of measurable exogenous input.

After data collection, the dataset can undergo further processing through algorithmic design. Algorithmic design refers to the systematic methodology and mathematical procedures used for problem-solving and engineering algorithms. In this research, the focus is on developing a data-driven dynamic mathematical model that characterizes the brain's cortical network by leveraging EEG data in conjunction with exogenous input. The model has been developed by exploring various methodological approaches and incorporating customized techniques based on the methodologies discussed in Chapter 2.

This chapter is structured into two sections, providing an in-depth examination of the experimental framework and the algorithmic design, forming the study's foundation for studying the brain's dynamic behaviour. The first section offers insights into the experimental setup that generated the dataset under consideration. This dataset not only serves as a critical testbed for evaluating the algorithm's performance but also lays the groundwork for potential extensions to more complex experiments in the future. The second section of this chapter explains the core of the study, presenting the four key steps that constitute the developed approach to modelling brain dynamics and cortical connectivity during simple stimulation tasks. The proposed algorithm is not restricted to the dataset introduced in the first section. It can be applied to various datasets, thereby enabling the exploration of brain connectivity in more complex tasks and providing insights into the intricacies of the brain's involvement in cognitive functioning and motor activities.

Overall, this chapter establishes the foundation for delving into the intricate dynamics

of the brain, enabling the modelling of cortical connectivity during the passive execution of tasks. This is achieved through the presentation of a developed algorithm and an in-depth description of the experimental approach employed to assess the algorithm's performance.

3.1. Experimental framework

This research utilises data from an earlier conducted experiment to assess the developed algorithm's efficacy in modelling human brain dynamics while performing a task. The dataset originates from an experiment conducted by Vlaar et al. [20], which focused on quantifying cortical activation in response to continuous robotic wrist joint manipulation after a stroke. Throughout the experiment, a robotic manipulator applied continuous periodic disturbances to the wrist to provide sensory stimulation. The signals from the robotic manipulator were recorded, while the simultaneous measurement of brain activity was achieved using EEG [20].

Within the scope of this research, the incorporation of measurable inputs is crucial for modelling brain dynamics. The input should comprise data that can be quantified, allowing for precise examination and easy integration into the computational model. This quantifiability simplifies the assessment of the brain's responses to particular stimuli or task demands while reducing the impact of extraneous noise or unexplained variables. To ensure the model's stability, it is advantageous to focus on studying SSR, which are evoked oscillatory responses to modulated or repetitive stimuli [20]. Hence, it is essential to ensure that the task exhibits a stable and synchronized rhythmic pattern. When aiming to model the dynamic behaviour of the brain during a specific task, it is necessary to distinguish the neural responses during task engagement from the baseline resting state. Therefore, selecting an experimental setup where the input stimulus remains persistently exciting is essential, meaning that its frequency spectrum encompasses a sufficiently broad range of harmonics. This choice also serves to maintain the model's stability and continuity.

It is important to emphasize that one of the foundational assumptions underlying open-loop subspace identification methods, as articulated in Section 2.2.2, is the necessity for a persistently exciting input. Subspace identification techniques are designed to estimate the parameters of dynamic systems, such as mathematical models describing input-output relationships. To achieve accurate parameter estimates, it is critical to have input signals that thoroughly explore the system's dynamic behaviour. A persistently stimulating input offers the requisite diversity and coverage of input patterns for practical parameter estimation [16].

The rationale for selecting the dataset from the experiments conducted by Vlaar et al. [20] aligns with the criteria to have a dataset where participants engage in a task involving an exogenously applied input that is easily quantifiable, continuously being applied and persistently exciting. Furthermore, the research demonstrated that the periodic angular disturbances during the passive task induce sustained oscillatory brain responses, supporting the successful elicitation of the intended SSR [20].

3.1.1. Participants and experimental setup

The research of Vlaar et al. [20] included the participation of ten unimpaired individuals who had not experienced any history of stroke, aged between 22 and 25 years, with a gender distribution comprising five men and five women. The local ethics committee granted ethical clearance for the study, and all participants provided written informed consent before involvement. A concise overview of the relevant aspects of the experimental protocol is provided here, while a detailed description of the protocol is available in Vlaar et al. [20].

A robotic manipulator applied continuous periodic angular disturbances to the wrist during the experiment. To facilitate this procedure, participants were seated in a manner

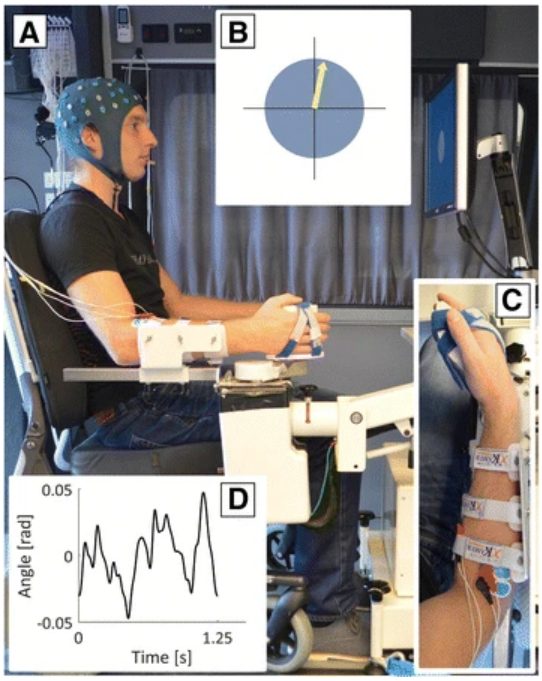


Figure 3.1: Experimental setup. (a) The forearm of the participant is strapped into an armrest, and the hand is strapped to the handle of the robotic wrist manipulator, requiring no force to hold this handle. (b) Visual feedback was presented to the participants. The yellow arrow is only visible when the subject applies torque. Therefore, it bore limited relevance in the context of the passive execution of the task. (c) Close-up of the arm in the robotic manipulator in which the arm is placed in the neutral angle of 20° wrist flexion. (d) The plot of one complete cycle of the angle of the disturbance signal applied to the wrist. Retrieved from [20].

where their dominant forearm was secured on an armrest, with their hands firmly attached to the handle of the robotic wrist manipulator known as the "Wristalyzer" (manufactured by MOOG Inc., headquartered in Nieuw-Vennep, The Netherlands). The experiment incorporated a passive task, which required no active participant involvement. Specifically, participants were instructed to maintain a relaxed state in their wrists and disregard the applied perturbations. Throughout the experiment, the robotic manipulator continued to apply continuous and periodic perturbations to the wrist, generating sensory stimuli. Figure 3.1 represents the experimental setup conducted in the experiment.

Each participant's involvement in the study encompassed 20 recorded trials. Within each trial, ten consecutive periods of the disturbance signal were administered and recorded for subsequent analysis. The cortical activity was sampled at 2048 Hz from 64 electrodes, arranged according to a subset of the extended 10/20 system, using the EEG amplifier Refa (manufactured by TMSi, headquartered in Oldenzaal, the Netherlands). The angle of the wrist manipulator was, via optical isolation amplifiers (manufactured by TMSi, headquartered in Oldenzaal, the Netherlands), recorded by the same EEG amplifier.

3.1.2. The input signal

The exogenous input signal for this study can be conceptualised as the periodic angular perturbations applied by the robotic manipulator at the wrist joint. This perturbation signal was constructed as a multisine signal characterised by random-phase attributes, combining several sinusoids, each exhibiting a random phase. This signal design aimed to elicit responses from the sensory system within a frequency range relevant to the control of movement [20].

The effectiveness of controlling wrist movements at high frequencies is inherently constrained by factors such as limb mass and the rapid contraction capacity of muscles. To

account for lower-frequency phenomena, the perturbation signal was configured with a periodicity of 1.25 seconds, resulting in a corresponding frequency resolution of 0.8 Hz. This choice represented a trade-off between frequency precision and the number of signal cycles that can be accommodated within a given measurement duration [20]. The multisine signal incorporated sinusoids spanning a wide frequency range from 0.8 Hz to 19.2 Hz, ensuring that the signal continually evolved over time and is, therefore, persistently exciting.

In the experimental setup, angular perturbations were applied at the neutral angle, defined explicitly as a wrist flexion of $20°$, as illustrated in Figure 3.1.c. These angular perturbations exhibited a root-mean-square value of 0.02 radians, indicating that the angular perturbations caused the wrist to deviate from its neutral angle by 0.02 rad on average. Figure 3.1.d illustrates a single cycle of 1.25 seconds of the multisine disturbance signal, serving as the input signal for subsequent analyses.

3.2. Algorithmic design

As a response to the limitations associated with traditional methods for assessing directed functional connectivity in the brain, as elaborated in Section 1.1, this section introduces an alternative approach to model brain dynamics by tailoring existing theoretical frameworks. The algorithm to perform the analysis comprises several steps. Each step is extensively elaborated upon in distinct paragraphs. This section terminates with a block diagram, providing an overview of the complete framework to model brain dynamics and connectivity. The Matlab code for implementing each step of the mathematical framework can be accessed through the link provided in Appendix D.

3.2.1. Processing EEG data

The initial processing phase entails converting raw data measured with EEG into a format more suitable for further analysis within two steps. The first step consists of preprocessing the raw data to eliminate extraneous noise, while the second averages the data to derive the SSR, aiming for accurate neural signal extraction. Following the formation of SSRs, an output designated as y is generated, serving as the fundamental basis for constructing a state-space model to represent the dynamic activities occurring within the brain. The input and output for the initial step are summarised in the block diagram of Figure 3.2.

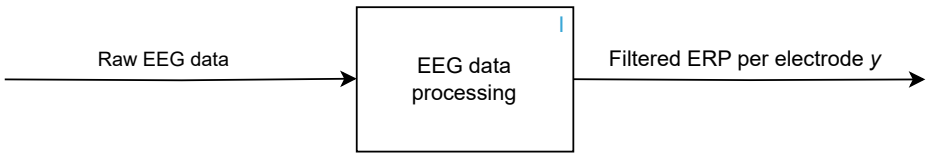


Figure 3.2: Block diagram of step 1, illustrating the input and output of the first stage of the algorithmic framework.

The preprocessing stage unfolds through a series of filtering techniques. The specific steps to be applied may vary depending on the dataset. Nonetheless, the suggested general approach for the algorithm encompasses the following steps:

1. Band-pass filter the EEG data to allow frequencies within a particular band to pass through. The band-pass filter settings will depend on the specific task performed.
2. Band-stop filter the EEG data in narrow bands to remove line noise and the harmonic of the line noise.
3. Reject bad channels with high-frequency noise and outliers for further analysis by examining the data's spectral characteristics.

4. Re-reference the remaining EEG signal to the common average, thereby rendering the electrical activity in relation to the average activity across all electrodes.
5. Perform independent component analysis (ICA) to remove artefacts embedded in the data without removing the affected data portions. These artefacts can consist of muscle activity, eye blinks or eye movement. The core idea of ICA is to assume that the observed signal is a linear combination of statistically independent source signals. In the case of EEG, these source signals correspond to the underlying brain activities and artefacts. These mixed signals should be separated for explicit artefact removal.
6. Partition the dataset into epochs corresponding to the recurrent intervals of the input signal.

The next phase involves generating SSRs. These SSRs represent the system's behaviour that emerges from evoked neural responses, typically induced by external or internal stimuli when it has reached a constant state. To estimate the SSR, the experimenter must conduct many trials, segment the trials aligning with the exogenous input into epochs, and average the results across these independent epochs [61]. This causes random brain activity to be averaged out and the relevant waveform to remain. Within the generation of SSRs, it is crucial to ensure that a sufficient number of datasets are averaged together to enhance the signal-to-noise ratio while concentrating on the neural responses of particular interest.

The combination of preprocessing the raw EEG data and determining the SSR for each electrode results in the output data y. This output data is ready for further analysis and modelling within the subsequent stages of the algorithm. This procedure ensures the dataset is appropriately prepared for additional modelling procedures and guarantees that the data quality is improved.

The first step can be performed using Matlab R2023A (The Mathworks, Inc., Natick, MA, USA) and the specialized Matlab toolbox for processing EEG data called EEGLAB [62].

3.2.2. Performing subspace identification

The algorithm's second step involves forming a state-space model using subspace identification. This approach should be applied to estimate the inherent system dynamics from the measured input and output data. In this context, the known components include the exogenous input u, related to the performed task, and the output y, which captures the cortical activity recorded from the EEG electrodes. The objective is to construct a state-space model that encapsulates the dynamic processes by identifying the state matrices A, B, C, D, the Kalman gain K and the initial state $x(0)$ using subspace identification. The process of the second step, involving the inputs and desired outputs, is outlined in the block diagram presented in Figure 3.3.

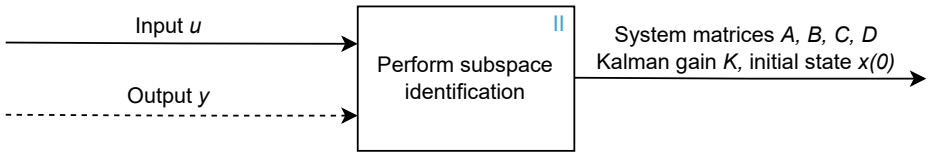


Figure 3.3: Block diagram of step 2, illustrating the input and output of the second stage of the algorithmic framework. A dashed line is used to signify that the input value is an output of a prior step.

The system of interest in this research is the human brain. A linear mathematical model

representing brain dynamics can be expressed as [63, 64]:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + w(k), \\ z(k) &= Hx(k) + \eta(k), \\ y(k) &= Lz(k) + Du(k) + v(k). \end{aligned} \tag{3.1}$$

where x denotes the latent variable, z represents the source activity corresponding to neural activity, and y corresponds to the output representing the electrical activity recorded by the electrodes placed at the scalp during EEG measurements. The matrices A, B, D represent the system matrices, while w, η, v are noise terms associated with the latent state, source signal, and output measurements, respectively. The matrix L is the lead-field matrix, facilitating the mapping of neural sources to EEG sensors on the scalp. The matrix H represents the mapping from the latent state x to the source activity z.

It is important to note that the brain source activity z is not directly identifiable through system identification. However, the latent variable x can be estimated. The latent variable, also referred to as the hidden state, is an unobservable variable in a mathematical model, inferred or calculated based on available data to capture underlying patterns or processes that are not directly observable.

To apply subspace identification, a modified state-space model is derived from Equation 3.1:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + Ke(k), \\ y(k) &= Cx(k) + Du(k) + e(k), \end{aligned} \tag{3.2}$$

where $C = LH$ and $e = L\eta + v$. The implementation of the Kalman gain K is derived from Equation 2.22. As a component of the algorithm, two subspace identification techniques are employed, specifically N4SID and PO-MOESP, to compare the performance. Both methods share the common goal of identifying the state space model of Equation 3.2, using the input and output data.

Both of the subspace identification methods involve a three-step process. The theoretical framework and working principles of these algorithms are extensively discussed in Section 2.2. In short, the procedure is initialised by data collection, followed by data embedding to construct the data equation, which consists of Hankel and Toeplitz matrices derived from the provided input and output data. Subsequently, subspace decomposition takes place using projection matrices and matrix factorization methods such as SVD and RQ factorization. Once the subspace decomposition is complete, the system matrices, the Kalman gain and the initial state can be identified. This, in turn, allows for the simulation of the hidden state sequence $\widehat{x}(k)$ and the output $\widehat{y}(k)$ for validation.

The parameters that need to be chosen are the model order n and the number of block rows in the Hankel matrices s. To determine the appropriate model order, an examination of the singular values derived from the subspace decomposition step in subspace identification can be conducted. This entails identifying those values of n that exhibit a significant gap between the n^{th} and $(n+1)^{th}$ singular value. In scenarios where multiple such values emerge, the selection process will hinge on a trade-off between achieving good performance, managing computational effort, and avoiding overfitting. For both N4SID and PO-MOESP, the same model order will be chosen for consistency in comparing the performance. The selection of s is critical, as it should satisfy the condition that $n < s \ll N$ [16].

When handling extensive datasets in this algorithm, which may encompass data from experiments conducted at different times involving, for example, distinct subjects interacting with the same device, it might become imperative to partition the data into separate batches.

Subdividing the dataset into smaller batches can be advantageous to train the model on more data, prevent excessive memory usage and enhance computational efficiency [65].

Subspace identification serves to identify the state-space model of Equation 3.2, using the input u and the output y. The state-space model represents the dynamics of the brain activity of a participant while engaged in a task. After determining system matrices using subspace identification, the output $\widehat{y}$ can be estimated and subsequently compared to the actual output for validation. Successful validation, signified by effective performance, indicates that the algorithm provides a dependable linear model of brain dynamics suitable for a wide range of analytical applications.

The N4SID method can be performed with the Matlab R2023A System Identification Toolbox (The Mathworks, Inc., Natick, MA, USA). The PO-MOESP method can be applied using the LTI System Identification Toolbox 2.4 [65].

3.2.3. Mapping from latent to source

The third step of the algorithm involves translating the latent state information obtained through subspace identification into details about the source signals, enabling the extraction of insights into the interconnections among brain sources and regions. This step closely corresponds to the traditional "forward problem", reflecting the objective to derive meaningful conclusions regarding the connectivity among distinct brain sources, as opposed to focusing on the connectivity among latent states.

The previous subspace identification step estimates the state matrix C. However, within the matrix C, the information of L and H, as represented in Equation 3.1, are ambiguously mixed via $C = LH$ [63]. The lead-field matrix, denoted as L, can be estimated from prior knowledge of the head model, source model and sensor positions of the electrodes at the scalp [66]. The inputs necessary for performing the third step, as well as the desired outputs, are summarized in the block diagram of Figure 3.4.

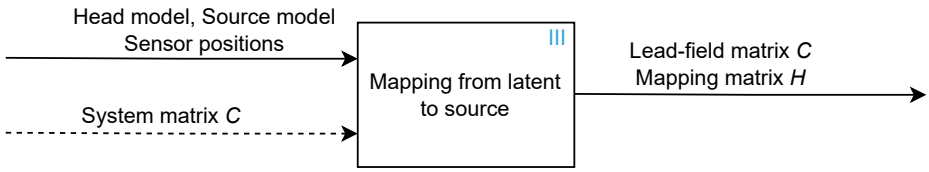


Figure 3.4: Block diagram of step 3, illustrating the input and output of the third stage of the algorithmic framework. A dashed line is used to signify that the input value is an output of a prior step.

In the context of EEG, the lead-field matrix L is a fundamental component of the forward modelling procedure. It operates as a linear operator, effectively establishing a mathematical relationship that maps source activations within the brain to the corresponding electrical potentials detected at sensor locations. To compute the lead-field matrix for each participant, five steps should be conducted [67]:

1. Create a mesh representing the head based on available magnetic resonance imaging (MRI) data defining the participant's head characteristics to create a head model. From the MRI data, geometric surface meshes are created that segment the head into distinct components, including the scalp, skull, cerebrospinal fluid, grey matter, and white matter. Electrical activity in the brain primarily originates from the grey matter.
2. Merge the geometrical (mesh) and electrical (electrodes placed) features to integrate the head model. Ensuring precise alignment of the electrodes with the MRI data is crucial in this phase to maintain accuracy.

3. Define the source model by selecting the location of sources within the grey matter compartment, as electrical activity in the brain primarily originates from the grey matter. The spatial resolution, in this context, is a parameter associated with the source locations and significantly impacts the computational requirements. In essence, the spatial resolution characterises the level of detail and granularity in the placement of source locations within the source space model. More densely sampled grids necessitate a greater allocation of computational resources and time.
4. To characterise brain sources within the grey matter, these sources can be mapped to established anatomical labels, as specified by a volumetric anatomical atlas. Such atlases provide a comprehensive spatial description of head geometry, with coordinates labelled according to specific schemes like Brodmann areas or the Talairach atlas. The mapping of these sources onto the atlas labels, including the choice of which atlas to employ and how it is interpolated onto the source model, is at the researcher's discretion.
5. Estimate the lead-field matrix from the results of the four steps above by solving the forward problem for each distinct source with, for example, the boundary element method.

The derivation of the transformation matrix H, representing the mapping from the latent to the source, becomes feasible via the equation $C = LH$. Since the C and L matrices are known respectively from subspace identification and the computation of the lead-field matrix, the H matrix can be calculated. An estimation of H from C with a known matrix L can be solved from a linear least-squares problem:

$$\min_{C} \frac{1}{2}||C - LH||_F^2. \tag{3.3}$$

Due to the under-determined nature of the forward problem, it is characterised by an infinite set of solutions capable of faithfully reproducing the observed signal [68]. In pursuit of obtaining a more precise representation of the mapping from latent to source, it is beneficial to impose constraints on the H matrix. These constraints can promote group sparsity in the rows of H, assuming that not all sources are synchronised. Only some sources are activated corresponding to EEG sensors in a brief time [63]. Consequently, by encouraging many source locations to be precisely zero, indicating inactive sources, this approach effectively localizes the more potent active sources within the brain. This results in a H matrix that contains specific zero rows corresponding to these inactive sources. To address this, the problem can be formulated and solved as a Group Lasso problem, incorporating an L1-norm penalty term [69]:

$$\min_{C} \frac{1}{2}||C - LH||_F^2 + \lambda \sum_{i=1}^{m} ||H_i^T||_2, \tag{3.4}$$

where m is the total number of sources. This problem is convex, given that the objective function comprises norm functions, which are inherently convex. In this equation, λ acts as a positive scalar that can be systematically adjusted to control the weight for penalizing certain rows of H to become entirely zero. Increasing λ drives more rows in H toward a state of complete zero. The objective is to identify an appropriate sparsity pattern among the rows in H. This can be achieved using the Bayesian information criterion (BIC) criterion, a trade-off between model fitting, the term $||C - LH||$ and model complexity. The BIC score of the problem of Equation 3.4 can be calculated as [13]:

$$\begin{aligned} \text{BIC}(\lambda) &= -nm - n \log\det \widehat{\Sigma}(\lambda) + k(\lambda)\log(n), \\ \widehat{\Sigma}(\lambda) &= \frac{1}{n}(C - L\widehat{H}(\lambda))(C - L\widehat{H}(\lambda))^T. \end{aligned} \tag{3.5}$$

Such that the H is chosen with the lowest BIC score. This strategy enables the identification of the matrix H, representing the most relevant and active source locations, ultimately improving source localisation accuracy in EEG analysis. After creating the transformation matrix H, it becomes possible to simulate the source activity z using the equation $z = Hx$, with the noise component being disregarded. This particular step can be executed with the purpose of establishing a relative measure of source activity.

The approach, encompassing both the computation of the lead-field matrix from the head model derived from MRI data and the calculation of the matrix H through the application of the Group lasso technique, facilitates the precise identification of the most relevant and active source locations, enhancing the interpretability and accuracy of source localization in EEG analysis.

The third step of the framework can be executed through the utilisation of Matlab R2023A (The Mathworks, Inc., Natick, MA, USA), in conjunction with the Matlab software toolboxes EEGLAB [62] for loading preprocessed EEG data and the Matlab toolbox FieldTrip [67] for conducting advanced computations related to the lead-field matrix.

3.2.4. Connectivity analysis

In the last step of the algorithm, the directed functional connectivity within the brain is assessed, with the goal of revealing statistical dependencies among various neural sources. This analysis relies on the state matrices identified via the subspace identification procedure in the second step. Additionally, it leverages the mapping from latent variables to source activity, accomplished in the third step through the lead-field matrix. To determine the functional connectivity, the process employs the Granger causality method for state-space models [11]. This method examines whether brain activity in one source can influence the activity in another source, quantified within the Granger causality measure F_{ij}. The inputs required for this final phase and the expected outcome are illustrated in Figure 3.5.

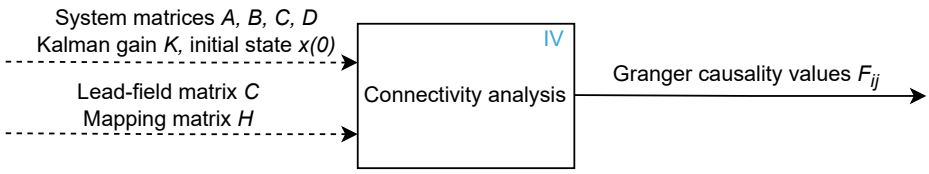


Figure 3.5: Block diagram of step 4, illustrating the input and output of the fourth stage of the algorithmic framework. A dashed line is used to signify that the input value is an output of a prior step.

As elucidated in Section 2.3.1, the Granger causality analysis requires access to the covariances of the noise related to the latent state, source and measurement noise components, none of which have been computed thus far. The research of Songsiri influences the computation of the covariance matrices [13], and it is further customized to suit the specific requirements of this study by incorporating the provided input data.

Considering the brain dynamics described in Equation 3.1, the covariances of the noise terms related to the latent state, the source, and the output, as denoted by Equation 2.78, are respectively represented by $\Sigma_w = \mathbb{E}[ww^T]$, $\Sigma_\eta = \mathbb{E}[\eta\eta^T]$, and $\Sigma_v = \mathbb{E}[vv^T]$. The state noise covariance Σ_w, where $w = Ke(k)$, can be determined by calculating the difference between the estimated state sequence $\widehat{x}(k)$ when including the Kalman gain and error term and when not, resulting from the state matrices and initial state from performing subspace identification. The covariance Σ_w can then be determined based on this difference. Utilising the identified state matrices, the covariance of the measurement noise Σ_e can be computed by assessing the covariance of the differences between the actual output y and a simulated

output $\widehat{y}$ generated from the determined state matrices obtained through subspace identification.

A mathematical relation exists between the noise characteristics in the EEG and source equations, following from $e = L\eta + v$, as outlined by Equation 3.2 [13]. This connection yields a relationship between noise covariances, which can be expressed as:

$$\Sigma_e = L\Sigma_\eta L^T + \Sigma_v. \tag{3.6}$$

Given that both Σ_e and the lead-field matrix L are known, the Σ_η and Σ_v within Equation 3.6 can be determined as part of an optimization problem. This optimization problem can be formally stated as [13]:

$$\begin{array}{ll} \text{minimize} & \left\|\Sigma_e - L\Sigma_\eta L^T - \Sigma_v\right\|_F \\ \text{subject to} & \Sigma_\eta \succeq 0, \quad \Sigma_v \succeq 0, \\ & \Sigma_\eta = \alpha I, \quad \Sigma_v \text{ is diagonal}. \end{array} \tag{3.7}$$

The optimization problem seeks to minimize the Frobenius norm between the covariance matrix Σ_e and the expression $L\Sigma_\eta L^T + \Sigma_v$, subjected to specific constraints. The two initial constraints, $\Sigma_\eta \succeq 0$ and $\Sigma_v \succeq 0$, are rooted in the property of covariance matrices being positive semi-definite. This property ensures that the covariance matrix accurately represents the variance and covariance of the underlying variables, preserving their statistical integrity. Two additional conditions are imposed on the variables to add further constraints and enhance problem uniqueness. The third constraint mandates that Σ_η must be a diagonal matrix scaled by a positive scalar α, denoted as $\Sigma_\eta = \alpha I$. This constraint is proposed assuming that the noise covariances exhibit a diagonal structure, signifying that the noise vectors η and v are uncorrelated. By enforcing these diagonal structures in the variables, the optimization problem introduces the possibility of obtaining a non-zero optimal value. Additionally, the problem adheres to the convexity criterion, with the cost objective and constraint set falling within the convex space. Consequently, this optimization problem can be efficiently solved for Σ_η and Σ_v.

Once the noise covariances have been computed, all the necessary information for calculating the Granger causality measure is available. The computation of the Granger causality measure using state-space models adheres to the details in Section 2.3.1. It consists of solving the DARE of Equation 2.21 for both the full model and the reduced model. For the full model, applied to the state-space representation of Equation 3.1, this results in:

$$P = APA^T - (APH^T + S)(HPH^T + \Sigma_v)^{-1} \times (HPA^T + S) + \Sigma_w. \tag{3.8}$$

Notably, S, representing the covariance between $w(t)$ and $v(t)$ as $S = \mathbb{E}[wv^T]$, is assumed to be zero. This assumption is grounded in the constraints that necessitate Σ_v to be diagonal [13]. Consequently, the prediction error covariance Σ for the full model is derived from:

$$\Sigma = HPH^T + \Sigma_\eta. \tag{3.9}$$

Upon transitioning to the reduced model, denoted by the superscript "R" in variables like H^R, the influence of brain source z_j is eliminated from the full model. Assuming that the source vector z corresponds to m distinct sources, an examination is conducted to determine whether a specific component, referred to as z_j, affects the remaining $m-1$ variables. This evaluation is performed by solving the DARE for the reduced model after removing z_j from the computations. The following series of steps, as described in [13], are executed:

1. Remove the state z_j from the original state-space model of Equation 3.1. This action is equivalent to deleting the j^{th} row of H and thereby obtaining H^R as the output matrix representing the mapping between the latent and source in the reduced model.
2. Determine the noise covariance of η in the reduced model, denoted as Σ_η^R, by removing the j^{th} row and column of the original Σ_η matrix.
3. Solve the DARE by implementing $(A, H, \Sigma_w, \Sigma_\eta^R)$ as input parameters to obtain the covariance error of the reduced model P^R.
4. Compute the reduced covariance error $\Sigma^R = H^R P^R (H^R)^T + \Sigma_\eta^R$.

After computing the prediction error covariance matrix Σ and the reduced prediction error covariance matrix Σ^R, the Granger causality measure F_{ij} can be calculated using the following formula:

$$F_{ij} \equiv F_{z_j \to z_i | \text{ all other } z} = \log\left(\frac{\det \Sigma_{ii}^R}{\det \Sigma_{ii}}\right). \tag{3.10}$$

The process of computing Σ and Σ^R, and filling in their values of the diagonal in F_{ij}, should be repeated for $i, j = 1, \ldots, m$ to systematically construct a comprehensive Granger causality matrix F. The resulting F values indicate causality strength between paired variables. Expressly, if $F_{ij} \neq 0$, it can be concluded that brain source j is a Granger cause to brain source i. If $F = 0$, it signifies an absence of causality between the variables. These F values can be represented in a heat map, visually representing the strength of connections between different sources.

Researchers can opt for a threshold value to determine the significance of the Granger causality matrix F. In this algorithm, the threshold is determined based on the maximum absolute value of the Granger causality measure F_{ij}. The causal relationships among different brain regions can also be visually represented in a network diagram. In this diagram, the brain sources are depicted as nodes, and the connections or causal influences between these sources are shown using arrows. These arrows are determined based on the values of F_{ij} and a predefined threshold. This network diagram provides a graphical representation of how different brain regions interact and influence each other, making it easier to visualize and understand the causal relationships within the brain. The threshold controls which connections are shown in the diagram, helping to highlight the most significant causal interactions while excluding weaker or less relevant connections.

The final step of the algorithm, utilizing the identified system matrices as well as the derived lead-field matrix and mapping matrix, encompasses several key components. This includes the computation of the noise covariances using optimization techniques, the calculation of the prediction error covariance for the full- and reduced model and the computation of the Granger causality measure. This fourth step ensures that the connectivity between brain sources can be determined to state something about the interaction between brain regions during a passively performed task stimulating the brain.

The fourth step can be executed using Matlab R2023A with the optimization toolbox (The Mathworks, Inc., Natick, MA, USA). The Granger causality analysis can be done manually based on the equations presented in this section. Alternatively, it can also be performed using the ssgc toolbox, a small Matlab toolbox to demonstrate state-space Granger causality computation [11].

3.2.5. Algorithm design overview

A detailed exploration of the step-by-step process within the algorithm used to ascertain brain dynamics and establish brain network connectivity has been provided. To conclude,

the available data first undergoes preprocessing and transformation to ensure it is appropriately structured. Utilising the filtered dataset in the subsequent step, a state-space model is constructed to circumvent the limitations associated with autoregressive models. Two subspace identification techniques will identify the state-space model. The third step entails converting state information obtained through subspace identification into discernible source signals. This transformation facilitates the derivation of insights into interconnections among brain sources and regions from the recorded EEG data. Finally, after the identification and mapping process, Granger causality analysis is applied within the framework of state-space models. This designed algorithm, incorporating these four steps, enables the analysis of the statistical relationships between brain sources, furnishing insights into the cortical connectivity of the brain during a simple passive task.

The algorithm's structural configuration is visually conveyed through a block diagram, as illustrated in Figure 3.6. In this diagram, each block represents a procedural step, distinguished by a blue number in the top right corner. Arrows, in turn, symbolically represent the transfer of information, either as input or output, for that specific step.

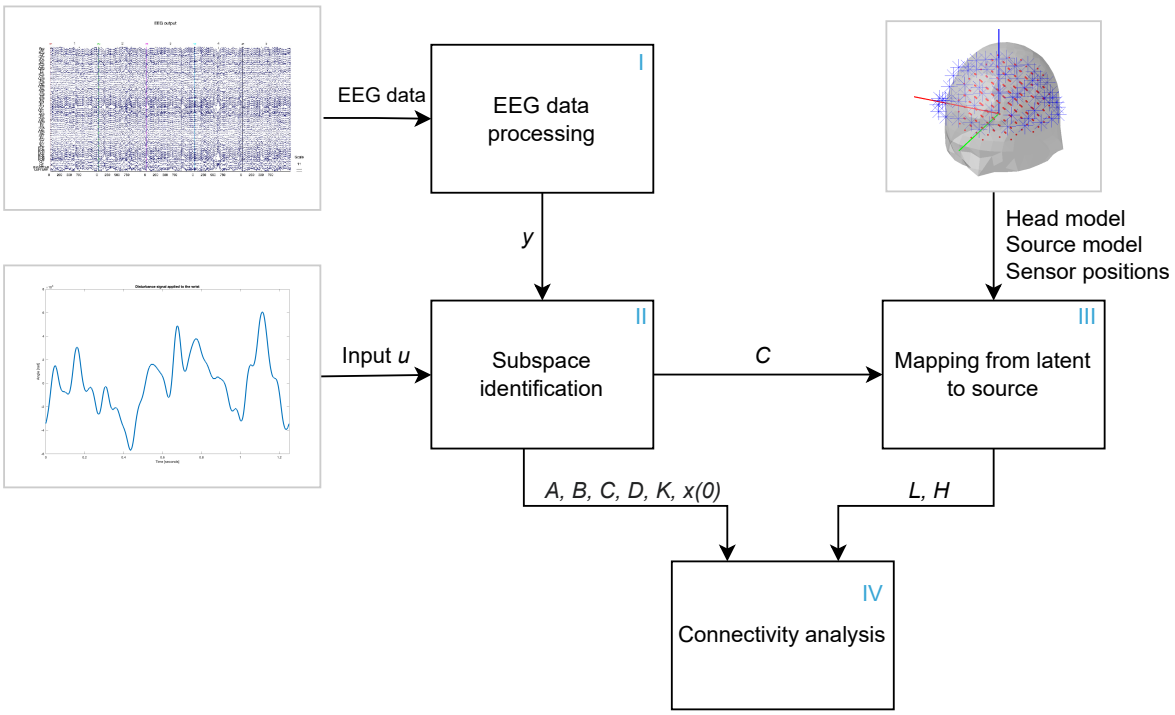


Figure 3.6: Block diagram illustrating the algorithmic framework for building a dynamic system and assessing functional connectivity in an individual subject. Each block represents a distinct stage within the framework, aligning with various subsections of this section.

4

Model validation and analysis

The analysis of the developed algorithm, as introduced in Section 3.2, involves its application to an established dataset derived from the experiment conducted by Vlaar et al. [20]. This dataset originates from experiments involving ten healthy participants subjected to continuous wrist perturbations delivered by a robotic wrist manipulator to elicit sensorimotor steady-state responses [20]. The exploration of the outcomes serves a dual purpose. The primary objective is to validate the robustness, performance, and fidelity of the model to real-world dynamics. The secondary aim is to unveil the dynamics of interactions among various brain regions during continuous wrist perturbations.

The upcoming sections will examine the analysis and results of applying the algorithm to the dataset. To facilitate clarity and coherence, these discussions will be structured around the four fundamental steps of the algorithm.

The first section engages in the initial stage of the algorithm, which involves data preprocessing and aligning the dataset with the developed algorithmic framework. This preparation optimises the dataset for in-depth analysis of neural responses to continuous perturbations applied to the participants. The second section centres on subspace identification using both the N4SID and PO-MOESP methods. These methods are applied to derive system matrices for constructing the dynamic brain model and subsequently reconstructing the latent state and output of the dataset. The simulated outputs are then compared to the previously generated SSRs to assess model performance. Validation and verification of both methods are conducted to identify the most appropriate approach for further analysis. In the third section, the study focuses on establishing the head and source models for participants involved in the experimental procedure. These models are crucial in creating the lead-field matrix L and computing the transformation matrix that links latent states to source states H. With the computed matrix H, the source activity z is reconstructed. The L and H matrices are the building blocks for the algorithm's final step: connectivity analysis. The connectivity analysis is conducted by computing Granger causality measures. These measures are essential for inferring directed functional connectivity among the brain sources. The results are visually represented through Granger causality heatmaps and network diagrams, highlighting the interactions between various brain regions during the experiment.

4.1. Data analysis

The dataset of Vlaar et al. [20] is characterised by two key components that play an integral role in forming the brain model. The first component comprises the EEG recordings collected via an array of 64 electrodes placed at the participants' scalp from a cohort of

ten healthy participants. The second component encompasses the angular disturbances inflicted upon the participants' wrists by a robotic wrist manipulator. The same EEG amplifier that records the EEG signals is employed to capture these disturbances. The data is sampled at a rate of 2048 Hz.

The procedure for analysing the data, outlined comprehensively in Section 3.2.1, operates through a two-phase algorithm. The first phase involves data preprocessing, where the aim is to filter the EEG signal, enhancing its fidelity to reflect the neural activity within the participants' brains primarily. Subsequently, the second phase entails the formation of SSRs, to enhance the understanding of neural responses to external stimuli. The results from these two phases applied to the dataset are elaborated in the following two paragraphs.

4.1.1. Data preprocessing

In the context of EEG data analysis, preprocessing is critical in transforming the raw EEG signals into a format suitable for subsequent analysis. This preprocessing involves the application of various used-defined parameters. It is noteworthy that for the application of the algorithm on the available dataset, the specific parameter values, such as filter frequencies, closely adhere to the approach utilised by Vlaar et al. [20] in their investigation of the identical dataset.

The EEG data is band-pass filtered between 0.8 and 120 Hz and band-stop filtered in narrow bands around 50 Hz to remove line noise and around 100 Hz to remove the harmonic of the line noise. Two channels, AF4 and F6, are rejected and excluded from further analysis by examining the channel data and the spectral characteristics [62]. In the data of all participants, it is observed that these two channels exhibited exceptional oscillatory behaviour compared to the others, likely suggesting the presence of high-frequency noise. A snippet of the electrical activity data, clearly illustrating the differences from the other channels, is presented in Figure C.1 in Appendix C.

After rejecting the bad channels, the data is re-referenced, whereafter an ICA is performed using the Infomax algorithm implemented in CUDAICA [70] to separate brain signals from artefacts without removing the affected data portions. The Infomax algorithm, a form of blind source separation, seeks to find a mixing matrix that transforms observed mixed signals into statistically independent source signals. Through an iterative optimization process, it adjusts this mixing matrix to maximize the statistical independence among the estimated sources, reducing mutual information between them [70]. The outcome of the ICA procedure is a set of independent components (ICs), some representing neural activity, while others may capture various artefacts. After the separation of these components, the ICs are visually inspected, and specific ICs that are identified as artefacts suggested by the CUDAICA algorithm, such as those representing eye blinks or muscle activity, are removed. The remaining signal is linearly combined to reconstruct the original EEG signal.

Subsequently, the dataset per participant is segmented into distinct epochs of 1.25 seconds, aligning with the periodicity of the perturbation signal considered as the input. The first and last epoch per trial are excluded from further analyses to minimize the impact of transient effects. This process yields 160 time periods for each participant engaged in the passive task. Two epochs of the remaining dataset for Participant 1 are displayed in Figure 4.1. Within this figure, one can observe the temporal dynamics, replete with distinctive waveform patterns of the electrical activity.

Of significance is the fact that the initial dataset included one participant whose data had already been subjected to epoching procedures, thereby necessitating distinct preprocessing protocols tailored to accommodate this particular participant's data. Consequently, this participant's data is excluded from the subsequent analyses. Additionally, two participants

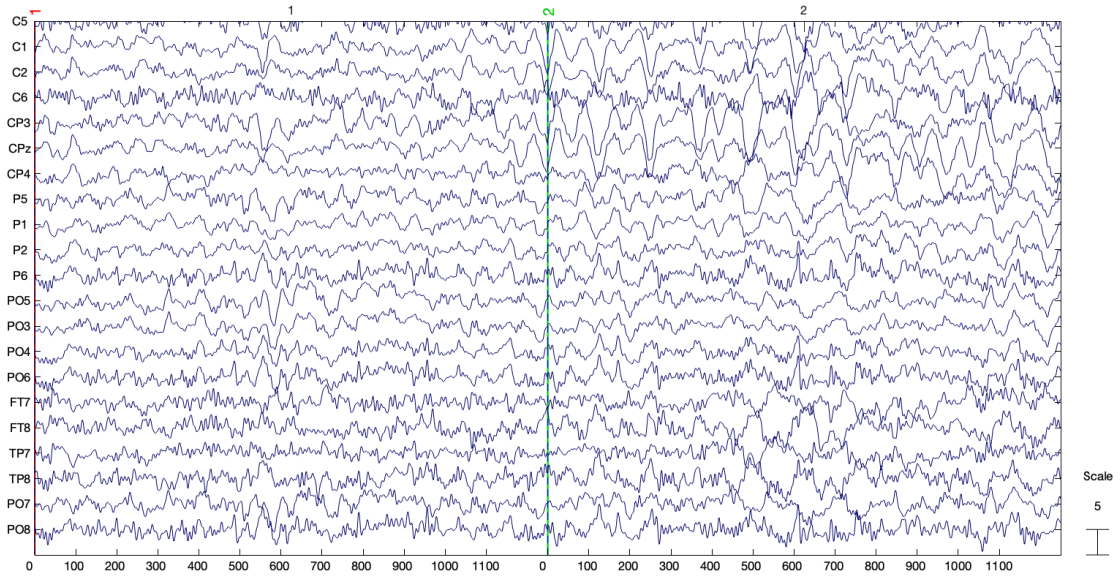


Figure 4.1: The preprocessed scalp recorded electrical activity from 20 electrodes over two distinct epochs for Participant 1. The x-axis represents time in milliseconds, with a green line indicating the boundary between the two epochs. The y-axis displays channel labels, adhering to the electrode labelling conventions of the 10/20 system.

exhibited discrepancies where the input signals were not synchronized temporally for each epoch. Therefore, the data from these participants is also omitted from further research. This results in a refined dataset comprising data from 62 electrodes collected from seven participants instead of the original dataset encompassing 64 electrodes from ten participants.

4.1.2. Steady-state response analysis

The SSRs are formed by averaging the electrical activity per electrode for each participant separately, allowing for the effective removal of outliers and enhancing the signal-to-noise ratio within the EEG signal. In relation to the dataset, the averaging is carried out across 160 epochs per participant, resulting in a 1.25-second signal for each electrode. With a sampling frequency of 2048 Hz, this process yields a dataset of 62×2560 dimensions for all seven participants.

Figure 4.2 provides a visual representation of the SSR recorded for Participant 1 for the electrodes labelled CP3 and CP4, as well as a scaled perturbation signal. The electrode CP3 is positioned at the Central Parietal region, situated above the primary somatosensory cortex, on the left side of the head. The electrode CP4 is located at the same height but on the right side of the head. The selection of these electrodes is grounded in the anticipation that joint manipulation on one side is expected to trigger the contralateral primary somatosensory cortex, a concept discussed in Section 1.4.

For Participant 1, the perturbations are applied to the right wrist [20]. The input signal, corresponding to the angle of perturbation applied at the wrist, has been adjusted to match the maximum electrical activity measured at the specific electrode. The perturbation signal is presented to investigate periodic patterns or relationships between the electrical activity and the input signal. Due to research decisions aimed at studying periodicity, the choice is made to adjust the scale factor for each electrode.

The results depicted in Figure 4.2 provide evidence that the electrical activity recorded at electrode CP3, situated contralateral to the perturbed wrist, exhibits notably higher amplitude and more prominent oscillations compared to the electrical activity at electrode CP4.

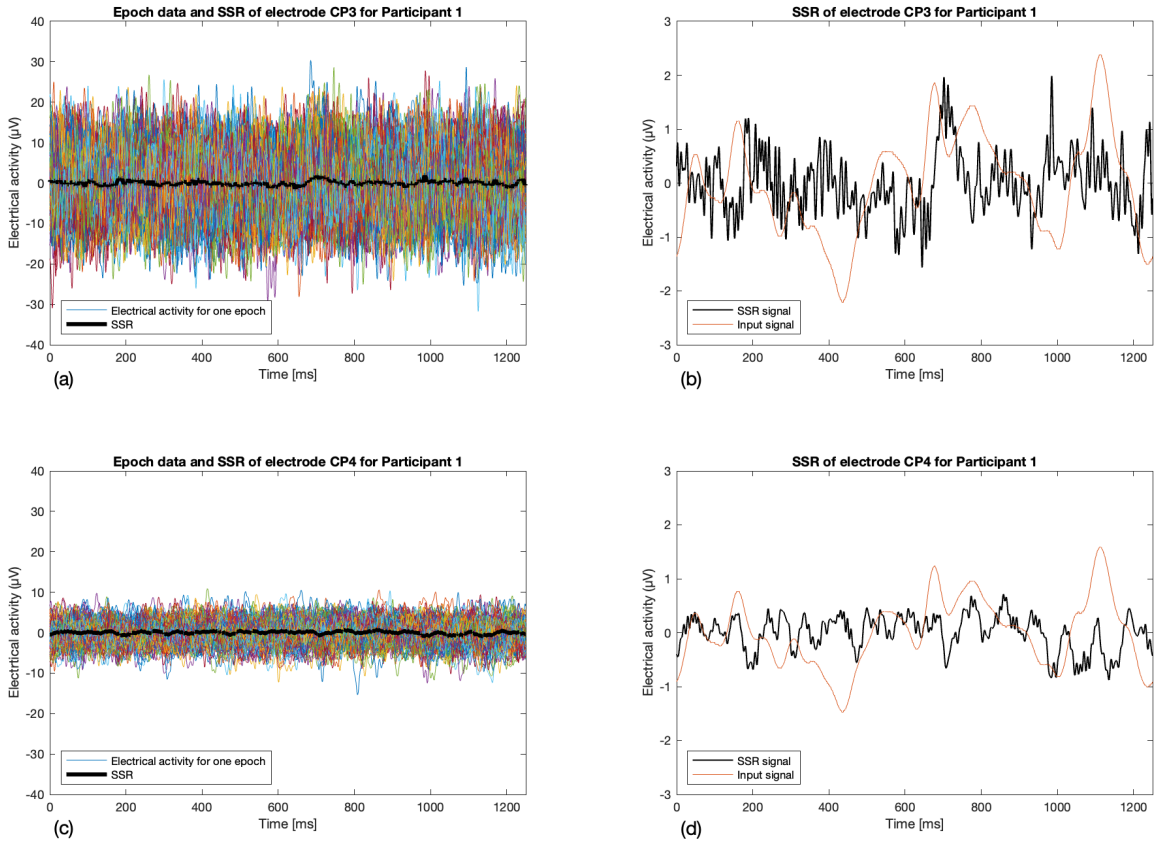


Figure 4.2: The SSR computed for electrode CP3 and CP4 of Participant 1. (a) Visualization of electrical activity across 160 epochs (coloured lines) and the calculated SSR (black line) at electrode CP3. (b) Zoomed-in version of the SSR signal at electrode CP3 and the scaled input signal. (c) Visualization of electrical activity across 160 epochs (coloured lines) and the calculated SSR (black line) at electrode CP4. (b) Zoomed-in version of the SSR signal at electrode CP4 and the scaled input signal.

When examining the electrical activity at electrode CP3 in relation to the perturbation signal, particularly at time points around 700 and 1000 milliseconds, it is apparent that the peaks, whether positive or negative, in the perturbation signal, align with corresponding peaks in the electrical activity. Interestingly, the peak occurring at 400 milliseconds does not seem to coincide with a simultaneous or slightly delayed peak in the electrical activity. Upon closer examination, a comparable pattern of peaks emerges in the electrical activity recorded by electrode CP4, although with lower amplitude and a sign reversal. A certain periodicity between the input signal and the SSR signal for these electrodes can be discerned.

4.2. Subspace identification performance

Within this section, the focus shifts to the performance of the second phase of the algorithm, where the aim is to construct a state-space model of the brain's dynamics for simplification. The proposed method in this research for identifying the state-space model relies on applying subspace identification techniques. Specifically, two distinct methods, N4SID and PO-MOESP, are employed in this study to model the brain. It is the objective to compare the performance of these two methods to determine the most suitable approach for further research endeavours. The method demonstrating superior performance will be selected for continued investigation in the proposed algorithm.

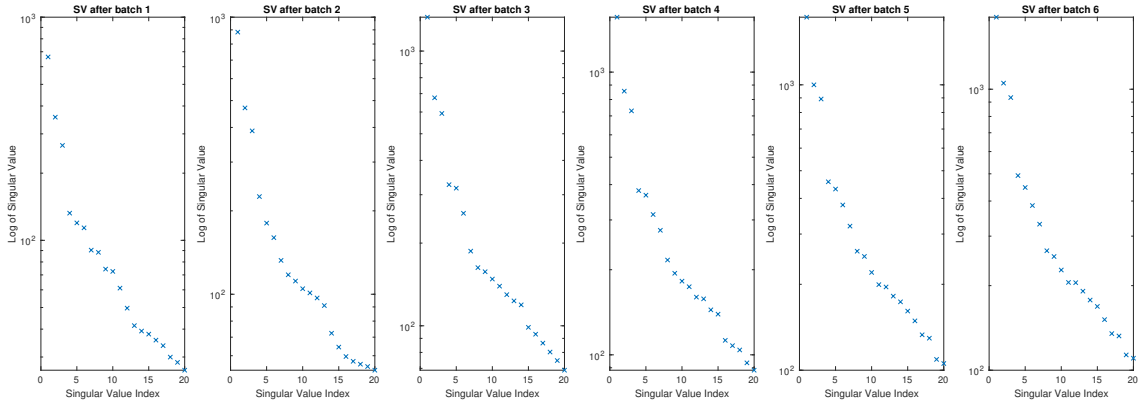


Figure 4.3: Singular values obtained by computing the SVD in the PO-MOESP method after adding each of the six batches, necessary for determining the model order.

4.2.1. Model parameter selection in subspace identification

In the context of conducting subspace identification using either N4SID or PO-MOESP, researchers are tasked with specifying two crucial model parameters: the model order n and the block size s that pertains to the Hankel matrices. To maintain an equitable and unbiased comparison between the two methods, the parameters open to user-defined selection are consistently configured with identical values for both N4SID and PO-MOESP.

The proposed strategy on data batches is employed to enhance the model's training process, as presented in Section 3.2.2. This involves concatenating the input and output arrays from six participants, enabling the application of subspace identification on data batches. Using data batches effectively leverages all the available information from the system. Notably, one participant's dataset is intentionally withheld for validation, ensuring the accuracy and integrity of the model evaluation process.

In the subspace decomposition phase for both methods, the critical task is determining an appropriate model order n. The choice of model order follows the methodologies outlined in Section 2.2.3. For N4SID, this is determined by assessing the singular values of $Y_{s,s,N}/U_{s,s,N}Z_N$ while PO-MOESP established the model order is through the analysis of the SVD of R_{32}. This process involves the examination of the gap between the singular values.

The PO-MOESP method is employed as the baseline reference. By computing the SVD of R_{32}, the corresponding singular values can be derived. These singular values, both the initial set and those obtained after each batch iteration, are visualized in Figure 4.3. Examining these singular values across iterations show a changing pattern in the gaps. Looking at the gaps in the singular values in the plot after the sixth batch, one could choose a model order of 3, 7, 10, 15, 16 or 18.

In line with the procedure elucidated in Section 3.2.2, when multiple significant gaps manifest in the plot of the singular values, the selection of the model order involves a balance between achieving optimal performance, managing computational resources, and preventing overfitting. The variance accounted for (VAF) is a measure of performance, that quantifies the proportion of the total variance in the predicted values that can be attributed to the actual values [16]. It is expressed as:

$$\text{VAF}(y(k),\widehat{y}(k)) = \max\left(0,\left(1-\frac{\frac{1}{N}\sum_{k=1}^{N}\|y(k)-\widehat{y}(k)\|_2^2}{\frac{1}{N}\sum_{k=1}^{N}\|y(k)\|_2^2}\right)\cdot 100\%. \tag{4.1}$$

In this research, the variance accounted for (VAF) is used to compare the SSR, denoted by $y(k)$, and the simulated output generated using the PO-MOESP method, denoted by $\widehat{y}(k)$.

These output values represent the electrical activity of a single participant for a specific electrode, measured at a given time instant, k. The VAF falls within the range of 0% to 100%, where higher VAF values indicate reduced prediction errors and superior model performance.

Validated on the data of Participant 7, Table 4.1 represents the VAF, which is the mean VAF over all the electrodes for different model orders based on the singular values. Notably, a substantial increase in VAF occurs at order $n = 15$ compared to lower orders. Elevating the order to 16 or 18 may lead to an increased risk of overfitting and extended computation time without a significant performance improvement. Consequently, for subsequent calculations in both subspace identification methods, the model order $n = 15$ is adopted.

Table 4.1: The average VAF for different model orders is based on the gaps observed in the singular values.

Order n	**VAF** [%]
3	69.29
7	86.31
10	88.95
15	92.04
16	92.65
18	93.67

The model parameter left to determine is the block size s, which should be larger than the expected system order and smaller than the number of data points, such that $n < s \ll N$. For the ensuing results, $s = 20$ meets the criterion.

4.2.2. State-space model validation

In the context of this research, linear state-space models are identified on the available dataset by employing both the N4SID and PO-MOESP methods. The identification process utilises data batches containing information from six participants. The chosen model parameters consist of a model order, determined as $n = 15$, and a block size, specified as $s = 20$. These model parameters lead to the estimation of system matrices A, B, C, D, the Kalman gain K, and the initial state $x(0)$. The identified matrices enable the estimation of the latent state $\widehat{x}$ and the output $\widehat{y}$.

A validation process is conducted to assess the accuracy and reliability of the identified models. The validation criterion measures the model's ability to accurately fit a part of the data not employed during the identification process. This validation entails comparing the predicted output, $\widehat{y}$, generated by reconstructing the state-space model with the identified state matrices, and the actual output y, corresponding to the SSR signals for all the electrodes. The evaluation of the model's accuracy is conducted using two key performance metrics, the VAF and the root-mean-square error (RMSE). The VAF is explained by Equation 4.1. The RMSE assesses the average magnitude of the error, signifying the differences between the model's predictions and the actual data points. The RMSE is calculated as:

$$\text{RMSE} = \sqrt{\frac{\sum_{k=1}^{N}(y(k) - \widehat{y}(k))^2}{N}}. \tag{4.2}$$

A smaller RMSE value indicates a more accurate model with lower average prediction errors. In practical terms, a lower RMSE signifies that the model offers more precise predictions that closely align with the observed data, making it a better match for the dataset.

To validate the models effectively, the "leave one out" cross-validation approach is applied. This involves training the model using data from the first six participants and validat-

ing its performance using the data from the seventh participant. This process is iteratively repeated for each of the seven participants, with each participant's dataset serving as the validation set once. The outcomes of the validation process are presented in Table 4.2, which includes the mean and standard deviation (std) of the performance metrics, expressing the deviations in the performance between the seven participants in the leave-one-out principle. The standard deviation is taken into account to test the robustness of the model.

Table 4.2: Evaluation of N4SID and PO-MOESP using performance metrics.

	VAF [%]		**RMSE** [μV]	
	Mean	Std	Mean	Std
N4SID	78.28	9.05	0.27	0.04
PO-MOESP	89.53	3.29	0.17	0.04

Figure 4.4 visualises the output signals by displaying the computed SSR for Participant 7 at electrode CP3, in conjunction with the SSR derived from the reconstructed outputs. These reconstructions are accomplished using system matrices acquired through the N4SID and PO-MOESP identification techniques.

As depicted in Table 4.2, the model validation measures indicate superior results for the PO-MOESP method. Furthermore, as shown in Figure 4.4, the red line, marking the reconstructed output from the system matrices identified with N4SID, shows a slight offset. In contrast, the offset is considerably reduced for the blue line, representing the simulated data from the system matrices estimated using the PO-MOESP method. The reduced offset between the PO-MOESP-derived output and the actual output suggests that this method is more effective in capturing the dynamics of the brain under investigation. This outcome highlights the potential advantages of the PO-MOESP approach for modelling and analyzing brain dynamics in similar experimental contexts.

The validation of both models, established through the utilisation of the two subspace identification techniques, affirms the suitability of a linear model for capturing brain dynamics.

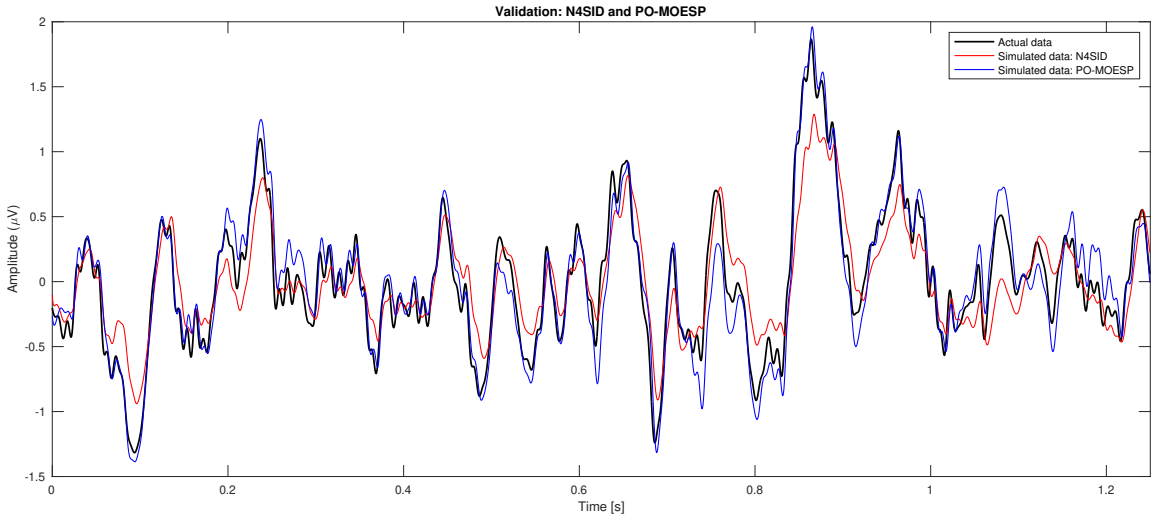


Figure 4.4: Comparison of the determined SSR at electrode CP3 for Participant 7 (highlighted in black), with the reconstructed output estimated by N4SID (highlighted in red) and the output reconstructed using the PO-MOESP method (highlighted in blue).

4.2.3. State-space model verification

Verification of state-space models is crucial for ensuring their reliability and validity. The verification process aims to confirm whether the assumptions described in Section 2.2.2 are met for both modelling approaches. Although subspace identification methods can estimate models, they do not automatically guarantee that the models meet essential assumptions [71]. If, for example, the model is found to be unobservable, it implies that the chosen inputs and outputs are not sufficient to determine the internal states of the system uniquely. In such cases, additional or different inputs or outputs may need to be considered, or alternative modelling approaches may be required to ensure observability.

In this section, the focus is on testing key assumptions for the reliability of open-loop subspace identification models, including linearity, asymptotic stability, observability, and controllability. For both approaches, it will be demonstrated whether these assumptions hold.

N4SID model verification

Figure 4.4 shows that the predicted output from the N4SID model follows the actual output, displaying a remarkable similarity, although some offsets are observed. Quantitative metrics further validate the model's performance. The average VAF, a measure of the model's explanatory power, is calculated at 78.38%, with a std of 9.05%. In addition, the RMSE, an indicator of the average prediction error, is computed at $0.27\mu V$, with a std of $0.04\mu V$. This RMSE value signifies that, on average, the model's predictions deviate from the actual data by $0.27\mu V$, further substantiating the N4SID model's predictive performance. These outcomes suggest that employing the N4SID approach to model brain dynamics as a linear time-invariant system is a reasonable assumption.

The second assumption for model verification focuses on assessing the system's stability, which is achieved by examining the eigenvalues of the matrix $A - KC$. The eigenvalues are visualized in Figure 4.5. The figure shows that the eigenvalues are positioned strictly inside the unit circle. This indicates that the model composed with the N4SID method is asymptotically stable. In practical terms, this implies that the system is internally stable, and with the appropriate control strategies, it can be effectively regulated to achieve the desired performance.

The final aspect of verification involves the assessment of observability and controllabil-

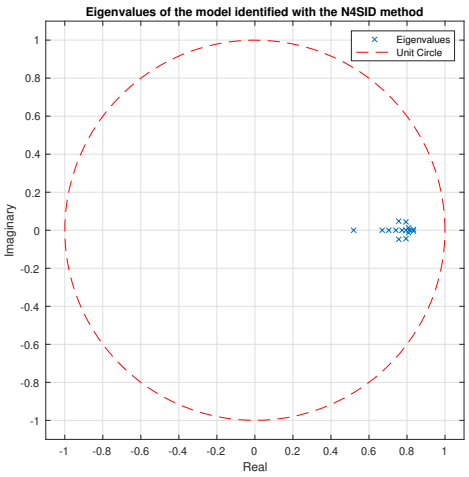


Figure 4.5: Eigenvalues of $(A - KC)$ identified with the N4SID method

ity. Mathematical calculations reveal that the system modelled using N4SID is observable and controllable. The rank for the observability and controllability matrix is equal to the model order, $n = 15$. This alignment with the fundamental requirements for observability and controllability underpins the model's effectiveness in capturing the intricate brain dynamics. It affirms that crucial state variables can be reconstructed from the available output measurements and that the system can be manipulated to meet specific performance objectives.

In conclusion, the N4SID model verification process demonstrates that the model represents the observed brain dynamics and meets fundamental mathematical criteria, establishing its suitability as a linear time-invariant representation of the underlying system.

PO-MOESP model verification

From Figure 4.4, it can be noticed that the predicted output from the PO-MOESP model closely tracks the actual output, displaying a high degree of similarity and fewer offsets when compared to the N4SID model. Quantitative metrics further substantiate the model's exceptional performance. The VAF is 89.53%, with a std of 3.29%. Additionally, the RMSE is determined to be $0.17\mu V$, with a std of $0.04\mu V$. The high VAF and low RMSE values signify that employing the PO-MOESP approach to model brain dynamics as a linear time-invariant system is a highly accurate assumption.

The system's stability is assessed by examining the eigenvalues of the matrix $A - KC$. As depicted in Figure 4.6, the eigenvalues are strictly inside the unit circle, verifying that the PO-MOESP model meets the stability requirement.

The final aspect of verification involves assessing observability and controllability. The mathematical analysis demonstrates that the system modelled using PO-MOESP is observable and controllable. The ranks of the observability and controllability matrices align with the model's order, which is $n = 15$. This adherence to the fundamental prerequisites for observability and controllability confirms that essential state variables can be reconstructed from available output measurements, and the system can be manipulated to meet specific performance objectives.

In summary, the PO-MOESP model verification process demonstrates that the model closely aligns with the brain dynamics and satisfies fundamental mathematical criteria. This establishes its suitability as a robust linear time-invariant representation of the underlying system.

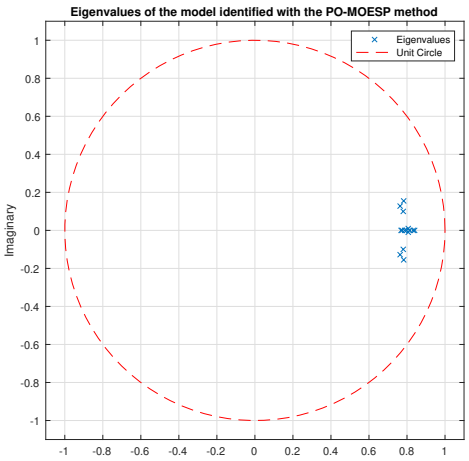


Figure 4.6: Eigenvalues of $(A - KC)$ identified with the PO-MOESP method

4.2.4. Model selection

A comparative examination of the two subspace identification methodologies, specifically N4SID and PO-MOESP, through the process of validation and verification, demonstrates the superior performance of the state-space model obtained through the PO-MOESP approach. This superiority is evident from the results presented in the tabulated data, encapsulating the performance measures and the graphical representation, representing the disparities between the actual and reconstructed outputs. The lower std observed in the VAF suggests that the various VAF values obtained when excluding individual participants exhibit less variation. Therefore, it can be inferred that the PO-MOESP method demonstrates greater robustness than the N4SID method. As a result, the PO-MOESP method is the preferred approach for all subsequent computations and analyses.

4.3. Mapping from latent to source

The subsequent step in the algorithm leads to the establishment of a mapping from the latent states denoted as x, reconstructed in the previous phase, to the source dynamics represented by z. This mapping holds significant implications for unveiling the brain connectivity patterns associated with actual neural sources distributed within the brain.

To derive the mapping from the latent to the source, the head model and the positions of the electrodes are acquired from the available dataset. The head model and electrode positions are synchronised for further investigations.

4.3.1. Construction of the lead-field matrix L

The derivation of the lead-field matrix is a direct outcome of the head model, source model and electrode positions. To obtain the source model in the context of the available dataset, a spatial resolution of 20 mm is employed. In practical terms, a resolution of 20 mm entails dividing the source space into small cubic regions or voxels, with each voxel measuring 20 mm along its sides. Each of these voxels represents a potential source location for estimating neural activity. The choice of a 20 mm resolution results in 204 candidate source locations for each participant.

The selection of this resolution primarily hinges on computational considerations and the desired number of source locations to ensure that the results remain understandable and meaningful while also ensuring that brain regions are not omitted from the analysis. Testing has confirmed that opting for a lower resolution, such as a 10 mm resolution, would result in the Granger causality heatmaps in the subsequent step displaying similar statistical dependencies. However, this choice also leads to an increased number of brain sources.

Afterwards, the atlas "ROI_MNI_V4" is interpolated on the source model. The atlas is a template within FieldTrip referring to a neuroimaging file in the NIfTI (Neuroimaging Informatics Technology Initiative) format. This file contains a region of interest (ROI) map spatially registered to the Montreal Neurological Institute (MNI) standard brain space. The mapping process is carried out to enable subsequent inferences regarding the functional connections between well-recognised brain sources. When the source locations are mapped onto the atlas, specific source points share the same tissue labels. This observation suggests that multiple source points pertain to one common anatomical region within the atlas space. In such cases, the elements within the lead-field matrix and the positions of these sources within the grid are averaged to create a unified representation, with each grid point in the source model representing an individual source corresponding to a recognized tissue label within the grey matter. This process results in 74 unique and distinct brain sources.

The catalogue of the anatomical labels of the brain sources defined in each hemisphere and their location within brain regions can be found in Table B in Appendix B. The order of

the sources is based on the automated anatomical labelling 2 (AAL2) atlas [72], following an alphabetical arrangement of the brain regions. A comprehensive and in-depth elucidation of the anatomical volumes of interest, along with the terminology for the regions specified in the atlas, is available in the research paper of Tzourio-Mazoyer et al. [72].

The resulting lead-field matrix possesses the dimension of 74×62, where 62 corresponds to the number of electrodes, and 74 reflects the number of unique source locations. The values in the lead-field matrix indicate the extent to which the electrical activity of each source influences the electrical measurements recorded at the scalp.

4.3.2. Construction of the transformation matrix H

Using the determined lead-field matrix L and the system matrix C obtained in the second step of the algorithm, the Group Lasso problem of Equation 3.4 is solved. This utilises the construction of the transformation matrix H, which facilitates the transition from latent states to source states. To choose the matrix from a finite set of potential outcomes, the BIC of Equation 3.5 balances the trade-off between model fitting, represented by the term $||C - LH||$, and model complexity.

The dimensions of the resulting H matrix are 74×15, wherein 74 accounts for the number of candidate source locations, while 15 represents the number of latent states. The H matrix constitutes the mapping from the latent space to the source domain, which is necessary to determine the functional connectivity from state-space models.

4.3.3. Simulation of the source activity

The source activity z is constructed from the computed matrix H, ignoring the noise component. The results for the source activity of Participant 1 are shown in Figure 4.7 for four significant brain regions. An observation drawn from this figure is the significant difference in electrical activity between the thalamus, located in the subcortical regions of the brain, and the other three sources in cortical regions. Based on prior research, it is anticipated that the left postcentral gyrus, located within the contralateral somatosensory cortex, displays heightened electrical activity during the stimulation task [32]. This expectation is based on the role of the somatosensory cortex in processing sensory information and supported by the observed high SSR values in Figure 4.2 for electrode CP3, which is customarily positioned in proximity to the postcentral gyrus. However, the electrical activity of the postcentral gyrus is not elevated compared to the other cortical sources, deviating from the initial expectations.

The simulated electrical activity of all the sources, according to the various brain regions, are presented in Appendix C. These visual representations allow straightforward comparisons of the relative electrical activity levels among the multiple sources. Upon reviewing these figures, it becomes apparent that specific brain regions, such as the frontal lobe, occipital lobe, and central sulcus, exhibit relatively high electrical activity. Notably, the left supplementary motor area, the left superior parietal gyrus, and the superior occipital gyrus stand out as sources with particularly pronounced electrical activity. In contrast, regions like the insular lobe, subcortical grey nuclei, temporal lobe, and cerebellum display considerably lower electrical activity. Overall, sources located deep in the brain, such as the thalamus, show low electrical activity.

4.4. Connectivity analysis

Utilising the subspace matrices extracted with the PO-MOESP method in the second step, along with the lead-field matrix L and the transformation matrix H from the third step, the final step of the algorithm is used to determine the functional connectivity of the brain for

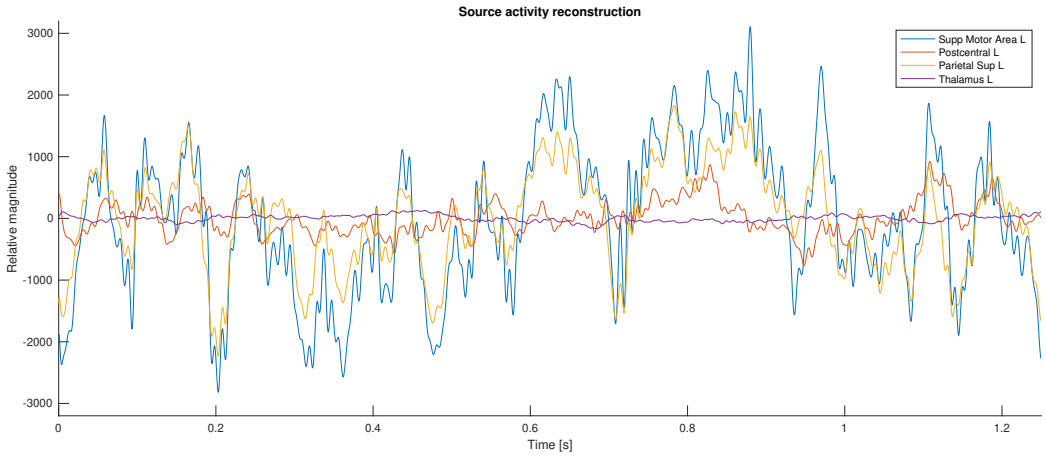


Figure 4.7: Simulated source activity in the left hemisphere, including the supplementary motor area, postcentral gyrus, superior parietal gyrus, and the thalamus for Participant 1.

participants conducting the experiment where the robotic wrist manipulator is continuously perturbing there wrist.

As outlined in Section 3.2.4, this stage commences with the computation of the covariance matrices. The state noise covariance Σ_w and the output noise covariance Σ_e are directly derived from the identified state matrices and reconstructed outputs of step 2. The remaining pair, Σ_η and Σ_v, are obtained by solving the optimisation problem of Equation 3.7.

Subsequently, the Granger causality analysis is performed. It initiates with the computation of the DARE for the entire model, allowing to derive Σ, which is the prediction error covariance. In the ensuing step, the DARE is computed for the reduced model by systematically eliminating the jth column and row within the matrices H and Σ_η, resulting in H^R and Σ_η^R respectively. The reduced prediction error covariance Σ^R can be computed following this.

Looping through all the different sources to span the entire spectrum, the Granger causality matrix F is computed. This measure utilises the calculated prediction error covariance to determine whether region j Granger causes region i. In essence, it enables quantifying the statistical dependency of brain source j on brain source i, thereby shedding light on the intricate web of functional connectivity within the brain. Here, each i and j corresponds to a specific brain source, identified by a tissue label within the atlas space.

4.4.1. Granger causality heatmap

The outcomes of the Granger causality analysis for Participant 1 are presented in the Granger Causality heatmap of Figure 4.8. This heatmap employs a colour-coding scheme to represent the connectivity strength between two sources, which depends on a specified threshold. The warmer colours, such as red, signify strong connectivity, while cooler colours, such as blue, suggest weaker or negligible connections. The heatmap contains two axes. The x-axis corresponds to the source from which the causal influence originates, referred to as the "cause", denoted as j in the algorithm. The y-axis represents the recipient of this influence, marked as the "effect", characterised as i in the algorithm. The white markings along the axes indicate the boundaries of a specific brain region to which the brain source is associated. A comprehensive list of the brain sources, along with their abbreviated notations and positions in the superior brain region, can be found in Table B. The ordering of the sources is based on an alphabetical arrangement of the respective brain regions, as outlined in the AAL2 atlas.

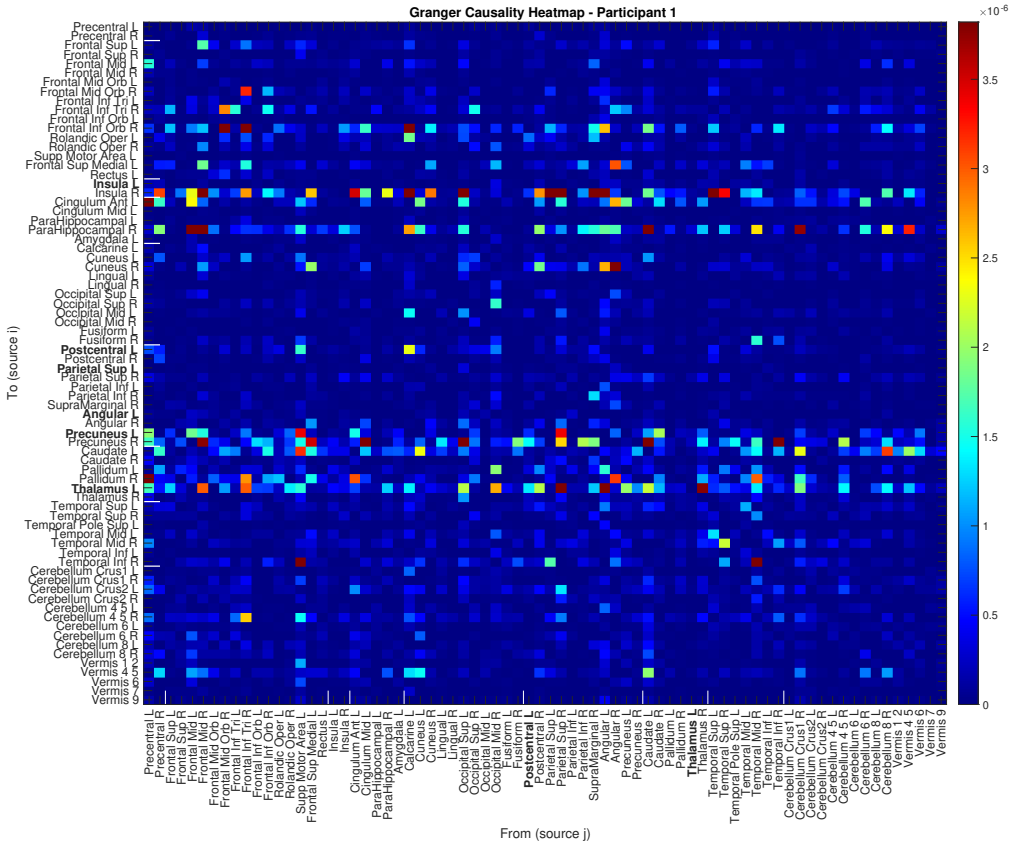


Figure 4.8: The Granger Causality heatmap of Participant 1, created from the Granger Causality matrix F. The heatmap shows how much region j (displayed on the x-axis) Granger causes region i (displayed on the y-axis). The bold text highlights sources with expected strong dependencies.

The diagonal of the Granger causality heatmap pertains to self-causal relationships for each brain source. It quantifies how much a particular brain source "causes" itself, which is not the primary focus of the connectivity analysis. Therefore, the diagonal elements are set to zero when interpreting the Granger causality heatmap to emphasize interactions between distinct brain sources.

Analysing the Granger causality measures for Participant 1 as depicted in Figure 4.8, it is evident that specific sources, particularly the right insula, left thalamus, right precuneus and the right parahippocampal, display significantly high Granger values in the rows. This indicates a more pronounced influence from other brain sources. On the contrary, sources such as the right lingual, vermis 7, right middle frontal gyrus, right precentral gyrus and the left supplementary motor area demonstrate minimal or no dependence on the activity of other sources.

During wrist perturbation, it is expected that constant joint manipulation triggers the transmission of proprioceptive and tactile sensory information to the contralateral primary somatosensory cortex via the thalamus [32]. The source corresponding to the contralateral primary somatosensory cortex is the left postcentral gyrus. Subsequently, this information is expected to be relayed to various regions, including the posterior parietal cortex [37], the insular cortex [38], and the secondary somatosensory cortices. Within Figure 4.8, the sources anticipated to exhibit strong statistical dependencies are identified in bold. Upon closer examination of these sources in the Granger causality heatmap, it becomes evident that

the left postcentral gyrus exerts a notable influence on other brain regions. However, it is worth noting that this influence lacks statistical significance when compared to the impacts of other sources. Contrary to expectations, the left thalamus exhibits a significant statistical dependency on other sources, whereas the original anticipation is that other sources would be dependent on the left thalamus. A somewhat surprising finding is that the right insula, as opposed to the left insula, displays a marked statistical dependence on other sources. Among the brain regions associated with the posterior parietal cortex, which includes the superior parietal gyrus, inferior parietal gyrus, supramarginal gyrus, angular gyrus, and precuneus, the precuneus stands out by demonstrating substantial statistical dependence on other sources. In contrast, the remaining regions do not manifest such dependencies. In summary, these results diverge from the initially predicted outcomes.

4.4.2. Network connectivity diagram

In the context of the brain connectivity analysis, the network connectivity diagram signifies the directed dependencies between different brain sources, determined with the Granger causality analysis. For Participant 1, the network connectivity diagram is presented in Figure 4.9.

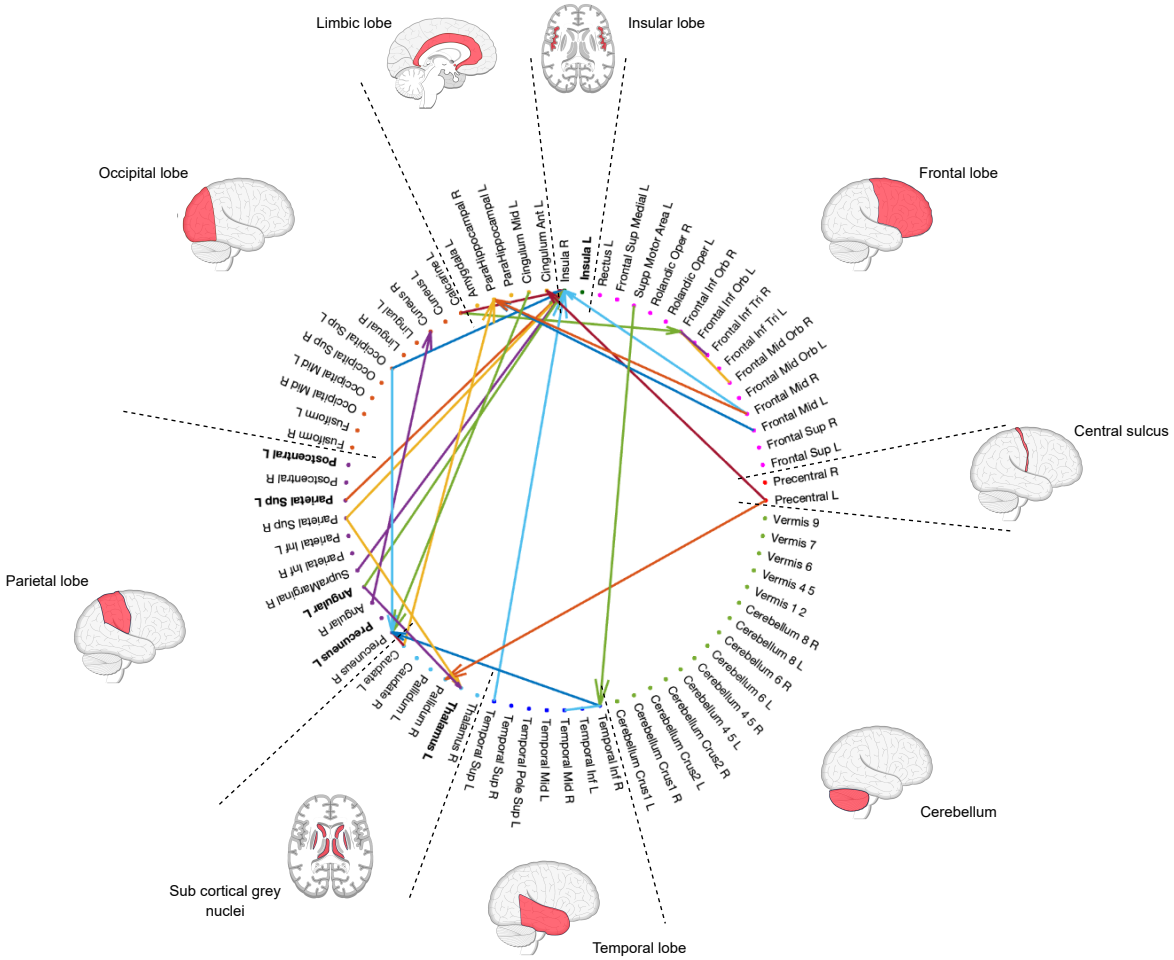


Figure 4.9: Network diagram of Participant 1, illustrating directed functional connectivity. Nodes represent individual brain sources; arrows show causal relationships between these sources. The dotted lines define brain regions, and the red area on the brain map indicates the locations of the regions. The bold text highlights sources with expected strong dependencies.

Figure 4.9 consists of nodes with edges representing the relationships between the various sources. Each brain source is defined as a node, and the links between these nodes are depicted as arrows, with the arrow indicating the direction from the cause, represented as brain source j, to the effect, symbolized as brain source i. Each brain source is associated with a specific region, delineated by the dashed lines. These brain regions are visually represented in red on a brain map. A map showing the overview of the distinct brain regions is provided in Figure B.1.

The network diagram of Participant 1 highlights that the right insula exhibits a statistically significant dependence on several other sources compared to the remaining sources. Furthermore, it reveals that the left precuneus and the right parahippocampal demonstrate substantial statistical dependencies on other sources, serving as the primary causal factors. These findings align with the observations from the heatmap.

The Granger causality heatmap and the network connectivity diagram employ the same threshold. This threshold is selected and adjusted per participant to ensure the diagrams remain informative, revealing 25 connections per participant.

4.4.3. Inter-participant connectivity evaluation

The Granger causality heatmaps and their associated network diagrams for all the participants are available in Section C.3 to analyze the results. This assessment includes the application of various threshold values, ranging from 0.1 to 0.35 times the maximum absolute threshold for each participant, to visualize 25 arrows within the network connectivity diagrams.

A recurring observation in the Granger causality heatmaps is the distinct prominence of colour patterns within the rows compared to the columns. This finding suggests a temporal hierarchy in the brain's information processing. It is plausible that specific cerebral sources demonstrating strong dependencies in the columns serve as central hubs or integration points where information converges, ultimately influencing other brain regions. Conversely, particular sources might primarily receive and process this information.

In examining network connectivity diagrams, specific connections between nodes exhibit a recurring pattern of occurrence. These recurrent directed connections observed more than twice across the seven participants, are documented in Table 4.3. In this tabulation, the initial column designates the causal component of the connection, represented as the brain source j, signifying the point of origin of the directed arrow. Meanwhile, the second column designates the effect aspect of the connection, identified as the brain source i, indicating the terminal point to which the arrow is directed inside the network connectivity diagrams. Overall, it is evident that the insular lobe, limbic lobe, and subcortical grey nuclei exhibit the highest statistical dependence on other regions compared to the remaining brain regions.

It can be acknowledged that while there are some variations in the network diagrams across different participants, there remains a commonality in the strong dependencies between the brain sources mentioned above. This shared observation suggests a consistent functional connectivity pattern in these brain regions across the study's participants who performed the same experiment.

4.4.4. Assessing functional connectivity in relation to source activity

When assessing the relation between the electrical activity of brain sources and the functional connectivity diagrams, specific patterns emerge. Specifically, it is noticeable that brain sources exhibiting high electrical activity during the experimental task significantly influence other sources but do not exhibit a corresponding statistical dependence, meaning

Table 4.3: Inter-participant analysis of recurrent connections within the network connectivity diagrams.

From (source j)	**To (source i)**	**Occurence**
Temporal Mid R	Temporal Inf R	5
Parietal Sup R	Thalamus L	4
Calcarine L	Insula R	3
Parietal Sup R	Insula R	3
Angular L	Insula R	3
Angular L	Thalamus L	3
Frontal Sup Medial L	Cingulum Ant L	3
Calcarine L	Thalamus L	3
Angular R	Pallidum L	3

they are causal but not affectable.

This observation is particularly clear when reviewing the columns in the Granger causality heatmap depicted in Figure 4.8 for the left supplementary motor area, the superior occipital gyrus, and the left superior parietal gyrus. In these columns, these sources show significant behaviour, demonstrating a substantial influence on other brain regions. When referring to the source activity plots in Figure C.2 and C.3, it is evident that these sources exhibit consistently high electrical activity. Conversely, specific brain sources, such as the thalamus and precuneus, show low electrical activity but strongly depend on other brain areas, as evidenced by the rows in the Granger causality heatmaps.

5
Conclusions

In the preceding chapters of this report, a data-driven mathematical framework is designed to study brain dynamics from EEG measurements, including an exogenous input. This framework is crafted through an in-depth exploration of the theoretical foundations of potential methodologies. Subsequently, the framework's structure is introduced, incorporating the studied methodologies. Comprising four consecutive steps, this framework collectively serves its ultimate objective: the revelation of functional connectivity within the human brain, particularly during the passive execution of a task, based on EEG measurements and an external input. The performance of the designed framework is validated by applying it to an available dataset.

This chapter serves as the conclusion of this report, providing a comprehensive overview of the essential findings and their alignment with the research's primary objective. It initiates with a summary of the main findings, followed by a discussion of the results. The discussion encompasses the significance of the developed framework, the implications drawn from the outcomes, and methodological considerations discussing its potential and limitations. The chapter concludes with recommendations for future research.

5.1. Final remarks

The multi-step mathematical framework has been developed and validated to fulfil the research objective, employing a dataset from Vlaar et al. [20]. The dataset includes EEG data and the angle of disturbance applied by the robotic wrist manipulator, serving as the exogenous input. The investigation into the application of the framework on the dataset aims to evaluate the efficacy and performance of the algorithm.

The first step of the algorithm involves preprocessing EEG data. This step serves as an essential data preparation measure, partly due to the susceptibility of raw EEG data to noise. The utilisation of filtering techniques, epoch extraction, and subsequent averaging leads to the generation of SSRs, optimising the dataset for further analysis and providing clarity regarding the brain's response to stimuli.

The second step of the framework employs a data-driven approach to model the dynamics of brain activity. This approach relies on subspace identification techniques, namely N4SID and PO-MOESP, to identify a state-space model. The assessment, guided by the VAF metric, reveals high success rates for both N4SID (78%) and PO-MOESP (90%). The VAF values, combined with results from the simulation and model verification, robustly support the effectiveness of both methods in representing EEG activity. Consequently, it suggests that linear models effectively represent brain dynamics, particularly in passively performed tasks stimulating the brain.

The third step of the designed framework facilitates the transmission from latent states to brain sources. The latent states are the states simulated from the identified state-space model in the second step of the algorithm. A head model and sensor positions for each participant can be extracted from the acquired data. This step is critical in enhancing the interpretability of the outcomes produced in the following stage by contributing to a better understanding of brain dynamics and connectivity. However, challenges arise when working with EEG data, mainly when analysing sources deep within the brain, such as the cerebellum and thalamus. When evaluating the reconstructed electrical brain activity, it is evident that lower activity is detected in the subcortical grey nuclei, insula and cerebellum. This observation aligns with the hypothesis that deeper brain regions exhibit reduced electrical activity due to measurement challenges, indicating the successful execution of the transition from latent to source state through this procedure.

The final step of the algorithm conducts the Granger causality analysis on the identified state-space model. This step allows the determination of the statistical dependencies between nodes in a network, in this context representing the functional connectivity between distinct brain sources. The process concludes in Granger causality heatmaps and network connectivity diagrams, offering insights into the magnitude and direction of statistical influence among sources, enabling directed functional connectivity assessment and participant comparisons.

In conclusion, the algorithm developed, consisting of four distinct steps, has successfully achieved the research objective. This framework provides a data-driven mathematical model representing the brain's cortical network using EEG data in the presence of an exogenous input.

5.2. Discussion of the results

The developed framework could have significant implications for neuroscience by paving the way for a deeper comprehension of brain dynamics and its interactions with external stimuli. The successful creation of the data-driven mathematical model that integrates EEG data and incorporates an exogenous input extends the existing body of knowledge in the field, addressing a gap in the literature. Additionally, the model holds promise for practical applications, including the realm of brain dynamics control, encompassing BCI, and its potential in diagnosing neurological disorders.

In data-driven modelling, the developed mathematical framework demonstrates its capability to effectively capture brain dynamics within a linear state-space model, employing subspace identification methods. This achievement highlights the potential for using similar data-driven approaches in modelling complex systems across diverse domains. The successful integration of exogenous inputs into the model presents a valuable case study for researchers working with statistical dependencies within dynamic systems influenced by external factors, providing valuable insights into modelling and analytical methodologies. Furthermore, the results suggest that choosing the PO-MOESP method over N4SID in comparable models can enhance overall performance.

As stated, the research undertaking yields valuable insights. Nonetheless, specific findings are noteworthy or warrant further examination. This section provides a thorough analysis of the research findings and their implications, comprising two primary subsections. The first part is dedicated to methodological considerations, enabling an exploration of the various methodological choices made throughout the research and pinpointing areas that may benefit from alternative approaches in future investigations. Subsequently, the section transitions to a comparative analysis, where the research findings are compared to the existing body of literature.

5.2.1. Methodological considerations

In the pursuit of understanding brain connectivity, this section delves into several methodological considerations. A noteworthy observation arises during the second step of dataset utilisation, indicating that the PO-MOESP method outperforms the N4SID method in terms of performance. This performance difference may be attributed to the distinct projection matrices employed by both methods. Revising the preliminaries on projection matrices in Section 2.1.4, the oblique projection of the row space of the future input Hankel matrix, used in N4SID, involves an additional correction term containing the Moore-Penrose pseudoinverse. While the Moore-Penrose pseudoinverse can help to deal with situations where the projection matrix is not orthogonal or perfectly aligned, it can also introduce some inaccuracy or offset in the results. In contrast, the orthogonal projection used in PO-MOESP obviates the necessity for such supplementary corrections when the projection matrices are ideally orthogonal. This indicates that the fundamental nature of the projection matrix used is inherently orthogonal. Further investigation could be conducted to confirm the validity of this deduction.

The third step of the framework involves transitioning from latent to source states by utilising an atlas encompassing various brain sources. When the source locations are mapped onto the atlas, specific source points share the same tissue labels. This observation suggests that multiple source points pertain to one anatomical region within the atlas space. Therefore, as a deliberate design choice within this framework, it is decided to average the grid locations and the entries within the lead-field matrix, which indicate the extent to which the activity can be measured at the scalp for each source. Methodologically, it might be necessary to revisit this choice and explore the incorporation of weighting matrices.

Upon generating the transformation matrix, the simulation of source activity is conducted with the exclusion of the noise component. It is essential to highlight that the simulation of source activity extends slightly beyond the central scope of this research, which primarily revolves around unravelling the connectivity patterns between brain sources rather than delving into the detailed activities of those sources. As a result, the exploration of the noise component has received limited attention. Nonetheless, there is room for further exploration in this domain to refine the precision in estimating source electrical activity. This would not only improve accuracy but also allow more robust conclusions about the relationship between source activity and connectivity.

During the implementation of the fourth step, the investigation into brain network connectivity patterns unveils notable connections with subcortical regions and sources situated in deeper brain structures. When assessing the relation between functional connectivity and electrical activity of brain sources, a remarkably heightened statistical dependence has been observed in the deeper brain sources, exemplified by the thalamus and insula. This heightened dependence can be attributed to their relatively lower levels of electrical activity. This lower electrical activity could render them more susceptible to external influences, making even minor perturbations significantly impact their neural activity. This revelation sheds light on the vulnerability of these regions to external factors and prompts a comprehensive reconsideration of the methodological approach.

In particular, questions arise regarding incorporating deeper brain sources in the analysis, given the inherent challenges in accurately measuring their electrical activity using EEG. Furthermore, the findings underscore the necessity for a more profound examination of the calculations behind Granger causality values based on correlations. This necessity encourages exploration into whether the observed low electrical activity correlation with high functional connectivity truly reflects a natural phenomenon or if it necessitates adjustments to the methodology to ensure a more precise interpretation of the results.

5.2.2. Comparative analysis with existing literature

When comparing the findings to relevant literature, similarities, differences, and novel insights arise. Applying the initial step of the framework on the dataset results in the generation of the SSR per electrode for each participant. Analysing these SSRs reveals that the side contralateral to the perturbation site, where the participant is being stimulated, exhibits higher activity than the ipsilateral side. This trend is clearly illustrated in the plots depicting the SSR for two electrodes, CP3 and CP4, positioned on the left and right sides of the head, respectively. These results are of particular interest because they align with the expected surge in activation during tasks that activate the contralateral somatosensory cortex [32].

Earlier research has successfully explored the determination of functional brain connectivity during resting state by employing the Granger causality within the framework of state-space models that do not incorporate an input [13, 14]. These investigations substantiate the potential of subspace identification as a pathway for modelling brain dynamics. However, it is notable that the literature lacks explicit mention of the specific subspace identification method employed, often defaulting to the use of N4SID [13, 14]. This is despite the findings from this study indicating that PO-MOESP outperforms N4SID.

The development of Granger causality for state-space models excludes the integration of an input component [11]. This exclusion results in state-space models and successive analyses that lack the incorporation of the B and D system matrices. Upon revisiting the theoretical framework and conducting an analysis of the DARE, it can be inferred that the estimation of the noise covariances, which are essential for solving the DARE and subsequently calculating the Granger causality measure F_{ij}, exhibits variations when the model incorporates an exogenous input. In contrast, the prediction error covariance Σ remains consistent even when an exogenous input is introduced into the state-space model.

Studies using the same datasets have demonstrated that sensorimotor processing activates various cortical regions, including the primary somatosensory cortex, which receives proprioceptive and tactile sensory information [20, 32]. The postcentral gyrus predominantly represents the somatosensory cortex in this study. Within the columns of the Granger causality heatmap, increased dependency is evident, signifying a statistical impact on other brain sources. This supports the idea that the postcentral gyrus functions as a central hub, influencing various network regions where different information streams converge, exerting a significant influence on these regions. However, it is essential to note that the columns of the postcentral gyrus do not exhibit the highest dependencies, which does not align with the expectations drawn from literature [32]. Furthermore, other areas involved in sensorimotor processing include the thalamus, posterior parietal cortex, and insular cortex [37, 38, 39]. Notably, the angular gyrus within the posterior parietal cortex, as well as the insula and thalamus, display distinctive statistical dependencies on other brain sources processing the information. These insights are derived from examining inter-participant connectivity networks, strengthening the expectation of heightened connectivity among the actively engaged sources.

5.3. Future research

Several topics lend themselves to further research. Firstly, the mathematical framework's performance should be evaluated across a range of datasets, including both simple sensorimotor processes and more complex tasks, to enhance the research reliability and robustness. While the obtained dataset was instrumental in addressing the research objectives, it is essential to acknowledge the limitations of relying solely on a single data source. Additionally, two critical considerations must be addressed for seamless dataset integration within the framework. First, the suitability of employing open-loop identification methods in conjunc-

tion with specific datasets or tasks should be thoughtfully assessed. Second, assigning values to the stimulus as the exogenous input is essential.

The designed framework identifies the brain dynamics within a linear state-space model, although existing literature frequently suggests modelling brain dynamics should incorporate nonlinear elements [32]. In the context of the specific task employed in this research, the linear approach has demonstrated excellent performance. However, for more complex tasks, there exists a promising avenue for future research. This involves expanding the framework to integrate nonlinear elements alongside the linear state-space model. Potential nonlinear model identification techniques, such as the Hammerstein-Wiener or a neural state-space model, can be considered to investigate whether including nonlinear components can further enhance the model's performance.

In this study, the framework utilises open-loop subspace identification for modelling brain dynamics, which is well-suited for the passively performed tasks under investigation. However, when the research is extended to encompass more complex tasks that involve active participant engagement and feedback mechanisms, the dynamics of the brain become more intricate. In such scenarios, employing closed-loop subspace identification methods becomes essential to identify the state-space model representing the brain dynamics.

Evaluating the framework's reliability through utilising multiple datasets, coupled with the addition of nonlinear components and closed-loop subspace identification, could potentially enhance the comprehension of brain dynamics and functional connectivity, particularly in the context of more complex tasks. In the long run, these advancements might facilitate the diagnosis of neurological disorders. When utilising the framework for diagnosing neurological disorders, it becomes imperative to examine the threshold values applied in the context of Granger causality heatmaps and network connectivity diagrams. Currently, these thresholds vary between participants. This approach can pose challenges in detecting diminished connectivity patterns. Adopting a fixed threshold value or range could provide a more consistent basis for detecting alterations in connectivity, which can be particularly significant in diagnosing neurological disorders.

It is well-established that during the initial stages of Alzheimer's disease, a decline in brain connectivity can be observed. If this framework, or an adapted iteration, proves effective in handling more complex tasks, future experiments could be designed to investigate processes reliant on the coordinated activity of multiple brain regions. Incorporating datasets from these experiments into the framework may unveil distinctive connectivity patterns, particularly in patient groups, that could potentially facilitate the early detection of Alzheimer's disease.

A

Appendix A - Mathematical definitions

The theoretical foundation of this report relies upon a set of defined concepts and terms, which serve as the basis for formulating assumptions regarding the state-space models and subspace identification, explained in Chapter 2. The materials of this section are borrowed from Verhaegen and Verdult [16].

Time-invariance
A discrete-time state-space system can be represented as:

$$\begin{aligned} x(k+1) &= f(k, x(k), u(k)), \\ y(k) &= h(k, x(k), u(k)). \end{aligned} \tag{A.1}$$

The first equation describes the time evolution of the state at time instant $k+1$, given the state and the input at time instant k. The second equation shows how the value of the output at time instant k depends on the values of the state and the input at that particular time instant.

A significant category of systems is the group known as time-invariant systems [16]:

Definition A.1 (Time-invariance)**.** The system of Equation A.1 is time invariant if the functions f and h are independent of time k for all x, u and y

In other words, a time-invariant system is described by functions f and h that do not change over time:

$$\begin{aligned} x(k+1) &= f(x(k), u(k)), \\ y(k) &= h(x(k), u(k)). \end{aligned} \tag{A.2}$$

Linearity
Another important class of systems is the class of linear systems. To define a linear system, it is important to define a linear function [16]:

Definition A.2 (Linear function)**.** The function $f : \mathbb{R}^m \to \mathbb{R}^n$ of Equation A.1 is linear if, for any two vectors $x_1, x_2 \in \mathbb{R}^m$ and any $\alpha, \beta \in \mathbb{R}$,

$$f(\alpha x_1 + \beta x_2) = \alpha f(x_1) + \beta f(x_2). \tag{A.3}$$

Definition A.3 (Linear system)**.** The state-space system of Equation A.1 is a linear system if the functions f and h are linear functions concerning $x(k)$ and $u(k)$.

Therefore, a "linear time-invariant" (LTI) state-space system (with additional noise) can be represented as Equation 2.1.

Dynamical system
A system transforms a particular set of input signals into another set of signals, referred to as output signals. Most systems are dynamic systems. The term "dynamical" refers to the system's memory, meaning that the current output signal is not only determined by the current input but also influenced by its time history and the time history of the input signal [16]. In mathematical terms, a dynamical system can be defined as follows:

Definition A.4 (Dynamical system)**.** A dynamical system is characterised by a set of state variables, denoted as x, representing the system's internal configuration or state at a given time. The system's evolution is governed by a set of equations, often called "dynamical equations" or "differential equations." These equations describe how the state variables change over time as a function of their current and past values and possibly external inputs.

Noise
In system modelling and estimation, it is crucial to account for various uncertainty or disturbances. Two familiar sources of uncertainty are process noise $w(k)$ and measurement noise $v(k)$, incorporated in the state-space model of Equation A.1. Process noise represents the inherent randomness or unpredictability in the evolution of a dynamic system, which is typically modelled as a stochastic sequence with specific statistical properties that capture the system's uncertainty. Process noise must be included in a model due to model approximations and model integration errors. Measurement noise accounts for the imperfections and inaccuracies in the measurement or observation process. Like process noise, measurement noise is also represented as a stochastic sequence, but it characterises the uncertainty associated with sensor readings or measurements.

One common assumption made when modelling process and measurement noise is the use of a zero-mean white noise sequence:

Definition A.5 (Zero-mean white noise)**.** A white noise sequence is a stochastic signal characterised by constant power spectral density across all frequencies. Zero-mean white noise specifically has an expected value or mean of $E[n(k)] = 0$, where $n(k)$ represents the value of the white noise sequence at time k.

Observability
Observability is a fundamental concept in system theory that assesses the ability to infer the complete internal state of a dynamic system based solely on its external output measurements. In other words, an observable system allows all its internal states to be uniquely determined from the observed outputs over a finite time horizon. This property is crucial in various applications, such as control design, estimation, and diagnosis, as it ensures that no essential information is hidden from external observations. For a system to be observable, the following definition should hold:

Definition A.6 (Observability)**.** The system of Equation 2.1 is observable if any initial state $x(k_a)$ is uniquely determined by the corresponding zero-input response $y(k)$ for $k_a \leq k \leq k_b$ with k_b finite.

To test the observability of an LTI system, the observability rank condition can be used Equation 2.33:

$$\text{rank}(O_n) = n, \tag{A.4}$$

making use of the observability matrix O_n:

$$O_n = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}. \tag{A.5}$$

Controllability
Similarly, controllability is a fundamental principle within system theory that examines the capacity to guide any state of the dynamic system through external input actions. A controllable system can be directed from any initial state to any desired state within a finite period by selecting appropriate control inputs. For a system to meet the criteria of controllability, consider the following definition:

Definition A.7 (Controllability)**.** The system described by a state-space model is deemed controllable if, given any initial state $x(k_a)$, there exists an input signal $u(k)$ for $k_a \leq k \leq k_b$ such that $x(k_b) = 0$ for some k_b.

The problem with controllability in discrete time is that specific systems are categorized as controllable, even though the input does not genuinely contribute to steering the state [16]. To illustrate, consider a system characterised by $A = 0$ and $B = 0$: in this scenario, every initial state converges to zero, but the input holds no sway over the system's behaviour. As a result, a more robust concept termed "reachability" is frequently employed:

Definition A.8 (Reachability)**.** The system described by a state-space model is reachable if for any two states x_a and x_b, there exists an input signal $u(k)$ for $k_a \leq k \leq k_b$ that will transfer the system from the state $x(k_a) = x_a$ to $x(k_b) = x_b$.

Analogous to the observability assessment, the reachability of the LTI system can be evaluated through the reachability rank condition:

$$\text{rank}(C_n) = n, \tag{A.6}$$

using the controllability matrix C_n:

$$C_n = \begin{bmatrix} B & AB & \dots & A^{n-1}B \end{bmatrix}. \tag{A.7}$$

B

Appendix B - Anatomical regions of interest

The brain sources employed for mapping from latent to source states are detailed in Table B. The brain sources are labelled according to the atlas "ROI_MNI_V4" of the Automated Anatomical Labeling 2 (AAL2) atlas.

The table lists the brain regions, the associated brain sources, and their corresponding labels as featured in the heatmaps and network connectivity diagrams. In these diagrams, the letter L or R is appended to the labels, indicating the placement in either the left or the right hemisphere. This letter designation is omitted from the table. Furthermore, a visual representation of the brain regions on a brain map is provided in Figure B.1, with each region distinguished by a unique colour.

Table B.1: List of the anatomical regions and sources according to the AAL Atlas ROI_MNI_V4.

Anatomical description	**Label ROI_MNI_V4**
Central sulcus	
Precentral gyrus	Precentral
Frontal lobe	
Superior frontal gyrus	Frontal Sup
Middle frontal gyrus	Frontal Mid
Middle frontal gyrus (orbital)	Frontal Mid Orb
Inferior frontal gyrus (triangular)	Frontal Inf Tri
Inferior frontal gyrus (orbital)	Frontal Inf Orb
Rolandic operculum	Rolandic Oper
Supplementary motor area	Supp Motor Area
Superior Frontal gyrus (medial)	Frontal Sup Medial
Rectus gyrus	Rectus
Insular lobe	
Insula	Insula
Limbic lobe	
Anterior cingulate gyrus	Cingulum Ant

Continued on next page

Table B.1: (continued from previous page)

Anatomical description	**Label ROI_MNI_V4**
Median cingulate gyrus	Cingulum Mid
Parahippocampal gyrus	Parahippocampal
Amygdala	Amygdala
Occipital lobe	
Calcarine cortex	Calcarine
Cuneus	Cuneus
Lingual gyrus	Lingual
Superior occipital gyrus	Occipital Sup
Middle occipital gyrus	Occipital Mid
Fusiform gyrus	Fusiform
Parietal lobe	
Postcentral gyrus	Postcentral
Superior parietal gyrus	Parietal Sup
Inferior parietal gyrus	Parietal Inf
Supramarginal gyrus	SupraMarginal
Angular gyrus	Angular
Precuneus	Precuneus
Sub cortical grey nuclei	
Caudate	Caudate
Pallidum	Pallidum
Thalamus	Thalamus
Temporal lobe	
Superior temporal gyrus	Temporal Sub
Temporal pole (superior)	Temporal Pole Sup
Middle temporal gyrus	Temporal Mid
Inferior temporal gyrus	Temporal Inf
Cerebellum	
Cerebellum	Cerebellum
Vermis	Vermis

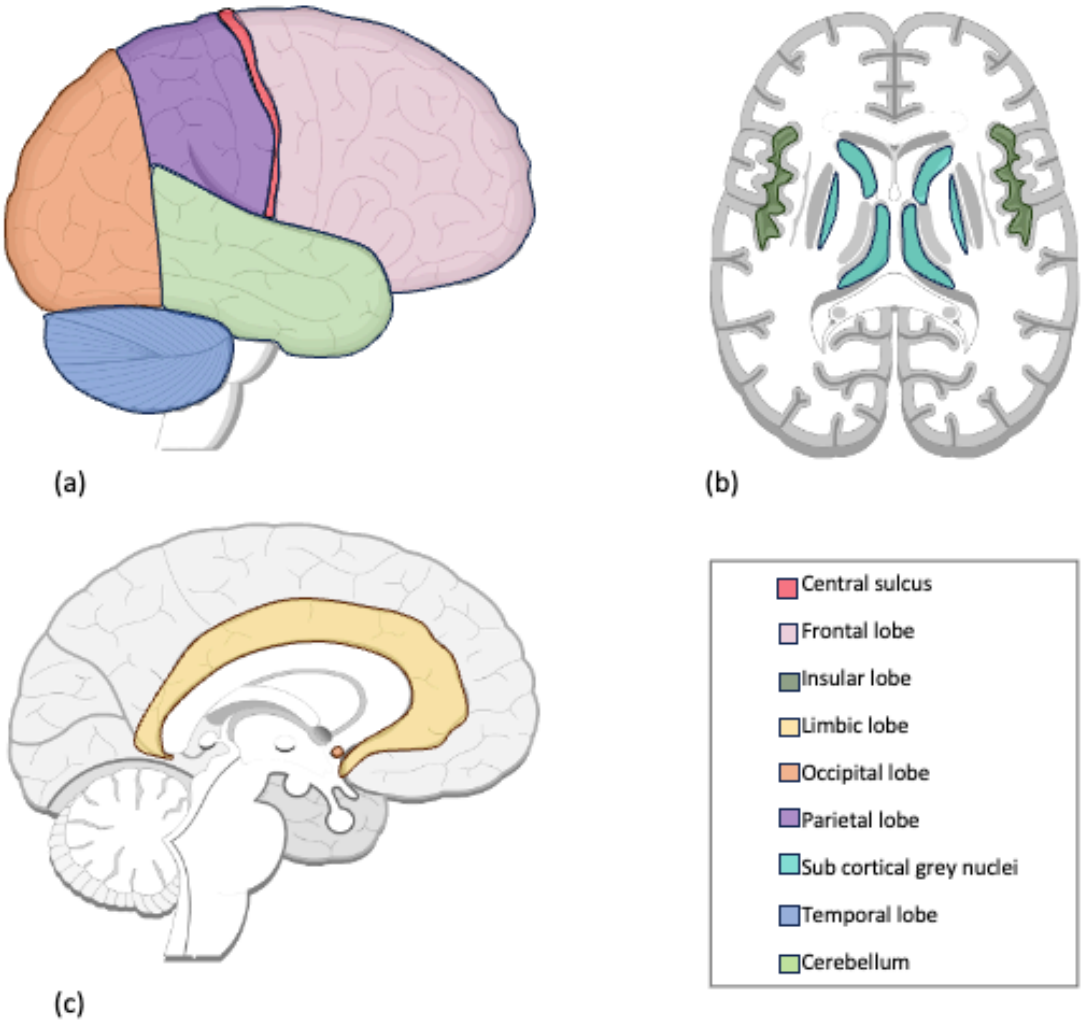


Figure B.1: Spatial orientation of the brain regions, to which the defined brain sources are associated, based on the Automated Anatomical Labeling 2 (AAL2) atlas. These regions are differentiated by various colours, as described in the accompanying legend. (a) The lateral external view of the brain. (b) The interior view of the brain. (c) The sagittal view of the brain.

C

Appendix C - Expanded visuals and tables for model analysis

Chapter 4 of this report detailed the model analysis. This appendix serves as a supplementary section, offering additional visual representations and tables that further elucidate the results and conclusions presented in that chapter.

This section includes visualisations of the raw EEG data, the reconstruction of source activity, and the Granger causality heatmaps, along with their corresponding connectivity diagrams. These supplementary materials provide a more comprehensive and in-depth understanding of the research findings and analysis.

C.1. Data analysis

In the initial preprocessing stage of the algorithm applied to the dataset from Vlaar et al. [20], the data from the electrodes labelled as AF4 and F6 are identified as problematic channels and subsequently excluded from further analysis. As depicted in Figure C.1, the electrical activity of these channels can be compared to that of six other channels, emphasising the irregular behaviour. It is worth noting that the electrical activity depicted in this figure is captured before the completion of the preprocessing phase.

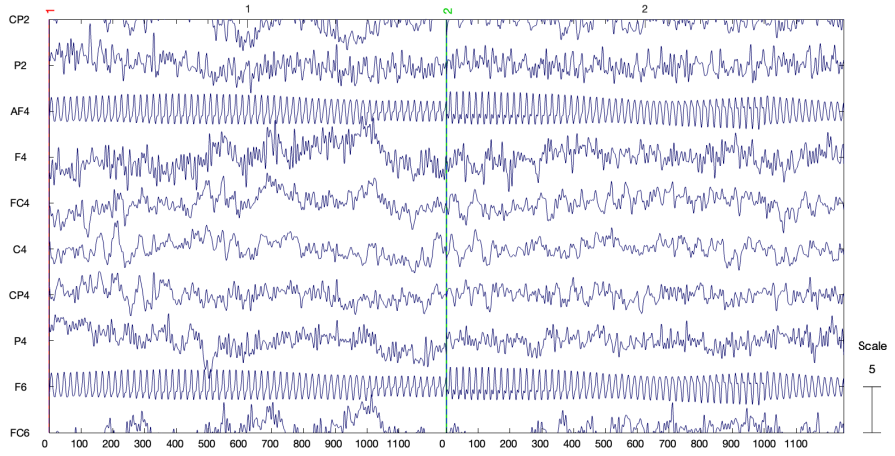


Figure C.1: The scalp recorded electrical activity from 8 electrodes over two distinct epochs for Participant 1, showing the distinctive behaviour of electrode AF4 and F6. The x-axis represents time in milliseconds, and the y-axis displays channel labels

C.2. Source activity reconstruction

After creating the transformation matrix H, representing the mapping from latent to source states, it becomes possible to simulate the source activity z using the equation $z = Hx$, with the noise component being disregarded.

The outcome of this critical step is the relative measure of source activity, which provides insights into the intensity of neural activity within specific brain regions. To facilitate a comprehensive understanding, the results of the relative source activity measurements are presented in the figures below. These figures offer a systematic data organisation, categorised according to the corresponding brain regions, shedding light on the varying levels of neural activity in each source.

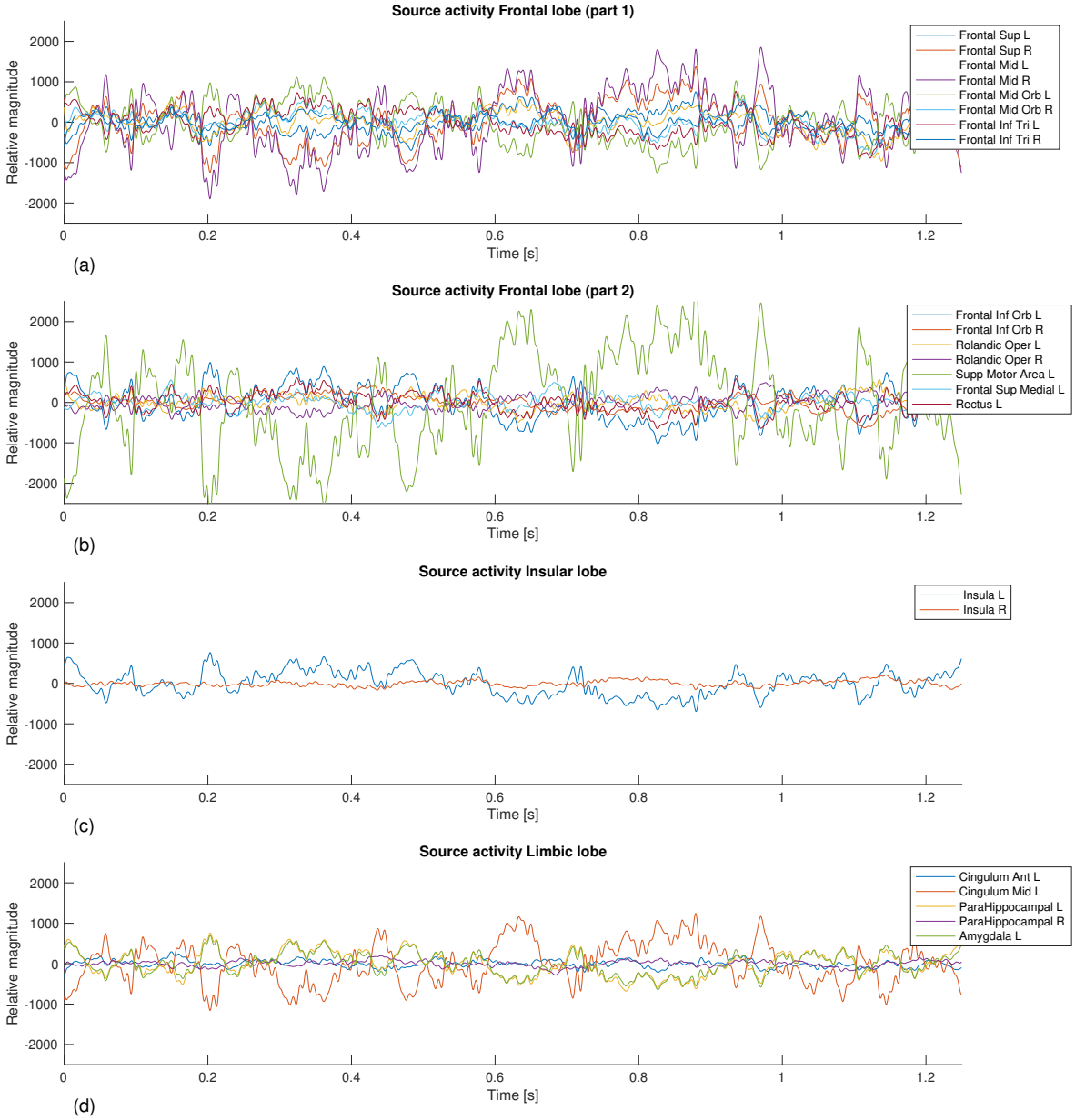


Figure C.2: Source activity reconstruction for the brain sources, part 1.

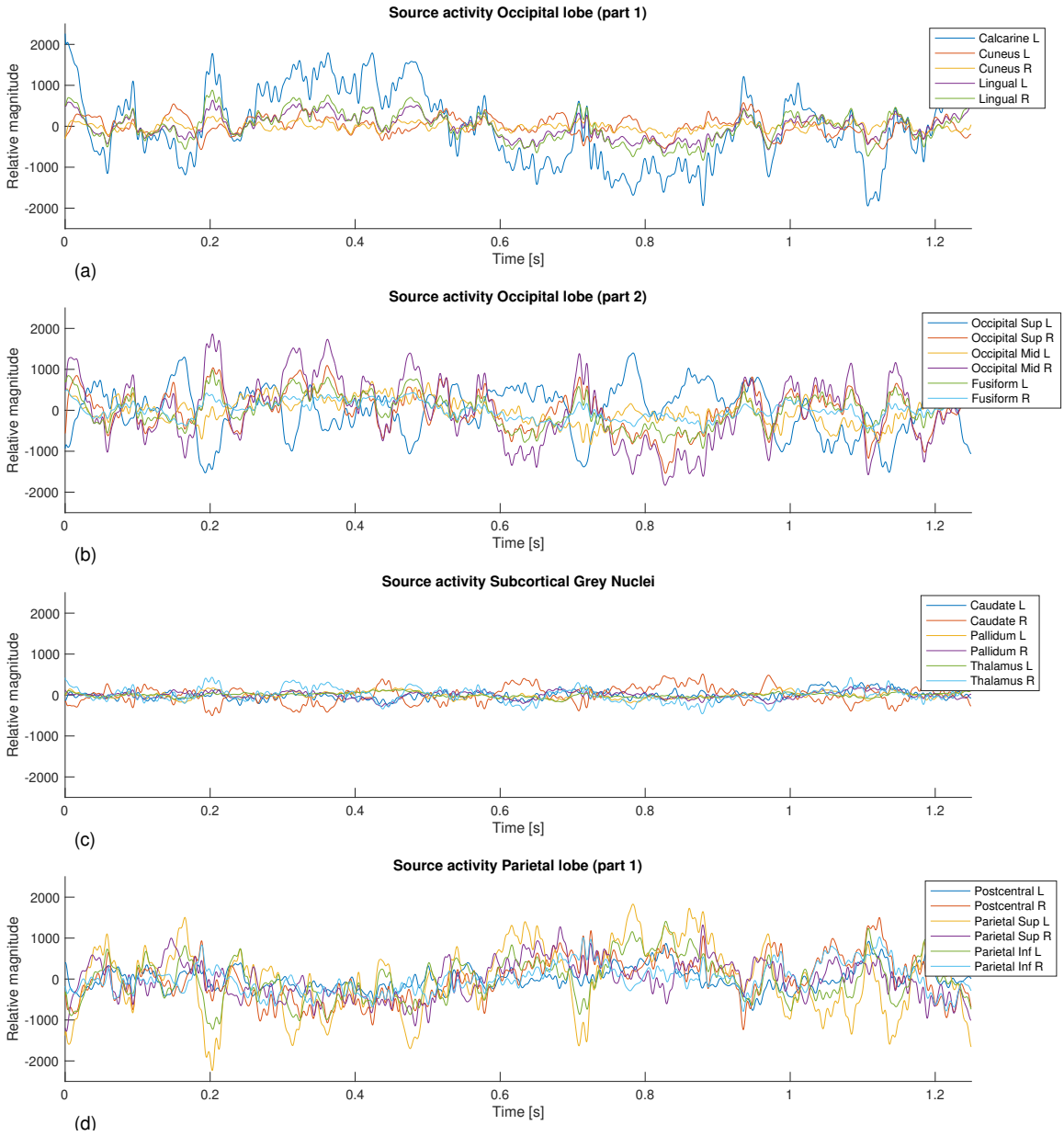


Figure C.3: Source activity reconstruction for the brain sources, part 2.

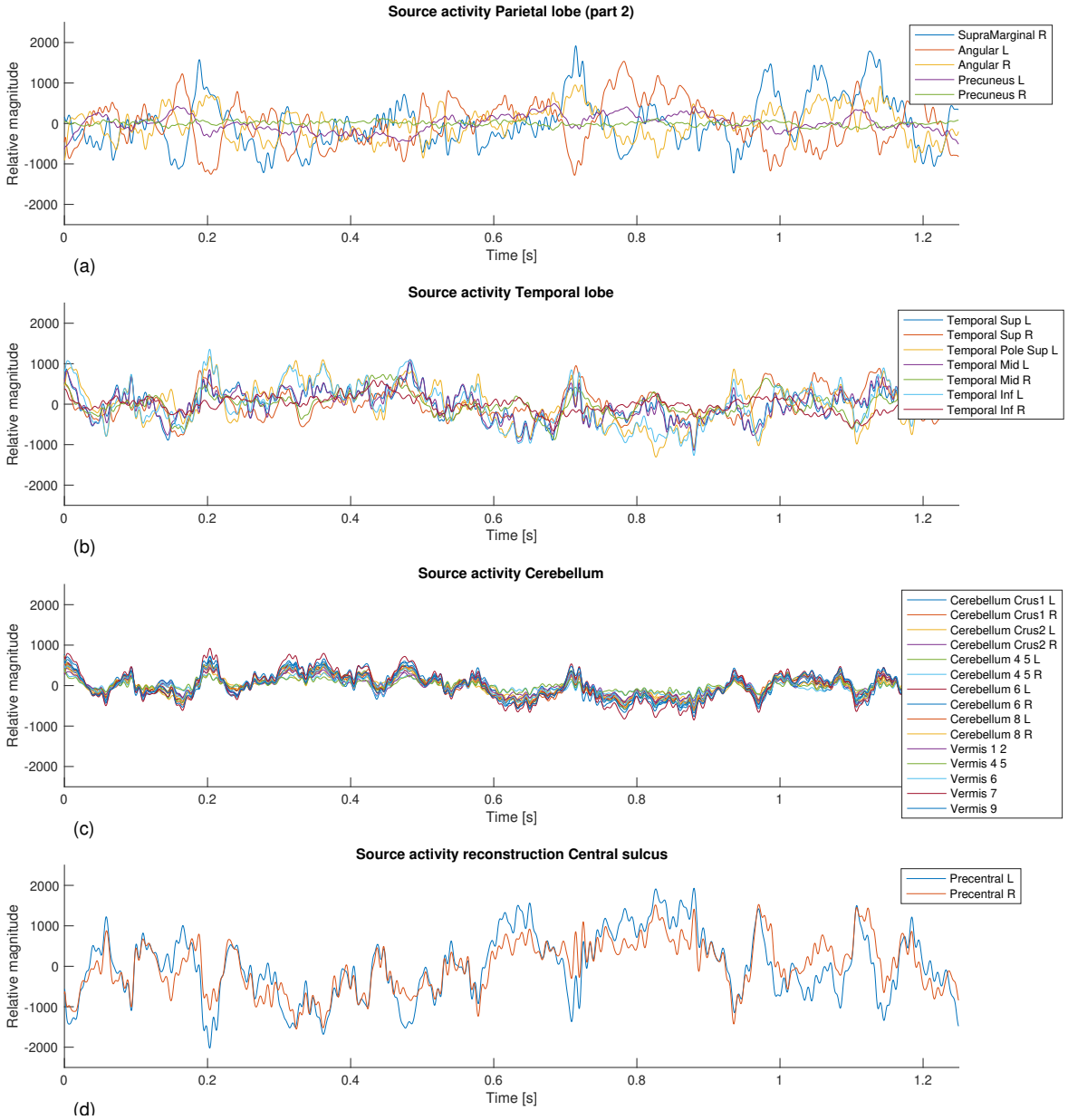


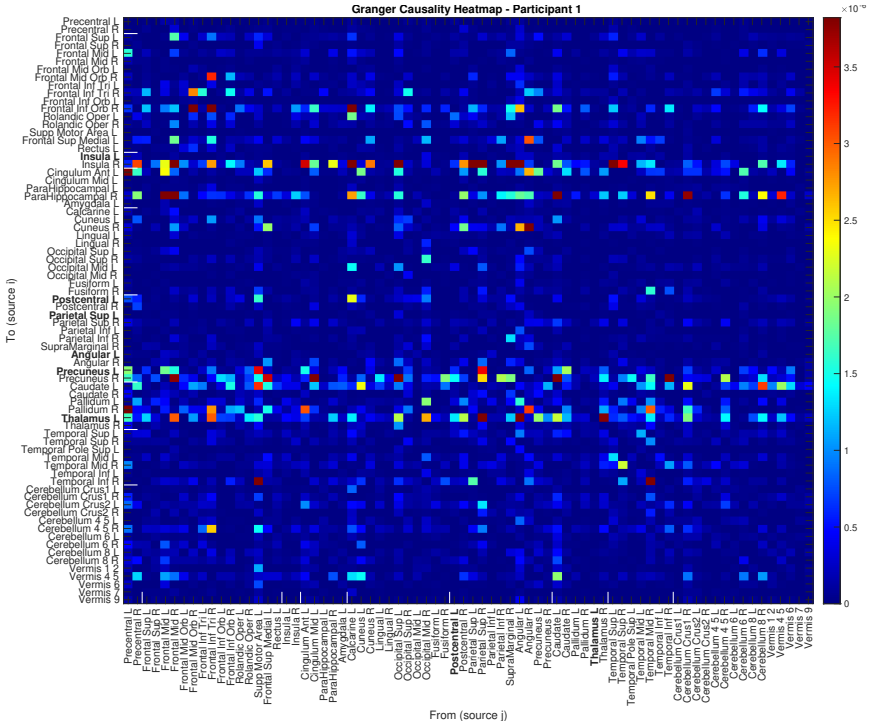
Figure C.4: Source activity reconstruction for the brain sources, part 3.

C.3. Brain connectivity

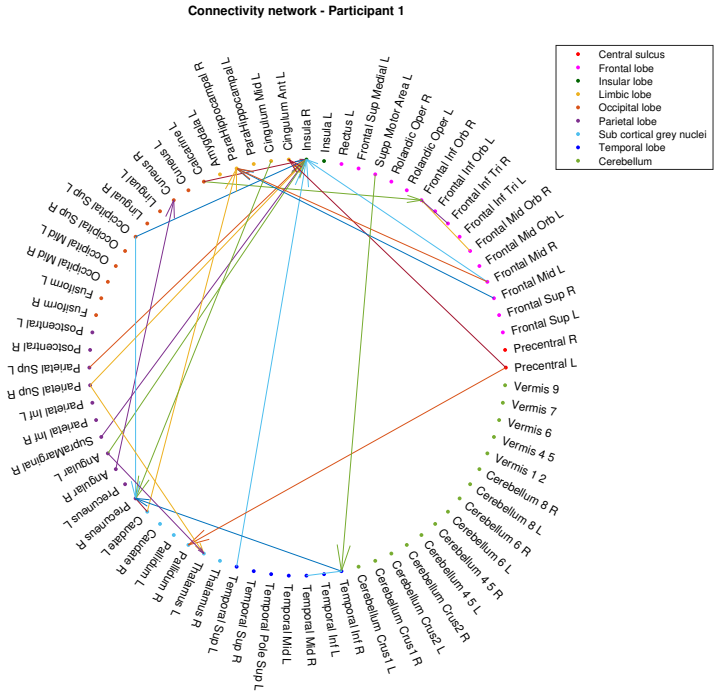
The functional connectivity results for all seven participants involved in this study are presented below. This collection of heatmaps allows for an insightful comparison of the differences and similarities observed in the functional connectivity patterns among the participants. To ensure completeness and comparability, the heatmap for Participant 1 is included, enabling a comprehensive analysis of the findings across all participants.

The outcomes of the Granger causality analysis are represented in the heatmaps. The heatmap employs a colour coding scheme to represent the strength of the connectivity, also denoted in the colour bar. From the Granger causality measures, the network diagrams are formed. The diagram consists of nodes representing brain sources and edges representing a directed statistical dependency between the nodes. In the figures, ”L” or ”R” designations indicate the hemisphere in which each element is situated. The brain sources are labelled according to the atlas ”ROI_MNI_V4”.

Functional connectivity of participant 1



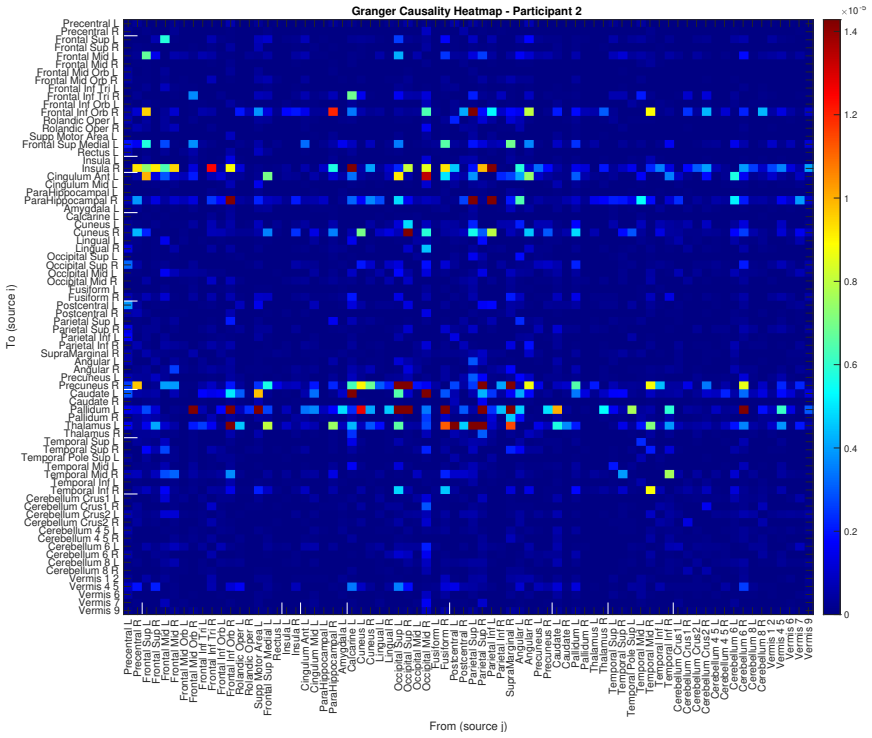
(a)



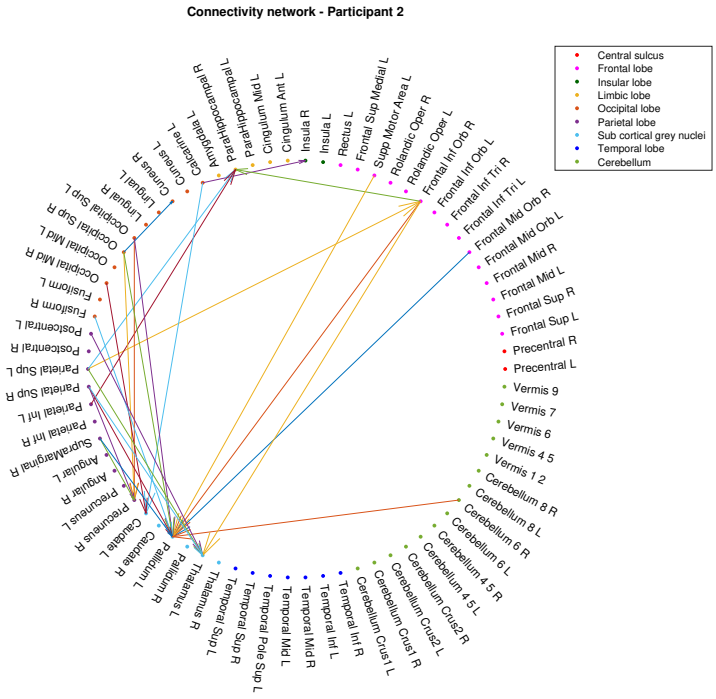
(b)

Figure C.5: The functional connectivity of the brain of Participant 1, determined from the Granger causality measure F_{ij}. (a) The Granger Causality heatmap. (b) The network connectivity diagram, representing the directed functional connectivity.

Functional connectivity of participant 2



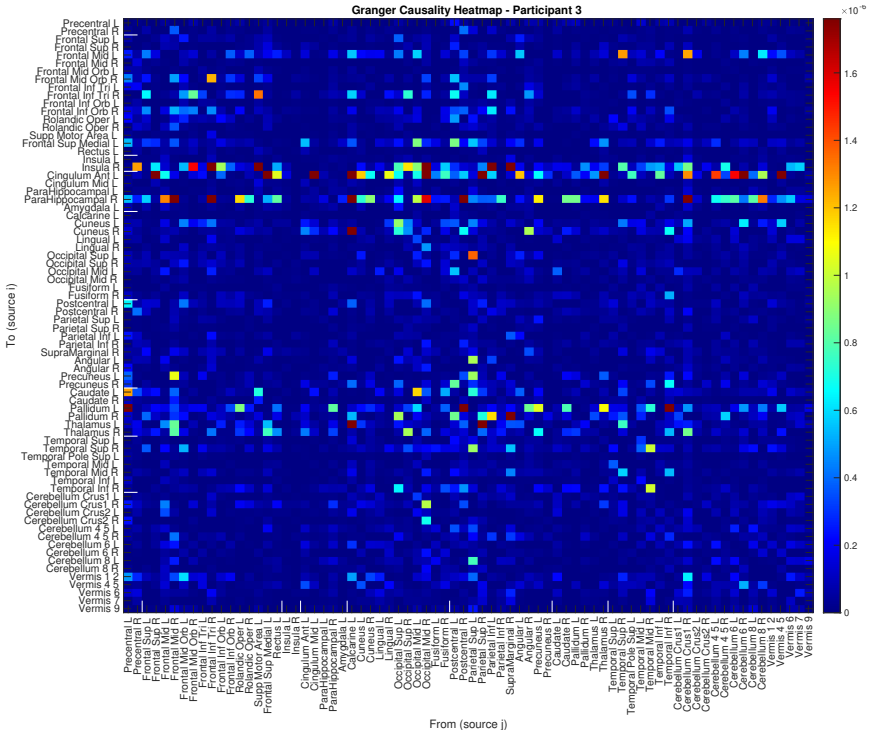
(a)



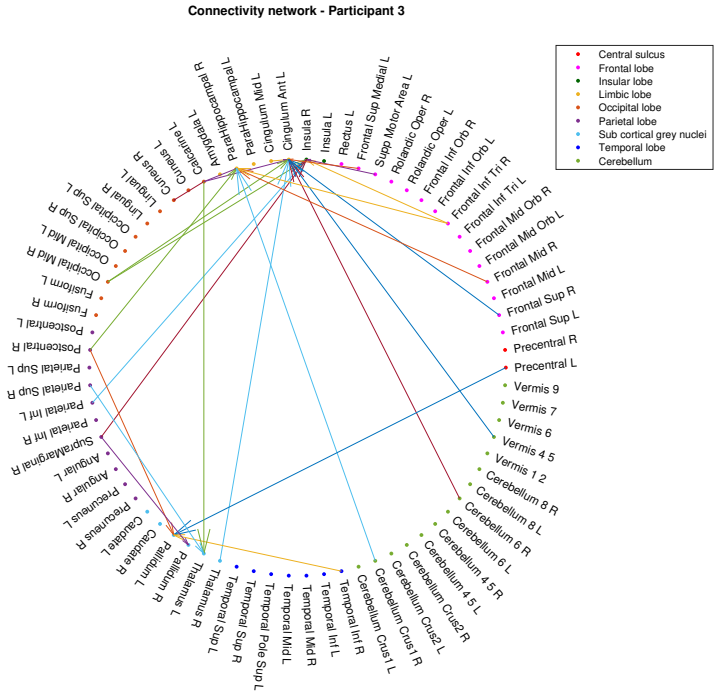
(b)

Figure C.6: The functional connectivity of the brain of Participant 2, determined from the Granger causality measure F_{ij}. (a) The Granger Causality heatmap. (b) The network connectivity diagram, representing the directed functional connectivity.

Functional connectivity of participant 3



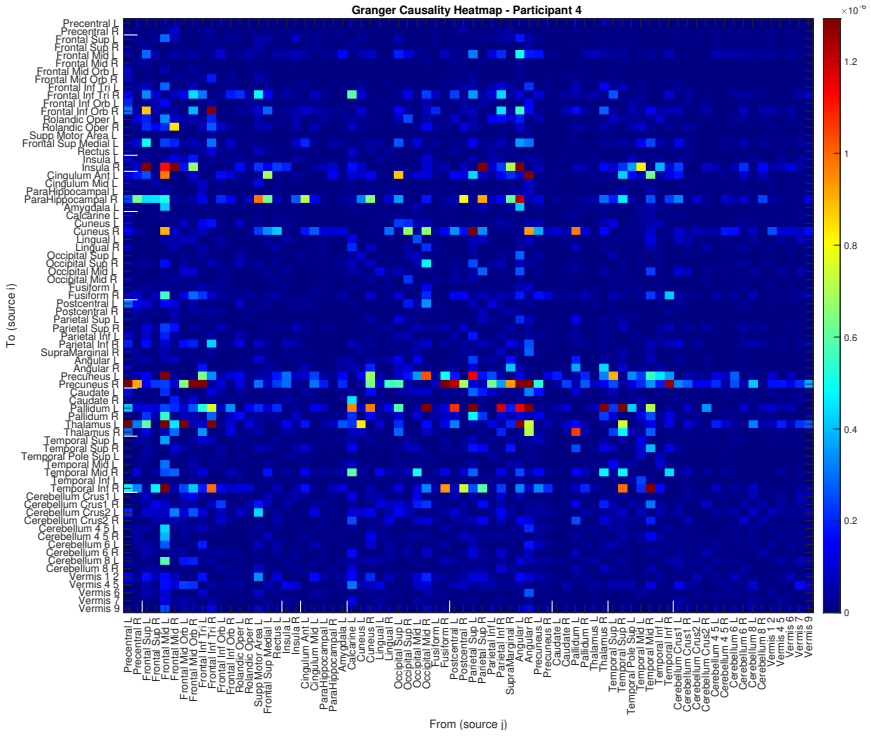
(a)



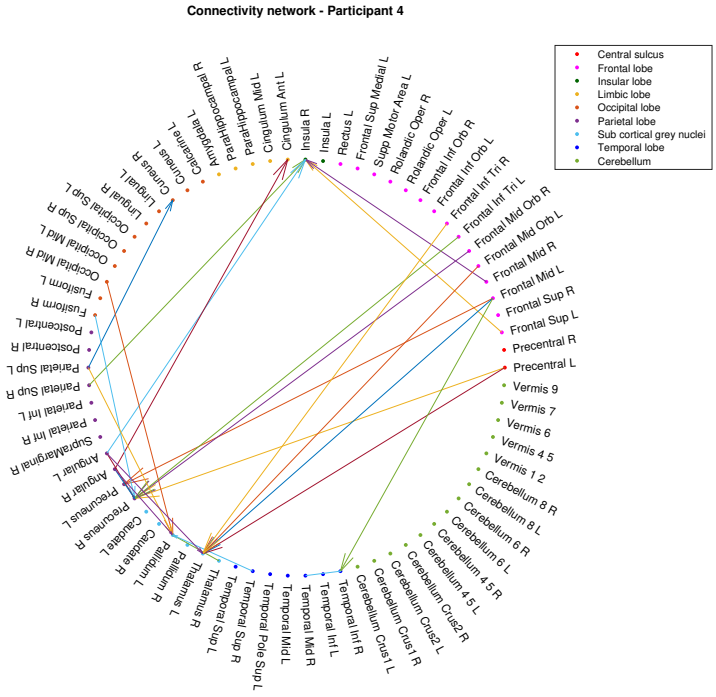
(b)

Figure C.7: The functional connectivity of the brain of Participant 3, determined from the Granger causality measure F_{ij}. (a) The Granger Causality heatmap. (b) The network connectivity diagram, representing the directed functional connectivity.

Functional connectivity of participant 4



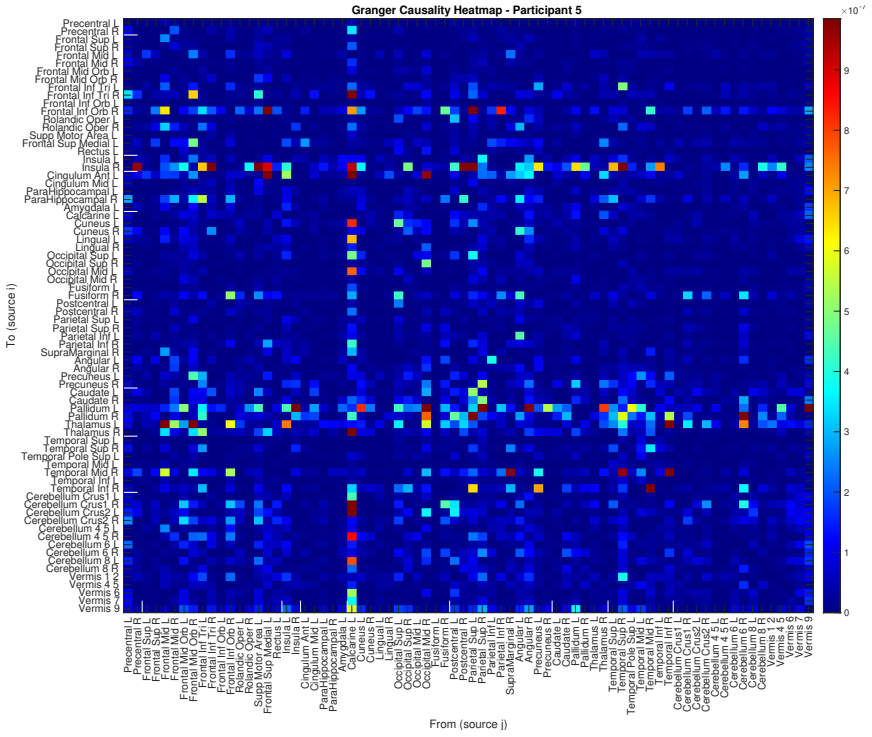
(a)



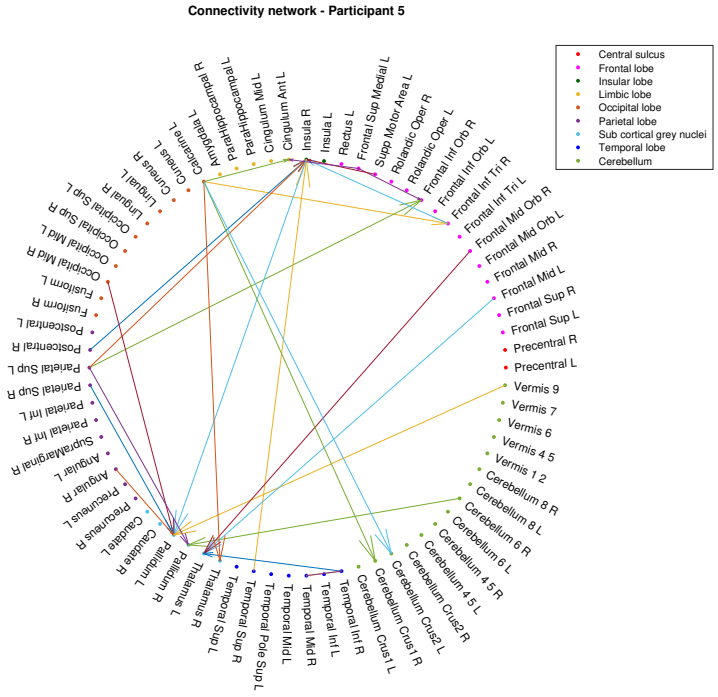
(b)

Figure C.8: The functional connectivity of the brain of Participant 4, determined from the Granger causality measure F_{ij}. (a) The Granger Causality heatmap. (b) The network connectivity diagram, representing the directed functional connectivity.

Functional connectivity of participant 5



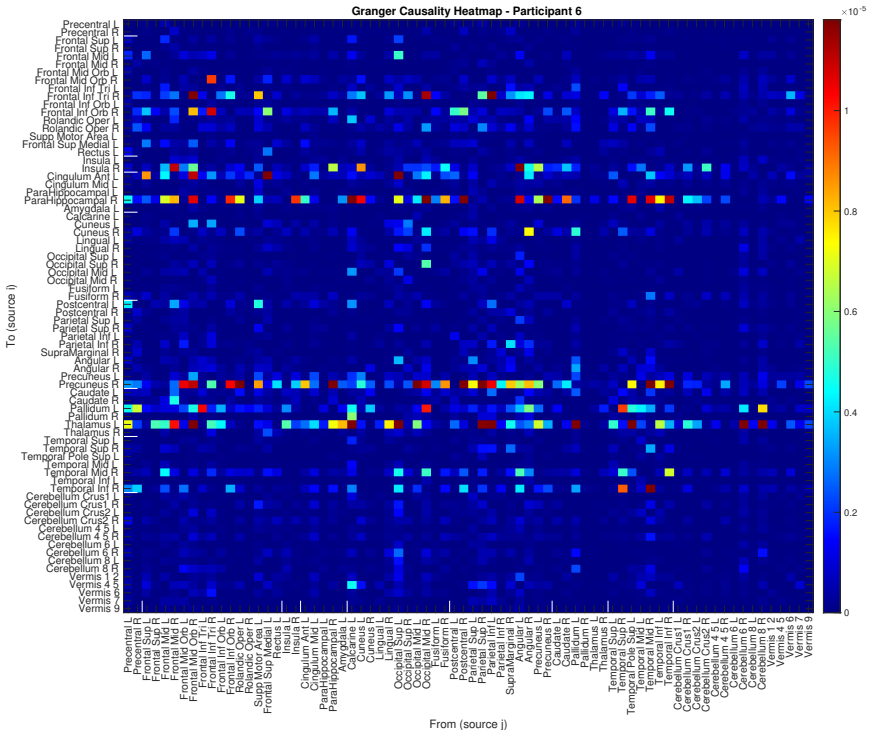
(a)



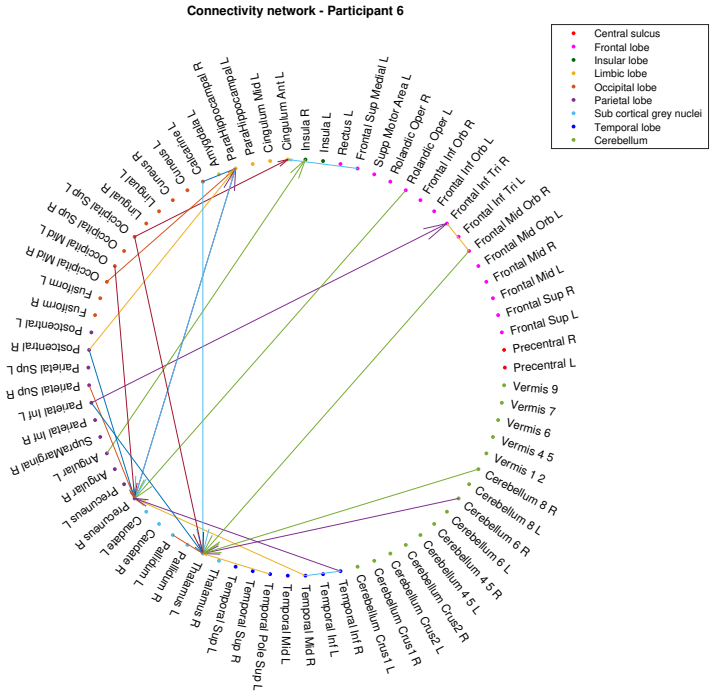
(b)

Figure C.9: The functional connectivity of the brain of Participant 5, determined from the Granger causality measure F_{ij}. (a) The Granger Causality heatmap. (b) The network connectivity diagram, representing the directed functional connectivity.

Functional connectivity of participant 6



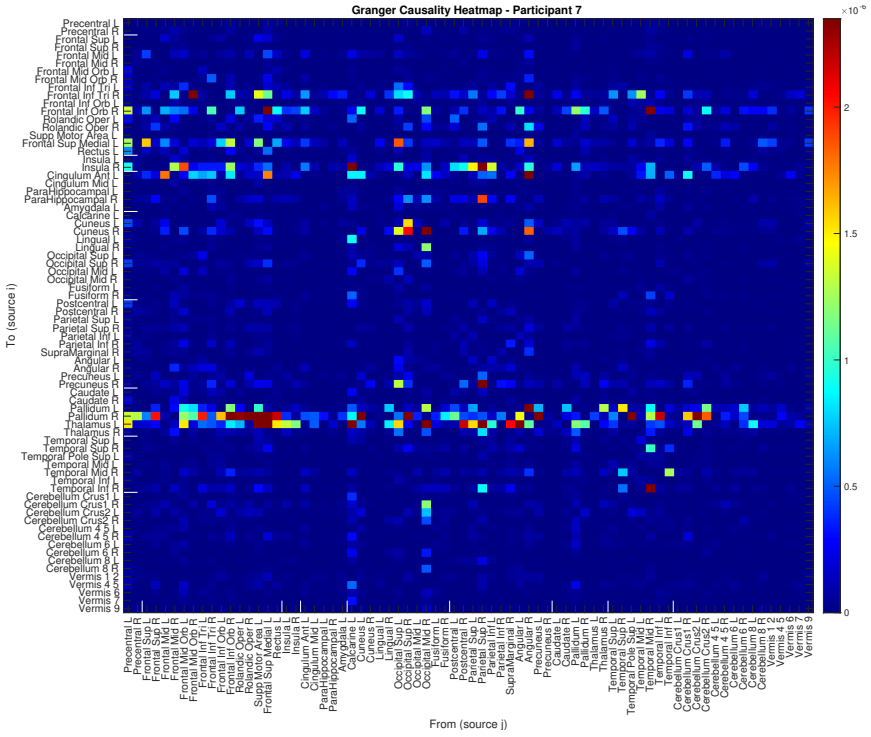
(a)



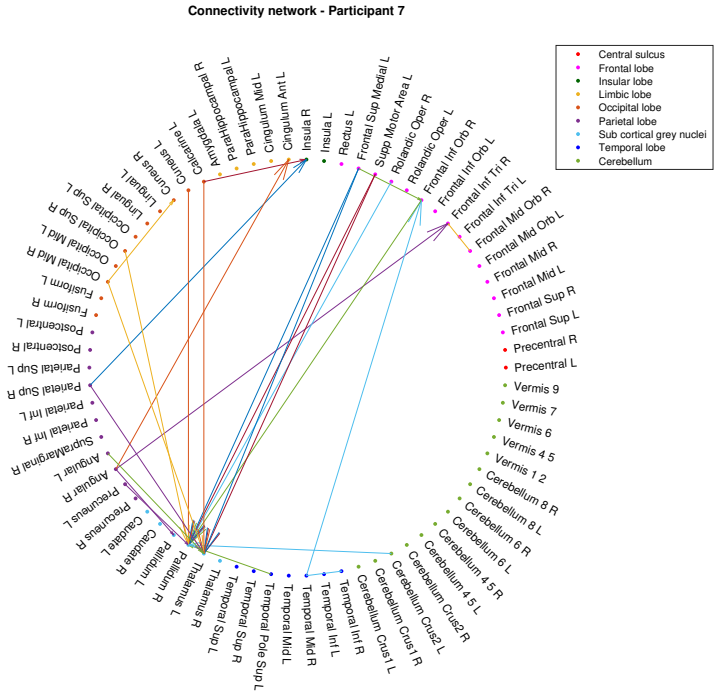
(b)

Figure C.10: The functional connectivity of the brain of Participant 6, determined from the Granger causality measure F_{ij}. (a) The Granger Causality heatmap. (b) The network connectivity diagram, representing the directed functional connectivity.

Functional connectivity of participant 7



(a)



(b)

Figure C.11: The functional connectivity of the brain of Participant 7, determined from the Granger causality measure F_{ij}. (a) The Granger Causality heatmap. (b) The network connectivity diagram, representing the directed functional connectivity.

D

Appendix E - Link to Matlab code

The steps of the designed mathematical model are programmed in distinct scripts utilizing Matlab R2023A (The Mathworks, Inc., Natick, MA, USA). The Matlab code of these steps can be found on a GitHub page. The link for this page is:

`https://github.com/slbakels/Thesis-files`

References

[1] S. P. Ellner and J. Guckenheimer. *Dynamic Models in Biology*. Princeton University Press, Dec. 2006.

[2] M. Breakspear. "Dynamic Models of Large-Scale Brain Activity". In: *Nature Neuroscience* 20.3 (Feb. 2017), pp. 340–352.

[3] J. M. Schoffelen and J. Gross. "Source Connectivity Analysis with MEG and EEG". In: *Human Brain Mapping* 30.6 (June 2009), pp. 1857–1865.

[4] K. Friston, R. Moran, and A. K. Seth. "Analysing Connectivity with Granger Causality and Dynamic Causal Modelling". In: *Current Opinion in Neurobiology* 23.2 (Apr. 2013), pp. 172–178.

[5] T. T. N. Do, C. T. Lin, and K. Gramann. "Human Brain Dynamics in Active Spatial Navigation". In: *Scientific Reports 2021 11:1* 11.1 (June 2021), pp. 1–12.

[6] C. M. Michel and M. M. Murray. "Towards the Utilization of EEG as a Brain Imaging Tool". In: *NeuroImage* 61.2 (June 2012), pp. 371–385.

[7] B. L. P. Cheung, B. A. Riedner, G. Tononi, and B. D. Van Veen. "Estimation of Cortical Connectivity from EEG Using State-Space Models". In: *IEEE Transactions on Biomedical Engineering* 57.9 (Sept. 2010), pp. 2122–2134.

[8] L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda. "Recipes for the Linear Analysis of EEG". In: (2005).

[9] C. W. Anderson, E. A. Stolz, and S. Shamsunder. "Multivariate Autoregressive Models for Classification of Spontaneous Electroencephalographic Signals during Mental Tasks". In: *IEEE Transactions on Biomedical Engineering* 45.3 (Mar. 1998), pp. 277–286.

[10] L. Chisci et al. "Real-Time Epileptic Seizure Prediction Using AR Models and Support Vector Machines". In: *IEEE Transactions on Biomedical Engineering* 57.5 (May 2010), pp. 1124–1132.

[11] L. Barnett and A. K. Seth. "Granger Causality for State-Space Models". In: (2015).

[12] E. D. Übeyli. "Least Squares Support Vector Machine Employing Model-Based Methods Coefficients for Analysis of EEG Signals". In: *Expert Systems with Applications* 37.1 (Jan. 2010), pp. 233–239.

[13] J. Songsiri. "Learning Brain Connectivity from EEG Time Series". PhD thesis. 2019.

[14] P. Manomaisaowapak, A. Nartkulpat, and J. Songsiri. "Granger Causality Inference in EEG Source Connectivity Analysis: A State-Space Approach". In: *IEEE Transactions on Neural Networks and Learning Systems* 33.7 (July 2022), pp. 3146–3156.

[15] C. O. Becker, D. S. Bassett, and V. M. Preciado. "Large-Scale Dynamic Modeling of Task-fMRI Signals via Subspace System Identification". In: *Journal of neural engineering* 15.6 (Sept. 2018).

[16] M. Verhaegen and V. Verdult. *Filtering and System Identification: A Least Squares Approach*. Cambridge University Press, Jan. 2007.

[17] M. Verhaegen and P. Dewilde. “Subspace model identification Part 1. The output-error state-space model identification class of algorithms”. In: *International Journal of Control* 56.5 (1992), pp. 1187–1210.

[18] P. Van Overschee and B. De Moor. “Subspace Identification for Linear Systems”. In: *Subspace Identification for Linear Systems* (1996).

[19] C. W. J. Granger. “Investigating Causal Relations by Econometric Models and Cross-spectral Methods”. In: *Econometrica* 37.3 (Aug. 1969), p. 424.

[20] M. P. Vlaar et al. “Quantification of Task-Dependent Cortical Activation Evoked by Robotic Continuous Wrist Joint Manipulation in Chronic Hemiparetic Stroke”. In: *Journal of NeuroEngineering and Rehabilitation* 14.1 (Apr. 2017), pp. 1–15.

[21] A. Biasiucci, B. Franceschiello, and M. M. Murray. “Electroencephalography”. In: *Current Biology* 29.3 (Feb. 2019), R80–R85.

[22] M. F. Bear. *Neuroscience Exploring The Brain, Enhanced Edition*. 4th ed. Wolters Kluwer, 2016.

[23] A. M. Grech, J. P. Nakamura, and R. A. Hill. “The Importance of Distinguishing Allocentric and Egocentric Search Strategies in Rodent Hippocampal-Dependent Spatial Memory Paradigms: Getting More Out of Your Data”. In: *The Hippocampus - Plasticity and Functions*. 2018.

[24] S. J. Qin. “An Overview of Subspace Identification”. In: *Computers and Chemical Engineering* 30 (2006), pp. 1502–1513.

[25] P. Van Overschee and B. De Moor. “N4SID: Subspace Algorithms for the Identification of Combined Deterministic-Stochastic Systems”. In: *Automatica* 30.1 (Jan. 1994), pp. 75–93.

[26] W. E. Larimore. “Canonical Variate Analysis in Identification, Filtering, and Adaptive Control”. In: *Proceedings of the IEEE Conference on Decision and Control* 2 (1990), pp. 596–604.

[27] M. Verhaegen and P. Dewilde. “Subspace Model Identification Part 2. Analysis of the Elementary Output-Error State-Space Model Identification Algorithm”. In: *International journal of control* 56.5 (1992), pp. 1211–1241.

[28] M. Verhaegen. “Application of a Subspace Model Identification Technique to Identify LTI Systems Operating in Closed-Loop”. In: *Automatica* 29.4 (1993), pp. 1027–1040.

[29] L. A. Baccalá and K. Sameshima. “Partial Directed Coherence: A New Concept in Neural Structure Determination”. In: *Biological Cybernetics* 84.6 (Jan. 2001), pp. 463–474.

[30] L. Timmermann et al. “The Cerebral Oscillatory Network of Parkinsonian Resting Tremor”. In: *Brain* 126.1 (2003).

[31] J. Liu, Y. Sheng, and H. Liu. “Corticomuscular Coherence and its Applications: A Review”. In: *Frontiers in Human Neuroscience* 13 (Feb. 2019), p. 442607.

[32] M. P. Vlaar et al. “Modeling the Nonlinear Cortical Response in EEG Evoked by Wrist Joint Manipulation”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26.1 (Jan. 2018), pp. 205–215.

[33] Y. Yang, T. Solis-Escalante, M. van de Ruit, F. C. van der Helm, and A. C. Schouten. “Nonlinear Coupling between Cortical Oscillations and Muscle Activity during Isotonic Wrist Flexion”. In: *Frontiers in Computational Neuroscience* 10 (Dec. 2016), p. 126.

[34] J. van Kordelaar et al. “The Cortical Response Evoked by Robotic Wrist Perturbations Reflects Level of Proprioceptive Impairment After Stroke”. In: *Frontiers in Human Neuroscience* 15 (Nov. 2021), p. 695366.

[35] M. C. Chang et al. “The Comparison of Cortical Activation Patterns by Active Exercise, Proprioceptive Input, and Touch Stimulation in the Human Brain: A Functional MRI Study”. In: *NeuroRehabilitation* 25.2 (2009).

[36] V. A. Vakorin et al. “Complexity Analysis of Source Activity Underlying the Neuromagnetic Somatosensory Steady-State Response”. In: *NeuroImage* 51.1 (2010).

[37] N. Forss et al. “Activation of the Human Posterior Parietal Cortex by Median Nerve Stimulation”. In: *Experimental Brain Research* 99.2 (May 1994), pp. 309–315.

[38] N. Gogolla. “The Insular Cortex”. In: *Current Biology* 27 (2017), R580–R586.

[39] H. Onishi et al. “Neuromagnetic Activation of Primary and Secondary Somatosensory Cortex following Tactile-on and Tactile-off Stimulation”. In: *Elsevier* 121.4 (2010), pp. 588–593.

[40] M. Rubinov and O. Sporns. “Complex Network Measures of Brain Bonnectivity: Uses and Interpretations”. In: *NeuroImage* 52.3 (Sept. 2010), pp. 1059–1069.

[41] T. Keviczky. *Lecture Slides 2 - SC42015 Control Theory (2022/23 Q1)*. 2022.

[42] T. Kailath. *Linear Systems*. Vol. 156. Englewood Cliffs: Prentice Hall, 1980.

[43] W. H. Durfee and A. S. Householder. “Principles of Numerical Analysis.” In: *The American Mathematical Monthly* 61.9 (1954).

[44] G. Strang. *Linear Algebra and its Applications*. 3rd. San Diego: Harcourt Brace Jovanovich, 1988.

[45] G. W. Stewart. “On the numerical analysis of oblique projectors”. In: *SIAM Journal on Matrix Analysis and Applications* 32.1 (2011).

[46] S. Hachicha, M. Kharrat, and A. Chaari. “N4SID and MOESP Algorithms to Highlight the Ill-Conditioning into Subspace Identification”. In: *International Journal of Automation and Computing* 11.1 (Feb. 2014), pp. 30–38.

[47] T. Keviczky. *Lecture Slides 4 - SC42015 Control Theory (2022/23 Q1)*. 2022.

[48] P. Torres, J. W. Van Wingerden, and M. Verhaegen. “PO-MOESP Subspace Identification of Directed Acyclic Graphs with Unknown Topology”. In: *Automatica* 53 (Mar. 2015), pp. 60–71.

[49] A. Chiuso and G. Picci. “On the ill-conditioning of subspace identification with inputs”. In: *Automatica* 40.4 (2004).

[50] V. Verdult. “Nonlinear System Identification: A State-Space Approach”. PhD thesis. Twente: University of Twente, 2002.

[51] K. De Cock, B. De Moor, and K. U. Leuven. “Subspace Identification Methods”. In: *Contribution to section* 5 (2003), pp. 933–979.

[52] J. Dauwels, F. Vialatte, T. Musha, and A. Cichocki. “A Comparative Study of Synchrony Measures for the Early Diagnosis of Alzheimer’s Disease based on EEG”. In: *NeuroImage* 49.1 (Jan. 2010), pp. 668–693.

[53] R. J. van Esch et al. “A Bayesian Method for Inference of Effective Connectivity in Brain Networks for Fetecting the Mozart Effect”. In: *Computers in Biology and Medicine* 127 (Dec. 2020), p. 104055.

[54] C. W. J. Granger. “Testing for Causality. A Personal Viewpoint”. In: *Journal of Economic Dynamics and Control* 2.C (1980), pp. 329–352.

[55] A. G. Dankers. “System Identification in Dynamic Networks”. PhD thesis. 2014.

[56] D. A. Dickey and W. A. Fuller. “Distribution of the Estimators for Autoregressive Time Series with a Unit Root”. In: *Journal of the American Statistical Association* 74.366a (June 1979), pp. 427–431.

[57] J. Lever, M. Krzywinski, and N. Altman. “Model selection and overfitting”. In: *Nature Methods* 13.9 (2016).

[58] J. Geweke. “Measurement of Linear Dependence and Feedback between Multiple Time Series”. In: *Journal of the American Statistical Association* 77.378 (1982), pp. 304–313.

[59] P. Lancaster and L. Rodman. “Algebraic Riccati Equations”. In: (1995), p. 480.

[60] M. Ruiz-Olazar, E. S. Rocha, C. D. Vargas, and K. R. Braghetto. “The Neuroscience Experiments System (NES)–A Software Tool to Manage Experimental Data and Its Provenance”. In: *Frontiers in Neuroinformatics* 15 (2022).

[61] P. Prado-Gutierrez, E. Martínez-Montes, A. Weinstein, and M. Zañartu. “Estimation of auditory steady-state responses based on the averaging of independent EEG epochs”. In: *PLoS ONE* 14.1 (2019).

[62] A. Delorme and S. Makeig. “EEGLAB: An Open Source Toolbox for Analysis of Single-Trial EEG Dynamics including Independent Component Analysis”. In: *Journal of Neuroscience Methods* 134.1 (Mar. 2004), pp. 9–21.

[63] N. Plub-In and J. Songsiri. “State-Space Model Estimation of EEG Time Series for Classifying Active Brain Sources”. In: *11th Biomedical Engineering International Conference (BMEiCON* (Jan. 2019), pp. 1–5.

[64] S. Chunnawong and J. Songsiri. “System Identification: An EEG subspace identification by using subspace method”. In: (2017).

[65] M. Verhaegen, V. Verdult, and N. Bergboer. “Filtering and System Identification: An Introduction to using Matlab Software”. In: *Delft University of Technology* 68.163 (2007).

[66] G. Nolte and G. Dassios. “Analytic Expansion of the EEG Lead Field for Realistic Volume Conductors”. In: *Physics in Medicine and Biology* 50.16 (2005).

[67] R. Oostenveld, P. Fries, E. Maris, and J. M. Schoffelen. “FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data”. In: *Computational Intelligence and Neuroscience* 2011 (2011).

[68] M. Lim, J. M. Ales, B. R. Cottereau, T. Hastie, and A. M. Norcia. “Sparse EEG/MEG Source Estimation via a Group Lasso”. In: *PLoS ONE* 12.6 (2017).

[69] H. Hallez et al. “Review on Solving the Forward Problem in EEG Source Analysis”. In: *Journal of NeuroEngineering and Rehabilitation* 4.1 (Nov. 2007), pp. 1–29.

[70] F. Raimondo, J. E. Kamienkowski, M. Sigman, and D. Fernandez Slezak. “CUDAICA: GPU Optimization of Infomax - ICA EEG Analysis”. In: *Computational Intelligence and Neuroscience* 2012 (2012).

[71] S. L. Lacy and D. S. Bernstein. “Subspace identification with guaranteed stability using constrained optimization”. In: *IEEE Transactions on Automatic Control* 48.7 (2003).

[72] N. Tzourio-Mazoyer et al. “Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain”. In: *NeuroImage* 15.1 (2002).

Glossary

List of Acronyms

AAL2	automated anatomical labelling 2
BCI	brain-computer interfaces
BIC	Bayesian information criterion
DARE	discrete-time algebraic Riccatti equation
EEG	electroencephalogram
fMRI	functional magnetic resonance imaging
ICA	independent component analysis
ICs	independent components
LTI	linear time-invariant
MOESP	Multivariable Output Error State sPace
MRI	magnetic resonance imaging
N4SID	Numerical algorithms for Subspace State Space System Identification
PI-MOESP	Past Input Multivariable Output Error State sPace
PO-MOESP	Past Output Multivariable Output Error State sPace
RMSE	root-mean-square error
ROI	region of interest
SSR	steady-state responses
std	standard deviation
SVD	singular value decomposition
VAF	variance accounted for
VAR	vector autoregressive
VARMA	vector autoregressive moving-average

List of Symbols

Π_B	The orthogonal projection of the row space of matrix B.
$\widehat{x}(k+1)$	Estimate of the state variable. In every instance, the symbol $\widehat{\ }$ signifies an approximation of a vector or matrix.
$A/_B C$	The oblique projection of matrix A along the row space of matrix B onto the row space of matrix C.
$A^\dagger$	The Moore-Penrose pseudoinverse of matrix A
$B^\perp$	The orthogonal complement of a subspace B

C_n	The controllability matrix, where the first subscript denotes the number of block columns
F_{ij}	The Granger causality measure, indicating whether state x_i is Granger caused by state x_j.
O_n	The observability matrix, where the first subscript denotes the number of block rows
S_s	The lower block triangular Toeplitz matrix for the triple (A, C, K).
T_s	The lower block triangular Toeplitz matrix for the quadruple (A, B, C, D).
$X_{i,N}$	Extended vector, where the first subscript denotes the time index of its top left entry and the second refers to the number of block rows
$Y_{i,s,N}$	Hankel matrix, where the first subscript denotes the time index of its top left entry, the second refers to the number of block rows, and the third refers to the number of columns