Causal Feature Selection in Neuroscience

Causal Feature Selection in Neuroscience

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Dipl.-Eng. Anastasia Atalanti Mastakouri
aus Athen, Griechenland

Tübingen
2020

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultt der Eberhard Karls Universitt Tbingen.

Tag der mündlichen Qualifikation:	16.10.2020
Stellvertretender Dekan:	Prof. Dr. József Fortágh
1. Berichterstatter:	PD. Dr. Dominik Janzing
2. Berichterstatter:	Prof. Dr. Bernhard Schölkopf
3. Berichterstatter:	Prof. Dr. Felix Wichmann

To my parents, my sister and to Alexandros

Abstract

Causal inference, at times correct and at times false, is fundamentally intertwined with the human nature. Humans tend to approach and explain the systems in the world and every day life via causal reasoning and causal statements, by unconsciously trying to recover the causal graph that underlies their observations. Nevertheless, causal reasoning based on observations of the real world is seldom equitable and precise. Particularly when the method that one uses is based on plain correlations, causal statements can be far from causal, first, because of the implicit assumption about linear relationships, and second, due to the major problem of hidden confounding.

One of the most complex and difficult systems for an applied scientist to explain is the human brain. The reason for that is threefold. First and foremost, because of the daedal and sophisticated manner that the human brain is constructed. Secondly, because of our limited means of observing its global functionality, which ultimately leads to the problem that no *causal sufficiency* can be assumed in such a system. In other words, hidden common causes (also termed hidden confounders) in our limited observations will be omnipresent. Finally, the significant heterogeneity that the human brain exhibits in some of its physiological functionalities, across subjects, hinders the problem even further. This, subsequently, justifies the lack of generalization of machine learning methods that try to predict biomarkers through the traditional approach of a non-causal model, across different brains. Hence, someone should be particularly careful with the methods that she or he selects to use and the causal statements that are made, to understand and interpret the brain functionality.

In this thesis, we focus on constructing theorems and algorithms for causal inference on real data, trying to understand the relationship between the human brain and motor function. More specifically, we target the problem of the identification of causes of a target variable, without assuming *causal sufficiency*. We tackle both the cases of non-sequential and of time series data, proving theorems for both cases accordingly. Our methods' applications have an immediate focus on the activity of the human motor cortex at the time it arises, first, naturally, and second, from non-invasive brain stimulation. We build experimental set-ups and conduct electroencephalographic (EEG) and stimulation experiments to study the functionality of the motor cortex across different subjects, during these two different cases, with an ultimate goal to explain the observed heterogeneity in the recorded activity.

The work presented in this thesis is both experimental –in its first part– with non invasive experiments on the human brain, contributing to the better understanding of the motor cortex, and theoretical, with contributions of four theorems in the field of causal inference, and two causal feature selection methods.

We first attempt to approach the brain activity from a purely machine learning perspective, analysing the data of the brain activity of 27 healthy subjects during an upper-limb reaching task. We introduce a multi-task regression method to build personalised models that predict movement stability from limited trials. We do so by taking into account information from other subjects as prior and updating -when necessary- the weights of the model with trials from the current subject. Although the original goal of this work was to show the superiority of this prediction method, a side-observation turned out to be the most fundamental key to define the next steps of the hereby presented research. The learnt features by the individual prediction models differed significantly across subjects, and although no causal claim can be made yet -since this is a correlation-based observation- it is the first hint of existing heterogeneity in the activity of the human motor cortex. Such a discrepancy, in frequency and location in the learnt features, could also imply a discrepancy in the response to non-invasive brain stimulation techniques, over the motor cortex.

To examine this possibility, a new series of electrophysiological experiments, with application of transcranial alternating current stimulation at 70 Hz over the motor cortex –as this has been considered to facilitate movement– , is conducted on twenty healthy participants. At this point, having observed a significant variability in the behavioural response, ranging from negative to positive responders, we decided to further investigate the reasons that could explain it. An incremental method with three steps is introduced to narrow down the causal model that can explain the aforementioned discrepancy in responses. With our method, we conclude that the beta oscillatory activity over the motor cortex could play a mediating role between the gamma stimulation and the motor performance, without being able to exclude the case that GABA activity could be a hidden common cause.

Having witnessed such a heterogeneity, both during natural movements and under brain stimulation, we stress the importance of taking steps towards personalisation of brain stimulation parameters. We conclude the experimental part of this work by constructing a pipeline, to predict from *resting state* EEG data the behavioural response of each subject to the stimulation treatment. Such a screening could avoid redundant or even harmful stimulation sessions. With two different stimulation studies, recruiting in total 42 healthy participants, we identify a biomarker that could be informative about the response of an individual to the aforementioned motor stimulation.

In the theoretical part of this thesis, we focus on the problem of the identification of direct and indirect causes of a target (e.g. motor performance) given a collection of possible candidates (e.g. brain activity in different locations, in different frequencies), allowing at the same time for latent common causes. First, we propose and prove a theorem which introduces sufficient conditions, under assumptions that can naturally be met, to decide for the causal role of a feature, with a single *conditional independence* test, and a single conditioning variable. Given the hardness of statistical testing of conditional independences in large and dense graphs (such as the brain), limiting the necessary tests to one, significantly boosts the statistical strength of the results. Application of our conditions on the aforementioned neurophysiological data supports further the validity of the method. Applying the proposed conditions independently on each individual, without prior knowledge, led to three groups of identified causal features, each one being related in a consistent manner with different quality of movements across subjects. We discuss how such a method could contribute in the selection of personalised brain stimulation parameters.

As a final step, we approach the brain signal as continuous time series data. Although time series are observed almost everywhere in nature, yet, causal inference on such data, in the presence of hidden confounders, has been an unsolved problem, with the widely known Granger Causality being the only approach for almost half a century. The final contribution of this thesis, are two theorems with which we introduce both necessary and sufficient conditions for the causal feature selection on time series, under some graph constraints, and a third theorem that relaxes one of the stricter assumptions of the aforementioned two. We demonstrate the validity of our method both on simulated and real data.

Kurzfassung

Kausales Schlussfolgern, manchmal korrekt und manchmal flschlich, ist grundlegend mit der menschlichen Natur verflochten. Menschen neigen dazu, sich den Systemen in der Welt und im tglichen Leben durch kausale Argumentation und kausale Aussagen zu nhern und diese zu erklren, indem sie unbewusst versuchen, den Kausalgraphen zu finden, der ihren Beobachtungen zugrunde liegt. Dennoch ist kausale Argumentation, die auf Beobachtungen der realen Welt beruht, selten angemessen und akkurat. Insbesondere wenn die verwendete Methode schlicht auf Korrelationen beruht, knnen die vermeintlich kausalen Aussagen alles andere als kausal sein, erstens wegen der implizit angenommenen linearen Beziehungen, und zweitens wegen des groen Problems unbercksichtigter Strfaktoren.

Eines der komplexesten und aus der Sicht eines angewandten Wissenschaftlers am schwierigsten zu erklrenden Systeme ist das menschliche Gehirn. Dafr gibt es drei Grnde. In erster Linie die geschickte und hochentwickelte Weise, in der das menschliche Gehirn aufgebaut ist. Zweitens, unsere eingeschrnkten Mglichkeiten, die globale Aktivitt und Funktionalitt zu beobachten, was letztlich zu dem Problem fhrt, dass fr ein solches System keine *causal sufficiency* angenommen werden kann. Mit anderen Worten, in unseren eingeschrnkten Beobachtungen werden unbeobachtete Konfundierungseffekte (oder unbercksichtigte Strfaktoren) allgegenwrtig sein. Zuletzt erschwert die erhebliche Heterogenitt, die das menschliche Gehirn in einigen der physiologischen Funktionen ber Probanden hinweg aufweist, das Problem noch weiter. Dies begrndet auch die fehlende Generalisierbarkeit von Methoden des maschinellen Lernens, die versuchen, Biomarker durch den traditionellen Ansatz eines nicht-kausalen Modells ber verschiedene Gehirne hinweg vorherzusagen. Daher sollte man besonders vorsichtig sein welche Methoden man verwendet und welche kausalen Aussagen man macht, um die Funktion des Gehirns zu verstehen und zu interpretieren.

In dieser Arbeit konzentrieren wir uns auf die Erarbeitung von Theoremen und Algorithmen zur kausalen Inferenz auf realen Daten und versuchen, die Beziehung zwischen dem menschlichen Gehirn und der Motorik zu verstehen. Genauer gesagt zielen wir auf das Problem ab, Ursachen einer Zielvariablen zu identifizieren ohne dabei von *causal sufficiency* auszugehen. Wir gehen sowohl die Flle von I.I.D. als auch von Zeitreihendaten an und beweisen entsprechend Theoreme fr beide Flle. Unser unmittelbarer Fokus fr die Anwendung unserer Methoden liegt auf der Aktivitt des menschlichen motorischen Kortex, wie sie erstens natrlich und zweitens durch nicht-invasive Hirnstimulation entsteht. Wir bauen Versuchsanordnungen auf und fhren elektroenzephalographische (EEG) und Stimulationsexperimente durch, um die Funktionalitt des motorischen Kortex bei verschiedenen Probanden whrend dieser beiden unterschiedlichen Flle zu untersuchen, mit dem endgltigen Ziel, die beobachtete Heterogenitt in der aufgezeichneten Aktivitt zu erklren.

Die in dieser Arbeit vorgestellte Arbeit ist sowohl experimentell - im ersten Teil - mit nicht-invasiven Experimenten am menschlichen Gehirn, die zum besseren Verstndnis des motorischen Kortex beitragen, als auch theoretisch, mit Beitrgen von vier Theoremen im Bereich der kausalen Inferenz und zwei Methoden zur Auswahl kausaler Variablen.

Zunchst versuchen wir, uns der Hirnaktivitt aus einer rein maschinellen Lernperspektive zu nhern, indem wir die Daten der Hirnaktivitt von 27 gesunden Probanden whrend einer Greif-Aufgabe analysieren. Wir fhren eine Multi-Task-Regressionsmethode ein, um personalisierte Modelle zu erstellen, die die Bewegungsstabilitt anhand weniger Versuche vorhersagen. Wir tun dies, indem wir Informationen von anderen Versuchspersonen in einer Priorverteilung bercksichtigen und, wenn ntig, die Gewichte des Modells anhand Versuche der aktuellen Versuchsperson aktualisieren. Obwohl das ursprngliche Ziel dieser Arbeit darin bestand, die berlegenheit dieser Vorhersagemethode zu zeigen, erwies sich eine Nebenbeobachtung als der grundlegende Schlssel zur Definition der nchsten Schritte der hier vorgestellten Forschung. Die von den einzelnen Vorhersagemodellen erlernten Merkmale unterschieden sich signifikant ber Probanden hinweg, und obwohl noch kein kausaler Anspruch erhoben werden kann, da es sich um eine korrelationsbasierte Beobachtung handelt, ist dies der erste Hinweis auf eine bestehende Heterogenitt in der Aktivitt des menschlichen motorischen Kortex. Eine solche Diskrepanz in Hufigkeit und Lokalisation in den erlernten Merkmalen knnte auch eine Diskrepanz in der Reaktion auf nicht-invasive Hirnstimulationstechniken ber den motorischen Kortex implizieren.

Um diese Mglichkeit zu untersuchen, wird eine neue Serie elektrophysiologischer Experimente mit der Anwendung transkranieller Wechselstromstimulation bei 70 Hz ber dem motorischen Kortex - da vermutet wurde, dass dies Bewegung untersttzt - an zwanzig gesunden Teilnehmern durchgefhrt. Da wir eine signifikante Variabilitt der Reaktion auf Stimulation beobachteten, von negativer bis zu positiver Reaktion, untersuchten wir weiter mgliche Erklrungen hierfr. Eine inkrementelle Methode mit drei Schritten wird eingefhrt, um das kausale Modell, das die beobachtete Variabilitt der Reaktion erklren kann, weiter einzugrenzen. Mit der vorgeschlagenen Methode gelingt es uns, eine Hirnfrequenz zu identifizieren, die die Reaktion auf die Stimulation vermitteln knnte. Mit unserer Methode kommen wir zu dem Schluss, dass die Beta-Oszillationsaktivitt ber dem motorischen Kortex eine modulierende Rolle zwischen der Gammastimulation und der motorischen Leistung spielen knnte, ohne den Fall ausschliessen zu knnen, dass die GABA-Aktivitt eine versteckte gemeinsame Ursache sein knnte.

Nachdem wir eine solche Heterogenitt sowohl whrend natrlicher Bewegungen als auch unter Hirnstimulation beobachtet haben, betonen wir, dass es wichtig ist, die Hirnstimulationsparameter zu personalisieren. Wir schliessen den experimentellen Teil dieser Arbeit ab, indem wir erarbeiten, wie von EEG Daten im *Ruhezustand* die Reaktion eines Probanden auf Stimulation vorhergesagt werden kann. Durch ein solches Screening knnten berflssige oder sogar schdliche Stimulationssitzungen vermieden werden. Anhand zweier Stimulationsstudien, mit insgesamt 42 gesunden Teilnehmern, identifizieren wir einen Biomarker, der ber die Reaktion eines Individuums auf die oben erwhnte motorische Stimulation Aufschluss geben knnte.

Im theoretischen Teil dieser Arbeit konzentrieren wir uns auf das Problem der Identifizierung von direkten und indirekten Ursachen einer Zielgre (z.B. der motorischen Leistung) aus einer Sammlung von mglichen Kandidaten (z.B. Hirnaktivitt an verschiedenen Orten, in verschiedenen Frequenzen), in Gegenwart von versteckten Strfaktoren. Zunchst schlagen wir ein Theorem vor und beweisen dieses, das unter Annahmen, die normalerweise erfllt werden knnen hinreichende Bedingungen vorstellt, um mit einem einzigen *Bedingte-Unabhngigkeit*-Test und einer einzigen bedingenden Variable ber die kausale Rolle eines Merkmals zu entscheiden. Angesichts der Schwierigkeit bedingte Unabhngigkeit in groen und dichten Graphen (wie dem Gehirn) statistisch zu testen, ist dieser Beitrag von erheblicher statistischer Bedeutung, da er die Anzahl der notwendigen Tests auf eins reduziert. Die Anwendung unserer Bedingungen auf die oben genannten neurophysiologischen Daten untersttzt die Gltigkeit der Methode weiter. Die Anwendung der vorgeschlagenen Bedingungen auf jedem Individuum unabhngig, und ohne weitere Vorkenntnisse, resultierte in drei Gruppen von identifizierten kausalen Merkmalen, die in konsistenter Weise mit unterschiedlichen Bewegungs-Qualitten ber Probanden hinweg einhergehen. Wir diskutieren, wie eine solche Methode zur Auswahl von personalisierten Hirnstimulationsparametern beitragen knnte.

Im letzten Schritt betrachten wir das Gehirnsignal als kontinuierliche Zeitreihendaten. Obwohl Zeitreihen fast berall in der Natur beobachtet werden, ist die kausale Inferenz auf solche Daten in Gegenwart von versteckten Strfaktoren ein ungelstes Problem, wobei die sogenannte Granger-Kausalitt seit fast einem halben Jahrzehnt der einzige Ansatz ist. Der letzte Beitrag dieser Arbeit sind zwei Theoreme, mit denen wir sowohl notwendige als auch hinreichende Bedingungen fr die kausale Merkmalsauswahl bei Zeitreihen unter einigen graphischen Einschrnkungen einfhren, und ein drittes Theorem, das eine der strikteren Annahmen der vorgenannten zwei lockert. Wir demonstrieren die Gltigkeit unserer Methode sowohl auf simulierten als auch auf realen Daten.

Acknowledgments

I deeply thank Bernhard Schölkopf for giving me the opportunity to do my PhD in his lab, and to be a member of the Max Planck Institute family. Thank you, Bernhard, for the support, the trust you showed me and for the excellent research environment which gave me the possibility to meet and work alongside accomplished researchers. It was an honour and an exciting experience working with you and having you as a mentor.

I deeply thank Dominik Janzing for introducing me in the world of Causality, for his mentorship, for believing in me and supporting me when things turned tough. Thank you, Dominik, for teaching me your high standards in research and for being an excellent example of a researcher.

I deeply thank Felix Wichmann and Zeynep Akata for their willingness, and the opportunity they gave me to defend my thesis in the University of Tübingen.

I thank Sabrina Rehbaum, Sabrina Jung, Camelia Fritz and Lidia Pavel for their constant support in handling daily administrative challenges.

Thank you, Karin Bierig and Bernd Battes, for your support in all the experimental settings and for always being so kind and willing to help. Thank you Sebastian Stark for being the best IT ever. Thank you Periklis Zisis, Bjarni Kjartansson, Telintor Ntounis. Without your technical support all this would not be possible.

Thank you, Eleni Sgouritsa, for being the first person to meet on my interviews' day, and for encouraging me in a rather stressful day that turned out to be the beginning of this journey. Thank you, Dimitris Tzionas, for brainstorming titles for my paper until late alongside Vasilis Choutas.

Thank you very much, Sebastian Weichwald, for being a true friend and a brilliant colleague. I learnt a lot from you.

I would also like to thank all the former and current colleagues of the Empirical Inference group for the interesting discussions and the fun we had: thank you Michel Besserve, Giambattista Parascandolo, Alexander Neitz, Amir Karimi, Niki Kilbertus, Paul Rubenstein, Diego Fioravanti, Julius von Kügelgen, Luigi Gresele, Dieter Büchler, Diego Agudelo, Sebastian Gomez, Kristof Meding, Simon Buchholz, Vincent Stimper, Matthias Bauer, Timm Meyer, Vinay Jayaram, Matthias Hohmann and Jonas Kübler, and anyone I may have by accident missed.

Finally, I am deeply grateful to my parents, my sister and my life partner Alexandros. Thank you for teaching me your high values and for believing in me and supporting me throughout this whole journey. Alexandros thank you for knowing me better than me and for always challenging me to become better. Without all four of you I would have never made it so far.

Symbols

X	Random variable X (usually used to denote observed variables)
U	Random variable U (usually used to denote un-observed variables)
Q	Random variable Q (usually used to denote any kind of variables)
x	Value of random variable X
P_X	Probability distribution of X
p	Probability mass function or probability density function
$p(x)$	Density of P_X calculated at value x
$p(y \mid x)$	Conditional density of $P_{Y \mid X=x}$ calculated at value y
$corr[X,Y]$	correlation of X,Y
$cov[X,Y]$	covariance of X,Y
Σ	Covariance matrix
$\mathbf{X} = (X^1, X^2, \cdots, X^d)$	Random vector of length d
$X \perp\!\!\!\perp Y$	Random variables X and Y are independent
$X \perp\!\!\!\perp Y \mid Z$	Random variables X and Y are independent after conditioning on random variable Z
$\mathcal{C}$	Structural causal model
$\mathcal{G}$	graph

Abbreviations

BCI	Brain computer interface
CDF	Cumulative distribution function
CFS	Causal feature selection
DAG	Directed acyclic graph
EEG	Electroencephalography
FCM	Function causal model
FDR	False discovery rate
FNR	False negative rate
FPR	False positive rate
HD-tACS	High definition transcranial alternating current stimulation
IGCI	Information Geometric Causal Inference
NARJ	Normalized average rectified jerk
NIBS	Non invasive brain stimulation
RMSE	Root mean square error
SCM	Structural causal model
SEM	Structural equation model
tACS	Transcranial alternating current stimulation
tDCS	Transcranial direct current stimulation
TES	Transcranial electric stimulation
TMS	Transcranial magnetic stimulation
TNR	True negative rate
TPR	True positive rate

List of Figures

List of Tables

List of Algorithms

Chapters dependency diagram

Here is a recommendation of how to read this dissertation, depending on your field of interest.

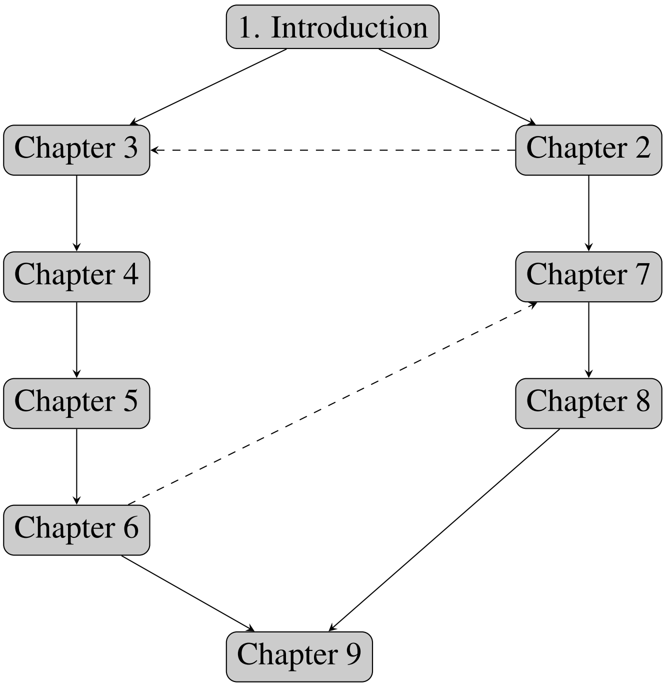


Figure 1: Chapters dependency diagram. This chart shows the three possible ways that someone can study this thesis, depending on the field of her/his interests. First way, by reading all the chapters in the listed order from 1 to 9 , in order to have a complete picture of all the work and contributions made in this dissertation. Second way, by reading Chapters 1, 3, 4, 5, 6 and 9 in that order, for someone who is interested only in the neuroscientific findings of this thesis. Third way, by reading Chapters 1, 2, 7 and 8 in that order, for someone who is interested only in the causal inference methods introduced in this thesis. Direct links (→) indicate hard dependency to the previous chapter. Dashed links (⇢) indicate that it is not necessary for the understanding, nevertheless it is recommended to study the previous chapter.

Contents

Chapter 1

Introduction

1.1 Why?

The motivation for this thesis comes from the observation that there is a lack of personalised treatment methods and, more specifically, lack of personalised brain stimulation treatment for motor rehabilitation. Although the human motor cortex has been a subject of intense research for centuries, starting with (Campbell, 1905), modern research still lacks personalisation of the parameters that are used for facilitation of movement in brain stimulation sessions. Taking a step back, the need for such personalisation arises foremost from the encountered heterogeneity in the behavioural response to non-invasive brain stimulation treatments accross subjects (López-Alonso *et al.*, 2014; Strube *et al.*, 2015; Yang *et al.*, 2020), ranging from positive, to negative and to no response at all, and the lack of knowledge of the exact cause of this discrepancy (Ridding and Ziemann, 2010). Although many different explanations for inter-subject variability have been given (Stecher *et al.*, 2017; Vosskuhl *et al.*, 2018), including individual differences in brain anatomy (Buch *et al.*, 2017; Datta, 2012; Parazzini *et al.*, 2015), brain-state dependent susceptibility to NIBS (Silvanto *et al.*, 2008; López-Alonso *et al.*, 2014; Strube *et al.*, 2015; Wiethoff *et al.*, 2014), prior activity (Rosenkranz *et al.*, 2007; Iezzi *et al.*, 2008), age (Moliadze *et al.*, 2015; Fujiyama *et al.*, 2014), attention (Kiers *et al.*, 1993), sex (Pitcher *et al.*, 2003), pharmacological effects(Ziemann *et al.*, 2008; Grundey *et al.*, 2012; Nitsche *et al.*, 2004), genetic variations (Voti *et al.*, 2011; Mori *et al.*, 2011), and time of day (López-Alonso *et al.*, 2014), still there is no known cause that can explain and predict response to NIBS.

Approaching this problem from a data scientist or mathematician's perspective, it seems that the problem could be pointed down to the lack of knowledge of causal brain factors of the human upper-limb movement. Identification of such causal features could also indicate targets for intervention, hence contributing to the personalisation of the stimulation parameters.

Not until very recently, was the importance of causal inference accepted in the field of neuroscience. In sensitive systems, like the human brain, the possibility of randomized interventions, as part of the process of the discovery of the causal graph, is limited for ethical and safety reasons. In such systems causal inference based on observations becomes very important. Of course, as the problem of causal inference itself is very hard, necessary assumptions need to be made.

Deriving motivation from the aforementioned problem, and as a step towards personalised brain stimulation, the work presented in this thesis focuses on the development of screening and causal methods for the better understanding of the human motor cortex and for the better explanation of the observed heterogeneity of responses. As the field of Causality itself is rather young, gradually, we realized, through this process, the lack of causal methods that could be applied on real, large-scale, complex data, as the brain signals. For that reason, a large part of this thesis is dedicated in the construction and proof of theorems that introduce new methods for causal feature selection on real data.

1.2 Problem Statement

The previous section gave the motivation behind the need for causal models that could lead to causal statements about the brain features and the upper-limb movement. In this section we try to break the overall question of the identification of causal brain features and of the personalisation of brain stimulation into sub-problems, which tackle more technical and precise questions.

PROBLEM 1

Description: Individualized motor-performance EEG-based prediction models.

Input: Observed brain signals from distinct electrode locations, and arm stability.

Question: Is it possible to build individualized models that predict arm stability from brain signals, with a limited number of recording trials? Are the features used by the predictor causal? Can they be used as targets of brain stimulation? Are they consistent across different subjects?

PROBLEM 2

Description: 70 Hz transcranial alternating current brain stimulation over contralateral motor cortex to facilitate arm speed.

Input: Observed brain signals from distinct electrode locations before and after the stimulation blocks, and arm speed.

Question: Is the arm speed facilitated significantly to healthy subjects with 70 Hz contralateral tACS compared to sham? Is the behavioural response consistent across subjects? If not, what can explain the heterogeneity?

PROBLEM 3

Description: Stratification of behavioural response to transcranial alternating current stimulation by resting-state electroencephalography.

Input: Observed brain signals from distinct electrode locations before and after the stimulation blocks, and arm speed.

Question: Is there any biomarker that can screen off in advance negative-responders and non-responders to the treatment, in order to avoid redundant and prevent harmful stimulation sessions?

PROBLEM 4

Description: Causal brain feature selection with latent variables

Input: Observed brain signals from distinct electrode locations and arm speed.

Question: Given the target variable (motor performance) and the brain activity on i.i.d. trials, is it possible to identify the causal features? If yes, then under which assumptions?

PROBLEM 5

Description:	Causal feature selection on time-series with latent variables.
Input:	Time-series data, with a sink node [1] target variable.
Question:	Given observed time series and a target time series, is it possible to identify its causes? Under which assumptions is this possible? Is it possible to propose both sufficient and necessary conditions for this problem?

1.3 Thesis outline

The motivation behind the work in this thesis is presented Section 1.1 of the current chapter. In Section 1.2, the overall question that this thesis aims to answer is broken into five sub-problems, each one of which is tackled in each of the six manuscripts that this dissertation is based on.

Chapter 2 addresses the basic concepts of causal inference that are necessary to follow the theoretical contributions provided by this dissertation. More specifically, Section 2.1 introduces basic graph notations and Section 2.3 focuses on the problem of *causal discovery* from observational data [2]. Different groups of methods for causal discovery and their limitations are presented and discussed in Sections 2.3.1, 2.3.2 and 2.3.4. In Section 2.4 we introduce the causal problem we try to answer in this thesis, a challenging sub-problem of causal graph discovery - that of *causal feature selection*. Finally, in Section 2.5 we introduce the same problem but for sequential (time-series) data, where we also discuss commonly used methods, such as Granger Causality.

We devote Chapter 3 to relevant basic concepts of motor cortex brain functionality (sections 3.1 and 3.2), as these will consist the main area of application of the causal detection methods proposed in this dissertation. Section 3.3 provides necessary information regarding the non-invasive recording technique we use for all of our experiments (Electroencephalography). Finally Section 3.4 provides information about the non-invasive brain stimulation techniques used in research, and with more detail, basic concepts and background on the state of the art research with transcranial alternating current stimulation (tACS), which we use in two of our experiments.

Chapter 4 focuses on the problem of motor performance prediction from electroencephalographic data (Problem 1). A multi-task regression based method is proposed and the algorithm is evaluated on twenty seven healthy participants with leave-one-out cross validation. The method and the findings of this work are part of the publication

[1] The term is going to be properly introduced in Section 2.1.1. Here we briefly explain that this means the target has no descendants

[2] Without the possibility of running randomized control trial interventional experiments

(Mastakouri *et al.*, 2017). The outcome of this analysis shed more light on the observed heterogeneity of the motor cortex activity, across subjects, motivating the non-invasive stimulation experiments introduced in chapters 5 and 6.

Chapter 5 presents the first of the two stimulation experiments of 70 Hz transcranial alternating current stimulation, over the contralateral motor cortex, that was performed in this thesis. This Chapter presents the analysis on the twenty healthy participants' EEG activity, and proposes a method to narrow down the possible brain factors that can explain the encountered discrepancy in the behavioural response of the subjects to the stimulation (Problem 2). The analysis presented here and in the publication (Mastakouri *et al.*, 2019a), deduces a possible mediating role of the beta oscillations, between the applied gamma stimulation and the measured motor performance.

The recruiting, stimulation and analysis of the brain activity of twenty two new healthy participants is presented in Chapter 6. Having encountered a significant variability in the response once more, this chapter is dedicated to the construction of a pipeline that will screen responders from non-responders based on their resting brain activity prior to the application of stimulation. This chapter tackles Problem 3 and includes and discusses the findings from the work under submission (Mastakouri, 2020).

Chapter 7 introduces and proves a theorem that tackles the problem of causal feature selection on non-sequential data with latent variables (Problem 4). Moreover, in this chapter we present the results and conclusions from the application of our method both to simulated and real EEG data. The theory, the algorithm and the experiments presented in this chapter are met in the publication (Mastakouri *et al.*, 2019b).

Chapter 8 expands the logic of the methodology presented in Chapter 7 for time series data. It focuses on the challenge of identifying direct and indirect causes of a target time series in environments with latent series (Problem 5). Here we propose two novel theorems that introduce necessary and sufficient conditions for causal feature selection on time series data with latent variables. We present the graph constraints and the assumptions under which our conditions identify direct and indirect causes, even in the presence of latent common causes. Furthermore we compare our method against the commonly used Granger Causality and other state of the art methods. The theory, the algorithm and the experiments presented here are part of the work under submission (Mastakouri *et al.*, 2020) and (Mastakouri and Schölkopf, 2020).

We conclude in Chapter 9, summarizing the contribution of this dissertation to causal inference on real data, and discussing how this could facilitate the personalisation of brain stimulation. We further discuss future directions that could be extensions of the work of this thesis.

This dissertation is built upon and provides results from the following publications and papers under submission:

- Mastakouri, A. A., Weichwald, S., Özdenizci, O., Meyer, T., Schölkopf, B., and Grosse-Wentrup, M. (2017). Personalized brain-computer interface models for motor rehabilitation. In *2017 IEEE International Conference on Systems, Man,* *and Cybernetics (SMC)*, pages 3024–3029. IEEE

- Mastakouri, A. A., Schölkopf, B., and Grosse-Wentrup, M. (2019a). Beta power may meditate the effect of gamma-tacs on motor performance. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5902–5908. IEEE

- Mastakouri, A. A., Schölkopf, B., and Janzing, D. (2019b). Selecting causal brain features with a single conditional independence test per feature. In *Advances in Neural Information Processing Systems*, pages 12532–12543

- Mastakouri, A. A., Schölkopf, B., and Janzing, D. (2020). Necessary and sufficient conditions for causal feature selection in time series with latent common causes. *arXiv preprint arXiv:2005.08543* (under submission)

- Mastakouri, A. A. (2020). Stratification of behavioral response to transcranial current stimulation by resting-state electrophysiology. *bioRxiv* (under submission)

- Mastakouri, A. and Schölkopf, B. (2020). Causal analysis of covid-19 spread in germany. *Advances in Neural Information Processing Systems*, **33**

Chapter 2

Causal Inference from observational data

2.1 Graph Terminology and Notations

Here we present the most important terminology and notations on graphs, that are going to be needed during the studying of this thesis.

2.1.1 Graph notions and concepts

A **graph** $\mathcal{G} = (\mathbf{V}, \mathcal{E})$ consists of a finite set of **nodes** (vertices) $\mathbf{V}$ and **edges** $\mathcal{E} \subseteq V^2$, with $(v, v) \notin \mathcal{E}$ for any $v \in \mathbf{V}$ (Peters *et al.*, 2017). Having defined the graph $\mathcal{G}$, a set of definitions and properties are associated with it. A graph $\mathcal{G}_1 = (\mathbf{V_1}, \mathcal{E}_1)$ is called a **proper subgraph** of $\mathcal{G}$ if $\mathbf{V_1} = \mathbf{V}$ and $\mathcal{E}_1 \subset \mathcal{E}$, and we write $\mathcal{G}_1 \leq \mathcal{G}$. For the case that $\mathcal{E}_1 \subseteq \mathcal{E}$ we say that $\mathcal{G}_1$ is a **subgraph** of $\mathcal{G}$. Two nodes v^i and v^j are **adjacent** if either $(v^i, v^j) \in \mathcal{E}$ or $(v^j, v^i) \in \mathcal{E}$. An edge between two adjacent nodes v^i and v^j is called **undirected edge** if $(v^i, v^j) \in \mathcal{E}$ and $(v^j, v^i) \in \mathcal{E}$. Accordingly, an edge between two adjacent nodes is called **directed**, and we denote this by $v^i \to v^j$ for $(v^i, v^j) \in \mathcal{E}$, if it is not undirected. The undirected graph $((V), \widetilde{\mathcal{E}})$ with $(v^i, v^j) \in \widetilde{\mathcal{E}}$, if $(v^i, v^j) \in \mathcal{E}$, that results if we ignore all the arrow heads in $\mathcal{G}$ is called **skeleton** of $\mathcal{G}$. $\mathcal{G}$ is a **fully connected graph** if all pairs of nodes are adjacent. $\mathcal{G}$ is called a **directed graph** if all its edges are directed.

A node v^i is called a **parent** of v^j if $(v^i, v^j) \in \mathcal{E}$ and $(v^j, v^i) \notin \mathcal{E}$ (or schematically if $v^i \to v^j$). If $(v^j, v^i) \in \mathcal{E}$ and $(v^i, v^j) \notin \mathcal{E}$ then v^i is called a **child** of v^j. We denote the set of parents of a node v^j by $PA_j^{\mathcal{G}}$, and the set of its children by $CH_j^{\mathcal{G}}$.

A sequence of distinct vertices $v^1, v^2, \cdots, v^n$ such that $(v^i, v^{i+1}) \in \mathcal{E}$ or $(v^{i+1}, v^i) \in \mathcal{E}$ for all $i = 1, \cdots, n-1$ is called a **path**. A **directed path** is a path $v^1, v^2, \cdots, v^n$ such that $(v^i, v^{i+1}) \in \mathcal{E} \forall i = 1, \cdots, n-1$. For any two nodes v^i and v^j, if there is a directed path from v^i to v^j, then v^i is called an **ancestor** of node v^j. Accordingly, a **descendant** of v^i is any v^j such that there is a directed path from v^i to v^j. We call a **source node** a node that has no parents, and a **sink node** a node without children.

A node v^i is called a **collider** relative to a path, if $v^{i-1} \to v^i$ and $v^{i+1} \to v^i$. A path from v^i to v^m is called a **directed path** if $v^i \to v^{i+1}$ for all i.

A graph $\mathcal{G}$ is called a **directed acyclic graph** (DAG) if all the edges are directed and there is no directed cycle, that is, if there is no pair (v^j, v^k) with directed paths from v^j to v^k and from v^k to v^j. Three nodes are called an **immorality** or a **v-structure** if one node is a child of the two others that are not adjacent themselves.

We now present one of the most fundamental definitions that we are going to use throughout this thesis; Pearl's **d-separation** ((Pearl, 2009)).

Definition 1 (Pearl's d-separation)**.** *In a directed acyclic graph (DAG) $\mathcal{G}$, a path between two nodes v^1 and v^m is blocked by a set $\mathcal{S}$ (v^1, $v^m \notin \mathcal{S}$) whenever there is a node $v^k, k = 2, \cdots, m-1$, such that one of the following two possibilities is true*

(i) $v^k \in \mathcal{S}$ and $v^{k-1} \to v^k \to v^{k+1}$ or $v^{k-1} \leftarrow v^k \leftarrow v^{k+1}$ or $v^{k-1} \leftarrow v^k \to v^{k+1}$

(ii) Neither v^k nor any of its descendants is in $\mathcal{S}$ and $v^{k-1} \to v^k \leftarrow v^{k+1}$.

In a DAG G, we say that two nodes v^i and v^j are d-separated by a third node v^k if every path between nodes v^i and v^j is blocked by v^k. We then write $v^i \perp\!\!\!\perp_{\mathcal{G}} v^j \mid v^k$.

2.1.2 Probability distributions and graphs

Bayesian Networks

Consider a finite set of random variables $\mathbf{X} = (X^1, \cdots, X^d)$ with index the set of $V := \{1, \cdots, d\}$, a joint distribution $P_{\mathbf{X}}$ and a density $p(\mathbf{x})$, with respect to some product measure. In a probabilistic graphical model (**Bayesian Network** (Pearl, 2014)) each node $v^i \in \mathbf{V}$ represents a random variable X^i, and each edge represent probabilistic relations between the nodes that it connects.

The following definitions are important to connect probability distributions to a DAG.

Definition 2 (Markov property)**.** *Given a DAG $\mathcal{G}$ and a joint distribution $P(X)$, this distribution is said to satisfy*

*(i) the **global Markov property** with respect to the DAG $\mathcal{G}$ if $\mathbf{A} \perp\!\!\!\perp_{\mathcal{G}} \mathbf{B} \mid \mathbf{C} \Rightarrow \mathbf{A} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{C}$ for all disjoint vertex sets $\mathbf{A}, \mathbf{B}, \mathbf{C}$ (the symbol $\perp\!\!\!\perp_{\mathcal{G}}$ denotes d-separation.)*

*(ii) the **local Markov property** with respect to the DAG $\mathcal{G}$ if each variable is independent of its non-descendants given its parents, and*

*(iii) the **Markov factorization** property with respect to the DAG $\mathcal{G}$ if $p(\mathbf{x}) = p(x^1, \cdots, x^d) = \prod_{j=1}^{d} p(x^j \mid \mathbf{pa}_j^{\mathcal{G}})$, where we assume that $P_{\mathbf{X}}$ has a density p. The factors in the product are referred to as causal Markov kernels describing the conditional distributions $P_{X_j \mid \mathbf{PA}_j^{\mathcal{G}}}$.*

Theorem 3.27 [Equivalence of Markov properties] in (Lauritzen, 1996) proves that if $P_{\mathbf{X}}$ has a density p, then all Markov properties are equivalent.

Definition 3 (Bayesian Network)**.** *A Bayesian Network over* $\mathbf{X}$ *is a pair* $(\mathcal{G}, P(\mathbf{X}))$ *such that the joint distribution* $P(\mathbf{X})$ *is Markov with respect to the DAG* $\mathcal{G}$*.*

The Causal Markov Condition relates d-separation statements on the graph to conditional independences. However this mapping is not $1-1$. It is possible that different graphs encode the exact same set of conditional independences.

Definition 4 (Markov equivalent classes)**.** *Two DAGs* $\mathcal{G}_1$ *and* $\mathcal{G}_2$ *are Markov equivalent (belong to the same Markov equivalence class) if the set of distributions that are Markov with respect to* $\mathcal{G}_1$ *coincides with the set of distributions that are Markov w.r.t.* $\mathcal{G}_2$*. This is the case if the Markov condition entails the same set of conditional independences. This happens if and only if the two graphs have the same skeleton and the same set of v-structures (Pearl* et al.*, 1991).*

For example, the DAGs $X \rightarrow Z \rightarrow Y$ and $X \leftarrow Z \leftarrow Y$ are Markov equivalent.

Causal Markov condition allows us to read off independences between probabilities from the graph structure. Causal Faithfulness, on the other hand, allows us to infer dependences from the graph structure.

Definition 5 (Causal Faithfulness)**.** *A distribution* $P_{\mathbf{X}}$ *is said to be faithful to the DAG* $\mathcal{G}$ *if* $\mathbf{A} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{C} \Rightarrow \mathbf{A} \perp\!\!\!\perp_{\mathcal{G}} \mathbf{B} \mid \mathbf{C}$ *for all disjoint vertex sets* $\mathbf{A}, \mathbf{B}, \mathbf{C}$*.*

Looking closer to the definition of Causal faithfulness we see that it is the opposite of the global Markov condition.

Definition 6 (Causal Minimality)**.** *A distribution* $P_{\mathbf{X}}$ *is said to satisfy causal minimality with respect to the DAG* $\mathcal{G}$ *if it is Markovian with respect to* $\mathcal{G}$*, but not to any proper subgraph of* $\mathcal{G}$*.*

Causal minimality is a weaker assumption than faithfulness.

An assumption that is often made in causal discovery algorithms, is that of **causal sufficiency**, as defined by (Spirtes *et al.*, 2000) in the following definition.

Definition 7 (Causal Sufficiency)**.** $\mathbf{X}$ *is a causally sufficient set of variables if and only if there is no variable* $H \notin \mathbf{X}$ *such that* H *is a cause of two or more variables in* $\mathbf{X}$.

In other words, when causal sufficiency is assumed, we assume that all the common causes of $\mathbf{X}$ are observed. In Chapters 7 and 8 of this dissertation we propose algorithms for causal feature selection that do not assume causal sufficiency, as we believe that it is an assumption hardly met in real datasets.

2.2 Interventions

When we are able to intervene on a variable X, meaning we can set its value to a specific number, we expect that this will lead to a change of the distribution of the system. By intervening on X we force the change of its causal parents, which are no longer the ones before the intervention. An intervention corresponds to modification of the structural causal model $\mathcal{C}$ and then calculating the new distribution. It is important to emphasize on the fact that the interventional and the observational distributions are two different objects.

Definition 8 (Structural Causal Model (SCM))**.** *A structural causal model (SCM) $\mathcal{C} := (\mathbf{S}, P_{\mathbf{N}})$ consists of a collection $\mathbf{S}$ of d (structural) assignments*

$$X_j := f_j(\mathbf{PA}_j, N_j), j = 1, \cdots, d \tag{2.1}$$

where $\mathbf{PA}_j \subseteq \{X_1, \cdots, X_d\} \setminus X_j$ are called parents of X_j ; and a joint distribution $P_{\mathbf{N}} = P_{N_1}, \cdots, P_{N_d}$ over the noise variables, which we require to be jointly independent; that is, $P_{\mathbf{N}}$ is a product distribution. The graph $\mathcal{G}$ of an SCM is obtained by creating one vertex for each X_j and drawing directed edges from each parent in $\mathbf{PA}_j$ to X_j, that is, from each variable X_k occurring on the right-hand side of equation ((2.1)) *to X_j. Hence, we assume this graph to be acyclic.*

*The elements of $\mathbf{PA}_j$ sometimes are also called **direct causes** of X_j, and we call X_j a direct effect of each of its direct causes. An SCM entails an observational distribution (entailed distribution, see definition 9) and additionally, SCMs entail intervention distributions (see definition 10).*

Definition 9 (Entailed distribution)**.** *An SCM $\mathcal{C}$ defines a unique distribution over the variables $\mathbf{X} = (X_1, \cdots, X_d)$ such that $X_j = f_j(\mathbf{PA}_j, N_j)$, in distribution, for $j = 1, \cdots, d$. We refer to it as the entailed distribution $P_{\mathbf{X}}^{\mathcal{C}}$ and sometimes write $P_{\mathbf{X}}$.*

Definition 10 (Interventional distribution)**.** *Consider an SCM $\mathcal{C} := (S, P_{\mathbf{N}})$ and its entailed distribution $P_{\mathbf{X}}^{\mathcal{C}}$. We replace one (or several) of the structural assignments to obtain a new SCM $\tilde{\mathcal{C}}$. Assume that we replace the assignment for X_k by $X_k := \tilde{f}(\tilde{\mathbf{PA}}_k, \tilde{N}_k)$. We then call the entailed distribution of the new SCM an **intervention distribution**. We denote the new distribution by $P_{\mathbf{X}}^{\tilde{\mathcal{C}}} =: P_{\mathbf{X}}^{\tilde{\mathcal{C}};do(X_k:=\tilde{f}(\tilde{\mathbf{PA}}_k,\tilde{N}_k))}$.*

*The set of noise variables in $\tilde{\mathcal{C}}$ now contains both some "new" $\tilde{N}$'s and some "old" N's, all of which are required to be jointly independent. To denote an intervention on a variable X_k to a real value α we write $P_{\mathbf{X}}^{\mathcal{C};do(X_k:=\alpha)}$. We call this **atomic intervention**. We call an **imperfect intervention** and denote with $\tilde{\mathbf{PA}}_k = \mathbf{PA}_k$ when the direct causes remain direct after the intervention. Imperfect intervention is a special case of a **stochastic intervention**, where the marginal distribution of the intervened variable has positive variance (Korb* et al.*, 2004). The new SCM $\tilde{\mathcal{C}}$ must have an acyclic graph, so that the set of allowed interventions depends on the graph induced by $\mathcal{C}$.*

When there is a possibility to actually intervene on a variable, then there is a possibility in some cases to perform randomized controlled trials experiments (RCTs). We define RCTs in the next section and we explain why this option for causal inference is a privileged one.

Before that, we need to define an important notion, that of total causal effect, as introduced by (Pearl, 2009), which we will also mention in one of our methods in Chapter 8.

Definition 11 (Total causal effect)**.** *Given an SCM $\mathcal{C}$, we say that there is a total causal effect from X to Y if and only if $X \not\perp\!\!\!\perp Y$ in $P_{\mathbf{X}}^{\mathcal{C};do(X:=\tilde{N}_X)}$ for some random variable $\tilde{N}_X$.*

The existence of a directed path between X and Y in the corresponding graph is necessary but not sufficient for a total causal effect.

2.2.1 Randomized controlled trials (RCTs)

According to Cartwright (2010), an ideal RCT is an experiment where "all factors that can produce or eliminate a probabilistic dependence between cause (C) and effect (E) are the same in both wings except for C, which each subject in the treatment group is given and no-one in the control wing is given, and except for factors that C produces in the course of producing E, whose distribution differs between the two groups only due to the action of C in the treatment wing. **An outcome in an RCT is positive if $P(E)$ in the treatment wing $>$ $P(E)$ in the control wing**." In practice this means:

- **Double-blind experiments**: The subjects should be unaware of whether they receive the cause or the placebo; the attendant physicians should not know;

- **Random assignment of subjects to the treatment or control wings**, to ensure that other possible reasons for dependencies and independences between cause and effect under test will be distributed identically in the treatment and control wings;

- **Careful choice of a placebo** [1] to be given to the control, to ensure that any **'psychological' effects** produced by the recognition that a subject is receiving the treatment will be the **same in both wings**.

If it is possible to run RCTs experiments, then, depending on the interventional distribution of the effect in the treatment and the control group, someone is possible to make causal statements.

In many cases, due to ethical constraints or limited resources, it may not be possible to run randomized trial experiments. For example, someone is not allowed to test all the possible different stimulation combinations in the human brain to conclude the most effective one. In such scenarios, it is important to still be able to make causal statements based on the observational distributions. In this thesis, in Chapters 7 and 8 we propose theorems and algorithms that tackle the problem of identification of causes of a target variable, when the SCM is not known, causal sufficiency cannot be assumed and RCTs experiments are not possible.

[1] Placebo is an item indistinguishable from the cause, for those associated with the experiment, apart from the fact that in contrast to the cause, it does not change anything with respect to the targeted effect.

2.3 Causal Discovery

When the graph is unknown, the procedures of trying to learn the underlying graph from observational data is called **causal discovery** (Spirtes *et al.*, 2000). In this section we make a brief review over the most known causal discovery methods (Glymour *et al.*, 2019; Peters *et al.*, 2017) and in the subsections 2.4 and 2.5 we introduce a sub-category of causal discovery, the causal feature selection.

Causal discovery, as it is purely based on observational data, suffers from some issues, which are characteristics of observational distributions with finite sample sizes. The most important ones include non-stationarity, missing values and selection bias (Glymour *et al.*, 2019). **Non-stationarity** is the phenomenon where the underlying mechanism that produces the data changes over time, or across different datasets. (Zhang *et al.*, 2017) and (Huang *et al.*, 2017) have been studying causal discovery under distribution shift and in non-stationary data. Regarding **missing data**, when missing data are not completely random but they follow some unknown mechanism then the causal sufficiency is violated in the dataset, and the distribution of the observed data might differ from the true one. (Tu *et al.*, 2018) proposed an algorithm based on PC (Spirtes *et al.*, 2000), which performs causal discovery in datasets with missing values, given certain assumptions. Finally, when the inclusion of a data point in the sample depends on some of its attributes and it is not random, we say that we have **selection bias**. (Zhang *et al.*, 2016) tackle a specific version of selection bias, that of outcome-dependent selection, however there is still a lot of room for research in this topic.

2.3.1 Constrained-based methods

Constrained-based methods are methods that mostly use conditional independences and d-separation statements (under the assumption of causal Markov property and of causal faithfulness that put constraints in the joint distribution) to infer the existence of an edge between two observed variables. Most of these methods refer to independent point data (not time-series). Two of the most known methods that aim at full causal graph discovery are **PC** and **FCI** (Spirtes *et al.*, 2000). PC assumes causal sufficiency, while FCI does not. Both methods have a two-step procedure.

The first step of **PC** determines the variables $X - Y$ that are adjacent in the graph. To do this, someone needs to test every pair of variables to check whether they are dependent given any other set of observed variables. When the graph is not sparse, this approach can lead to very large conditioning sets in the conditional independence tests (up to $d-1$ variables, for d observed variables in total). The statistical strength of tests with such large conditioning sets is particularly low and unreliable. The second step of PC determines for each pair, whether X and Y are independent of each other, given a set C with an edge connected on either of them. If they are found independent, then the edge is eliminated. Then they decide for the orientation of the edges based on some orientation propagation rules (for details see (Spirtes *et al.*, 2000)). Although conditional independences are the main key of this algorithm, it should not be confused with the so-called conditional independence graphs (Lauritzen, 1996), in which two variables are not adjacent if and only if they are conditionally independent given all the remaining variables.

One of the most known variation of PC is the Fast Causal Inference (**FCI**) algorithm. This method allows for hidden common causes and it is proven to asymptotically converge to the correct causal graph. In the first step, FCI prunes the undirected graph with conditional independence tests similar to PC. In the second step, it orients edges with a procedure similar to PC, with the difference that it does not assume that every edge is necessarily directed the one way or the other. There are many variants of PC and FCI that try to speed up the computationally very expensive search, such as RFCI (Colombo *et al.*, 2012).

None of these algorithms is able to distinguish between Markov equivalent graphs (as these entail the same set of conditional independences).

The causal discovery methods that are based on conditional independencies have the advantage that they are not limited to linear relationships (non-linear dependence tests can be used for the detection of a dependence, such as kernel based conditional independence tests (Fukumizu *et al.*, 2008; Gretton *et al.*, 2008; Zhang *et al.*, 2012)). However, causal faithfulness is a very strong assumption, which is not possible to test, and can lead to false conclusions when it is violated.

2.3.2 Score-based methods

A different family of algorithms tries to infer the causal graph $\mathcal{G}$, given data $\mathcal{D}$ from a vector $\mathbf{X}$ of variables, trying to assign a score $\mathcal{S}(\mathcal{D},\mathcal{G})$ to each graph $\mathcal{G}$. Following, it searches over the space of DAGs for the graph with the best score. What the $\mathcal{S}$ function can be differs between the different versions of algorithms. Often a parametric model is used for $\mathcal{S}$. Depending on whether a Bayesian or a frequentist approach is used, the score function can be defined as (2.2) and (2.3) accordingly.

$$\mathcal{S}(\mathcal{D},\mathcal{G}) = \log p_{prior}(\mathcal{G}) + \log p_{(\mathcal{D}|\mathcal{G}} = \log p_{prior}(\mathcal{G}) + \int_{\theta\in\Theta} p_{(\mathcal{D}|\mathcal{G},\theta} \cdot p_{prior}(\theta). \tag{2.2}$$

or

$$\mathcal{S}(\mathcal{D},\mathcal{G}) = \log p(\mathcal{D} \mid \hat{\theta},\mathcal{G})\frac{d}{2}\log n. \tag{2.3}$$

where n is the sample size.

In the Bayesian approach (equation ((2.2))) the highest-score graph $\tilde{\mathcal{G}}$ is the maximum a posteriori estimator (MAP). The selection of the prior over the parameters is studied by (Heckerman *et al.*, 1995). In the frequentist approach (equation ((2.3))) the maximum likelihood estimator $\hat{\theta}$ is used alongside the Bayesian Information Criterion (BIC).

To tackle the problem of the super exponential grown number of DAGs, with the number of nodes, Chickering et al. proposed a greedy score-based approach, named the Greedy Equivalence Search (**GES**) algorithm (Chickering, 2002). Instead of starting with a fully connected undirected graph, it starts with no edges and gradually adds the ones that are currently needed, eliminating in the end the unnecessary ones until a local maximum is reached. At each iteration of the algorithm, the addition of an edge to the graph is decided based on whether this increases the score function.

Ogarrio *et al.* (2016) et al. proposed a combination of GES and FCI, in their algorithm GFCI, where they are using the first to find a supergraph of the skeleton and the latter for pruning and for orientation of the edges. This algorithm has been shown to be more accurate than the original FCI.

Nevertheless, none of the aforementioned algorithms that aims to recover the full graph can distinguish among Markov equivalent classes.

2.3.3 Methods based on SCMs

Many recent causal discovery methods aim to find the cause-effect direction. These algorithms are based on SCMs (also mentioned as Functional causal models (FCMs)), where the target variable Y is a function f of the direct causes $\mathbf{X}$ and some noise (that also includes unobserved variables) which is assumed to be independent from $\mathbf{X}$. Such methods assume that the transformation from $(\mathbf{X}, N)$ to $(\mathbf{X}, Y)$ is invertible, such as the noise term N can be uniquely recovered form $\mathbf{X}$ and Y (Glymour *et al.*, 2019). The main idea behind these methods is that the noise N is independent from the input X for only one direction ($X \rightarrow Y$ or $Y \rightarrow X$). Moreover, the main assumption is that there are no hidden confounders in the data. Under these constraints the methods first fits an SCM in both directions, and then calculates the independence of the noise in both cases. (Hyvärinen and Pajunen, 1999) and (Zhang *et al.*, 2016) have proved that under no further assumptions on the function class, it is impossible to identify the causal direction, as it will always be possible to find independence between the noise and the input.

Linear non-gaussian additive models

For the bivariate case, restricting the function class to linear functions, the SCM can be written as $Y = bX + N$, where as explained above $N \perp\!\!\!\perp X$ and additionally X and N follow a normal distribution. If non-gaussian noise is assumed, then it is possible to identify the SCM. (**?**). (Shimizu *et al.*, 2006) proved this theorem for multivariate inputs as well, using Independent Component Analysis (ICA) (Hyvärinen and Oja, 2000). The model proposed by (Shimizu *et al.*, 2006) is called linear non-Gaussian acyclic model (LiNGAM) and was later improved in (Shimizu *et al.*, 2011) (DirectLiNGAM).

Non-linear methods

(Hoyer *et al.*, 2009) extended the LiNGAM model to a non-linear additive noise model. A different model (post-nonlinear transformation PNL) is introduced by (Zhang and Chan, 2006) and (Zhang and Hyvarinen, 2012) where $Y = f_2(f_1(X) + N)$, and f_1, f_2 are non linear functions. Furthermore, f_2 is assumed to be invertible. In contrast to LiNGAM, PNL has been shown to be identifiable in the generic case, with exception of specific cases mentioned in (Zhang and Hyvarinen, 2012). The general conclusion form the aforementioned models is that non-linear SCMs methods are not computationally as efficient as in the linear case (Glymour *et al.*, 2019).

2.3.4 Markov blanket detection methods

Here we give the definition of a Markov blanket, as given in (Pearl, 2014) and (Peters *et al.*, 2017) and discuss algorithms that try to detect it.

Definition 12 (Markov blanket). *Consider a DAG $\mathcal{G} = (\mathbf{V}, E)$ and a target node Y. The Markov blanket of Y is the smallest set $\mathbf{M}$ such that $Y \perp\!\!\!\perp_{\mathcal{G}} \mathbf{V} \setminus (\{Y\} \cup M)$ given M. If $P_{\mathbf{X}}$ is Markovian with respect to $\mathcal{G}$, then $Y \perp\!\!\!\perp \mathbf{V} \setminus (\{Y\} \cup M)$ given M.*

According to (Pearl, 2014), for DAGs we know that the Markov Blanket contains not only the parents of Y, but also children and parents of children $M = \mathbf{PA}_Y \cup \mathbf{CH}_Y \cup \mathbf{PA}_{\mathbf{CH}_Y}$.

Having defined this object MB, we can see that it can play a useful role in feature selection. Markov blanket detection methods take a more local approach on the problem of caual discovery, trying to identify those variables that are conditionally independent of a target given the remaining variables (Guyon *et al.*, 2019). Fu and Desmarais (2010) give a review of algorithms that try to identify the Markov blanket in unknown graphs, with the goal of optimal feature selection, for the time period 1996 (KS) to 2008 (IPC-MB). Some of the most known Markov blanket algorithms are IAMB (Tsamardinos and Aliferis, 2003), HITON-MB (Aliferis *et al.*, 2003) and MMMB (Tsamardinos *et al.*, 2006) which include a forward selection phase during which variables are required to display some dependency with the target in order to be included in the conditioning set. HITON-MB is an algorithm that checks for univariate dependencies between the candidate feature and the target, while IAMB and MMMB test the conditional dependency of the feature and the target Y given a growing conditioning set of previously selected variables. As we can see, such forward selection methods suffer from two fundamental drawbacks: first, they fail if causal sufficiency is violated, and, second, they may identify an existing dependence with the target only when all the variables of the Markov blanket are added in the conditioning set. Moreover, HITON-MB and MMMB depend of PC and may also identify wrong variables, because as pointed out by Peña *et al.* (2005), under certain conditions variables not in the Markov blanket of Y can enter $MB(Y)$.

Another category of methods, such as BAHSIC (Song *et al.*, 2007a,b) and K-CDM (kernel based conditional dependence measures) (Strobl and Visweswaran, 2019) uses a different approach, that of backward elimination in combination with the Hilbert Schmidt Independence Criterion (HSIC) (Gretton *et al.*, 2005a) to measure dependence between two kernels. In (Song *et al.*, 2007a) the BAHSIC algorithm is proposed, where the target Y is embedded in the first kernel, and the rest of the variables in the second kernel. In the second step, backward elimination is used to remove variables from the second kernel that maximize HSIC.

2.4 Causal feature selection

In Chapter 7 we introduce another sub-problem of causal discovery, which we name **causal feature selection**, and we describe a theorem with sufficient conditions to solve it, under limited assumptions. More specifically, by causal feature selection we refer to the problem of the identification of the direct and indirect causes of a target variable Y, given a pool of candidate features $\mathbf{X}$, that may or may not be dependent with the target, under no causal sufficiency assumption (Mastakouri *et al.*, 2019b).

Many algorithms mentioned in the Markov Blanket Section 2.3.4 could be used to solve this problem if Y was a sink node, but all of them will fail if causal sufficiency is violated. Furthermore, most of the aforementioned methods in Section 2.3.4 will return a ranked list of the features based on some prediction criterion. In a very naive, yet very commonly used approach, also Lasso regression (Santosa and Symes, 1986; Tibshirani, 1996) and non-linear variations of it (HSIC-Lasso, (Yamada *et al.*, 2014)) have been used for feature selection (not causal). As these methods introduce a bias due to the regularization, methods such as Double ML (Chernozhukov *et al.*, 2017) and improvements upon it (Raj *et al.*, 2020) have been proposed to perform causal feature selection by mitigating this bias. These methods make use of the concept of orthogonalization to overcome the bias introduced due to regularization. Although these methods do not require faithfulness they do rely fully on the strong assumption of causal sufficiency.

2.5 Causal discovery on time series

Causal inference from time series is a fundamental problem in data science, with applications in the fields of economics, biology, earth science and machine monitoring. It is also a problem that has not overall been solved yet.

In this section we deal with the problem of causal feature selection in time series data, without causal sufficiency being assumed. What makes this problem different from the one mentioned in Section 2.4 is the nature of the input data itself. The structure and time order dependency imposed by a time series creates both an advantage, as it help us to know the direction, and a challenge to come up with both sufficient and necessary conditions for the identification of the causes of Y. We define the problem we try to answer in Chapter 8 as follows (Mastakouri *et al.*, 2020): We are given observations from a target time series $Y := (Y_t)_{t \in \mathbb{Z}}$ whose causes we wish to identify, and observations from a multivariate time series $\mathbf{X} := ((X_t^1, \dots, X_t^d))_{t \in \mathbb{Z}}$ of potential causes (candidate time series). Moreover, we allow an unobserved multivariate time series $U_t := (U_t^1, \dots, U_t^m)$, which may act as common cause of the observed ones. The system consisting of $\mathbf{X}$ and Y is not assumed to be causally sufficient, hence we allow for unobserved variables U_t.

Having defined the problem, we now need to give the definition of one of the most used methods that tries to tackle this problem: Granger Causality.

Definition 13 (Bivariate Granger Causality). *Under the assumption of Causal Sufficiency, X influences Y whenever the past values of X help in predicting Y from its own past. Formally, we write*

$$X Granger\text{-}causes Y \iff Y_t \not\!\perp\!\!\!\perp X_{past(t)} \mid Y_{past(t)} \tag{2.4}$$

Definition 14 (Multivariate Granger Causality). *X_j Granger causes X_k if*

$$X_t^k \not\!\perp\!\!\!\perp X^j_{past(t)} \mid \mathbf{X}^{-j}_{past(t)} \tag{2.5}$$

Granger emphasized that proper use of Granger causality would actually require to condition on all relevant variables in the world. Nevertheless, Granger causality is often used in its bivariate version or in situations in which clearly important variables are unobserved. Such a use can yield misleading statements when interpreting the results causally. (Peters et al.*, 2017)*

Although Granger Causality has been the most widely used approach for causal inference in time series for the last fifty years (Wiener, 1956; Granger, 1969, 1980), violations of its strict assumptions, such as causal sufficiency, and no instantaneous effects, lead to serious issues and to incorrect causal conclusions (Peters *et al.*, 2017). During the last decades, several extensions have been proposed to address these issues (Hung *et al.*, 2014; Guo *et al.*, 2008). Despite the fact that the time order of variables renders it an easier problem than the typical ‘causal discovery problem’ of inferring the causal DAG among n variables without any prior knowledge on causal directions (Pearl, 2009; Spirtes *et al.*, 2000), there is no doubt that causal inference in time series is still a challenging, non trivial task. Peters *et al.* (2017) et al. showed that Granger causality can be derived from d-separation (see, e.g., Theorem 10.7 in (Peters *et al.*, 2017)). In addition, beyond Granger's method, several authors showed how to conclude that one time series Granger-causes another one, based on d-separation criteria. For instance, Entner and Hoyer (2010a) and Malinsky and Spirtes (2018a), were inspired by the FCI algorithm (Spirtes *et al.*, 2000) and the work from Eichler (2007) aiming at the discovery of the full causal graph in time series, without assuming causal sufficiency (for an extended review see (Runge, 2018; Runge *et al.*, 2019b)). As their method can give reuslts up to Markov Equivalent Classes, their method will not identify all the causal relations. Runge *et al.* (2019a) et al. proposed PCMCI as an extension of PC and although lower rates of false positives are reported compared to classical Granger causality, the method still relies on the assumption of causal sufficiency.

Some of the biggest challenges when working with time series data, even outside the scope of causal inference, include the lack of knowledge of the generating process which may be non-linear, the missing data, as well as the finite samples that may not be able to capture slow dynamics, and non stationarity [2]. Finally a big problem is that of hidden confounding, which however, remains a big problem even with i.i.d. measurements.

Having presented a quick overview over the state of the art literature on this topic, we postpone the presentation of our method (Mastakouri *et al.*, 2020) for Chapter 8, pointing out that the last mentioned problem (that of hidden confounding) is not an issue for our method, under proper assumptions.

[2] The dynamics and the mechanisms of the system may not remain the same over time, which means that the probability distributions of the variables may change over time.

Chapter 3

Motor Cortex , Electroencephalography and Non-invasive brain stimulation

3.1 Brain rhythms

Brain rhythms or neural oscillations refer to distinct patterns of massed neuronal activity that happens in different frequency ranges (or bands) and which is linked with different behaviours, sleep states, as well as arousal levels (Frank, 2009). Human brain functions cover a frequency range from 0.1 Hz up to 200 Hz, separated into the following five canonical bands: “delta” (1–4 Hz), “theta” (4–8 Hz), “alpha” (8–13 Hz), “beta” (13–35 Hz) and “gamma” waves ($\geq$ 35 Hz).

3.1.1 Mechanisms that produce neocortical brain rhythms

Brain rhythms in the neocortex are the product of three mechanisms: first, intrinsic membrane properties of different classes of neurons, second, intracortical and thalamocortical network interactions, and finally, modulation by arousal circuits (Steriade, 2005). There are two basic modes of activity that many neurons in the cortex and thalamus are changing in between; those of tonic firing and intrinsically bursting. This change is caused by ionic membrane currents that are differentially activated and inactivated due to neuronal hyperpolarization (Steriade, 2005; Dickson *et al.*, 2000). Therefore, brain rhythms are a type of brain signal, in contrast with the EEG signals, which, as we will explain later, are a type of epiphenomenon of the ongoing brain activity.

3.1.2 Basic characteristics of each frequency band

While these neural oscillations can co-exist in different locations in various patterns, there are some very basic characteristics (not exclusive) which studies have associated with every frequency range (Miller, 2007). The brain areas in which those rhythms have been studied extensively are the neocortex, the hippocampus and the thalamus.

For example, “delta” (δ) waves are mostly associated with deeper sleep stages, while even slower signals (<1 Hz) in the neocortex have been found to coordinate de- and hyper-polarization in intra-cortical and thalamocortical networks (so called “up” and “down” states) (Destexhe *et al.*, 2007; Steriade, 2005). Surprisingly enough, the same frequency range in young children, has been found in waking EEG in the occipito-parietal and occito-temporal cortex with an amplitude < 100 uV (Swaiman *et al.*, 2017). Finally, it has been suggested that delta waves are part of the homeostatic mechanism (Benington and Frank, 2003).

“Theta” (θ) rhythms are normally measured in the frontal and central cortical regions and have been found to be related to emotions, drowsiness and memory. At the same time, it has been associated with various clinical conditions such as epilepsy. Theta oscillations are produced by the hippocampal pyramidal cells which have similar orientations and fire periodically in synchrony (Kropotov, 2010a). It has been reported that their role is significant for temporal coding and decoding and for the modification of synaptic weights (Kryger *et al.*, 2017).

“Alpha” (α) activity is most prominent in the occipital regions and tends to decrease when eyes are open, compared to when they are closed, and during visual attention. Alpha rhythms are also known as posterior dominant rhythms (PDR). Their power tends to systematically decrease with age in healthy subjects, with its highest values around 20 years old. As the frequency of alpha rhythms decreases with age, there is a big controversy about whether alpha activity could be a neuro-marker of cognitive function regulation (Kropotov, 2016a). Regarding its relation to movement, increased alpha-power over ipsilateral sensorimotor cortex has been associated with preparation and selection of movement (Brinkman *et al.*, 2014).

There are two categories of “beta” (β) rhythms: the Rolandic and the frontal ones. It has been reported that the Rolandic beta appears when the corresponding neuronal system in the sensory-motor strip is relaxing after a strong activation phase, and it is considered a postactivation indication (Kropotov, 2010b). Frontal beta rhythms have lower amplitude and in contrast to Rolandic rhythms that appear during motor tasks, they are associated with cognitive functions such as decision making and assessment of stimulus. Beta power deviates from its normal levels in cases of clinical conditions. In ADHD beta is decreased in resting state (Kropotov, 2016b), while in Parkinson’s disease it is elevated. In particular, beta activity has been found significantly elevated in patients with motor disorders (tremors, slowed movements) such as Parkinson’s disease (McAllister *et al.*, 2013; Brown, 2007; Khanna and Carmena, 2017). Furthermore, in healthy subjects, elevated beta power has been found to play an antikinetic role (Khanna and Carmena, 2017).

Finally, “gamma” (γ) activity due to its relatively low amplitude and the fact that it can easily be contaminated by muscular artefacts, is very underestimated and not as much studied as the rest of the frequency bands (Malik and Amin, 2017). It has been associated with a wide spectrum of functions such as working memory, movement and attention. Gamma waves in the neocortex have been suggested to play a role in synchronizing different cortical modules in cognitive tasks and in consciousness (Steriade, 2005). Moreover, hippocampal gamma activity may be important for encoding information and memory formation (Axmacher *et al.*, 2006). More importantly regarding the motor cortex, increased γ activity has been associated with large ballistic movements (Muthukumaraswamy, 2010; Nowak *et al.*, 2018). It has also been suggested to be prokinetic, given that it is increased during voluntary movement (Brown, 2003).

In this dissertation we will examine the important role of beta (see Chapter 5), gamma (see Chapter 6) and alpha (causal findings in Chapter 7) rhythms in the motor cortex during reaching tasks, as well as before and after non invasive brain stimulation sessions. Although there is some consensus about the basic role of each band, there is considerable controversy about the functional role when it comes to more complex/sophisticated processes. The role of each brain rhythm is still not fully established.

3.2 Motor cortex

In eutherian mammals (Lillegraven *et al.*, 1987), the large structure in the front of the cerebral cortex is called motor cortex. According to the definition given by Kandel *et al.* (2000), motor cortex "is the main source of motor fibres of the pyramidal tract, which synapse directly with motor neurons in the brainstem and spinal cord, thus, enabling voluntary movements".

This part of the human brain has been a topic of research for centuries, and yet the exact connection between complex behavioural outputs, such as reaching movements, and its functionality is elusive (Hatsopoulos and Suminski, 2011). Many research evidence implicate motor cortex in the volitional control of movement (Ebbesen and Brecht, 2017), for both its initiation and its suppression.

3.2.1 First studies on motor cortex

In 1870 Fritsch and Hitzig applied current to specific sites in the frontal cortex of dogs, observing that they evoked movements that varied with the cortical location of the stimulation (Fritsch, 1870). This was the first modern experiment to result in some kind of motor cortex mapping. Ferrier (1875) identified a similar cortical motor map in monkeys, and the stimulation experiments that followed revealed activation of complex motor patterns according to the location. The first mapping of the motor cortex in human brains was performed by Penfield and Rasmussen (Penfield and Rasmussen, 1950), during cortical stimulations in parallel with surgery in awake patients. This interactive experiments confirmed that the motor cortex in humans has a somatotopic map (called **homunculus**). Furthermore, these experiments lead to the conclusion that the motor cortex stimulation not always excites the activity but also causes movement inhibition and suppression (Ebbesen and Brecht, 2017).

3.2.2 Motor cortex functionality and heterogeneity

Motor cortex is part of a set of complex circuits that not only controls movement but also receives sensory inputs (Hatsopoulos and Suminski, 2011). The neurons of the motor cortex start firing up to many hundreds of milliseconds before limb movement is initiated (Georgopoulos *et al.*, 1982). Several groups of neurons that act in different stimuli and tasks have been identified. One of these groups are the "mirror" neurons (Di Pellegrino *et al.*, 1992) (Rizzolatti *et al.*, 1996) in the ventral premotor cortex of non human primates, which are a group of neurons that discharge similarly in response to overt motor action. Some other groups of neurons fire predominantly during voluntary movement but not during visual playback or passive movement, and another group does exactly the opposite. Finally, others respond to combinations of the aforementioned stimuli. Hatsopoulos et al. (Hatsopoulos and Suminski, 2011) propose that this heterogeneity may explain in part the lack of a unified theory of the function of motor cortex.

It has been proposed that as the motor cortex is not solely dedicated in movement but also in sensory processing, the same way the somatosensory cortex also contributes to motor control (Matyas *et al.*, 2010). Hatsopoulos and Suminski (2011) report rich heterogeneity in motor cortex response properties, including strong visual and somatosensory effects. They also review the kinematics of human movement and how these are encoded by individual motor cortex neurons and how they are related to the motor cortex activity. Amongst force and torque (Kalaska *et al.*, 1989; Cabel *et al.*, 2001), arm position (Georgopoulos *et al.*, 1984; Paninski *et al.*, 2004), acceleration (Stark *et al.*, 2007), distance (Fu *et al.*, 1993) and velocity (Moran and Schwartz, 1999), the most robust variables were found to be direction of movement (Georgopoulos *et al.*, 1982) and arm speed (Moran and Schwartz, 1999).

3.3 Electroencephalography

Electroencephalography –in short EEG– is a non invasive recording technique that is used to measure the electrical cortical activation, using electrodes affixed to the scalp. More specifically, EEG measures difference in electrical potential between two points where electrodes are placed (Bales *et al.*, 2018). EEG is a non-invasive recording technique with very high temporal resolution (ms) but with poor spatial resolution. As a result, it prevails other non invasive brain imaging techniques (such as PET, fMRI) in terms of temporal resolution, but it has a significantly lower spatial resolution. This low spatial resolution is justified by the fact that the activity recorded by the EEG electrodes are superposition of underlying brain activity. It is widely believed that the primary source of the electrical signals that are recorded by EEG is the current flow in the apical dendrites of pyramidal cells in the cerebral cortex. What is being recorded then, is the coherent activation of a large number of pyramidal cells in a small area. Such a small area of the cortex can be modelled as a current dipole (Okada *et al.*, 1997). Because of this nature of EEG signals, they are highly sensitive to the conductivity of the brain, skull, and extracranial tissue.

The EEG equipment consists of a set of scalp electrodes coupled to high-impedance amplifiers and a digital data acquisition system (Darvas *et al.*, 2004).

In contrast with other brain imaging techniques like fMRI and PET, EEG signals are the direct extracranial demonstration of neuronal activation; where "direct" refers to the nature of the signal and not to each exact location.

3.4 Non invasive brain stimulation

3.4.1 Definition of non-invasive brain stimulation

Non invasive brain stimulation (NIBS) is the broader family of methods which aim to modulate neural activity, behaviour, and brain plasticity through the creation of forced electrical current flows inside the brain by non-invasive means (Wagner *et al.*, 2007; Dayan *et al.*, 2013). NIBS modifies brain function through interaction with multiple neurotransmitters and networks (Obeso *et al.*, 2016). There are two main categories of NIBS: Transcranial Magnetic Stimulation (TMS), which uses external magnetic fields to force the creation of electrical potentials in the cortex depolarizing neurons and triggering action potentials (Di Lazzaro *et al.*, 2004), and Transcranial Electrical Stimulation (TES) (Bestmann and Walsh, 2017), which applies weak electrical direct (tDCS) or alternating (tACS) currents on the scalp (Nitsche and Paulus, 2000). In contrast to TMS, only a fraction of this current enters the brain and causes a membrane potential change of the affected neurons, which is sufficiently strong to change their probability of generating action potentials (Antal and Herrmann, 2016).

3.4.2 Applications of NIBS

Although the neural mechanisms of how NIBS modulates brain activity are not yet fully understood (Vosskuhl *et al.*, 2018), NIBS applications spread in a very broad field of research and treatment. Applications of NIBS can be divided into three main categories: studies that probe neurophysiology (e.g., how neural oscillations are causally related) (Shafi *et al.*, 2012; Polanía *et al.*, 2012b; Filmer *et al.*, 2014; Sehm *et al.*, 2012; Keeser *et al.*, 2011; Hampson and Hoffman, 2010; Anand and Hotson, 2002), studies that investigate how brain activity gives rise to cognition (Polania *et al.*, 2018; Vosskuhl *et al.*, 2018; Pogosyan *et al.*, 2009; Joundi *et al.*, 2012; Neuling *et al.*, 2012; Cecere *et al.*, 2015; Lustenberger *et al.*, 2015; Vosskuhl *et al.*, 2015; Polanía *et al.*, 2012a; Santarnecchi *et al.*, 2013; Sela *et al.*, 2012), and studies that attempt to use NIBS for rehabilitation (Schulz *et al.*, 2013; Tortella *et al.*, 2014; Fregni *et al.*, 2005; Palm *et al.*, 2014; Veniero *et al.*, 2019). In all three categories, however, NIBS studies report large variations in effect sizes across individual subjects, often resulting in small or even statistically insignificant group-level effects (Hashemirad *et al.*, 2016; Pogosyan *et al.*, 2009; Joundi *et al.*, 2012; Moliadze *et al.*, 2010; Triccas *et al.*, 2016; López-Alonso *et al.*, 2014; Strube *et al.*, 2015). Large percentages of non-responders (up to 55%) (López-Alonso *et al.*, 2014; Strube *et al.*, 2015), as well as a large variance in the direction of the response, from significantly positive to significantly negative (López-Alonso *et al.*, 2014) in the remaining population of responders, have been reported as outcome of the same stimulation set-up.

NIBS in neurophysiology

The first category examines network communication through stimulation-based modulation. Shafi et al. (Shafi *et al.*, 2012) gave evidence that tDCS paired with neuroimaging can be a powerful tool for identifying and describing functional brain networks. Anodal tDCS over the left motor cortex was found to increase functional connectivity between the left motor cortex and the ipsilateral thalamus, caudate nucleus, and parietal association cortex, whereas cathodal tDCS was found to decrease connectivity between the left motor cortex and the contralateral putamen (Polanía *et al.*, 2012b; Filmer *et al.*, 2014). Sehm et al. (Sehm *et al.*, 2012) reported that bilateral tDCS of motor cortex induces widespread changes in functional connectivity, predominantly modulating changes in primary and secondary motor as well as prefrontal region. Keeser et al. (Keeser *et al.*, 2011) found that tDCS over prefrontal cortex induces alterations in both the default mode (DMN) and fronto-parietal networks (FPN). In (Filmer *et al.*, 2014) are reviewed studies that provided evidence of causal changes in oscillatory activity in the theta, alpha, beta and gamma ranges, induced by tDCS. In the same field of application of NIBS, several studies have used TMS for accessing and altering neural dynamics in networks that are widely distributed anatomically (Hampson and Hoffman, 2010). As a conclusive statement of this category, Anand et al. in (Anand and Hotson, 2002) pointed out that TMS is an appropriate method for measuring neural conduction and processing time, activation thresholds, facilitation and inhibition in brain cortex, and neural connections.

NIBS in cognition and behaviour

The second category aims to understand how experimentally altered neural activity causally affects behaviour (Polania *et al.*, 2018). Vosskuhl et al. in their review (Vosskuhl *et al.*, 2018) reported the role of NIBS in research of lower and higher cognitive functions. Lower-level cognitive functions like voluntary movement (Pogosyan *et al.*, 2009), (Joundi *et al.*, 2012), audition (Neuling *et al.*, 2012) and vision (Cecere *et al.*, 2015) have been successfully modulated by Transcranial Alternating Current Stimulation (TACS). The same holds for higher cognitive functions, such as creativity (Lustenberger *et al.*, 2015), memory (Vosskuhl *et al.*, 2015; Polanía *et al.*, 2012a), intelligence (Santarnecchi *et al.*, 2013) and risk taking (Sela *et al.*, 2012).

NIBS studies in rehabilitation

The third category uses NIBS as a clinical treatment tool for higher cognitive functions and motor rehabilitation. More specifically, NIBS has been used in neurological diseases to enhance adaptive processes and prevent potential maladaptive ones (Schulz *et al.*, 2013). Representative clinical applications of NIBS include treatment of major depressive disorders (Tortella *et al.*, 2014), motor deficits after stroke or spinal cord injury (Schulz *et al.*, 2013), Parkinson's disease (Fregni *et al.*, 2005), therapy of multiple sclerosis (Palm *et al.*, 2014) and many more (Veniero *et al.*, 2019).

3.4.3 Effect-sizes of NIBS in the various application domains

The variations in effect sizes across individual subjects are a fundamental challenge for translating NIBS into clinical applications, because consistent positive effects of NIBS across individual subjects are essential to avoid harm to a patient during a stimulation treatment (Raffin and Siebner, 2014).

However, this consistency has not yet been established: A large variation is being observed in the reported effects of NIBS in behavioural (Hashemirad *et al.*, 2016; Pogosyan *et al.*, 2009; Joundi *et al.*, 2012; Moliadze *et al.*, 2010) and treatment studies (Triccas *et al.*, 2016; López-Alonso *et al.*, 2014; Strube *et al.*, 2015). The aforementioned variation in NIBS effects has been expressed either as contradictive outcomes from similar stimulation setups (Moisa *et al.*, 2016; Antal *et al.*, 2008; Joundi *et al.*, 2012) or as small insignificant effect-sizes (Triccas *et al.*, 2016; Buch *et al.*, 2017). Large percentages of non-responders (up to 55%) (López-Alonso *et al.*, 2014; Strube *et al.*, 2015), as well as large variance in the direction of the response, from significantly positive to significantly negative (López-Alonso *et al.*, 2014) in the remaining population of responders, result in very small across-subjects average effect-sizes. In either case, the generalisation of NIBS has not yet been achieved, and particular concern has been expressed about the importance of addressing the variability across subjects (López-Alonso *et al.*, 2014; Wiethoff *et al.*, 2014; Lafon *et al.*, 2017). The question whether these small effect-sizes could be justified due to external factors and not by fundamental neurophysiological differences across individuals, has lately been a topic of scientific research (Cocchi and Zalesky, 2018; Stecher *et al.*, 2017; Strube *et al.*, 2015; Ridding and Ziemann, 2010).

3.4.4 Reasons for NIBS variability

Several suggestions have been proposed as possible explanations of the observed across-subjects variability of NIBS effects. One such neurophysiological explanation suggests that different populations of cortical neurons are stimulated more easily, or are more excitable in different people, at different times (López-Alonso *et al.*, 2014; Strube *et al.*, 2015; Wiethoff *et al.*, 2014). Therefore, the question "state or trait?" arises. Hence, the variability might be caused by individual differences in the recruitment of cortical neurons. Another opinion focuses on the importance of the task-paradigm on the stimulation response, as it has been shown that hidden confounders in the experimental task can lead to variations in the brain-networks recruitment (Stecher *et al.*, 2017; Buch *et al.*, 2017; Vosskuhl *et al.*, 2018). As a result, the observed effects may be highly dependent on the specific context in which stimulation is applied. Finally, another explanation that has been proposed, is the fact that studies that include negative findings are not always published, because of the existing bias (Vannorsdall *et al.*, 2016). Yet, not always reporting negative effects of stimulation may mislead the conclusion about it, i.e. a stimulation that leads to negative effects most of the times, with no published studies that point to it, and, on the contrary, published studies that point to a few random cases with positive effects, create a fictitious variability, which wouldn't exist otherwise. Concluding, the cause of this heterogeneity is multi-factorial and, to some degree, still unknown (Wiethoff *et al.*, 2014).

Chapter 4

Personalised Multi-task Regression Models for Motor Performance

In this Chapter, we propose a multi-task learning-based method that builds from only a few EEG trials, personalised decoding models that relate the global EEG configuration of brain rhythms in individual subjects to their arm movement smoothness during 3D reaching task. Our models exhibit substantial heterogeneity across subjects, which we argue that could potentially also reflect limited effect sizes observed in brain stimulation studies that focus on enhancing motor performance. The problem presented in this Chapter is tackled in the author's publication (Mastakouri *et al.*, 2017), alongside Sebastian Weichwald, Timm Meyer, Bernhard Schölkopf and Moritz Grosse-Wentrup.

4.1 Problem statement

We try to answer the question: "Is it possible to build individualised models that predict arm stability from brain signals, with a limited number of recording trials?" (see Problem 1). If so, are the features used by the predictor consistent across different subjects? Are they causal and can they be used as targets of brain stimulation?

4.2 Motivation

Motor deficits are one of the most common outcomes of stroke. According to the World Health Organization, fifteen million people suffer a stroke each year, worldwide. Five million of them are permanently disabled afterwards. For this percentage of people, upper limb weakness and loss of hand function are among the most common types of disabilities, which affect the quality of their daily life (Organization, 2002). Although there is a wide range of rehabilitation therapies, including medication treatment (Walker-Batson *et al.*, 1995), conventional physiotherapy (Green *et al.*, 2002), and robotic-assisted physiotherapy (Lum *et al.*, 2002), only $\sim 20\%$ of patients achieve some form of functional recovery in the first six months (Kwakkel *et al.*, 2003; Nakayama *et al.*, 1994).

It has been found that post-stroke alterations of cortical networks are correlated with the severity of motor deficits (Sharma *et al.*, 2009; Grefkes *et al.*, 2008). For that reason, current research on novel therapies focuses on neurofeedback training based on brain-computer interface (BCI) technology and transcranial electrical stimulation (TES) (see Section 3.4.1). The first approach usually employs a robotic exoskeleton that is congruent to movement attempts, with the goal to support cortical reorganisation as decoded in real-time from neuroimaging data, by providing haptic feedback (Grosse-Wentrup *et al.*, 2011; Gomez-Rodriguez *et al.*, 2011). The latter type of research aims to inhibit or excite cortical areas in a way that supports motor performance. While initial evidence suggested that both approaches (Ramos-Murguialday *et al.*, 2013), (Hummel *et al.*, 2005) have a positive impact, there has not been recorded a consistent significance of these methods over conventional physiotherapy (Ang *et al.*, 2015), (Ang *et al.*, 2014), (Butler *et al.*, 2013).

One potential reason for the difficulty in replicating the initially promising findings is the heterogeneity of stroke patients' cortical networks. Different stroke-induced structural changes are likely to result in substantial across-patient variability in the functional reorganisation of their affected brain areas. As a result, not all patients may benefit from the same stimulation or neurofeedback protocol. Therefore, in this Chapter, we propose to learn personalised models that relate the configuration of cortical networks to each subject's motor deficits in search of evidence of such heterogeneity.

Using a multi-task learning framework developed in our group (Jayaram *et al.*, 2016), we build personalised decoding models that relate the EEG of healthy subjects during a 3D reaching task to their arm movement smoothness in single trials. The resulting models seem to capture substantial heterogeneity of the relevant features across subjects. This finding further supports our argument about the need for personalised models. We conclude by reviewing our findings in the light of brain stimulation studies that aim to facilitate motor performance in healthy subjects.

4.3 Methods - EEG dataset acquisition

4.3.1 Subjects

Twenty-six healthy, right-handed, male participants (mean age of 28.3 ± years) were recruited for this study. The study was approved by the ethics committee of the Max Planck Society, and all subjects gave informed consent after a detailed description of the experimental task.

4.3.2 Experimental Set-up

The experimental set-up consists of the following parts:

A real-time motion tracking system

The Impulse X2 Motion Capture System (PhaseSpace, San Leandro, CA, U.S.) was used for the real-time tracking of the subject's arm position (x, y, z-coordinates), with a sampling frequency of 960 Hz. A customised glove with three infrared LEDs was worn by the subject on their right arm, and the system's four infrared cameras were positioned around them.

Visual feedback screen

During the task, subjects are seated approximately 1.5 meters in front of a screen. The arm position is tracked and presented on the screen as a striped sphere, which the subject is able to control. A 3D stripe pattern and a visible shadow are assigned to the sphere object for a better adaptation of the subject to the 3D space on the screen.

EEG acquisition system

The electroencephalogram of the subjects is recoded using an active 121-channel cap and a BrainAmp DC amplifier (BrainProducts, Gilching, Germany), with 500 Hz sampling frequency. The electrodes were positioned according to the 10-5 system for high-resolution EEG (Oostenveld and Praamstra, 2001). The reference electrode was placed at the TPP9h location.

4.3.3 Experimental Paradigm

The experimental paradigm was developed using the BCPY2000 software, an extended Python version of BCI2000 (Schalk *et al.*, 2004). The experimental phases are described subsequently.

Calibration phase

At the beginning of each experimental recording, the subjects are instructed to place their arm in a comfortable position next to their leg. This position is assigned to be the "starting position" of the sphere on the screen. Afterwards, a calibration period follows. During this phase, the subject is instructed to move her/his arm in cyclic motions, in order to store in the system the individualised area of comfortable movements. During this "exploration" period, possible targets inside the limits of the subject's reaching area are computed.

Resting phase

Five minutes of baseline EEG are recorded, during which the participant is asked to relax focusing on a fixation cross shown on the screen, without moving.

Visuo-motor experiment phase

This phase consists of two blocks of 50 trials each. Each block is followed by a 5 min resting-state recording. Each trial begins with a 5 s "task baseline", during which subjects are asked to relax. During the following "planning" phase (duration was uniformly chosen between 2.5–4 s) a white and a yellow patterned sphere appears on the screen. The former reflects the subject's arm position, and the latter shows the randomly chosen target position to be reached. Subjects are asked to mentally plan their reaching movement but not yet move. Once the target sphere turns green, the "go" phase starts, and the subject is expected to move their arm and try to reach the target position. A trial is considered "failed", and a black screen with red bar is shown, if subjects move more than 4 cm during the "planning" phase, or if the subject does not manage to reach the target (to overlap the end-effector sphere and the target by less than 3.5 cm) within the 10 seconds "go" phase. If the trial is successful, a score that reflects their motor performance appears on the screen (cf. Section 4.3.4). Finally, the "return" phase begins. During this period, there is no time constraint, and the subjects should move their arm back to their starting position, which is now indicated by a green sphere. Once their arm sphere position overlaps by less than 1 cm with the green sphere, the trial is considered completed. The whole trial sequence which is depicted in Figure . 4.1.

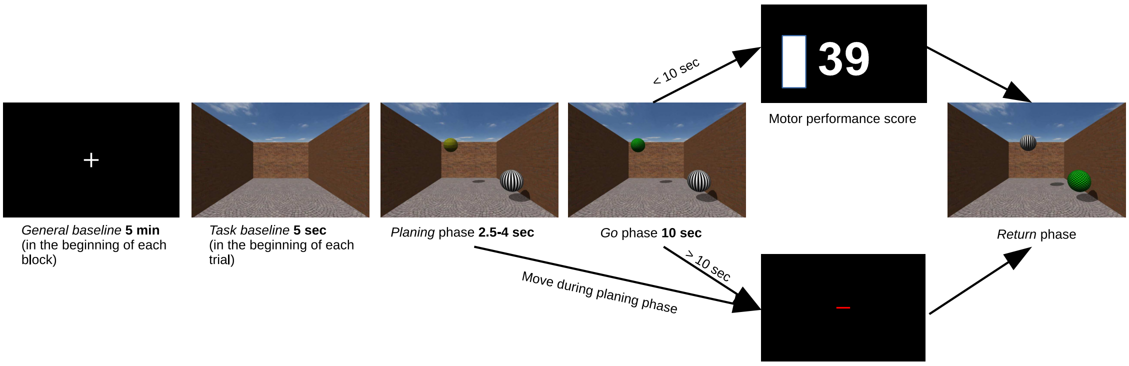

+
General baseline 5 min (in the beginning of each block)
Task baseline 5 sec (in the beginning of each trial)
Planing phase 2.5-4 sec
Go phase 10 sec
< 10 sec
39
Motor performance score
> 10 sec
Move during planing phase
Return phase

Figure 4.1: Trial sequence of the visuo-motor reaching task.

4.3.4 Index of Motor Performance

The normalised averaged rectified jerk (NARJ) (Cozens and Bhakta, 2003) was used as an index of motor performance. This metric reflects the *smoothness* of a movement and has been shown to correlate with the Fugl-Meyer Assessment of Motor Recovery after Stroke (FMA) (Wade *et al.*, 2011).

One NARJ value is calculated for each trial. First, we compute the jerk value Jerk$_{.,t}$ (the second derivative of the speed), at each time step t in each of the three dimensions

x, y, z as follows.

$$\text{NARJ} = T^3 \frac{1}{T} \sum_t \sqrt{\text{Jerk}_{x,t}^2 + \text{Jerk}_{y,t}^2 + \text{Jerk}_{z,t}^2}$$

where T is the duration of the reaching movement. The score feedback that was presented to the subjects by the end of every successful trial was the inverted NARJ value, fitted between 0 and 100, so that a higher score can be interpreted by the subjects as a "better" movement.

4.4 Methods - EEG Analysis

4.4.1 Preprocessing

After removing the *"failed"* trials, the remaining ones range between 89 and 98 per subject. We restrict our analysis to the 118 channels that were consistently recorded with low noise for all subjects, excluding those that were noisy in at least one of the recordings. The remaining channels were re-referenced to common average reference. The period of the trial that corresponds to the maximum arm movement duration is kept for all channels (the time window 7.5–17.5 s of each trial, where 7.5 corresponds to the earliest possible start of the "go" signal). In order to attenuate non-cortical artefacts, we perform an Independent Component Analysis (ICA) and only reproject those components that, by visual inspection of the topographies and source frequency spectra, correspond to cortical sources (cf. Section 2.3 of (Grosse-Wentrup and Schölkopf, 2012) for a description of this procedure).

4.4.2 Feature computation

For each trial and EEG channel we compute the log-bandpower in the following five frequency bands: *delta* (δ, 1–4 Hz), *theta* (θ, 4–8 Hz), *alpha* (α, 8–13 Hz), beta (β, 13–30 Hz) and *high gamma* (γ, 60–90 Hz) (the frequency range 30–60 Hz is excluded because of the 50 Hz power line noise). This results in 118 channels × 5 logarithmic bandpowers × number of trials feature space for each subject.

4.4.3 Multi-task learning regression

We want to predict the logarithmic NARJ value from the log-bandpower features of the "Go" phase of the same trial, for each trial individually. We adapt the multi-task learning algorithm presented in (Jayaram *et al.*, 2016) in order to perform linear regression, as follows:

$$\mathbf{w_s} = \arg\min_{\mathbf{w_s}} \frac{1}{\lambda} \|\mathbf{X_s w_s} - \mathbf{y_s}\|^2 + \frac{1}{2}[(\mathbf{w_s} - \mu)^T \Sigma^{-1} (\mathbf{w_s} - \mu)] \tag{4.1}$$

where λ is the variance of the original noise model of subject s and in the loss function it controls the ratio of the importance assigned to the prior probability of the learned weight vector versus how well the learned vector can predict the labels in the training data. Therefore, the higher the variance of the noise in the model the less we can trust our training data. Σ, μ are the covariance matrix and the mean of the weight vectors of the prior subjects. The regularizer term is modelled as a Gaussian distribution with mean and variance the mean and variance of the prior weights. This way it punishes overfitting on the current subject's s data. Σ and μ are initialized as I and 0, and following w_s, Σ and μ are updated in parallel until convergence. For more details see Jayaram *et al.* (2016). $\mathbf{X_s}$ are the input brain features for subject s, $\mathbf{w_s}$ are the weights for model we want to learn and y_s are the targets (NARJ index) that we want to predict from brain activity.

This enables us to leverage the data of 25 subjects when training a model for the 26th subject. More specifically, for every subject s we train a predictive model with features from all the trials of the remaining 25 subjects. This is the *prior* (also called generic) model of subject s. Then, we update the prior model's weights with the data from the first 20 trials of subject s. We call this model the *updated* or *personalised* model for subject s. This *personalised* model is then used to predict the remaining trials of this subject. We use leave-one-subject-out cross validation in order to evaluate our model.

4.4.4 Statistical Tests

To evaluate the predictive power of our models, first, we assess their ability to predict the average NARJ value in the final 50 trials of each subject. To do this, we compute the across-subject correlation coefficient between the predicted average NARJ values and the observed ones. To estimate the p-value under the null hypothesis that the predicted and observed average NARJ values are uncorrelated, we permute the subject-order of the predicted average NARJ values 10^4 times and count how many times the modulus of the resulting correlation coefficient exceeds the modulus of the correlation coefficient with the subject-order intact.

To evaluate the performance of single-trial NARJ predictions, we use the magnitude square coherence between the predicted and the observed NARJ values for each subject. Coherence is commonly used to estimate the power transfer between the input and the output of a linear system; it quantifies the extent to which one signal can be predicted from another by an optimum linear least squares function (Bendat and Piersol, 2011). We then randomly permute, within-subjects, the trial-order of the predicted NARJ values 10^4 times and compare the resulting magnitude square coherence with the one measured for the correct trial-order. This yields a p-value for each subject.

By definition, the p-values are drawn from a standard uniform distribution if the null-hypothesis is true. To quantify the deviation of the empirical cumulative distribution function (CDF) of the above p-values from the CDF of a standard uniform distribution we create 100 equally sized bins between 0 and 1, and then sum, across all bins, the absolute differences between the empirically observed CDF and the one generated by drawing the same number of samples from a standard uniform distribution. Sampling this test statistic 10^3 times gives us a p-value that reflects how likely it is that the subjects' p-values are drawn from a standard uniform distribution.

4.5 Results

4.5.1 Adaptation of Motor Performance over Time

As it can be seen in Figure 4.2 (left), there is a strong adaptation period during the first 20 trials regarding the stability of the arm movement, as this is measured by the NARJ. After about 50 trials, the mean NARJ values have almost converged to their final value. The distribution of the movement smoothness towards the end of the task (averaged over the last 50 trials) exhibits a substantial heterogeneity across subjects (see Fig. 4.2, right).

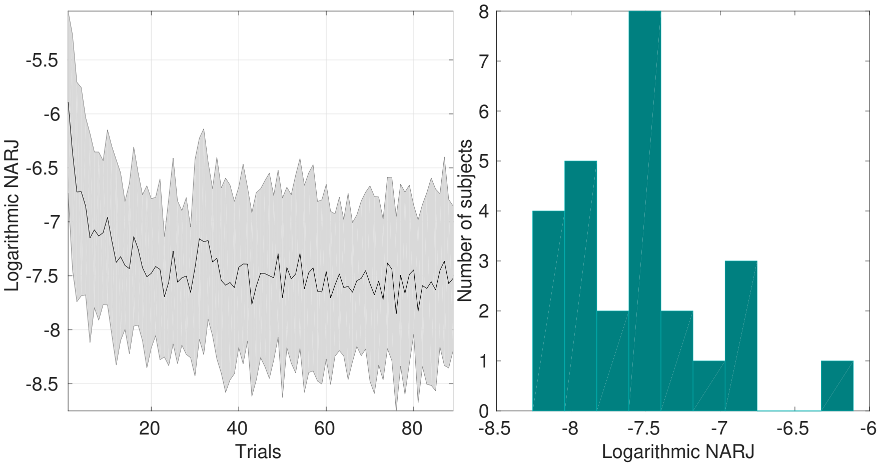


Figure 4.2: Left: Average across subjects arm NARJ over time (mean ± one standard deviation). Right: Histogram of the arm movement smoothness (NARJ) towards the end of the experiment (averaged over last 50 trials).

4.5.2 Model Validation

Prediction of subjects' final mean motor performances

Using leave-one-subject-out cross validation, we predicted the movement smoothness on a single trial level. In this section, we present the predicted movement smoothness towards the end of the task. Figure 4.3 depicts the observed vs the predicted average NARJ values in the final 50 trials, for the personalised (left column) and the prior model (right column). A significant correlation between the model predictions and the observed true NARJ values is observed only for the updated (personalised) models ($\rho = 0.52,\ p = 0.008$). On the contrary, no sufficient evidence to reject the null-hypothesis of chance-level performance is measured for the prior (generic) models ($\rho = 0.31,\ p = 0.139$).

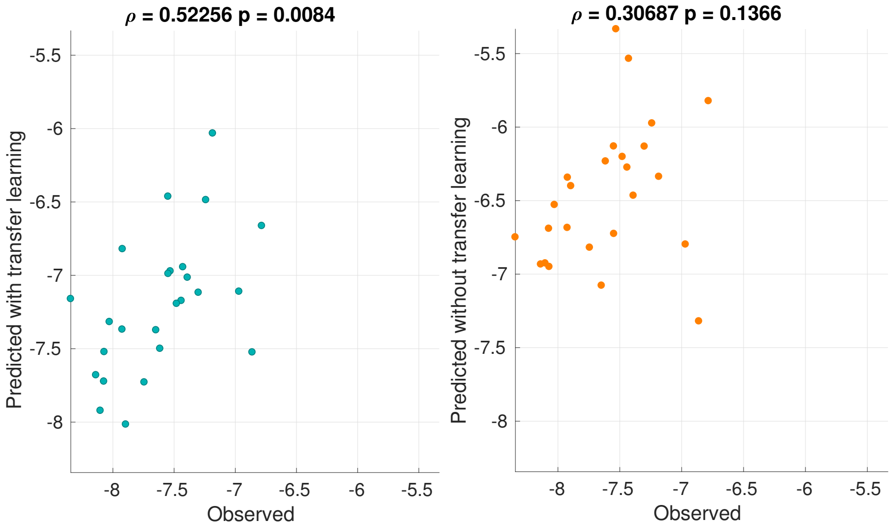


Figure 4.3: Predicted and observed average logarithmic NARJ across the last 50 trials for the 26 subjects with ($p = 0.0084$, left plot) and without ($p = 0.1366$, right plot) the use of transfer learning regression. Each point corresponds to one subject.

Prediction of motor performance in individual trials

We assess the ability of the personalised models - built with the multi-task regression framework- to predict movement smoothness for individual trials, calculating the magnitude square coherence (cf. Section 4.4.4) as shown in Figure 4.4 for every subject and model type (personalised and generic). While both models achieve a similar mean coherence across subjects—0.36 and 0.35 for the updated and prior model respectively—the comparison of the distribution of p-values across subjects with the CDF of a standard uniform distribution reveals a notable difference. The p-value under the null-hypothesis of standard uniformly distributed subject p-values is marginally significant for the updated model ($p = 0.06$), while the prior model returns a p-value of 0.63.

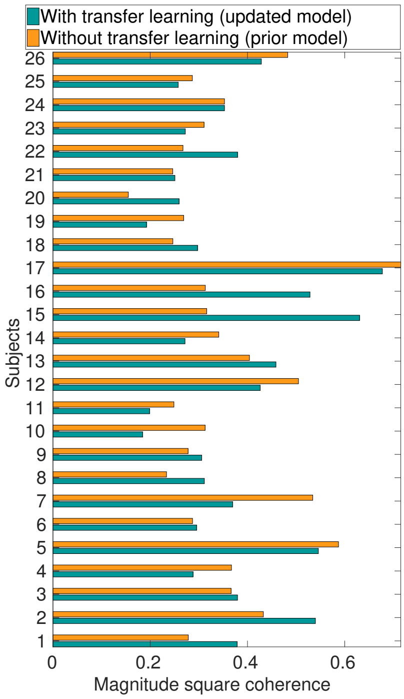


Figure 4.4: Magnitude-squared coherence between observed and predicted NARJ values for each subject.

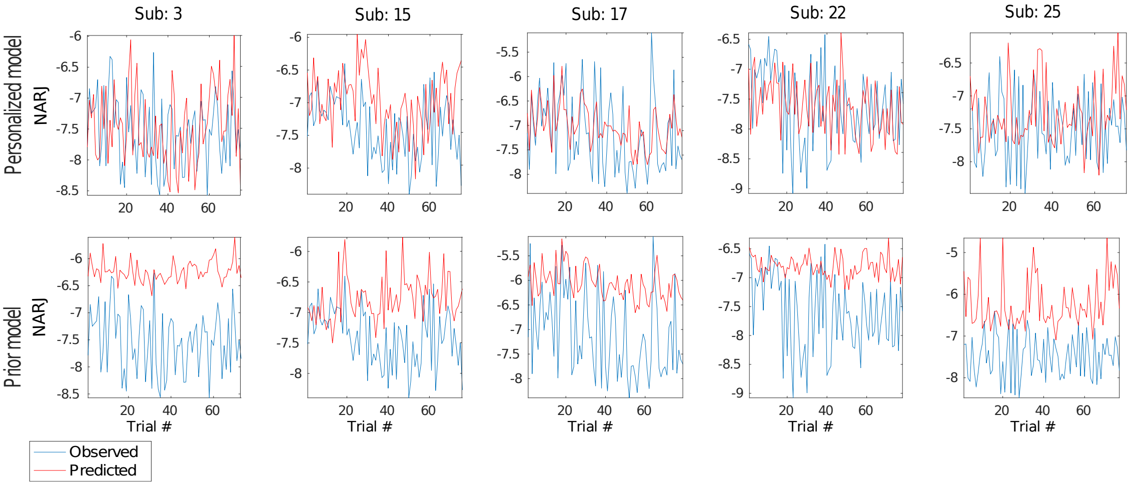


Figure 4.5: Predicted (red) and true (blue) arm NARJ values across trials for five representative subjects. Top row: predictions using the personalised model. Bottom row: predictions using the prior model.

Figure 4.5 shows the observed and predicted NARJ values across trials for five representative subjects for the personalised (top row) and the prior model (bottom row). As it can be seen, only the updated model captures meaningful variations in movement smoothness across trials.

4.5.3 Model Interpretation

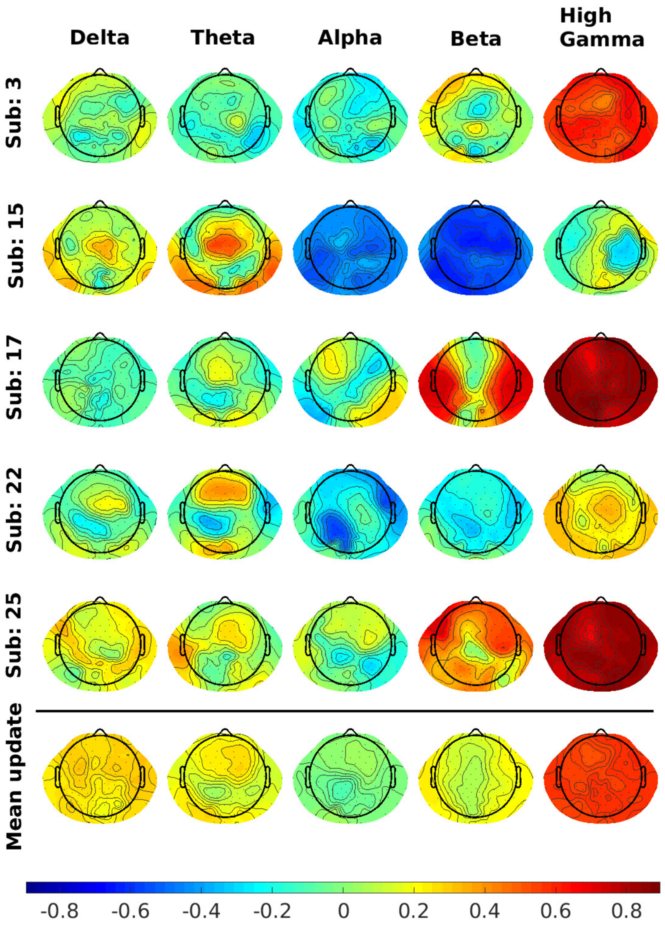


Figure 4.6: Correlation of each feature with the predicted NARJ values for the personalized models, for five representative subjects (first five rows), for the five frequency bands (five columns). The last row shows the correlation model averaged over all subjects for each frequency bands. Similar results are observed for the rest of the subjects. As we can see, gamma power seems to be dominant across a large cortical area in some subjects. Hadn't we performed the artefact correction step, we would be sceptical about the source of this activity. However, having removed all the muscular artefacts and all the non-cortical source signals, we can be optimistic that this is a cortical signal.

To better understand the cortical processes used for prediction, we compute the correlation coefficients between the personalised models' predictions ($\hat{y}_s$) and the individual electrode bandpower features $\mathbf{X_s}$ (also called by Haufe *et al.* (2014) an encoding model derived from the decoding model). This way, we quantify how much each channels' bandpower contributes to the prediction.

Figure 4.6 shows the resulting correlation topographies (5 representative subjects and mean topography) are shown. Red/blue colour indicates a positive/negative correlation between electrode bandpower and the logarithmic NARJ, i. e., bandpower at blue coloured electrodes is positively associated with smoother movements. Notice that there is a qualitative difference between the average model and the personalised correlation models. Moreover, notice that the personalised models exhibit substantial heterogeneity. Same features in same location for some subjects correlate positively and for some subjects negatively in the predicted smoothness. Alpha, beta and high gamma ranges exhibit strongest correlations —but with inconsistent signs— across subjects, while correlations in the delta and theta range are comparably small.

4.6 Discussion

4.6.1 Personalised Models with low sample size

In this Chapter, we showed a way to build "personalised" models that relate the global configuration of EEG rhythms to motor performance using multi-task learning. Such models allow us to cope with the heterogeneity of motor activity across subjects, as well as, to extend previous work (Meyer *et al.*, 2014) to the harder task of single-trial prediction (Meinel *et al.*, 2016). Because of the high dimensionality (590-D) of the feature space, building personalised models based on subject-specific training data would require several hundreds of training trials, resulting in a calibration time of several hours. Using the multi-task linear regression framework enabled us to learn each individual model from only 20 trials of that very subject.

4.6.2 TES and Model Heterogeneity

While most TES motor studies consistently focus on the contralateral motor cortex M1, not all reported findings are consistent with each other, inasmuch as they exhibit contradicting roles for the different frequency bands: Some studies report an inhibiting effect of 20 Hz tACS over the contralateral motor cortex on movement, but no significant effect of stimulation in the γ-range (Pogosyan *et al.*, 2009; Joundi *et al.*, 2012; Moliadze *et al.*, 2010). Others, at the same time describe significant effects in the γ-range, but they do not find significant evidence for inhibiting effects of stimulation in the β-range (Moisa *et al.*, 2016). In frontoparietal areas, γ oscillations have been found to correlate with reaction times in a motor task (Gonzalez Andino *et al.*, 2005), in contrast to stimulation studies that found improvements in implicit motor learning only after applying 10 Hz AC (α range stimulation) but for neither 1, 15, 30 or 45 Hz (Antal *et al.*, 2008).

The heterogeneity in the organisation of subjects' cortical networks could explain such inconsistent results. One of the most important findings of this study support this line of argument by evidencing a substantial heterogeneity amongst subjects: Features in the α, β, and γ range, used by the predictors, turn out to correlate sometimes negatively and sometimes positively with the measured motor performance. This line of thought could be extended in a stimulation setting, as lack of personalised stimulation parameters could lead to inconsistent group-level effects and non-significantly improved motor performance, due to the network differences across subjects.

Decoding models like the ones trained in this study *do not* reflect causal relationships immediately, and as such, they are not meant to directly indicate optimal stimulation parameters for each subject (Haufe *et al.*, 2014; Weichwald *et al.*, 2015) (see also interpretation rules R3 and R4 in (Weichwald *et al.*, 2015)). However, although association maps as the ones computed in Figure 4.6 cannot indicate causes of behavioural response, they do allow us to rule out EEG features that are not causal (cf. rule R2 in (Weichwald *et al.*, 2015)). Hence, we argue that a decoding model that is able to predict well and efficiently single-trial motor performance is a necessary prerequisite for personalised stimulation protocols as it can rule out subject specific irrelevant stimulation areas. At the same time, such models revealed a considerable across-subject heterogeneity in feature relevance which constituted the main motivation for the four following chapters, where we will conduct tACS studies and will propose new causal feature selection methods, towards the personalisation of brain stimulation protocols.

Chapter 5

Beta activity may mediate the response to 70Hz contralateral tACS

This Chapter is dedicated to a series of combined EEG with tACS visuomotor experiments that were conducted by the author, as part of the studying of the relation between motor cortex and motor performance, from a causal inference perspective. In this Chapter, we apply 70 Hz tACS over the contralateral motor cortex of twenty healthy participants, expecting to facilitate arm speed, according to the literature. Observing a significant variability in the behavioural response of each subject, we try to identify potential causes that could explain it, using only data of brain activity from before and after the stimulation blocks. We propose three statistical tests applied stepwise, with which we show that beta power could be a potential causal factor that explains the discrepancy in response across subjects. The analysis and the results presented in this Chapter come from the author's publication (Mastakouri *et al.*, 2019a), alongside Bernhard Schölkopf and Moritz Grosse-Wentrup.

5.1 Problem statement

First, we try to validate research studies that recommend 70 Hz contralateral tACS for the facilitation of movement, by replicating the spatial and amplitude parameters, incorporating them in our crossover experimental set up. So we ask whether the arm speed is facilitated significantly with 70 Hz contralateral tACS compared to sham. In other words, is the behavioural response to stimulation consistent across subjects (see Problem 2)? Finally, if that is not the case, what can explain the heterogeneity?

5.1.1 Physical limitation

One fundamental physical limitation is imposed by the signal of the tACS itself. The injected current creates a voltage signal which is orders of magnitude higher than the level of the EEG signals. As a result, during the stimulation blocks, it is impossible so far to recover the ongoing brain oscillations, due to the saturation of the amplifier. This problem limits the possible information to the brain activity right before and right after the stimulation.

5.2 Motivation

TACS gradually becomes more and more prominent as a motor rehabilitation method, because of its ability to influence non-invasively the ongoing brain oscillations at arbitrary frequencies and intensities. However, many studies report substantial variations in its effect across individuals, rendering tACS currently unreliable as a treatment tool. One reason that could explain this variability is the lack of knowledge about the exact way with which tACS interacts with ongoing brain activity. A reason that explains the latter is the difficulty in acquiring contemporaneous brain oscillations during stimulation. The present crossover stimulation study tries to shed light on the way that tACS entails the ongoing brain oscillations, by contributing to the understanding of the cross-frequency effects of gamma tACS - 70 Hz - over the contralateral motor cortex. Performing stimulation and EEG recording to 20 healthy participants, we provide empirical evidence which is consistent with a mediating role of low- (12–20 Hz) and high- (20–30 Hz) beta power between gamma-tACS and motor performance.

TACS modulates the neural activity and the behaviour through the creation of an electric field inside the brain (Herrmann *et al.*, 2013; Bestmann and Walsh, 2017). More specifically, a weak electrical alternating current is applied on the scalp (Nitsche and Paulus, 2000), which has been reported to cause changes to the membrane potential of the affected neurons (Antal and Herrmann, 2016), at least at non-human primates. TACS has been used widely in behavioural studies (López-Alonso *et al.*, 2014; Strube *et al.*, 2015) as well as for the treatment of neurological disorders (Schulz *et al.*, 2013; Fregni *et al.*, 2005), although its exact neurophysiological effect on brain networks has not yet been fully understood (Vosskuhl *et al.*, 2018).

TACS studies that target the motor cortex have reported considerable variability in stimulation response across individual subjects, with large percentages of non-responders (López-Alonso *et al.*, 2014; Strube *et al.*, 2015). Although it has been proposed that tACS in the γ- ($\sim$ 70 Hz) and β- ($\sim$ 20 Hz) range facilitates and respectively inhibits the movement (Joundi *et al.*, 2012; Pogosyan *et al.*, 2009; Moliadze *et al.*, 2010; Wach *et al.*, 2013), there have been reported contradictory findings regarding the significance of these effects (Moisa *et al.*, 2016; Gonzalez Andino *et al.*, 2005; Antal *et al.*, 2008). A lot of researchers have focused on the role of physiological γ- (Nowak *et al.*, 2018; Muthukumaraswamy, 2010) and β-oscillations in movement (Espenhahn, 2018; Gulberti *et al.*, 2015; McAllister *et al.*, 2013). However, the exact mechanism that tACS entrains these ongoing brain oscillations is not yet fully understood (Davis and Koningsbruggen, 2013; Helfrich *et al.*, 2016).

5.2.1 Role of physiological beta and gamma oscillations

In order to construct our main argument, it is essential to first give a short overview of the role of the physiological β- and γ- brain rhythms in movement. The brain activity in the γ-range has been associated with cued and self-paced transient finger movements (Muthukumaraswamy, 2010). Furthermore, relatively large ballistic movements of higher movement amplitude were associated with increased motor cortex γ-power (Muthukumaraswamy, 2010). Furthermore, (Gaetz *et al.*, 2013) gave evidence for a motor γ-band network that is associated with response selection and maintenance of planned behaviour. The selection of 70 Hz for stimulation of the contralateral motor cortex in the present study is based on the aforementioned findings and reports.

β activity, on the other hand, has been found to be significantly elevated in patients with motor disorders, such as Parkinsons disease, with symptoms like tremors, slowed movements, trouble initiating movements, (McAllister *et al.*, 2013; Brown, 2007; Khanna and Carmena, 2017). Furthermore, in studies with healthy subjects, it was reported that movements preceded by a reduction in β-power exhibited significantly faster reaction times than movements preceded by an increase in β-power (Khanna and Carmena, 2017). Two other studies have also proposed that β-activity represents the status quo (Engel and Fries, 2010), suggesting that enhanced β-activity prevents change from the current state (Schnitzler and Gross, 2005; Davis *et al.*, 2012). Moreover, evidence has been given that β-oscillations are the summed output of principal cells temporally aligned by GABAergic interneuron rhythmicity (Yamawaki *et al.*, 2008), (McAllister *et al.*, 2013). The power of both spontaneous and movement-related oscillatory β-activity in human M1 has been shown to be GABA-A dependent (Jensen *et al.*, 2005), (Hall *et al.*, 2011).

5.2.2 Our hypothesis

Taking into account the existing knowledge about the role of β-oscillations in the inhibition of movement speed (McAllister *et al.*, 2013), and about the effect of high stimulation frequencies on the decrease of β-power (Gulberti *et al.*, 2015), we hypothesise that the modulation of the ongoing β-activity may mediate the effect of γ-tACS on the observed behavioural motor response. This hypothesis consists of two empirically testable implications: First, we expect that the arm speed will be affected by the stimulation. We examine the effect of γ-tACS on movement response in Section 5.4.2. Second, we expect that the recorded β-power will be affected by the stimulation. In this regard, we examine if this is true and, if so, in which brain areas a modulation of β-power can be observed (cf. Section 5.4.5). We expect that any changes in β-power induced by γ-tACS will be significant only for the subjects that exhibit a significant behavioural response to the stimulation.

Having found evidence about the above points, as a final step we want to approach the relation between the modulation of beta activity and the motor response through a cause-effect perspective (Methods section 5.4.6). We thus perform an additional causal analysis, using three pairwise cause-effect tests, to examine the effect of β-power change on motor performance (Section 5.4.6). The results of these tests support further a potentially causal role of β-power in the response to γ-tACS, given the necessary assumptions. In Section 5.6.3, we discuss a plausible neurophysiological mechanism that could explain our findings. As an additional corroboration of our statement, we present evidence that the subjects that respond with faster arm movements to stimulation are the ones whose high-β-power significantly decreases during the γ stimulation.

5.3 Methods - EEG/tACS dataset acquisition

The study conformed to the regulations of the Declaration of Helsinki. The experimental procedures involving human subjects described in this Chapter were approved by the Ethics Committee of the Medical Faculty of the Eberhard Karls University of Tübingen.

5.3.1 Experimental Setup

Subjects

In this study, twenty healthy, right-handed subjects were recruited. One of the subjects (ID 10) did not participate in the second session of recordings, hence was excluded from the analysis. The remaining 19 healthy participants (nine female, ten male) were 28.36 ± 8.57 years old.

Stimulation parameters

We designed a crossover study in which both real- and sham stimulation were delivered to each subject in a randomised order. High-definition-tACS (HD-tACS) set-up was used for the stimulation (DC Stimulator Plus, Neuroconn). More specifically, the HD 4×1 montaging was preferred over the common two-electrode setup, in order to increase the focality of the current flow delivered by the stimulation, on the preferred motor area (Dmochowski *et al.*, 2011). The 4×1 set up was accomplished using the equalizer extension box of the same company, which extended the bipolar setting into five round rubber electrodes of 20 mm diameter, with one electrode on the region to be stimulated and four electrodes in a square around it. Each electrode was placed at 7.5 cm from the central one (Sreeraj *et al.*, 2018). Following the instructions described in (Sreeraj *et al.*, 2018), the central electrode was placed on channel C3 (primary motor cortex – M1) and the four surrounding ones on Cz, F3, T7, and P3. Both real- and sham stimulation was delivered in two blocks, 15 min each (Strüber *et al.*, 2015).

For the real stimulation, a 70 Hz sinusoidal signal, 1 mA peak-to-peak amplitude was used. For the sham stimulation, a sinusoidal signal at 85 Hz with 1 uA peak-to-peak was selected.

Paradigm

Each participant attended two sessions, separated by a one-day break. The task was the same on both days: participants performed a visuomotor target-reaching task with their right arm, consisting of three blocks of trials. The subject was seated on a chair in the middle of four infrared motion tracking cameras (PhaseSpace), in front of a screen, wearing a customized glove with three LEDs in order to track on real time their arm position, which was depicted in real-time as a 3D sphere, as shown in Fig. 5.1. The motor task on each trial was similar to the one we described in section 4.3.3.

On the first session (day 1) only their brain activity was recorded with EEG during the task. Each block was separated by a 5-minutes resting-state period, during which the participant had been asked to focus on a white cross shown on the black screen in front of them, trying to relax. On the second session (day 2), either real or sham tACS was applied in a randomised order in a single-blind fashion, during the second and third block. The order of the application of the real-/sham stimulation was randomised across subjects to compensate for unknown factors such as learning effects or tiredness over time. In the second session, a 20-minute break was introduced between the second and the third block to avoid carry-over effects between the two blocks. Each block consisted of as many random reaching–trials as the subject could complete in 15 minutes.

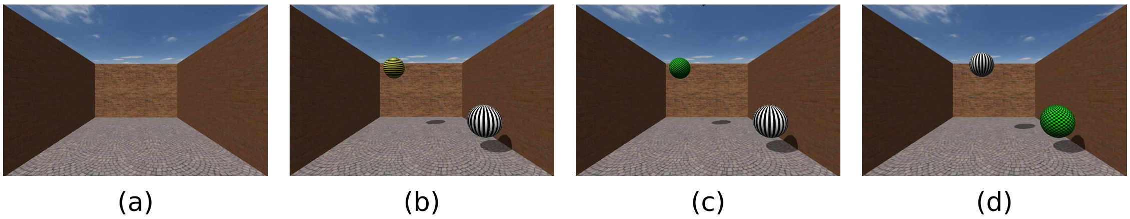


Figure 5.1: Paradigm: The white sphere represents the real-time position of the subject's right arm. Phases of a trial: a) Subjects wait for the next target. b) A yellow target appears at a random location. Subjects wait for the go signal, with their current hand position indicated by a white ball. c) A change of target colour to green instructs subjects to initiate the reaching movement. d) After the subject successfully reaches the target, a green sphere appears back to their starting arm position, indicating to return their hand there to complete the trial.

Experimental data

The experimental data consist of motion tracking data of the subjects' arm position, recorded with $f_s = 960$ Hz, and EEG data from high-density EEG (128 channels, $f_s = 500$ Hz, Brain Products GmbH).

5.3.2 Motor Response to γ-tACS

Here we focused on the analysis of the data recorded in the stimulation day (day 2). We further analyse the data of both days in Chapter 6. We choose the arm speed to quantify the behavioural response to tACS. We discuss the reasons for selecting this metric in Section 6.5.2. Therefore, for each trial of each subject, we calculate the mean arm velocity. The trials in which arm speed exceeded three standard deviations were excluded from the analysis as outliers.

5.4 Methods - Statistical analysis

5.4.1 Problem formalization

The variables of our problem are the following.

- S: the 70Hz tACS event
- $\Delta\beta_x$: the difference in beta log bandpower after and before the x block of stimulation, where x may denote *real stimulation block* or *sham block*.
- P: behavioural response (arm speed)
- h: a hidden non-measured variable

We assume that there cannot exist backwards arrows in time ($\not\leftarrow P$). In addition, because S is a randomised treatment, we can infer the direction $S \rightarrow \ldots$. Finally, we assume Causal Markov Condition (see definition 2) and Causal Faithfulness (see definition 5), in order to be able to relate probabilistic relationships with the causal graph. According to causal graph theory, if loops are not allowed, n vertices can result into $\binom{\binom{n}{2}}{n-1} * 2^{n-1}$ possible graphs with $n-1$ edges. For $n = 4 \Rightarrow 128$ possible graphs. To reduce the graphs further, we propose and perform **3 tests**, as described below in detail.

5.4.2 Division into responders and non-responders based on arm speed

Based on the behavioural response (arm speed) of each subject to tACS, the subjects were categorised into two groups. To do so, for each subject the null hypothesis that the average arm speed over the stimulation block is the same as the arm speed during the sham block was tested, with a permutation t-test. To build the null-distribution, we concatenated the arm speeds of the two blocks and permuted them 10000 times, calculating the average velocity of each of the two blocks after every permutation. The p-value was calculated as the frequency at which the absolute difference between the mean velocity during real- and sham stimulation was found larger than when drawing from the null-distribution (two-sided test). Setting a threshold $\alpha = 0.05$, we categorised subjects into two groups: *responders* if $p < \alpha$ and the average arm speed during the stimulation block was greater than during the sham block, and *non-responders* otherwise, i.e., subjects who did not show a significant increase or who exhibited a decrease in arm speed during the real tACS.

5.4.3 Effect size

We quantified the effect size of γ-tACS over contralateral M1 for each subject i as the difference between the average arm speed during the real and during the sham stimulation block, normalized by the standard deviation during the sham block.

$$\text{Effect size}(i) = \frac{\mu_{\text{stimulation}}(i) - \mu_{\text{sham}}(i)}{\sigma_{\text{sham}}(i)}, \tag{5.1}$$

where

$$\mu_{\text{stimulation}}(i) = \frac{\sum_{n_{\text{stim}}=1}^{N_{\text{trials stimulation}(i)}} |vel(i, n_{\text{stim}})|}{N_{\text{trials stimulation}}} \tag{5.2}$$

$$\mu_{\text{sham}}(i) = \frac{\sum_{n_{\text{sham}}=1}^{N_{\text{trials sham}}} |vel(i, n_{\text{sham}})|}{N_{\text{trials sham}}} \tag{5.3}$$

5.4.4 Effect of γ-tACS on β-power

To attenuate non-cortical artefacts in the EEG data we followed a common pre-processing step for EEG signal artefact removal (McMenamin *et al.*, 2010). The EEG signals of each subject's resting state were concatenated and then high-pass filtered with a Butterworth filter with cut-off frequency at 3Hz. Then, a common-average reference filter was applied, followed by SOBI Independent Component Analysis (ICA) and manual rejection of non-cortical components (McMenamin *et al.*, 2010). The reason why we set a cut off at 3 Hz, instead of lower, was because we were not interested in keeping information from lower than beta rhythms (such as delta).

To examine the effect of γ-stimulation on β-activity, and the relation between β-power and arm speed response, we calculated the log-bandpower of the 116 z-scored EEG channels for each subject (after having removed the channels used for stimulation) for the low- (12–20) Hz and high- (20–30) Hz β-range. For visualization purposes, we calculated the group averages of the difference between β-log-bandpower after and before the stimulation for the real- as well as for the sham tACS block.

5.4.5 Statistical analysis of β-power modulations

To test within each response group (*responders* and *non-responders*) whether the changes observed in β-power are statistically significant, we performed a permutation, paired, two-sided t-test with 10000 permutations: For each channel, we tested the null hypothesis that the neurophysiological changes in β-log-bandpower during the real stimulation come from the same distribution (across subjects) as those during sham. FDR-correction for multiple testing at significance level $\alpha = 0.05$ was applied (Benjamini and Hochberg, 1995). For further analysis, we kept only the channels that were found significant after the FDR correction.

For the kept channels, we performed two tests: first, a within-group statistical test (permutation mean test, one-sided) to examine the null hypothesis that the β-log-bandpower at these channels come from the same distribution before and after the stimulation (Fig.5.5). Second, we performed an across-group statistical test (permutation mean-test, one-sided), to examine the null hypothesis that for these channels the neurophysiological changes happening in β-range as a result of the tACS, come from the same distribution for the groups of “responders” and “non-responders” (fig. 5.6).

5.4.6 Causal-effect relationship between β-power and arm speed

At this point, making some necessary assumptions, we wanted to test if and which channels of each group exhibit a causal relationship between the recorded neurophysiological changes in β-power and the observed behavioural response [1]. To do so, we tested three different pairwise cause-effect tests: 1. IGCI assuming deterministic relationships Daniusis *et al.* (2012), 2. Additive noise models for non-linear causal relationships Hoyer *et al.* (2009), and 3. Pairwise LINGAM Hyvärinen and Smith (2013) assuming that cause and effect are non-Gaussian. More specifically, we examined which channels express the causal relationship

$$\Delta\beta_{\mathrm{BP}_{\mathrm{stimulation}}} \rightarrow \text{Effect size}$$

where

$$\Delta\beta_{\mathrm{BP}_{\mathrm{stimulation}}} = \beta_{\mathrm{BP}_{\mathrm{after\ stimulation}}} - \beta_{\mathrm{BP}_{\mathrm{before\ stimulation}}}{}^{2}$$

For the IGCI test, we assumed a deterministic relationship between the beta activity and the motor performance, although most likely this assumption is violated, and as such we should be careful on how much we trust the outcome of the specific test. The IGCI inference algorithm is based on the assumption that if $X \rightarrow Y$, then the distribution of X and the function f that maps X to Y are independent, i.e., it assumes that the mechanism and the data that it processes are not co-adapted. It is not a very unrealistic assumption

[1] Note that in case that exists the h node shown in Fig. 5.9, if h is GABA, then we can exclude the possibility that the relationship between h and β is deterministic, because from the litterature we know that there is a correlation coefficient of 0.6-0.7 between them.

[2] β_{BP} is an abbreviation we use for beta log band power.

to make that the mechanism that produces the beta brain activity is independent of the mechanism that uses part of this frequency range for motor purposes. In the real noisy dataset we have here, we should be very careful with the interpretation of the results from the IGCI method, as it was shown in Mooij *et al.* (2016) that the method had poorer performance in the real data benchmarking. To quantify the independence between the function f and the distribution of X, the relative entropy distance $D(.,.)$ is used. Then the following difference is calculated:

$$C_{X\to Y} = D(p_X||E_X) - D(p_Y||E_Y).$$

Based on the sign of $C_{X\to Y}$, the IGCI algorithm decides which causal direction is more likely. If $C_{X\to Y} > 0$, then X is inferred as the cause of Y.

The additive noise models-based (ANM) test and the pairwise LINGAM test are explained in detail in Sections 2.3.3 and 2.3.3. As explained there, for the ANM we had to assume independent Gaussian noise, and for the LINGAM that the cause and effect variables are non-Gaussian.

5.5 Results

5.5.1 Motor response to γ-tACS

Out of the original population of 19 subjects, only six responded significantly positively to γ-tACS over contralateral M1. Figure 5.2 shows the effect size of each subject, with the colour indicating whether the subject was a responder or a non-responder. Although contralateral M1 γ-tACS has been proposed as a stimulation setup that facilitates movement, our findings exhibit a quite small overall effect size (0.2366). This small effect size can be justified by the co-existence of responders and non-responders with effect sizes 0.9073 and -0.1547, respectively. Based on these observations, we next examined the effect of γ-tACS on β-power for each of the two groups individually.

Table 5.1 shows the p-value and the difference in mean velocity between the blocks of stimulation and sham, which were used for categorizing the participants into the groups of responders and non-responders.

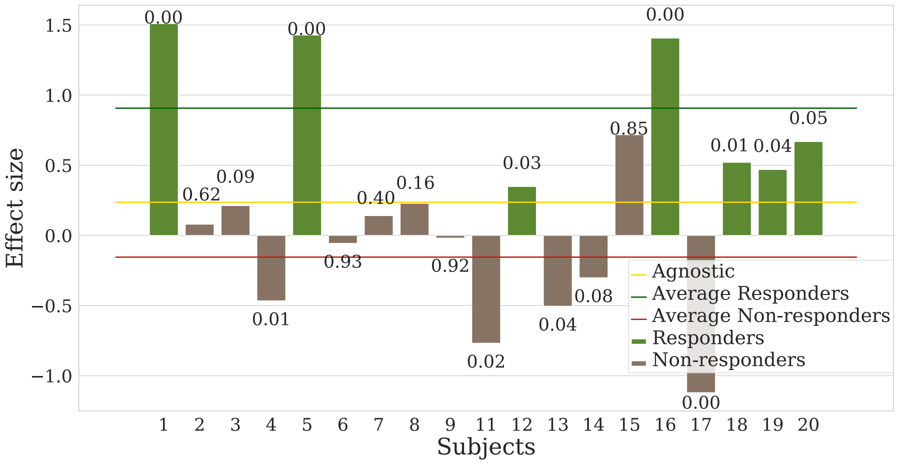


Figure 5.2: Effect size of real 70Hz tACS vs sham for the 19 recruited subjects, as measured by changes in movement velocity. The p-value (rounded up to two decimals) for each subject is shown on top of each bar (cf. 5.4.2). Green bars: Subjects that performed significantly better during stimulation (responders). Brown bars: Subjects who either did not respond to the tACS, or who performed significantly worse compared to sham (non-responders). Green line: Average effect size of responders. Red line: Average effect size of non-responders. Yellow line: Overall effect size of the whole population.

Table 5.1: Division of subjects into “responders” and “non-responders”

Subject	P-value	$\Delta Velocity > 0$	Response
1	0.0005	1	Responder
2	0.6217	1	Non-Responder
3	0.0902	1	Non-Responder
4	0.0121	0	Non-Responder
5	0.0000	1	Responder
6	0.9318	1	Non-Responder
7	0.3979	1	Non-Responder
8	0.1628	1	Non-Responder
9	0.9220	0	Non-Responder
11	0.0154	0	Non-Responder
12	0.0251	1	Responder
13	0.0384	0	Non-Responder
14	0.0800	0	Non-Responder
15	0.8484	1	Non-Responder
16	0.0000	1	Responder
17	0.0000	0	Non-Responder
18	0.0077	1	Responder
19	0.0370	1	Responder
20	0.0455	1	Responder

5.5.2 Effect of γ-tACS on β-power

Our hypothesis was that subjects who exhibit a larger decrease in β-log-bandpower over the contralateral motor cortex are those that respond positively to the stimulation, i.e., with faster movements. In Figure 5.3, we see that indeed the group of subjects that responded positively to the γ-stimulation, exhibited a larger decrease of β-power, mostly in the high β-range [20 30] Hz and spread out over the contralateral motor cortex, compared to the group of non-responders.

Moreover, for the sham condition in the group of responders, lower as well as non-significant change of β-log-bandpower over contralateral motor cortex was measured. In contrast, as we can see from Figure 5.3, the group of non-responders exhibited a bilateral increase of β-power.

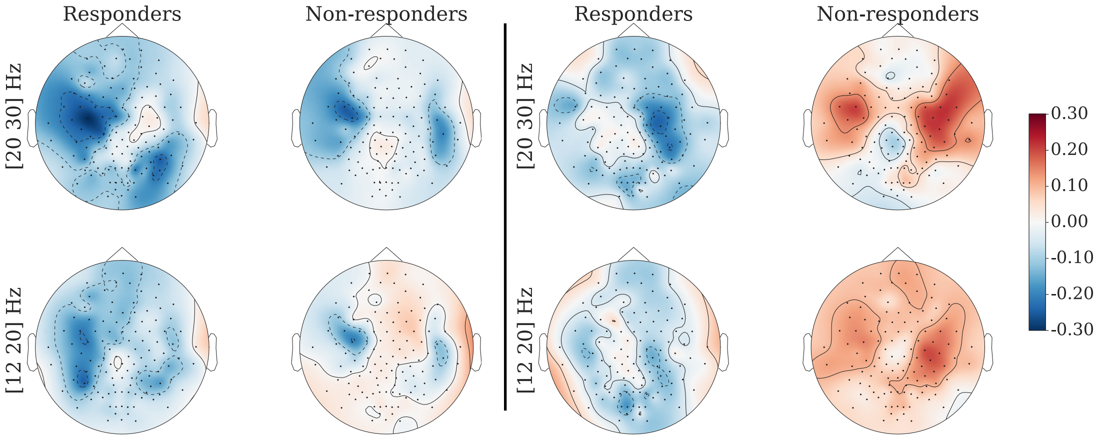


Figure 5.3: Group-average difference in β-power levels between after and before the real (left) and sham (right) block, in the low- (12–20) Hz and high- (20–30) Hz β-range, for the groups of responders and non-responders.

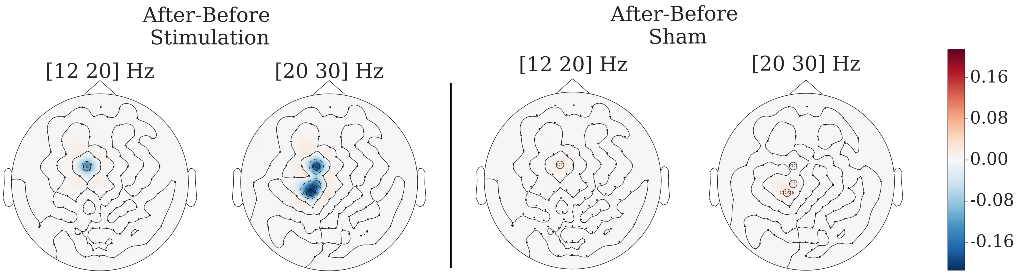


Figure 5.4: Difference of β-power after and before stimulation for the real- and sham condition, in the low (12–20 Hz) and high- (20–30 Hz) β-range, for the group of responders, for the channels that were found to exhibit significant difference between the two conditions (sham and real). FDR-corrected channels, that do not exhibit significant neurophysiological differences, are set to zero.

Figure 5.4 depicts the channels that exhibit FDR-corrected significant difference in β-power, between the conditions of real- and sham stimulation. For the responders group, these channels are found to be located over the contralateral motor cortex, ‘FC1’, ‘C1’ and ‘CCP3h’ for the high β-range and ‘FC1’ for the low β-range. For the non-responders, in contrast, we found no channel with a significant difference between the two conditions.

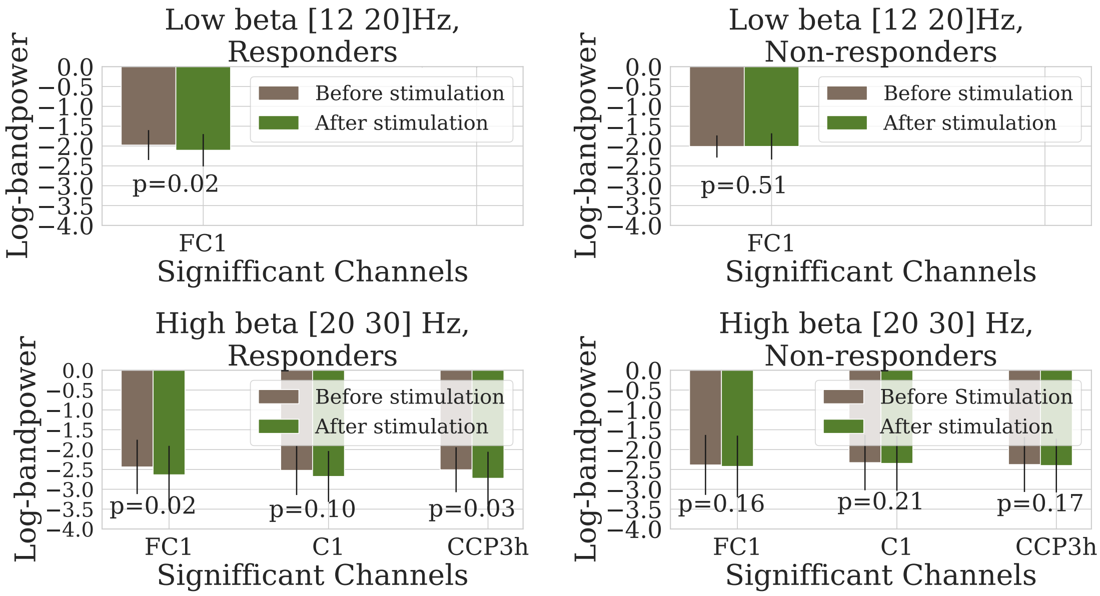


Figure 5.5: β activity after versus before stimulation, in low [12 20]Hz and high [20 30]Hz β range for the significant channels above. Left: for the group of “responders”, and right: for the group of “non-responders”.

For the channels found to be significant above, we examine for the groups of “responders” and “non-responders” whether they also exhibit significant neurophysiological changes because of the real stimulation. As shown in fig. 5.5, in the group of “responders”, β log-bandpower has significantly decreased after stimulation for the channels ’FC1’ ($p = 0.02$), ’CCP3h’ ($p = 0.03$) for the high β range, and ’FC1’ ($p = 0.02$) for low β. In contrast, in the group of “non-responders”, no significant decrease of β log-bandpower is observed ($p = 0.16$, $p = 0.21$, $p = 0.17$ for high β, $p = 0.51$ for low β).

5.5.3 Across-groups difference in neurophysiological changes on low and high β band during 70 Hz tACS

Finally, for the same channels, we examine whether the neurophysiological changes in the β power $\Delta\beta_{\mathrm{BP}_{\mathrm{stimulation}}}$ due to tACS, are significantly different between the two groups. High β range was found to be significantly lower in the group of “responders” compared against the groups of “non-responders’, for the channels of left motor cortex (’FC1’, $p = 0.03$ and ’CCP3h’, $p = 0.03$). No significant difference between the two groups was found for the low β range ($p = 0.08$).

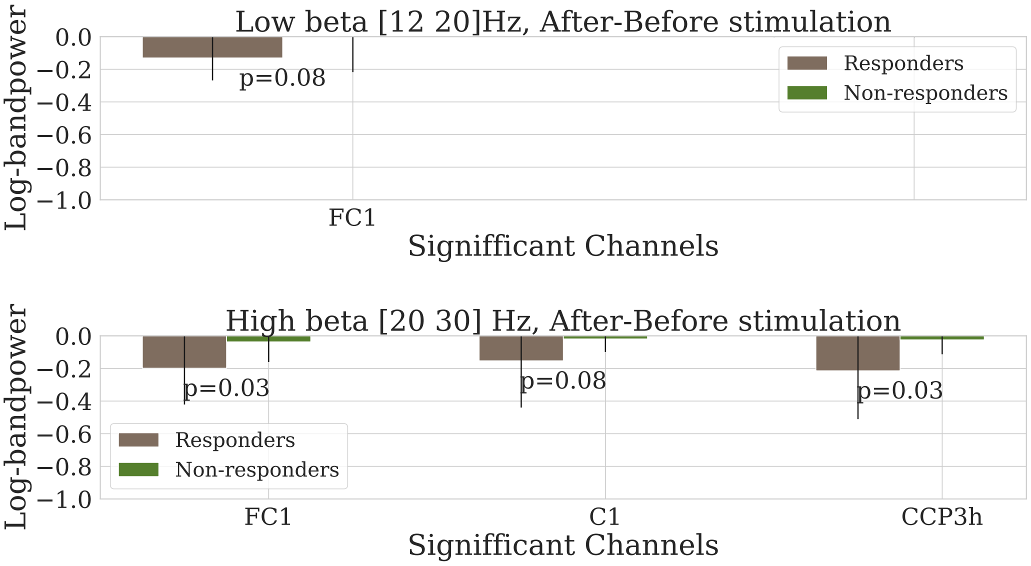


Figure 5.6: Neurophysiological changes for the groups of “responders” vs “non-responders”, in low (12 20) Hz and high (20 30) Hz β range, for the significant channels found above. For each channel, the significance of the difference between the two groups is depicted with the p-value of the corresponding test.

5.5.4 Identification of channels that exhibit the cause-effect relationship $\Delta\beta_{\mathrm{BP}_{\mathrm{stimulation}}} \rightarrow$ Effect size

At this point, we wanted to examine which electrodes in each group exhibit the causal relationship $\Delta\beta_{\mathrm{BP}_{\mathrm{stimulation}}} \rightarrow$ Effect size. To examine this, we tested three different pairwise cause-effect tests: 1. IGCI assuming deterministic relationships Daniusis *et al.* (2012), 2. Additive noise models for non-linear causal relationships Hoyer *et al.* (2009), and 3. Pairwise LINGAM Hyvärinen and Smith (2013). From the above methods, the additive noise model did not conclude to a single direction of the causal relationship. IGCI found that in the group of non-responders fewer channels satisfied the above relationship compared to the non-responders. The channels that did not satisfy the condition were set to zero for visualization purposes. Figure 5.7 shows the channels that were found to exhibit the above relationship as measured with the IGCI test. The channels are shown colour-coded according to the difference between high β-log-bandpower after and before stimulation. We observe that for the group of responders the left motor cortex exhibits in its majority the above causal relationship. In contrast, for the group of non-responders the majority of the EEG channels were not identified to satisfy the aforementioned relationship by the IGCI test.

Finally, pairwise LINGAM, which does not assume deterministic relationships - which is more likely to not be the case in such noisy environments- found very similar results to the IGCI. Figure 5.8 depicts that more electrodes both in the high- and low- beta range were found to exhibit this causal relationship in the group of responders, compared with the group of non-responders.

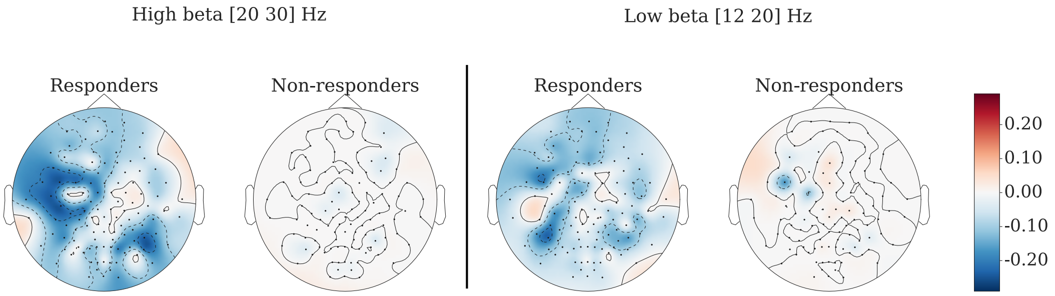


Figure 5.7: Difference of β-power after and before stimulation, in the low- (left) and high (right) β-range. The channels that were not found to satisfy the causal relationship $\Delta\beta_{\mathrm{BP}_{\mathrm{stimulation}}} \rightarrow$ Effect size by the IGCI test were set to zero.

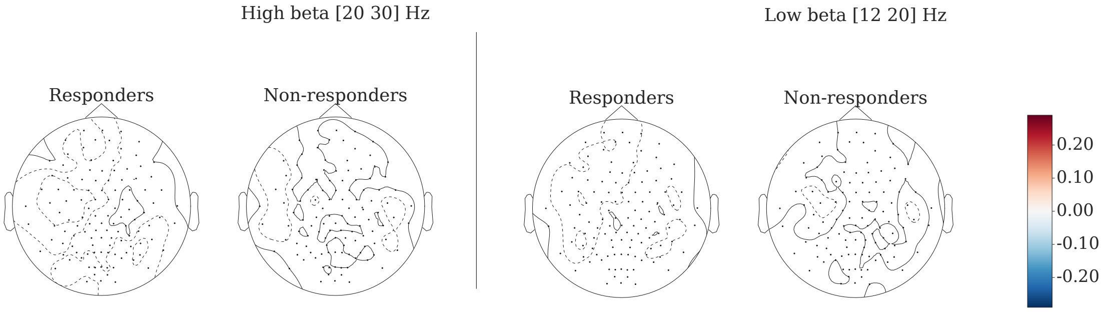


Figure 5.8: Difference of β-power after and before stimulation, in the low- (left) and high (right) β-range. The channels that were not found to satisfy the causal relationship $\Delta\beta_{\mathrm{BP}_{\mathrm{stimulation}}} \rightarrow$ Effect size by the pairwise LINGAM were set to zero.

IGCI and LINGAM tests, although they do not eliminate the possibility of hidden confounders and rely on some strong assumptions, show that contralateral motor cortex exhibits a path between $\Delta\beta_{\mathrm{BP}_{\mathrm{stimulation}}}$ and P.

The above statistical tests indicate two things: first, that there is a path which connects the S and P, and, second, that exists a path that connects $\Delta\beta_{\mathrm{BP}_{\mathrm{stimulation}}}$ and P. Note, these paths do not imply direct cause and it is possible to contain other h variables.

In combination with the above findings, the two possible graphs that describe the set of our variables are shown in fig. 5.9.

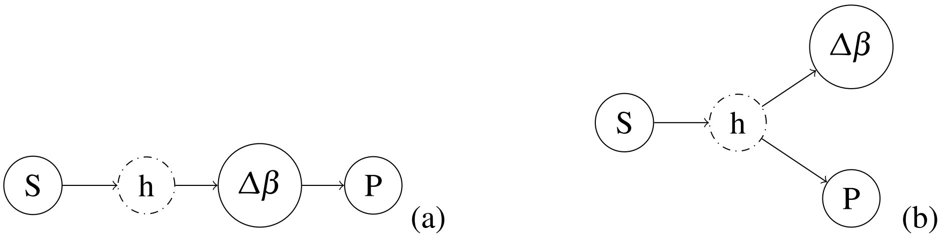


Figure 5.9: The two possible DAGs that describe the set of our variables, after the elimination of the rest with the statistical tests and the IGCI test. Hidden variable h may or may not exist in the paths.

Figure 5.9 depicts the two possible DAGs that describe the variables in our system. In graph (a), $\Delta\beta$ is a mediator of γ tACS to motor performance. In graph (b), γ tACS is a common parent of $\Delta\beta$ and motor performance, meaning that $\Delta\beta$ and P are dependent only because of their common parent and not otherwise. In the simple case, where there is no hidden variable in the path, the way to eliminate the possibility of graph (b) is by examining if conditioning on the common parents makes $\Delta\beta$ and motor performance P independent. However, in order to be confident that the (b) possibility can be excluded, the above should be explicitly tested from a causal inference point of view, in a future work. It has been shown (McAllister *et al.*, 2013) that there is a close dependency of beta activity and motor performance –therefore between $\Delta\beta$ and P–, even without stimulation. In the next Section we discuss why both models (a) and (b) could be supported by the neuroscientific literature if the role of the node h was played by the GABAergic activity. At this point, we can only be cautiously optimistic that (b) may not be the case, based on the extended literature about

5.6 Discussion

5.6.1 Background of hypothesis

Application of 70 Hz HD-tACS over the contralateral motor cortex on 19 healthy subjects, led to increase of arm speed in 36% of the original population. Our findings of a large number of non-responders was in line with the results in (López-Alonso *et al.*, 2014). Considering that γ-stimulation is believed to facilitate movement (Joundi *et al.*, 2012; López-Alonso *et al.*, 2014), as well as that an increase of γ-oscillatory activity has been associated with large ballistic movements (Muthukumaraswamy, 2010), we decided to investigate the underlying modulation of the antikinetic β-oscillations (Brinkman *et al.*, 2014) as a potential cause of this variability. We hypothesised that if γ-stimulation is affecting the ongoing β-activity, then the subjects that exhibit a larger decrease in β-power should be those that respond positively to the stimulation. The findings from our EEG analysis support this hypothesis.

5.6.2 Findings

Our findings provide evidence for a potential role of β-power as a mediator of γ-tACS on motor performance. In particular, the results reported in Section 5.5.1 establish that γ-tACS (S) has an effect on movement (P), in terms of arm speed. Because S was delivered in a randomized order (sham/real), it can be considered as a randomised treatment. Therefore, we can infer the direction of this relation as $S \rightarrow P$. The results in Section 5.5.2, on the other hand, demonstrate an effect of γ-tACS on β-power, i.e., a causal path $S \rightarrow \Delta\beta$. It then remains to distinguish between the two causal models $S \rightarrow \Delta\beta \rightarrow P$ (with potentially an additional path $S \rightarrow P$ that does not pass through $\Delta\beta$) and $\Delta\beta \leftarrow S \rightarrow P$, both of which are consistent with our evidence to this point. The results of the IGCI and LiNGAM in Section 5.5.4 indicate that in the stimulation condition $\Delta\beta \rightarrow P$, which is consistent with the former and not with the latter causal model. Nevertheless, the result of these two tests should be treated with caution as the first assumes deterministic relationships and the second causal sufficiency and non-Gaussian variables. Assuming we can trust the last two test, we argue that our empirical results are in favour of the causal model $S \rightarrow \Delta\beta \rightarrow P$, i.e., that β-power may mediate the effect of γ-tACS on motor performance. We stress, however, that the analysis presented here can not prove but only provide empirical results consistent with causal relationships.

5.6.3 Plausible neurophysiological explanation

Here we discuss why, even though we cannot conclude to a single graphical model, both (a) and (b) models are neurophysiologically plausible. In the context of neurophysiological mechanisms underlying the effect of γ-tACS on β-power and on the observed motor behaviour, a possible explanatory factor could be the modulation of γ-aminobutyric acid (GABA) concentration. We support this claim with the following argument: First, β-oscillations have been shown to be the summed output of principal cells temporally aligned by GABAergic interneuron rhythmicity (Yamawaki *et al.*, 2008). Specifically, GABA levels have been found to strongly correlate with β-power and to exhibit elevated values in bradykinesia and in Parkinson's disease (McAllister *et al.*, 2013). Secondly, high-γ deep brain stimulation in motor cortex has been reported to cause a significant decrease in β-power (Gulberti *et al.*, 2015), supporting our finding of the inhibitory effect of γ-stimulation on the ongoing β-oscillations. Combining these two pieces of knowledge from the literature, we argue that the behavioural response to γ-tACS may be explained by a decrease of the GABA levels modulated by the stimulation and hence of β-power. Therefore, it is conceivable that whenever γ-tACS leads to the inhibition of human movements, this may be caused by an increase in GABAergic drive, which hinders the decrease of β-power.

Chapter 6

High gamma activity over motor cortices for screening gamma tACS response

The content of this Chapter deals with the detection of a brain-biomarker that could be used for an early screening of non-responders to motor cortex stimulation treatments. As it has become obvious from the previous chapters, the problem of the large discrepancy in response to tACS over motor cortex is prominent. Although the attributes of tACS have rendered it into a widely used technique in cognitive neuroscience, the considerable heterogeneity of its behavioural effects hinders its conversion into a safe treatment tool.

In this Chapter, we present a machine learning pipeline for predicting the behavioural response to 70 Hz contralateral motor cortex-tACS from EEG activity, recorded during a resting period preceding the stimulation. Using the EEG/tACS data from our crossover study on 20 healthy participants in Chapter 5, we show that high-gamma (90–160 Hz) resting-state activity predicts arm-speed response to the stimulation in a concurrent reaching task. Moreover, we perform another EEG/tACS crossover stimulation study with 22 new healthy participants on whom we validate our screening tool. Finally, a plausible neurophysiological mechanism is presented and discussed, that links high resting-state gamma power in motor areas to the response to stimulation. Therefore, a method is proposed that can distinguish responders from non-responders to tACS, prior to the stimulation treatment. This contribution could help to render tACS a safe and effective clinical treatment tool. The work presented in this Chapter is part of the author's manuscript (Mastakouri, 2020).

6.1 Problem statement

Having encountered significant variability in the behavioural response to motor- cortex tACS on our previous experiment, we want to learn a model that can screen off the non-responders. Therefore, the question we try to tackle in this Chapter is the following: Given resting-state EEG activity preceding and following the real and sham stimulation blocks (see limitation imposed by the nature of the stimulation signal in subsection 5.1.1), is it possible to identify a biomarker that can screen off in advance negative-responders and non-responders to the treatment? (see Problem 3) If so, does this biomarker generalize properly on newly recruited subjects?

6.2 Motivation

As we have extensively presented in Section 3.4.2, although the neural mechanisms of NIBS are not yet fully understood (Vosskuhl *et al.*, 2018), NIBS applications spread in research and treatment in the fields of neurophysiology (Shafi *et al.*, 2012; Polanía *et al.*, 2012b; Filmer *et al.*, 2014; Sehm *et al.*, 2012; Keeser *et al.*, 2011; Hampson and Hoffman, 2010; Anand and Hotson, 2002), behavioural and cognitive neuroscience (Polania *et al.*, 2018; Vosskuhl *et al.*, 2018; Pogosyan *et al.*, 2009; Joundi *et al.*, 2012; Neuling *et al.*, 2012; Cecere *et al.*, 2015; Lustenberger *et al.*, 2015; Vosskuhl *et al.*, 2015; Polanía *et al.*, 2012a; Santarnecchi *et al.*, 2013; Sela *et al.*, 2012), and the field of rehabilitation (Hashemirad *et al.*, 2016; Pogosyan *et al.*, 2009; Joundi *et al.*, 2012; Moliadze *et al.*, 2010; Triccas *et al.*, 2016; López-Alonso *et al.*, 2014; Strube *et al.*, 2015).

In all the aforementioned categories, NIBS studies report substantial variations in stimulation response across individuals (López-Alonso *et al.*, 2014; Strube *et al.*, 2015). While non-responders decrease the statistical power of NIBS studies, this sub-group is unproblematic from an ethical point of view. Given the reported variances in effect sizes, however, it is not unreasonable to assume that there also exist subjects with negative stimulation responses (Hashemirad *et al.*, 2016; Moliadze *et al.*, 2010; Triccas *et al.*, 2016; López-Alonso *et al.*, 2014; Strube *et al.*, 2015). Subjects with negative stimulation responses would be highly problematic, because their existence would imply that NIBS studies may violate the principle of doing no harm. This ethical concern is particularly relevant in clinical settings, where NIBS is used to cause possibly permanent changes (Di Pino *et al.*, 2014; Cirillo *et al.*, 2017). In related work, Yang *et al.* (2020) and Kasten *et al.* (2019) emphasize the importance of pre-stimulation screening and of individualizing stimulation protocols. In addition, Kong *et al.* (2019) has reported the existence of individual-specific cortical networks and their importance in the prediction of human cognition. These reports emphasize even more that serious consideration of potential negative stimulation effects in individual subjects is required, to assure the ethical use of NIBS. Finally, if such adverse effects are noticed, an implementation of a screening procedure that reliably identifies negative responders before the stimulation treatment could be the first step towards the personalisation of NIBS treatments.

While there is a large body of literature on negative side-effects (Kadosh *et al.*, 2012; Davis and Koningsbruggen, 2013; Matsumoto and Ugawa, 2017; Yang *et al.*, 2020), NIBS studies typically only report variability across subjects in terms of positive and non-responders, without focusing on the negative responders. This is a strong indication of the limited understanding of the causes of inter-subject variability in NIBS. Factors that have been studied as potential causes of the encountered inter-subject variability are discussed in detail in Section 1.1. Nevertheless, up to now, there is no known set of factors that could enable a reliable screening of subjects, who do not respond to or may even be harmed by NIBS (Ridding and Ziemann, 2010).

6.3 Methods - EEG/tACS dataset acquisition

The study conformed to the Declaration of Helsinki, and the experimental procedures involving human subjects described in this Chapter were approved by the Ethics Committee of the Medical Faculty of the Eberhard Karls University of Tübingen. Informed consent was obtained by all participants, prior to their participation in the study.

6.3.1 Experimental paradigm and data

The experimental paradigm is the same 3D-reaching task, as described in Chapter 5.3.1. Nevertheless, for clarity here, we briefly describe it one more time, with more detailed diagrams (see Figure 6.1 and 6.2) for each session. Each participant attended two sessions, to which we also refer as days, separated by a one-day break. On each day, participants were seated on a chair in the middle of four infrared motion tracking cameras (Phase Space, San Leandro, California, USA), 1.5 meters away from a visual feedback screen (35″), while wearing a customized glove with three LEDs on its top for real-time tracking of their arm location. The position of the arm was depicted on the screen in real-time as a 3D sphere (cf. Figure 5.1).

In the beginning of each trial, a target sphere appeared at a random location in the simulated 3D space depicted on the screen. The subject was instructed to move their arm in order to reach and overlap with the target. Each trial started with a baseline of 5 s, followed by 2.5–4 s preparation period during which the target appeared on the screen as a yellow sphere. During this period, subjects had been instructed to plan but not yet initiate their movement. From the moment the target sphere turned to green, the subjects had 10 s to reach the target. After a successful reach, a score indicating the movement's quality, appeared for 2 s on the screen. This score was computed as an inverse mapping of their movement's normalized averaged rectified jerk score to a scale from 0 to 100 (Cozens and Bhakta, 2003; Meyer *et al.*, 2014), as explained in 4.3.4. Finally, in the last part of each trial, the target sphere appeared at the original starting position of the subjects' wrist. The trial was considered completed when subjects returned their wrist to the original starting position. If the accidentally subject moved before the sphere turned green, or if the target was not reached within 10 s, the trial was excluded from further analysis.

Session 1

During the first day of the experiment, only electroencephalographic (EEG) (124 active electrodes at 500 Hz sampling rate, BrainAmp DC, Brain Products, Gilching, Germany) and motion tracking (sampling rate of 960 Hz) data were recorded during the reaching task. The detailed phases of this session are depicted in Figure 6.1, showing the three blocks of 50 trials that consisted the first day. Before and after each block, 5 min of resting-state EEG was recorded, during which subjects were asked to relax, focus their eyes on a cross on the screen, and keep their arm in a comfortable position on top of their leg without moving.

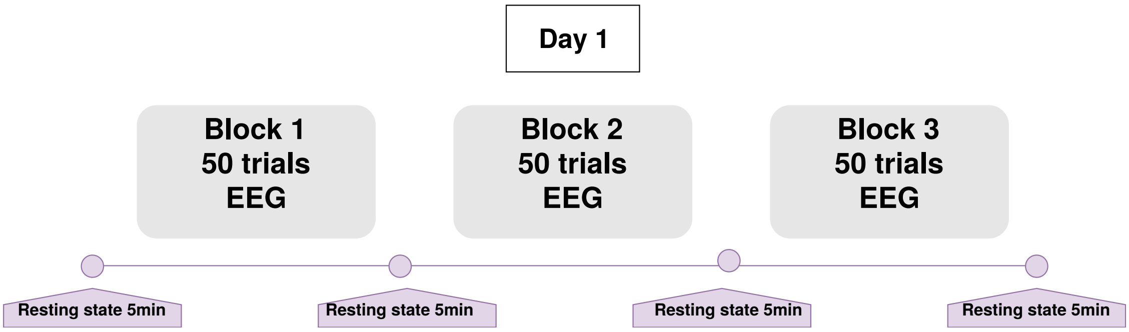


Figure 6.1: Experimental setup for the first recording session (no stimulation).

Session 2

The second recording day consisted of three blocks of reaching trials of 15 min each (cf. Figure 6.2). During the first block, EEG and motion tracking data were recorded, but subjects were not yet stimulated. During the second and third block real and sham high-definition (HD) transcranial alternating current stimulation (tACS) was applied in a single-bind randomized order. Between the two stimulation blocks, a break of 20 min was introduced in order to avoid carry-over effects. Each trial was similar to the ones described in Session 1. At the end of the session, the participant completed a questionnaire to evaluate the sensation of the stimulation and potential side effects. Before and after each block, a 5 min resting-state EEG was recorded, as described in Session 1.

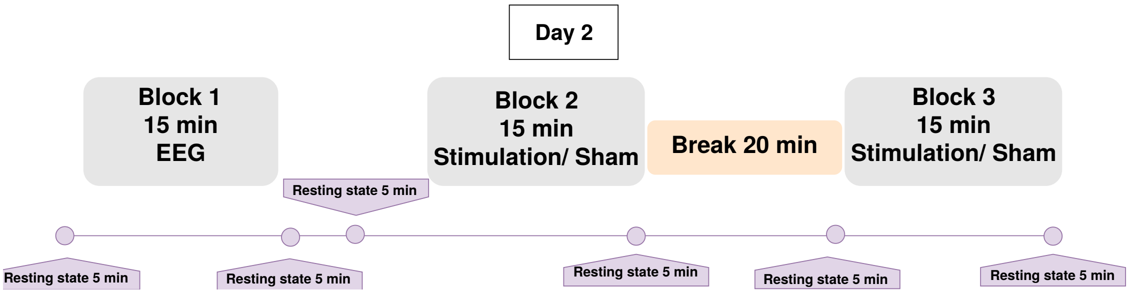


Figure 6.2: Experimental setup for the second recording session (with stimulation).

Stimulation setup

The stimulation parameters were the same as the ones we described in Chapter 5.3.1.

Subjects

The recordings/stimulations for the first part of this study (experiments of the first three months) were performed in Chapter 5 and the characteristics of the twenty recruited subjects are presented in Section 5.3.1. In the second part of the study (experiments of the last three months) twenty-two new subjects were recruited and stimulated following the exact same protocol. The second group of participants was only used for validation of our screening procedure, as described in detail in subsection 6.3.5].

6.3.2 Analysis of behavioural data

To quantify the behavioural effect size of γ-tACS over contralateral M1, we calculate the difference between the average reaching speed during the real and the sham stimulation block (see equation (5.1)). The reason for the selection of 70 Hz for the real stimulation was the fact that γ-tACS over motor cortex has been reported to influence the movement velocity (Joundi *et al.*, 2012; López-Alonso *et al.*, 2014; Muthukumaraswamy, 2010). Regarding our behavioural metric, we decided to focus on the arm velocity, as it has been considered one of the most stable variables related to the motor cortex activity (Moran and Schwartz, 1999; Hatsopoulos and Suminski, 2011).

To compute the effect sizes on the level of individual subjects, we first computed, for every trial and subject, the trial-averaged reaching speed. This was done by extracting the x, y and z coordinates of the subject's arm, from the frames of the camera during the moving period of the trial (from the "Go" phase until the reaching of the target), and then calculating the mean velocity as the amplitude of the discrete positional derivative (Meyer *et al.*, 2014). For each subject, we computed the block-average arm velocities by averaging the trial-averaged velocities within the real and the sham stimulation block. If a trial-averaged velocity deviated from the block-averaged velocity by more than three standard deviations, the trial was excluded as an outlier. Finally, to obtain the subject-level behavioural effect sizes, we computed the difference between the block-average velocities of the real and the sham stimulation blocks and normalized the difference by the standard deviation of each subject's sham stimulation block. To compute the group-level behavioural effect size, we averaged the subject-level effect sizes and normalized by their standard deviation.

To test for a statistically significant behavioural stimulation effect on the level of individual subjects, we performed, for each subject, a two-sided t-test on the trial-averaged arm velocities of the real and the sham stimulation block. To built the null-distribution, we randomly permuted 10.000 times the assignment of trials to the real and sham stimulation block. After every permutation, we re-computed the subject's average arm speed difference between the real and the sham stimulation block. The p-value was calculated as the frequency at which samples from the null-distribution exceeded the original absolute average speed difference between the real and the sham stimulation block. Subjects with $p < 0.05$ and larger average speed during the real compared to sham block were subsequently termed *responders*. The subjects with non-significant p-values and those with a significant reduction of arm speed combined were termed *non-responders*.

Finally, to test for a statistically significant behavioural stimulation effect on the group-level, we performed a two-sided, paired t-test on the single-subject effect sizes. The null-distribution was built by randomly flipping 10.000 times every subject's block-average velocities between the real and sham blocks. After every random flipping, we re-computed the group-level behavioural effect size as described above. The p-value was calculated as the frequency at which samples from the null-distribution exceeded the original absolute group-level effect size.

6.3.3 Analysis of EEG data

As a typical pre-processing step in EEG analysis, first, each subject's EEG data were "cleaned" from non-cortical artefacts using Independent Component Analysis (ICA) (McMenamin *et al.*, 2010). For each subject and session, we concatenated the raw data of all resting-state recordings (the stimulation blocks were removed as the amplifier was saturated by the stimulation current), high-pass filtered the data with a Butterworth filter at 3 Hz, and re-referenced the data to common-average reference. We then used the SOBI algorithm (Belouchrani *et al.*, 1993) to extract 64 independent components (ICs). The extracted 64 IC topographies were manually inspected, discarding those ICs that did not exhibit a cortical topography according to the rules by (McMenamin *et al.*, 2010). The remaining cortical ICs were re-projected to the scalp level, and the individual resting-state recordings were reconstructed. For each subject, resting-state, and electrode, the data were normalized by z-scoring. The reason for the few kept ICs (between four and eighteen) was that during the manual artefact cleaning we were particularly cautious to exclude any component that could have the slightest contamination. Therefore, we were conservative by keeping only IC components that looked cortical, rejecting all ambiguous ones.

After this step, the logarithmic bandpower of each of the following canonical EEG frequency bands (θ (4–8 Hz), α (8–12 Hz), β (12–25 Hz), γ_1 (25 –45 Hz), γ_2 (45 –65 Hz), γ_3 (65 –90 Hz), and γ_4 (90–160 Hz)) was calculated for the resting-state periods. For the bandpower calculation, a Hamming window was applied on the data, computing the Discrete Fourier Transform, taking the average of the absolute values of all frequency components within each of the eight frequency bands, and finally taking the natural logarithm.

6.3.4 Training of the stimulation response predictor

The development of our screening pipeline was based on the dataset that was collected from the first 19 healthy participants. A linear discriminant analysis (LDA) classifier was trained to predict subjects' response group (*responder* vs *non-responders*) from their resting-state EEG. Due to the small sample size, only three channels were selected as input features. The position of those channels was selected, such they are on top of key brain regions. Therefore, two channels were selected over left (CCP3h) and right motor cortex (CCP4h) and one channel over parietal cortex (Pz) (channel C3 directly over left motor cortex was blocked by the stimulation electrodes, hence for symmetry we did not use C4 as well). For each of the eight frequency bands (cf. Section 6.3.3) and the three resting-state EEG recordings preceding the stimulation on the second day, the prediction accuracy of the classifier was evaluated by leave-one-subject-out cross-validation. The statistical significance of each of the 24 settings (eight frequency bands times three resting-states) was tested by a permutation test with 1000 permutations. To build the null-distribution, we randomly permuted the labels on the training set of each cross-validation fold, retrained the classifier, and classified the subject in the test set. The p-value was calculated as the frequency at which samples from the null-distribution exceeded the original prediction accuracy. Then, the best performing combination of frequency-band and resting-state was selected in order to train the final stimulation response predictor (SRP) on all 19 subjects.

6.3.5 Validation of the stimulation response predictor

To validate the SRP, 22 new subjects (eleven female, average age of 26.81 years with a standard deviation of 6.32 years) were recruited. The same EEG processing pipeline as for the first group of subjects (cf. Section 6.3.3) was employed, except that the manual artefact cleaning was only performed on the EEG data of the first session (recorded on day one). To clean the EEG of the stimulation session (recorded on day two) from non-cortical artefacts, we applied the spatial filters derived on day one to the EEG data of day two, and re-projected only those ICs that corresponded to cortical sources on day one. This minimal-intervention pre-processing on the data that would be used for the validation of SRP was chosen in order to minimize the probability that any manual selection of ICs on day two confounds the predictions of the SRP. Then, the trained SRP, as described in Section 6.3.4, was applied to the first resting-state recording of every subject in the validation group. The predicted classes were then compared against the true response group derived from the behavioural analysis described in Section 6.3.2. The true response group (*responders* and *non-responders*) of the subjects from the validation recording period is shown in Table 6.2.

To test for a statistically significant difference in the behavioural stimulation effect between the predicted responder and non-responder group, a one-sided permutation-based t-test was employed: The predicted assignments of subjects to the responders- and non-responders groups were randomly permuted 10.000 times. After every permutation, the group-level effect size within each group (responders vs non-responders) was re-computed, as described in Section 6.3.2. Finally, the p-value was calculated as the frequency at which the permuted difference in effect sizes exceeded the original one.

6.3.6 Association of stimulation response with external factors

To examine the possibility that our findings are confounded by some external factor other than the brain activity, we tested the stimulation response of individual subjects (*responders* vs *non-responders*, cf. Section 6.3.2) for associations with four external factors: gender, order of the sham and real stimulation block, block of the strongest reported sensation reported by the subjects, and baseline motor performance during the first session. For all analyses, we pooled all 41 subjects from the training- and the validation group.

To test for an association of the stimulation response with sex (female, male) and with the order of the stimulation block, respectively, we used Fisher's exact test. To test for an association of the stimulation response with the block of the strongest reported sensory sensation, a Chi-squared test was used. Finally, to test for an association between the average reaching speed during the first session with subsequent stimulation response in the second session, we performed an ANOVA test for the average movement speed across all three blocks on day one. For the tests, a significance level of $\alpha = 0.05$ was chosen.

6.3.7 Signal to noise ratio of EEG in high gamma range

As high gamma frequency has been considered a heavily noisy brain band, we examined the possibility that our findings in the high gamma range are coincidental or mostly noise. To exclude this scenario, we performed the following analysis: To make sure that we do not introduce any bias during the manual artefact correction process, we tested the statistical association between the number of kept IC components and the true response group. Finally, we compared the spectrograms of the three channels used by the SRP with a noise floor which we recorded by submerging the electrodes in a saline solution for the same time length as the EEG resting-state period that was used for the prediction (5 min). Results are presented in Figure 6.9 and discussed in subsection 6.5.3.

6.4 Results

6.4.1 Positive- and negative stimulation effects in tACS

The predictability of the behavioural response to tACS based on the resting-state electroencephalographic activity preceding the stimulation was examined through a crossover study. During the first group of recordings, nineteen healthy subjects performed 15-minute blocks of 3D reaching movements with their right arm (Figure 5.1).

Figure 6.3 depicts the histogram and estimated probability density function (Gaussian kernel density estimate with a kernel width of 0.5 (Turlach, 1993; Scott, 2015)) of the individual effect sizes. While the group-level effect size of 0.33 is not statistically significant ($p = 0.1018$, subject-level permutation-based paired t-test assuming normal distribution of the effect sizes, cf. Section 6.3.2), a substantial variation in the effect sizes of individual subjects was observed, ranging from -1.12 to 1.51 with a standard deviation of 0.72. Statistical tests for significant effect sizes at the individual subject level revealed seven subjects with a statistically significant *positive* and four subjects with a statistically significant *negative* effect size (at significance level $\alpha = 0.05$, two-sided trial-level permutation-based t-test). The remaining eight subjects did not show a statistically significant effect at the individual subject level (individual p-values are shown in Table 6.1.

These findings indicate that γ-tACS can have positive- as well as negative behavioural effects on motor performance, which poses an ethical challenge to tACS studies [1]. The following results demonstrate how to address this challenge by predicting individual subjects' stimulation response from their resting-state configuration of brain rhythms.

[1] Even after multiple test correction on the p-values there are subjects in both positive and negative responders. After performing a holm-sidak multiple test correction and threshold 0.05, on the 19 p-values reported in Table I, there are still subjects with significant positive (subjects 1, 5, 16) and significant negative response (subject 17). The remaining subjects did not have significant difference between sham and real stimulation. The corrected p-values: 0.0079 0.9795, 0.4867, 0.1567, 0.0018, 0.9965, 0.9208, 0.6556, 0.9965, 0.1827, 0.2629, 0.3394, 0.4867, 0.9965, 0.0018, 0.0018, 0.1094, 0.3394, 0.3423.

Table 6.1: Categorization of subjects into *responders* and *non-responders*, first group of recordings. ΔVelocity > 0 refers to subjects with a higher movement speed in the real vs the sham stimulation block.

Subject	p-value	ΔVelocity > 0	Category
1	0.0005	1	responder
2	0.6217	1	non-responder
3	0.0902	1	non-responder
4	0.0121	0	non-responder
5	<0.0001	1	responder
6	0.9318	1	non-responder
7	0.3979	1	non-responder
8	0.1628	1	non-responder
9	0.9220	0	non-responder
11	0.0154	0	non-responder
12	0.0251	1	responder
13	0.0384	0	non-responder
14	0.0800	0	non-responder
15	0.8484	1	non-responder
16	<0.0001	1	responder
17	<0.0001	0	non-responder
18	0.0077	1	responder
19	0.0370	1	responder
20	0.0455	1	responder

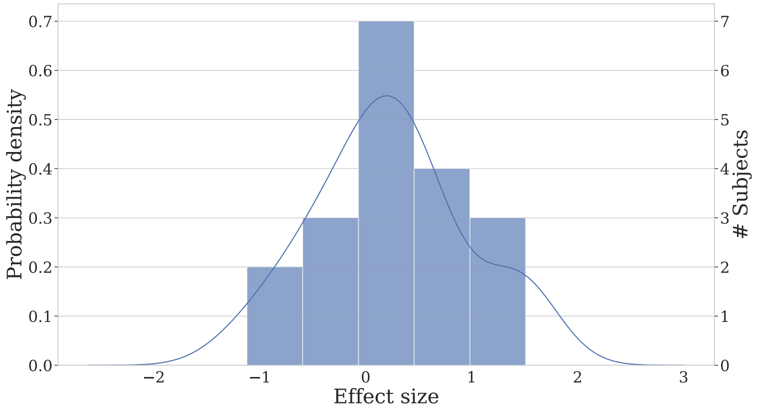


Figure 6.3: Histogram and estimated probability density function of stimulation response.

6.4.2 Resting-state EEG predicts tACS stimulation response

As described in detail in Section 6.3.2, the subjects were separated based on their behavioural response to the tACS into two groups: those with a statistically significant positive stimulation response, subsequently called the *responders*, and the remaining subjects (both negative responders and subjects with no significant response), subsequently termed the *non-responders*. The reason why we did not use three different groups in our prediction pipeline ("positive", "negative" and "non-responders") was the limited number of subjects for training such a classifier [2]. Therefore, the analysis was based on the aforementioned two groups.

The group-average topography of the bandpower in the γ-range (90–160 Hz), recorded prior to the first (real or sham) stimulation block, revealed strong resting-state γ activity over the contralateral motor cortex in the *responders* but not in the *non-responders* (Figure 6.4). This result suggests that only those subjects, who already had strong γ-power over contralateral motor cortex before the start of the stimulation, showed a subsequent positive behavioural response to contralateral γ-tACS. To systematically evaluate the predictive value of this frequency range for stimulation response, we trained a machine learning algorithm to predict individual subjects' responses to γ-tACS from their resting-state brain rhythms. The predictability of all the canonical brain bands is depicted in Figure 6.5, where also the different resting-states are presented.

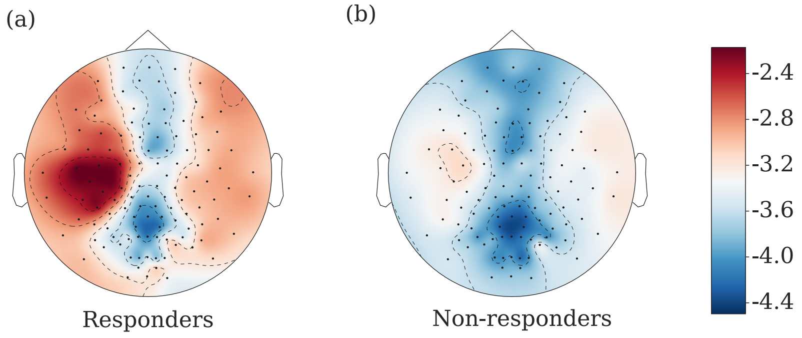


Figure 6.4: Group-averages for responders (a) and non-responders (b) of high γ (90–160 Hz) log-bandpower during the resting-state recorded at the end of the first block of the second session (prior to stimulation blocks).

The details of the prediction pipeline are described in Section 6.3.4. We found the prediction accuracy to increase with frequency, peaking at 89.47% in the band from 90–160 Hz ($p < 0.001$, permutation test, cf. Section 6.3.4 for details).

[2] In a preliminary analysis we did not manage to reach satisfactory classification accuracy for three groups, for none of the frequency bands.

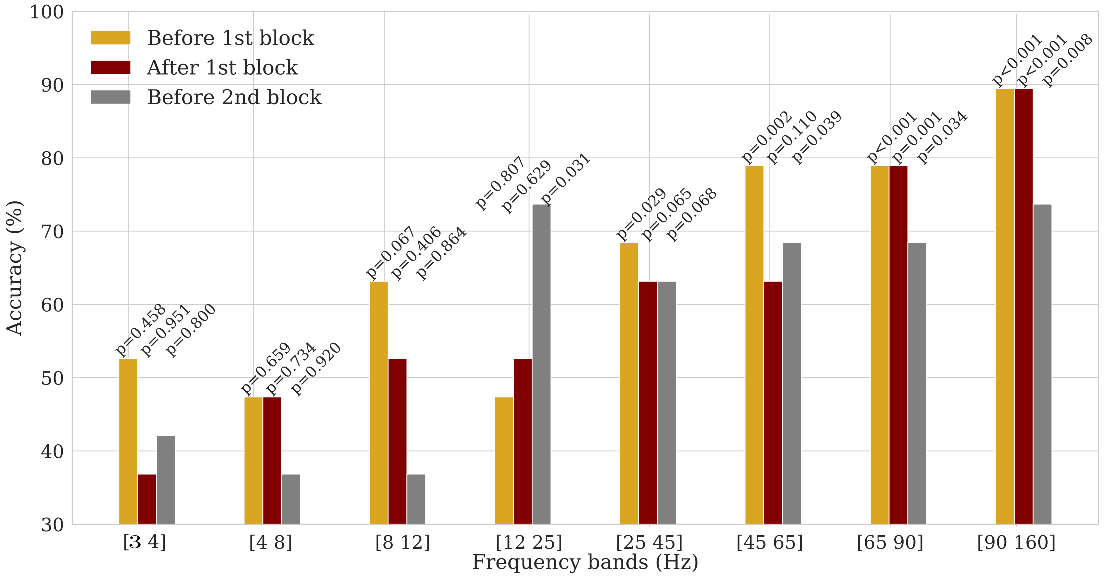


Figure 6.5: Leave-one-subject-out cross-validated prediction accuracy of stimulation response in the first group of subjects across canonical frequency bands and resting-states, cf. Section 6.3.4 for details. Stimulation day (day 2). From the δ-band only the range [3–4]Hz was tested because the data are high-pass filtered at 3 Hz. The band 45–60 Hz is not tested because we wanted to avoid the strong contamination from the line power (50 Hz).

To test the robustness of the prediction pipeline, we repeated the machine learning procedure for all three resting-state recordings preceding the first stimulation block. We found that all resting-state recordings enabled above chance level prediction in the 90–160 Hz band. Prediction accuracies of the remaining bands varied across the different resting-states. These results establish that distribution of γ-power across motor areas, as shown in Figure 6.4, is an accurate predictor of subjects' behavioural response to γ-tACS over contralateral motor cortex.

6.4.3 Validation group - Subject stratification by resting-state EEG enhances effect sizes

Here we present the application of the response stratification pipeline described in the previous section, on the new 22 healthy participants (validation group). Based on the results described in the previous section, the classifier that was trained on the resting-state recorded after the first block in the 90–160 Hz frequency band was chosen, as it exhibited the best accuracy in the training set. This classifier was trained on the first group of subjects and then used out-of-the-box to predict the stimulation response for each subject in the validation group from a resting-state EEG recorded prior to the first block of trials (see Section 6.3.5 for details).

In the validation group, group-level behavioural effect size of 0.12 was calculated ($p = 0.2847$, subject-level permutation paired t-test, assuming normal distributions), with subject-level effect sizes ranging from -0.94 to 1.19, as shown in Figure 6.6. The true response groups of the validation subjects based on their performance can be found in Table 6.2.

Table 6.2: Categorization of subjects into *responders* and *non-responders*, second (validation) group of recordings. ΔVelocity > 0 refers to subjects with a higher movement speed in the real vs the sham stimulation block.

Subject	*p*-value	ΔVelocity > 0	Category
21	0.0001	0	non-responder
22	0.0004	1	responder
23	0.0119	1	responder
24	0.0093	1	responder
25	0.0353	1	responder
26	0.5704	1	non-responder
27	0.0001	1	responder
28	<0.0001	0	non-responder
29	0.5334	0	non-responder
30	0.5449	0	non-responder
31	0.6537	1	non-responder
32	0.8046	0	non-responder
33	0.8770	1	non-responder
34	0.0055	0	non-responder
35	0.0678	0	non-responder
36	0.6119	1	non-responder
37	0.2996	0	non-responder
38	0.0048	1	responder
39	0.4444	1	non-responder
40	0.7515	1	non-responder
41	0.5464	0	non-responder
42	0.0075	0	non-responder

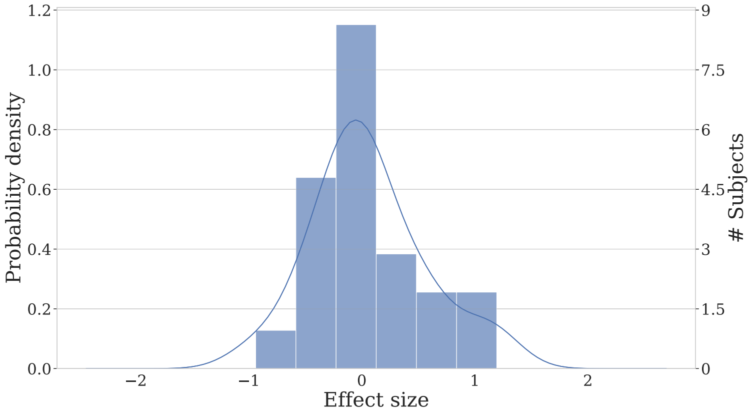


Figure 6.6: Histogram and estimated probability density function of stimulation response in the validation group.

The EEG-based stratification of subjects resulted in group-level effect sizes of 2.46 and -0.17 for the predicted responders and non-responders, respectively, a statistically ($p = 0.0048$, one-sided permutation-based t-test, assuming normal distributions) and practically highly significant difference.

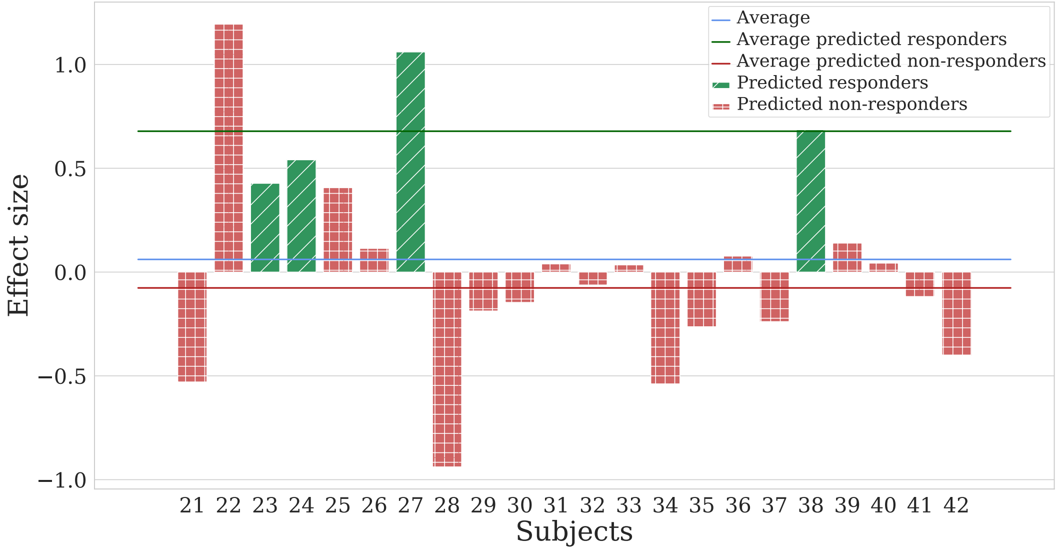


Figure 6.7: Individual effect-sizes and predicted stimulation responses in the validation group. Green colour depicts the effect sizes of the subjects that were predicted as “responders”, and the red colour depicts the effect sizes of the subjects that were predicted as “non-responders”. Subjects with id 22 and 24 were mis-classified.

In particular, all four subjects with a statistically significant negative- and all 12 subjects with no statistically significant stimulation response were correctly assigned to the group of non-responders. Further, only two subjects with a statistically significant positive stimulation response were misclassified as non-responders. These behavioural results are summarized in Figure 6.7.

The group-averaged topographies of log-bandpower in the γ-range of the predicted responders and non-responders, which closely resemble those observed in the training group shown in Figure 6.4, are displayed in Figure 6.8.

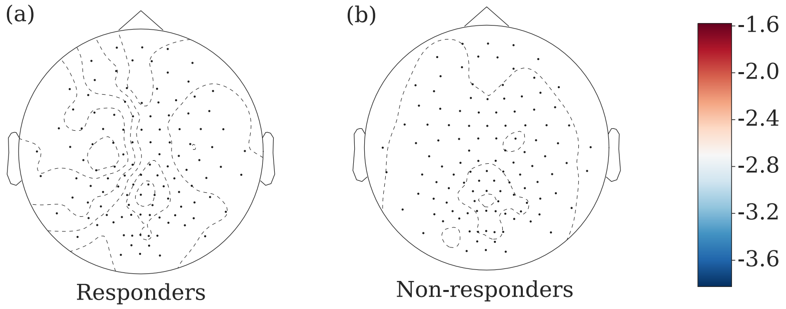


Figure 6.8: Group-averages for predicted responders (a) and non-responders (b) in the validation group of high γ (90–160 Hz) log-bandpower during the resting-state recorded at the beginning of the stimulation day (prior to stimulation blocks).

6.4.4 Stimulation response is contingent on brain state

Here the validated prediction pipeline was employed to test whether stimulation response is a state or a trait, i.e., whether subjects' response to γ-tACS changes or remains invariant over time. To test this, all 41 subjects were pooled and used to train the proposed prediction pipeline on the γ-power (90–160 Hz) of each of the four EEG resting-states of their first session, i.e., two days before the stimulation session, and predicted the stimulation response in the second session with leave-one-subject-out cross-validation (all other settings were identical to those described in Sections 6.3.3 and 6.3.4). A statistically significant prediction accuracy in this setting would imply that the configuration of subjects' brain rhythms is also predictive for their stimulation response two days later. However, no evidence in favour of this conclusion was found. Instead, training on brain activity of the first recording session resulted in statistically non-significant prediction accuracies between 62.5% and 68.3% (p-values of 0.23, 0.24, 0.27 and 0.73 for the four resting-states of day one). This observation could be an indication that stimulation response is contingent on subjects' brain-state directly prior to the stimulation, i.e., subjects' stimulation response is a state and not a trait.

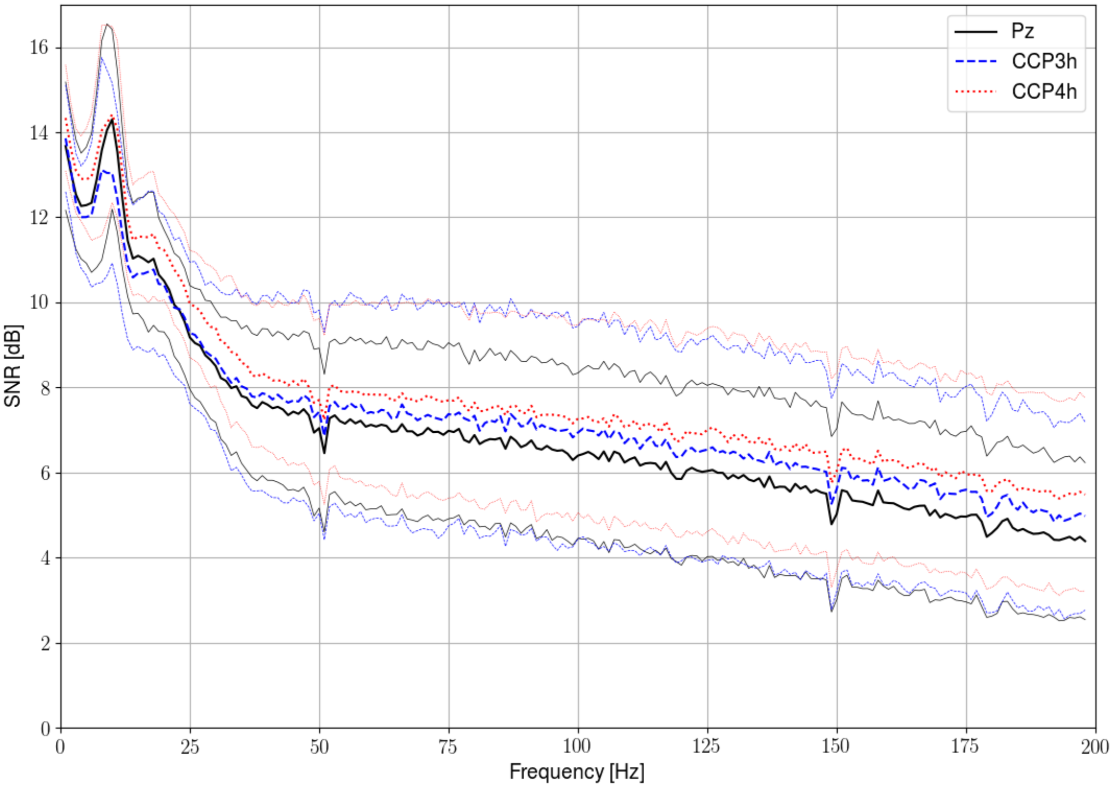


Figure 6.9: Mean SNR plus, minus one standard deviation across subjects' raw resting-state data (i.e., no common average and no ICA cleaning) during the 5 min of resting-state used by our SRP classifier (beginning of 2nd block, Session 2), with respect to the noise floor measured for the same time length, by submerging the electrodes in a saline solution. The three different colours/textures refer to the three channels used by the SRP classifier (CCP3h, CCP4h and Pz). The bold lines refer to the mean SNR across subjects, while the faded lines depict one standard deviation. As we can see, the actual EEG has more power than the noise floor in the high gamma range, which is another indication that our classifier is not based on noise.

Applying the original stimulus-response predictor, as described in Section 6.3.5, to the resting-state recordings of the first day, we estimate that out of the 28 subjects, who did not respond positively to the stimulation on day two, five subjects would have responded positively on day one (prediction results for individual subjects are shown in Table 6.3). As such, the percentage of subjects who can benefit from tACS may increase if they are stimulated at the right time.

Table 6.3: Subjects' predicted behavioural response from resting-state EEG data recorded on day one versus actual behavioural response measured on day two.

Subjects predicted as responders on day one	15, 18, 19, 22, 23, 24, 27, 33, 34, 36, 38, 42
Actual non-responders on day two	2, 3, 4, 6, 7, 8, 9, 11, 13, 14, 15, 17, 21, 26, 28, 29, 30, 32, 33, 34, 35, 36, 37, 39, 40, 41, 42
Intersection	15, 33, 34, 36, 42

6.5 Discussion

6.5.1 Screening of non-responders to motor tACS from resting EEG

Our results demonstrate that resting-state human brain activity in the high gamma range, recorded prior to NIBS, can distinguish responders from non-responders with high accuracy. This screening is of high importance for a safe and ethical application of NIBS in research and treatment. As NIBS has been shown to have behavioural effects of opposite polarity relative to the intended stimulation effect in individual subjects (cf. Section 6.4.1), a reliable exclusion criterion for subjects with a negative or non-significant stimulation response ensures that no subjects are harmed and that no redundant stimulation sessions are performed. This issue is of particular relevance in clinical settings, where NIBS is employed to cause long-term and, possibly, permanent changes.

6.5.2 The motor task

Here the rationale for the selection of the specific task is briefly discussed. The selected visuo-motor task allows for a free 3D arm movement that mimics natural reaching movements, which we would like eventually to enhance and facilitate through stimulation. We focused on the arm speed to be our behavioural metric, because this metric has been found to be the most robust variable related to the motor cortex (Hatsopoulos and Suminski, 2011), with Moran and Schwartz (1999) having even proposed a canonical model for it. In addition to the aforementioned reason, we preferred the arm speed over the normalized average rectified jerk, which we report as a score to the subjects, because NARJ is the second derivative of the speed, which already accumulates a lot of error in the measurement, starting from the position recorded from the infrared cameras. Therefore, this experimental set up was selected as it could help us measure the arm speed in a 3D movement that seems natural. For the present study, we are not interested in the performance of the subjects in terms of successfully reached targets as a metric, as this would potentially include a more complex cognitive procedure. This is the reason why we focus on the relation between the motor cortex activity and the arm speed.

6.5.3 High frequency band carries information about the response to gamma motor stimulation

At first, we were sceptical to find that such a high frequency (90–160 Hz) power was the most predictive of tACS response. Here we discuss, based on our analysis, why we can be optimistic that this findings are not artefact-driven. First, a particularly conservative artefact rejection procedure was employed, following the recommendations by McMenamin *et al.* (2010), to minimize the probability of retaining artifactual sources in the measurements. This approach also explains the small number of kept components per subject. To ensure that no bias was introduced during this manual process, we tested the statistical association between the number of kept ICs and the response group of each subject. Kruskal-Wallis test yielded no significant association ($p = 0.23$). Moreover, no association was found between the type of the kept ICs (i.e. exhibiting a horizontal or vertical dipole) and the response group. In the group-average topoplots of Figure 6.4, the power of [90–160] Hz is mostly located over the contralateral motor cortex. Nevertheless, our prediction algorithm uses the channels over left (CCP3h) and right motor cortex (CCP4h) and one channel over parietal cortex (Pz) as input to the classifier, to avoid focusing only on the stimulation area and to limit the possibility of picking stimulation artefacts.

In addition, we compared the spectrograms of the three channels used by the classifier with a noise floor which we recorded with the electrodes on a salty water solution as described in subsection 6.3.7. Figure 6.9 depicts the mean ± one standard deviation signal-to-noise ratio (SNR) across subjects' raw resting-state data during the 5 min of resting-state used by our SRP classifier, with respect to the noise floor measured for the same time length, by submerging the electrodes in a saline solution. Figure 6.9 indicates an average SNR between 6 and 7 dB in the 90 to 160 Hz range, i.e., the recorded signal in the high gamma range is at least four times stronger than the inherent measurement noise. This observation is in line with recent work that indicates that the feasibility to measure human gamma power in scalp EEG is not limited by recording hardware but rather depends on subjects' morphology (Butler *et al.*, 2019).

Another point that eliminates the probability that our findings are confounded by stimulation artefacts, is the fact that the periods used for the training and the testing of the classifier are resting-state periods. No stimulation, neither real nor sham, was applied during the resting-state periods, which also limits the possibility that the EEG data are contaminated with stimulation artefacts. Stimulation was delivered only during the reaching blocks, as shown in Figure 6.2. Therefore, it seems plausible that high gamma range indeed carries such significant information.

6.5.4 Plausible neurophysiological mechanism

In our experiments a strong resting-state γ-power over the contralateral motor cortex was found to be indicative of a positive stimulation response to tACS (in terms of arm speed). This finding is in line with our current understanding of the neurophysiological effects of γ-tACS and the role of γ-power in fronto-parietal networks for motor performance (Gonzalez Andino *et al.*, 2005). Resting-state γ-power in primary motor cortex has been shown several times to positively correlate with γ-aminobutyric acid (GABA) levels (Chen *et al.*, 2014; Muthukumaraswamy *et al.*, 2009; Bartos *et al.*, 2007; Wang and Buzsáki, 1996; Brunel and Wang, 2003). Because γ-tACS over motor cortex decreases GABA levels (Nowak *et al.*, 2017), and decreases in motor cortex GABA levels correlate with increased motor performance (Stagg *et al.*, 2011), high resting-state γ-power may signal a brain state in which motor performance can be improved through tACS-induced reduction of GABA levels. In contrast, low resting-state γ-power would indicate a brain state in which GABA levels are already low, thus limiting the extent of potential further reduction by γ-tACS. We note that this explanation is also in line with our finding that stimulation response is contingent on the current brain state (cf. Section 6.4.4). While our argument is consistent with the neurophysiological GABA activity, it is worth mentioning that there are studies on macaques, that exhibit enhancing of the ongoing gamma activity by gamma tACS (Krause *et al.*, 2019; Johnson *et al.*, 2019). Nevertheless, the protocols of these studies differ significantly from ours. Krause et al. (Krause *et al.*, 2019) do not target the motor cortex, but the hippocampus and visual cortex of anaesthetized animals, which could explain the discrepancy with our findings. Finally, Johnshon et al. (Johnson *et al.*, 2019) place the bipolar electrodes over the left and right temples, which could also induce different effects from our setup.

6.5.5 Evaluation of external factors' role in behavioural response to stimulation

To further probe the state vs trait hypothesis, and to examine the possibility that behavioural response might be affected by possible differences in sensation between sham and real stimulation, we tested a range of factors, including sex, order of real/sham stimulation, block of strongest-reported stimulation sensation and behavioural performance on the first day of the experiment, for associations with stimulation response. None of these factors reached a statistically significant association (see Section 6.3.6). Of course, besides the gamma activity there could be other unobserved factors that play a hidden confounding role. Methods that will focus on this particular problem are presented in Chapters 7 and 8. Without being able to exclude this possibility now, we can be confident that the measured behavioural response is not confounded by the different sensation levels between the two conditions for the following reason: Subjects evaluated a range of possible sensations, such as tingling, burning, phosphates, itching etc., for both conditions, without knowing which condition was applied, and the statistical analysis showed no significant association of the response group neither with the block of strongest reported sensation nor with any of the reported types of sensations.

6.5.6 Outlook

The results presented here indicate that the percentage of subjects who can benefit from NIBS, may be increased when subjects are stimulated at the right time. Concurrently, the neurophysiological interpretation of our results raises the question whether the effects of stimulation lie within the range of normal variations in behavioural performance, i.e., whether NIBS induces a beneficial state of mind that can also occur spontaneously, or whether NIBS can enhance behavioural performance beyond subjects' natural limits. Either way, a natural extension of our stimulation response prediction pipeline would be to consider multiple stimulation settings that vary in parameters such as spatial and spectral focus, paving the way for personalised NIBS. Being motivated by the exhibited difference in gamma power between the two groups – as depicted in Figures 6.4 and 6.8– another future extension would be to combine this pipeline with a pre-stimulation step of gamma-modulation, to study whether a self-modulated rise of the gamma activity could ensure a positive response to the tACS.

6.6 Conclusion

The identification of responders and non-responders prior to the application of stimulation treatment is a considerable first step towards personalised brain stimulation. This Chapter showed evidence that resting-state high-gamma power prior to stimulation enables this differentiation. Specifically, this was demonstrated in a first experimental group of 19 participants that subjects' resting-state EEG predicts their motor response (arm speed) to gamma tACS over the contralateral motor cortex. It was then ascertained in a prospective stimulation study with twenty-two new subjects that our prediction pipeline achieves a reliable stratification of subjects into a responder and a non-responder group.

Chapter 7

Selecting causal brain features of motor performance

This chapter is dedicated to the development of a theory and its corresponding algorithm for causal feature selection (see definition given in Section 2.4) even under no causal sufficiency assumption (see definition 7). More specifically, inspired by settings where a temporal parent of each candidate cause is known, we develop conditions which we prove to be sufficient for the detection of direct and indirect causes of a target variable. We prove that if we can observe a cause for each candidate cause, then a single conditional independence test with only one conditioning variable is sufficient to decide whether a candidate associated with the target is indeed its cause. Thus, we improve upon existing methods by significantly simplifying statistical testing and requiring a weaker version of causal faithfulness. We demonstrate the successful application of our method to simulated graphs and in encephalographic data of twenty-one participants. The detected brain causes of motor performance are in accordance with the latest consensus about the neurophysiological mechanisms and can provide new insights into personalised brain stimulation. The theory and the findings presented in this chapter belong to the publication (Mastakouri *et al.*, 2019b) of the author of this dissertation, alongside Bernhard Schölkopf and Dominik Janzing.

7.1 Problem statement

Here, first, we tackle the generic problem of causal feature selection with latent variables in i.i.d. nodes, and subsequently, we focus on the more specific problem of causal brain feature selection (see Problem 4). Given the target response variable (motor performance) and the brain activity from distinct electrode locations, is it possible to identify the causes of the observed motor performance? If so, what are the necessary assumptions?

7.2 Motivation from the causal perspective

Conditional independence relations have been an important tool in the field of computational statistics (Shah and Peters, 2018), (Koller and Friedman, 2009) and play a significant role in causal inference (Pearl, 2009). However, conditional independence-based causal inference in real datasets is a challenging task, since testing them is (Shah and Peters, 2018), particularly when the number of conditioning variables is large. PC (Spirtes *et al.*, 2000), FCI (Spirtes *et al.*, 2000) and CPC (Ramsey *et al.*, 2012) are three of the most prominent conditional independence-based causal discovery methods (see Section 2.3.1 of Chapter 2 for descriptions of these methods). To discover the underlying causal graph from the data, they require some assumptions, which, however, are often violated. These include the *causal Markov condition*, *faithfulness* and, in addition for PC method, also *causal sufficiency* (def. 7). Although the FCI algorithm (Spirtes *et al.*, 2000) does not need this assumption, it becomes unreliable due to the many statistical tests that it requires, if the connections between the features are not sparse. Moreover, faithfulness is a rather problematic assumption, as in causal models with many variables, typical parameter values may yield distributions that are close to being unfaithful (Uhler *et al.*, 2013). For all the aforementioned reasons, we focus on developing a causal feature selection method with a minimum number of conditioning sets and tests.

7.3 Motivation from the neuroscientific perspective

The field of non-invasive neuroimaging, such as Electroencephalography (EEG), is one typical case where the discovery of causal features is required. In such setups the activity of billions of neurons is recorded as noisy mixtures of underlying activity, traversing several layers of cortex, skull and skin. Therefore, it is not realistic to assume causal sufficiency. In addition, the dimensionality of the data is large, often comparable with or even larger than the sample size. Despite these limitations in such datasets, the need for causal inference often arises, in order to differentiate a set of causal brain features from a large number of correlations between the brain activity and the observed behavioural response (Weichwald *et al.*, 2015).

The motivation for the proposed method emanates from the field of non-invasive brain stimulation, which aims, among others, at the rehabilitation of motor functions, for patients with motor deficits. As we have mentioned before, one fundamental problem is the lack of exact knowledge of the mechanism that the stimulation entrains the ongoing brain oscillations (Davis and Koningsbruggen, 2013; Helfrich *et al.*, 2016). Subsequently, until now, the selection of the exact stimulation parameters (frequency, intensity and target location) is based on collected observations, instead of the patterns of the individual's brain activity. For instance, stimulation at γ-range frequencies (70Hz) has been proposed to facilitate movement (Nowak *et al.*, 2018; Muthukumaraswamy, 2010), while frequencies in β-range have been reported to inhibit it (Espenhahn, 2018; Gulberti *et al.*, 2015; McAllister *et al.*, 2013). However, similar stimulation parameters that focus particularly on motor tasks have resulted in very heterogeneous responses across subjects, which span from positive to negative (Mastakouri *et al.*, 2019a; Wiethoff *et al.*, 2014). In our previous studies (Mastakouri *et al.*, 2017), we have argued that the reason for this discrepancy of responses to NIBS may originate from the extensive variability of each brain's activity during movement, and as such, personalised stimulation parameters may be beneficial to ensure a positive response. A better understanding of the motor cortex activity during voluntary movements, for each individual independently, could contribute to the identification of such personalised parameters.

7.4 Methods

7.4.1 Definitions and notations

For the understanding of this chapter, the studying of Chapter 2 is required first. Here, we briefly recap some fundamental definitions in Causal Bayesian Networks (Pearl, 2009), which we will use to present our methodology and prove our theorem below. The notions of *faithfulness* (see def. 5) and of *causal Markov condition* (see def. 2) are fundamental, in order to be able to relate properties of the distributions of interest to the causal graph. Markov condition enables us to read off *independences* from the graph structure, while faithfulness allows us to infer *dependences* from the graph (Peters *et al.*, 2017). In other words, a distribution P is faithful to a directed acyclic graph (DAG) G if there are no other conditional independence relations other than the ones entailed by the Markov property. Another important notion for causal discovery is the *confounding path* between two variables. Here, we define that a variable is a confounder (observed or unobserved) if it is a common ancestor of two other variables.

The following notation is going to be used from now on for the description of our mehtod:

- $\dashrightarrow$: denotes a directed path with observed variables or a direct link.
- $\rightarrow$: denotes a direct link.

Now we will briefly introduce the environment of our methodology. The problem of causal features selection was inspired from brain datasets, where the candidate causes are brain features i; i.e. activity in different brain locations and frequencies, and the target variable R is some behavioural response metric that we measure on the subject. We also assume that each candidate feature i has an observed previous state P^i and a current M^i state. The variables' names read as "Plan", "Move" and "Response" respectively. An example of such a structure is given in Figure 7.1, where (brain) feature M^1 has an ancestor P^1 and is a cause of R, while feature M^2 does not cause R but connects with a confounding path that includes M^1. Without knowing the structure, our theorem is able to differentiate the true causes (M^1,M^n) from the ones that are dependent to the target due to confounding paths (M^2,M^3).

7.4.2 Formal problem description

Given the random variables P^i, M^i $i = 1,2...n$ and R, we assume the class of DAGs in which there can exist instantaneous acyclic effects between the P^i variables, between the M^i variables, as well as forward effects between the P^i, M^i and R. Section 7.5 provides further explanation about how the assumptions described below are commonly met in real datasets where candidate causes can be measured in two consecutive time stamps, and as such, a causal arrow from the previous to the current state can be assumed. Such a case is a brain dataset. Below we present the assumptions of our methodology.

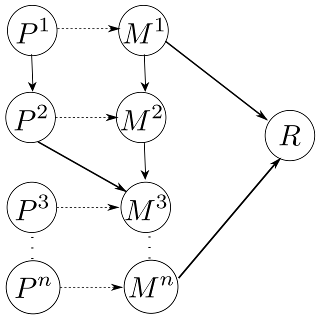


Figure 7.1: An example of a possible DAG that includes the random variables P^i, M^i $i = 1,2...n$ and R, assuming As1-As5. Each candidate causal feature M^i has a cause P^i and may have other acyclic edges with the other candidates. Some of the M^i features cause the target R.

7.4.3 Assumptions

As1 Causal Markov condition

As2 Causal Faithfulness

As3 $P \not\dashleftarrow M \not\dashleftarrow R$: There can be no backwards arrows in time. Variable R is measured after M, which is measured after P.

As4 $P^i \dashrightarrow M^i$ exists: Variables P^i and M^i represent two consecutive states of the same brain feature i. We assume that the state P is always a cause of state M for the same feature i.

As5 (R,M^i,P^i) are independently drawn from some distribution (i.i.d).

7.4.4 Theorem

Theorem 2. *Given the variables P^i, M^i $i = 1,2...n$ and R, and assuming As1-As5, if*

$$M^i \not\perp\!\!\!\perp R \tag{1}$$

and

$$P^i \perp\!\!\!\perp R \mid M^i \tag{2}$$

then

$$M^i \dashrightarrow R$$

Proof. We prove the claim by contradiction. Assume As1-As5 and that M^i and R are dependent (condition 1), but there is no directed path from M^i to R. Then there is a confounding path $p_1 := M^i \dashleftarrow C \dashrightarrow R$ with some common cause C (hidden or observed). Now consider some path $p_2 := P^i \dashrightarrow M^i$ (it exists due to Assumption As4). If p_1 and p_2 have only M^i in common, M^i is a collider and thus P^i and R are not d-separated by M^i. If p_1 and p_2 share more nodes, assume first they have P^i in common, that is, P^i lies on p_1. Then P^i and R are not d-separated by M^i because the sub-path of p_1 connecting P^i and R does not contain M^i and p_1 is collider-free. Assume now that P^i does not lie on p_1, and p_1 and p_2 share some node X other than M^i and P^i. Then either (i) $X = C$, or (ii) X is a node between C and R, or (iii) X is a node between C and M^i. For (i) and (ii), we have a directed path from P^i to R (that does not contain M^i). In case (iii), X is a collider and M^i a descendent of this collider, hence M^i unblocks the path from P^i to R. In all three cases, M^i does not d-separate P^i and R, which contradicts $P^i \perp\!\!\!\perp R \mid M^i$ (cond. 2) due to faithfulness. Hence there must be a directed path $M^i \dashrightarrow R$. □

Based on Theorem 2 we build an algorithm, which, given the previous (P) and the current state (M) of each candidate feature, as well as the target variable (R) for each independent observation (trial), returns a vector with the indices of the input features that were found to be causes of R. Note that our algorithm requires only one conditional independence test for each node. Therefore, it accelerates the causal feature selection as it scales linearly with the number of nodes in the graph; hence its complexity is $\mathcal{O}(n)$.

Algorithm 1: Find causes of R

```
Input: $P^i, M^i, R, \forall i = 1, ..., n.$
Output: $CausesR$
for $i \leftarrow 1$ to $n$ do
    $pvalue1 \leftarrow ind\_test(M^i, R)$
    if $pvalue1 < threshold1$ then
        $pvalue2 \leftarrow cond\_ind\_test(P^i, R, M^i)$
        if $pvalue2 > threshold2$ then
            $CausesR \leftarrow [CausesR, M^i]$
        end
    end
end
```

Note that Theorem 2 provides sufficient but not necessary conditions for M^i to be a cause of R. That means that some of the causes of R may not be identified. Note that our assumptions do not include causal sufficiency. Therefore, even in the case of unobserved common causes, if the conditions described in 2 are met, then we know that the dependency between the M^i and R is because of a directed path and not due to a confounder variable. [1]

7.5 Experiments

We apply our method on data produced from simulated graphs, as well as on an EEG dataset consisting of twenty-one healthy participants. All EEG experiments and recordings were performed in the Max Planck Institute for Intelligent Systems and were approved by the Committee of the Eberhard Karls University of Tübingen. Informed consent was given by all participants, prior to their participation to the study. For the implementation, to make sure that in practice Assumption As4 in the data is not violated, we check the dependence between P^i and M^i for the same i, with an independence test, and in case it is not significant we reject the candidate without further checking. Both for the simulated graphs described below and for the EEG data, an HSIC (Gretton *et al.*, 2005b) and a conditional HSIC (Fukumizu *et al.*, 2008; Zhang *et al.*, 2012) test was used to check for the independencies and conditional independences. A Gaussian kernel and the usual heuristic bandwidth was used in (Gretton *et al.*, 2005b). Thus, our algorithm also accounts for non-linear relationships between the features. For the statistical testing we examine the null hypothesis $H0_1 : M^i \perp\!\!\!\perp R$. We consider to have rejected the null hypothesis (hence consider to have found M^i and R to be dependent) if $p < \alpha_D = 0.05$. Then,

[1] Someone needs to check that the relationship between M^i and P^i is not too deterministic. Obviously this would amount to the conditional independence $P^i \perp\!\!\!\perp R \mid M^i$ even in the presence of counfounding. If M^i and P^i are too close in time it could result in violation of faithfulness.

we examine the null hypothesis $H0_2 : P^i \perp\!\!\!\perp R \mid M^i$ and accept it (hence the conditional independence) if $p > \alpha_{CI} = 0.25$ (usual values for accepting conditional independence in EEG datasets include thresholds above 0.25 (Grosse-Wentrup *et al.*, 2016)).

7.5.1 Simulated graphs

Given the variables P^i, M^i $i = 1,2...n$ and R as described in 7.4.2, and assuming As1-As5, we build simulations of possible DAGs and apply our Theorem 2.

Construction of simulated graphs: We sample the noise terms of P, M and R variables from a Gaussian distribution whose variance is randomly sampled from a Uniform distribution. We then define the adjacency matrix of the subgraph that consists of all P^i variables as an $n \times n$ matrix A_P, whose elements are independently drawn from a Bernoulli(p) distribution, denoting the existence of an edge between the different P^i variables, forbidding any self-cycles ($a_{P_{i=j}} = 0$). We update the P^i values by adding a function f_1 of each parent P^j variable:

$$P^i = P^i + \sum_{j=1}^{k_P} f_1(P^j_{a_{P_{ij}}==1}) \tag{7.1}$$

for the k_P parent P^j variables of P^i. Then, we define the adjacency matrix of the subgraph that consists of all P^i and M^i variables as a $n \times n$ matrix A_{PM}, whose elements are values independently drawn from a Bernoulli(p), indicating the existence of an edge between the different P^i and M^i variables, making sure that for $i = j$ the edge exists ($a_{\text{PM}_{i=j}} = 1$). We update the M^i values by adding a function f_2 of each parent P^j variable:

$$M^i = M^i + \sum_{j=1}^{k_{\text{PM}}} f_2(P^j_{a_{\text{PM}_{ij}}==1}) \tag{7.2}$$

for the k_{PM} parent P^j variables of M^i. To avoid creation of cycles, we only generate the following types of arrows: (1) $P^i \rightarrow M^j$ for $i \leq j$, (2) $P^i \rightarrow P^j$ for $i < j$, (3) $M^i \rightarrow M^j$ for $i < j$, (4) $P^i \rightarrow R$ and (5) $M^i \rightarrow R$. As a third step, we create the adjacency matrix of the subgraph that consists of all M^i variables as a $n \times n$ matrix A_M, whose elements are values independently drawn from a Bernoulli(p), denoting the existence of an edge between the different M^i variables, forbidding any self-cycles ($a_{M_{i=j}} = 0$). We update the M^i values by adding a function f_3 of each parent M^j variable:

$$M^i = M^i + \sum_{j=1}^{k_M} f_3(M^j_{a_{M_{ij}}==1}) \tag{7.3}$$

for the k_M parent M^j variables of M^i. Finally, we create the vectors A_{MR} and A_{PR} with n elements independently drawn from a Bernoulli(p), denoting the existence of an edge from M to R and from P to R. We update the R values by adding a function f_4 of each M^i that is a parent and a function f_5 of each P^i that is a parent:

$$R = R + \sum_{i=1}^{k_{\mathrm{MR}}} f_4(M^i_{a_{\mathrm{MR}_i}==1}) + \sum_{i=1}^{k_{\mathrm{PR}}} f_5(P^i_{a_{\mathrm{PR}_i}==1}) \tag{7.4}$$

for the k_{MR} parent variables M^i and the k_{PR} parent variables P^i of R. We sample the coefficients for the five linear functions f_1, f_2, f_3, f_4, f_5 from a Gaussian distribution. We test the performance of our algorithm for different number of nodes n for the P and M variables, varying sparsity of edges and sample sizes. For each combination, we examine 20 random graphs and report the percentage of the false positives and false negatives, calculated on the number n of features i.

Comparison with Markov Blanket methods and LASSO: LASSO or Markov Blanket (MB) discovery methods require causal sufficiency, let alone the curse of dimensionality. Furthermore, with high dimensional data, any algorithm using conditional independence tests has to condition on large variable sets. In that case, conditional independence testing is hard (Shah and Peters, 2018) and cannot be reliable unless sample sizes are huge. Finally, even if causal sufficiency was to hold, the known MB detection algorithms as well as LASSO do not *detect* variables but they rank them, and gradually evaluate the prediction accuracy by including more variables, according to the ranked order the algorithm returned. This requires a heuristic *hyperparameter* to define what is the right acceptable number of variables to be included in the MB, which subsequently affects the false positive and the false negative rates. For completeness, however, we provide comparison results of our method against the following three available algorithms (average for 10 random graphs): HSIC LASSO (Yamada *et al.*, 2014), Backwards elimination with HSIC, and Forward selection with HSIC for MB discovery (Song *et al.*, 2007b). We present the most optimistic for the other algorithms case, that of large sample size (800) and two cases of small (20) and large graphs (125 nodes), for sparse (0.2) and dense (0.5, more true causes) edges. We report the % of false positives and false negatives in the number of variables.

7.5.2 Identifying brain causes of motor performance from EEG data

Our motivation behind the development of this method was the identification of causal brain features of upper limb movement from brain activity during a reaching task. Such a detection could eventually help to identify targets of personalised non-invasive brain stimulation. Here, we apply our method to EEG data (no brain stimulation applied), independently for each subject. Our causal candidate variables are bandpower in different frequency bands and electrode locations.

Twenty-one healthy participants were recorded with high density EEG (128 electrodes, Brain Products), during a motor task. Our paradigm consisted of 150 trials and is described in detail in Section 6.3.1.

Each trial k consisted of a planning phase (p_k^i) followed by a moving phase (m_k^i). Trials in which the subject did not reach the target within the $10s$-window are excluded from the analysis. As an input to our causal discovery algorithm, we examine the bandpower of four brain frequency bands ($\alpha:\ (8-12)$Hz, $\beta:\ (12-25)$Hz, low-γ: $(25-45)$Hz and γ: $(60-80)$Hz) and thirty-eight electrodes over the left and right primary motor cortices, the supplementary motor areas and the central sulcus. That results in $n = 4\times 38 = 142$ features. We calculate each feature i as the log-bandpower during a window of 1 s in the end of the planning phase (P^i) and in the beginning of the moving phase (M^i) for the aforementioned four canonical brain frequency bands (larger interval between the period of P^i and M^i calculation was also examined, which led to less detected causes). Finally, we quantify the response R as the natural logarithm of duration of the reaching movement in seconds (see Figure 7.1). Assumption A3 and As4 arise in a natural way from an EEG set-up: There is a time ordering between the brain states P^i, M^i and R; that is why the measured response R cannot affect the preceding brain state (Assumption A3). In addition, we assume that the previous state of brain feature i (P^i) is a cause of its current brain state M^i (Assumption As4).

Preprocessing of EEG data: Before the bandpower calculation, to attenuate non-cortical artifacts in the EEG data we followed a standardised procedure often applied in this field (Grosse-Wentrup and Schölkopf, 2012; Frølich and Dowding, 2018). We filtered the EEG signal with a Butterworth 3 Hz high-pass filter, performed common average reference filtering on all electrodes, and then performed SOBI (Belouchrani *et al.*, 1993) Independent Component Analysis (ICA) followed by manual rejection of non-cortical sources (McMenamin *et al.*, 2010), which then we re-projected on the raw signal.

7.6 Results

7.6.1 Simulated data

Figures 7.2 and 7.3 depict the percentage of false positives and false negatives over twenty random graphs, for each combination of number of M^i nodes n, samples and sparsity of edges. The existence of an edge between the nodes of our simulated graphs is defined by a Bernoulli distribution with probability $p = 0.2, 0.3, 0.4$ and 0.5 respectively. As shown in detail in Fig. 7.2, the false positives occurring due to statistical error in the computation of the dependencies and conditional independences are very few (never exceeds 4%), tending to decrease with more samples. Clearly, the probability of false positives increases with the number of nodes. The number of false negatives (Fig. 7.3) appears inflated because we consider as true causes both the direct and the indirect ones. Therefore, if only the direct cause is correctly identified by our algorithm, then its ancestors which are indirect causes will be counted as false negatives. That explains why the number of false negatives increases with the number of features n and the density of the graph.

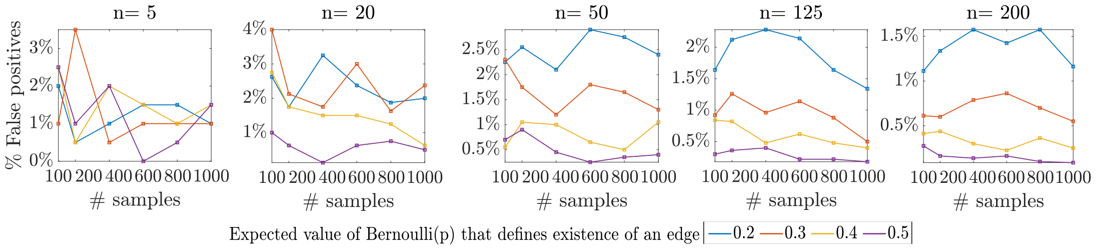


Figure 7.2: Percentage of false positives calculated on the number of n candidate features, over 20 random simulated graphs, for different number of i candidate features ($n = 5, 20, 50, 125, 200$), different Bernoulli probability to define sparsity and different number of samples (100, 200, 300, 400, 600, 800, 1000).

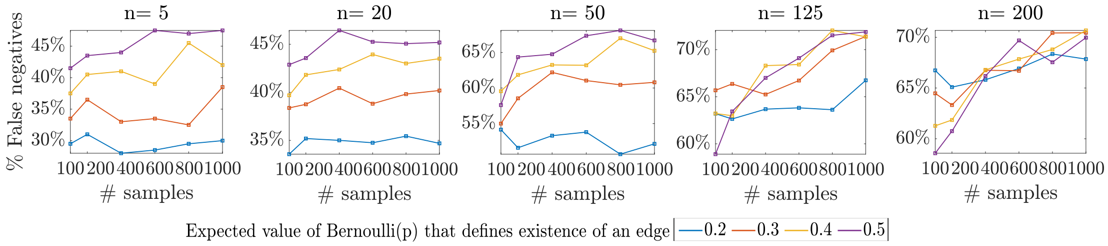


Figure 7.3: Percentage of false negatives calculated on the number of n candidate features, over 20 random simulated graphs, for different number of i candidate features ($n = 5, 20, 50, 125, 200$), different Bernoulli probability to define sparsity and different number of samples (100, 200, 300, 400, 600, 800, 1000).

Comparison with Markov Blanket methods and LASSO non-linear regression: In the simulated data, in sparse large graphs Forward Selection gave more false positives (table 7.1). LASSO non-linear regression and Forward Selection gave more false positives in small sparse and dense graphs. Backward Elimination performed worse in small sparse graphs. Overall, our method managed to keep the false positive rate very low ($\sim 2.1\%$) for all dense/sparse, small/large graphs, while other algorithms' performance varied with the case. Optimal parameters based on the true number of causes was selected for LASSO. In terms of computational time, Backward Elimination and Forward Selection took significantly long. Furthermore, we stress that in these simulations no hidden variables exist, which is an extra advantage for the algorithms with which we compare our methods.

Table 7.1: Comparison of false positive and false negative rates calculated in 10 random simulated graphs, among our method and Forward Selection, Backward elimination for Markov blanket detection and HSIC LASSO.

	FP(%)	FN(%)	FP(%)	FN(%)	FP(%)	FN(%)	FP(%)	FN(%)
(nodes, sparse)	**Our method**		**Hsic LASSO**		**BE hsic**		**FS hsic**	
(20,.2)	**3.5**	31.5	9.5	22.5	11	23	6	25
(20,.5)	**2**	80	5.5	47.5	1.5	79	7.5	26
(125,.2)	**2.9**	70.3	1.1	77.4	1.4	77.9	7.8	47.4
(125,.5)	**0**	80.8	0	84.8	0	97.6	1.1	14.5

7.6.2 Electroencephalographic data

Table 7.2: Detected causes for six representative subjects; two subjects for each of the three categories of detected causes: 1. β-range detected causal electrodes for inhibition of performance (AB and DC), 2.γ-range detected causal electrodes for improvement of performance (KK and II), 3. α-range detected causal electrodes over ipsilateral hemisphere (HH and JJ).

Subject	Alpha	Beta	Low Gamma	Gamma	Above Group Average	Performance
AB	-	FC2, CCP1h, CPP1h, CP6	-	-	False	Full inhibition
DC	FCC5h	CPP2h, CP5, CPz	C2, CCP2h	FC5, CCP2h	False	Inhibited but then improved
KK	C6, CP2	CCP4h, CCP3h, FCC3h, FCC5h, FCC6h, CP3	FCC2h, CP3, CP1, CP2	FC5, CCP4h, C6, CCP6h, FC6, FCC3h	False	Full improvement
II	-	-	-	FC2, FC4	False	Full improvement
HH	FC2, FCz	-	-	-	False	Improvement but then inhibited
JJ	FC4, FC6	-	-	FCC5h, CP1	False	Full improvement

We applied our method on the preprocessed EEG data described in 7.5.2, individually for each subject. In total, our algorithm identified causes in seventeen out of twenty-one subjects. Here we present results for six representative subjects in Table 7.2 and visualisation for three subjects in Figures 7.4, 7.5 and A.4.

Our causal findings are consistent across all subjects and form three categories that couple detected causes with subjects' performance: 1. γ-power is detected when subjects improve their performance, 2. β-power is detected when subjects worsen or do not improve their performance, and finally 3. α-power is detected in the ipsilateral hemisphere. The three groups are discussed in Section 7.7.

Subject AB (Fig. 7.4) and DC in Table 7.2 are two representative subjects who worsened or did not improve their movement duration throughout the sequence of trials (they needed longer times for completing the trial). Subject AB performed on average (green line) worse than the median performance of all subjects (pink line). Our algorithm detected causes over motor channels in the β-range (2nd head-plot), as well as a few in gamma range (for subject DC in table). Subjects KK and II (Fig. 7.5) improved their performance, decreasing the duration of their reaching movements throughout the trials. Our algorithm detected causes over motor channels in the γ-range (4th head-plot), for both subjects. Finally, HH and JJ (see fig. A.4) are two representative subjects for whom our algorithm detected causes over ipsilateral motor channels in the α-range. Results for each subject are presented and explained based on their performance in Section A.2 in Appendix A.

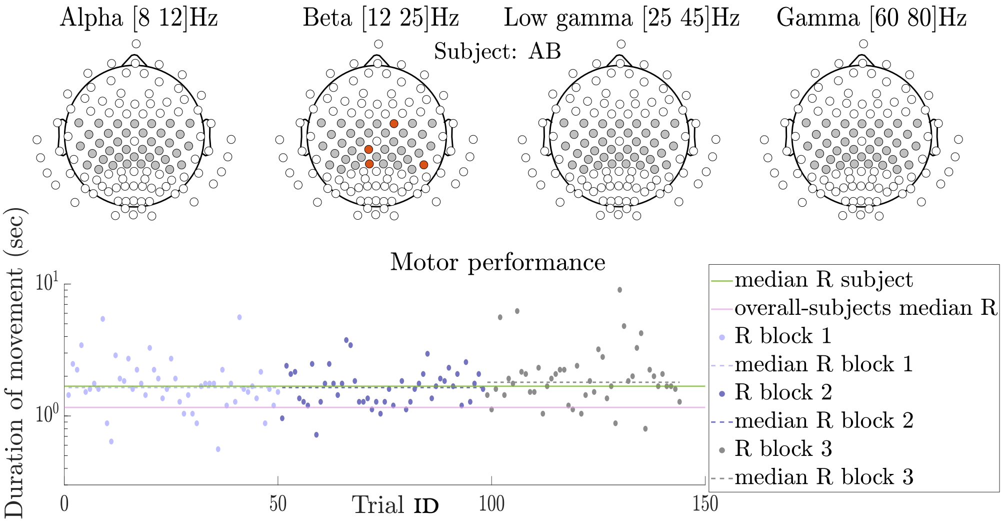


Figure 7.4: Electrodes over contralateral motor and parietal cortex in the β-range (colored red, 2nd plot) are detected as causal features from our algorithm, for subject AB, who worsened her movement duration during the reaching trials. Findings are in line with literature about the inhibitory role of beta power. Grey color depicts the motor channels we examine. The y-axis is in logarithmic scale.

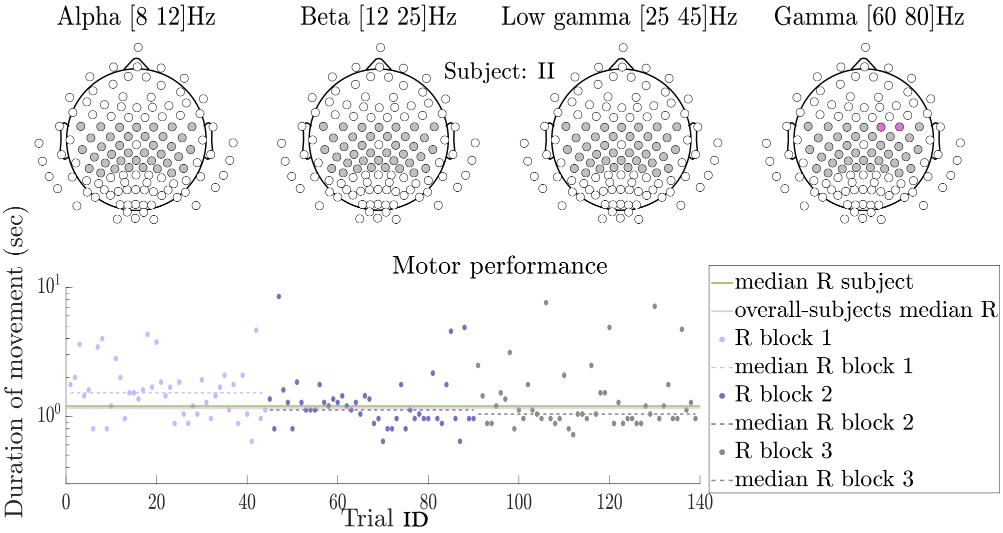


Figure 7.5: Electrodes over motor cortex in the γ-range (colored pink, 4th plot) are detected as causal features from our algorithm, for subject II, who improved her movement duration over the trials. Findings are in line with literature, as alpha activity particularely over the ipsilateral hemisphere has been associated with preparation of movement. The y-axis is in logarithmic scale.

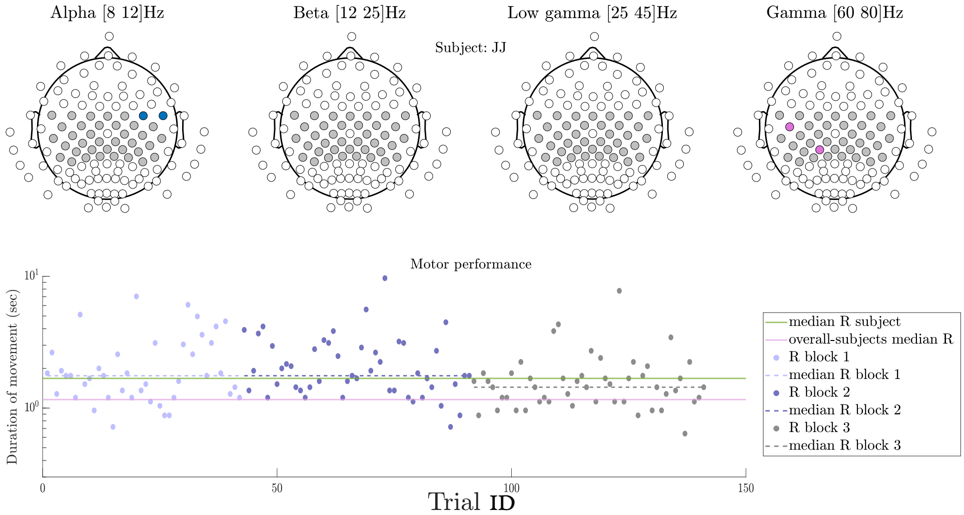


Figure 7.6: Electrodes mainly over ipsilateral motor cortex in the α-range (coloured blue, 1st plot), are detected as causal features from our algorithm in some subjects. Findings are in line with literature about the prokinetic role of gamma power. Grey color depicts the motor channels we examine. Y-axis is in logarithmic scale.

7.7 Discussion

7.7.1 Improvements upon previous methods

To the best of our knowledge, this is the first constraint based algorithm that scales linearly with the number of candidate features. Previous methods which are based on conditional independences grow exponentially in time with the number of variables, (if sparse data, then they grow polynomially), as they require more than one conditional independence test per variable. Therefore, we greatly reduce the computational complexity. Moreover, our algorithm requires only one variable as a conditioning set for each test. With this improvement, the statistical strength of our inference is superior compared to algorithms with larger conditioning sets. Furthermore, due to this improvement, we require a weaker notion of faithfulness (Uhler *et al.*, 2013), as we only assume one triplet of variables per candidate cause. Finally, our method does not assume causal sufficiency - a common assumption which is, however, often violated in real datasets.

7.7.2 Sufficient conditions for fast causal feature selection in large datasets

Our causal discovery theorem imposes assumptions that can easily be met in real datasets where candidate variables have one known cause. We proved that our proposed conditions, under Assumptions As1-As5, are sufficient for the identification of direct or indirect causes of a target variable. Thus, we can rule out the possibility that the measured dependency between the causal variable and the response is due to a confounding path, even due to a hidden variable. However, our procedure may not identify all causes (see Fig. A.1 in Section A.1 of Appendix A). Simulations yielded successful application of our algorithm with very low percentages of false positives in dense and large graphs. The robustness of our algorithm against confounders, alongside the linear scaling of complexity, renders it suitable for causal feature selection in large datasets, where false acceptance is considered much more harmful compared to false rejection.

7.7.3 Not an instrumental variable approach

Note that although our assumption about the existence of a path from P to M (Assumption As4) resembles part of the definition for instrumental variables (IV) (Pearl, 2013; Greenland, 2000), it is not. To apply our method, in contrast to IV, we do not assume any independence of variable P from unobserved variables that may affect M and R as hidden confounders, nor do we assume the lack of a directed path from P to R that does not include M ("exclusion restriction").

7.7.4 Neurophysiological validity of results

In the following paragraph, we discuss how our proposed method yields neurophysiologically plausible results in real-world data. These conclusions are derived a posteriori, and do not constitute an a-priori phrased complex hypothesis about the role of the three frequency ranges. The application of our proposed method on our EEG data gave performance-specific causes across subjects, which are consistent with the known roles of physiological α, β and γ brain rhythms in upper-limb movements. In particular, β activity has been found significantly elevated in patients with motor disorders (tremors, slowed movements) such as Parkinsons disease (McAllister *et al.*, 2013; Brown, 2007; Khanna and Carmena, 2017). Furthermore, in healthy subjects, elevated β-power has been found to play an anti kinetic role (Khanna and Carmena, 2017). Our findings support this conclusion, as we found channels in the β power to play a causal role for subjects that did not improve their motor performance. On the other hand, increased γ activity over the motor cortices has been associated with large ballistic movements (Muthukumaraswamy, 2010; Nowak *et al.*, 2018). It has also been suggested to be prokinetic, given that it is increased during voluntary movement (Brown, 2003). Our findings appear to be in accordance with this conclusion, since our method detected causal motor channels in the γ band, in subjects who managed to reduce the time length of their reaching movement. Moreover, our detected causal channels in the ipsilateral hemisphere at α-band are consistent with neurophysiological studies that report increased α-power over ipsilateral sensorimotor cortex during selection of movement (Brinkman *et al.*, 2014). Yet, no association of α-power and motor performance has been reported. Although there is no ground truth for comparing our neurophysiological results, the findings appear at least plausible given the current understanding of the aforementioned physiological brain rhythms in movement. Therefore, our method contributes to the more precise localisation of causal cortical electrode-areas.

Finally, we want to emphasise on the appropriate way of interpreting the neurophysiological results after the application of our causal feature selection method. Since EEG electrodes record mixtures of the underlying neuronal activity coming from neighbour sources, and, therefore, are macro-variables, one could argue about their adequacy for causal inference (Rubenstein *et al.*, 2017). In order to consider EEG electrodes appropriate causal candidates, we assume that the power recorded on the electrode level mostly depicts the cortical activity right underneath. We can then interpret our causal findings as the brain activity, which plays a causal role for the motor performance we observe. This causal feature detection sheds more light on the underlying cortical mechanism that acts during upper-limb movements. However, it is crucial to point out that there is not a one-to-one mapping between the causal brain features and the stimulation targets, as it is not yet fully understood how the stimulation current in a specific frequency interacts with ongoing brain oscillations. For example, as it has been shown in (Mastakouri *et al.*, 2019a) β-rhythms may act as a mediator of γ stimulation to motor performance. We can consider the problem of selecting personalised stimulation targets and frequencies as a two-step procedure: first, understand the effect of stimulation on brain activity, and second, detect the link between brain activity and motor response. In the logical graphical chain *stimulation parameters → brain activity → behavioural response*, our causal method contributes to the second link. Thus, it narrows the original question of personalised stimulation to the new problem: *stimulation parameters → detected causal brain activity*. Hence, the search for personalised stimulation parameters can now be reduced to the detection of those that up- or down-modulate accordingly, the causal brain features which our algorithm identifies.

7.7.5 Contribution

In this chapter, a new algorithm is proposed, alongside its theoretical foundation, that allows identifying direct or indirect causes of a response variable, tailored to problems in which a cause of a candidate cause is known. This can naturally happen in setups where two nodes constitute consecutive timestamps of a variable's state in a system, and an edge from the previous to the present state can be assumed. The number of required conditional independence tests is reduced to one targeted conditional independence test per variable with one conditioning variable. Therefore, the complexity of the proposed algorithm scales linearly with the number of variables. Finally, applying our algorithm on EEG data exhibited results with rigid consistency with current neuroscientific conclusions, helping to step closer towards personalised stimulation.

Chapter 8

Systematic Path Isolation (SyPI): Causal feature selection on time series

In this chapter, we present a new method for causal inference on time series. We focus on the problem of causes identification in setups where each variable is a time series, and propose both necessary and sufficient conditions for causal feature selection, in datasets with latent variables. Time series are a particular type of data due to their temporal structure, which although dominate almost every real dataset (EEG signals are one example), they have not been deeply studied from the causal point of view. Our theoretical results and estimation algorithm require two conditional independence tests for each observed candidate causal time series to decide whether it is a cause of an observed target time series or not. We provide experimental results in simulated graphs, where the ground truth is known, as well as in real datasets. Our results show that our method yields essentially no false positives and relatively low false negative rates, even in confounded environments, outperforming the widely used method of Granger causality (see Definition 14 and Section 2.5) and two more methods. Finally, we propose and prove a theorem that relaxes one of the stricter assumptions of our method, rendering it more easily applicable on real datasets. The theorems and the results presented in this chapter belong to the publication (Mastakouri *et al.*, 2020) of the author of this dissertation, alongside Bernhard Schölkopf and Dominik Janzing, and to the publication (Mastakouri and Schölkopf, 2020) of the author alongside Bernhard Schölkopf.

8.1 Problem statement

Here we work on the generic problem of causal feature selection in time series datasets (see description in Section 2.5). We try to propose and prove conditions that are necessary as well as sufficient to identify direct and indirect causal time series of a target sink node (see definition in Subsection 2.1.1) time series. We search under which assumptions this is possible (see Problem 5).

8.2 Motivation

Causal inference on time series is a fundamental problem in data science, with applications in many fields, such as economics, machine monitoring, biology and climate research. It is also a problem for which no overall solution has been found yet.

While Granger causality (Wiener, 1956; Granger, 1969, 1980) (see definition in 14) has been the standard approach to causal analysis of time series data during the last fifty years, several issues caused by violations of its assumptions, including causal sufficiency and no instantaneous effects, have been described in the literature (Peters *et al.* (2017) and references therein). Several approaches (Hung *et al.*, 2014; Guo *et al.*, 2008) addressing these problems have been proposed during the last decades. Nevertheless, it is fair to say that causal inference in time series is still a challenging problem – despite the fact that the time order of variables provide an additional information about the causal direction (Pearl, 2009; Spirtes *et al.*, 2000). Causal discovery is largely based on the graphical criterion of d-separation by formalizing the set of conditional independences to be expected based on causal faithfulness and the causal Markov condition (Spirtes *et al.*, 2000), see definition 1. Theorem 10.7 in (Peters *et al.*, 2017) shows how Granger causality can be derived from d-separation. Furthermore, several authors showed how to derive d-separation based causal conclusions in time series beyond Granger's work. Entner and Hoyer (2010a) and Malinsky and Spirtes (2018a), for instance, are inspired by the FCI algorithm (Spirtes *et al.*, 2000) and the work from Eichler (2007) without assuming causal sufficiency, aiming at the full graph causal discovery (for an extended review see (Runge, 2018; Runge *et al.*, 2019b)). However these methods can become unreliable in large graphs due to the heavy statistical testing with large conditioning sets. In (Runge *et al.*, 2019a) the PCMCI method is proposed and, although lower rates of false positives are reported compared to classical Granger causality (see definition 14 in Appendix B), the method still relies on the assumption of causal sufficiency. We give an extensive comparison of the aforementioned methods, as well as the more relevant (Pfister *et al.*, 2019) one in Section 8.6.2.

In the current chapter the problem of causal feature selection in time series is being studied. By this term, we mean the detection of direct and indirect causes of a given target time series. We construct and present conditions which, subject to appropriate graph connectivity assumptions, we prove to be sufficient for the identification of direct and indirect causes and necessary for direct causal time series of a target, even in the presence of latent variables. In contrast to approaches inspired by conditional independence based algorithms for causal discovery (like PC and FCI (Spirtes *et al.*, 2000)), our method directly constructs the right conditioning sets of variables, without *searching* over a large set of possible combinations. This is achieved by a pre-processing step that identifies the nodes of the time series that enter the previous time step of the target node. This way it avoids statistical issues of multiple hypothesis testing.

We provide experiments with simulated data, examining scenarios with different number of time series, density of edges, number of hidden variables, noise levels and sample sizes. Our results demonstrate that our method leads to essentially no false positives and relatively low false negative rates, even in confounded environments, thus outperforming Granger causality. We refer to our method as *SyPI* as it performs a Systematic Path Isolation approach for causal feature selection in time series.

8.3 Methods

8.3.1 Formal problem description

We are given observations from a target time series $Y := (Y_t)_{t\in\mathbb{Z}}$ whose causes we wish to find, and observations from a multivariate time series $\mathbf{X} := ((X_t^1,\ldots,X_t^d))_{t\in\mathbb{Z}}$ of potential causes (candidate time series). Moreover, we allow an unobserved multivariate time series $\mathbf{U_t} := (U_t^1,\ldots,U_t^m)$, which may act as common cause of the observed ones. The system consisting of $\mathbf{X}$ and Y is not assumed to be causally sufficient, hence we allow for unobserved variables $\mathbf{U_t}$. We introduce the following terminology to describe the causal relations between these variables:

8.3.2 Terminology & Notation

T1 “full time graph”: the infinite DAG having X_t^i, Y_t and U_t^j as nodes.

T2 “summary graph” is the directed graph with nodes $(X^1,...,X^d,U^1,...,U^d,Y) =: \mathbf{Q}$ containing an arrow from Q^j to Q^k for $j \neq k$ whenever there is an arrow from Q_t^j to Q_s^k for $t \leq s \in Z$. (Peters *et al.*, 2017)

T3 “$Q_t^i \to Q_s^j$” for $t \leq s \in Z$ means a directed path that does not include any intermediate observed nodes in the full time graph (confounded or unconfounded).

T4 “$Q_t^i \dashrightarrow Q_s^j$” for $t \leq s \in Z$ in the full time graph means a directed path from Q_t^i to Q_s^j.

T5 “confounding path”: A confounding path between Q_t^i and Q_s^j in the full time graph is a path of the form $Q_t^i \dashleftarrow Q_{t'}^k \dashrightarrow Q_s^j$, $t' \leq t, s \in Z$ consisting of two directed paths and a common cause of Q_t^i and Q_s^j.

T6 “confounded path”: an arbitrary path between two nodes Q_t^i and Q_s^j in the full time graph which co-exists with a confounding path between Q_t^i and Q_s^j.

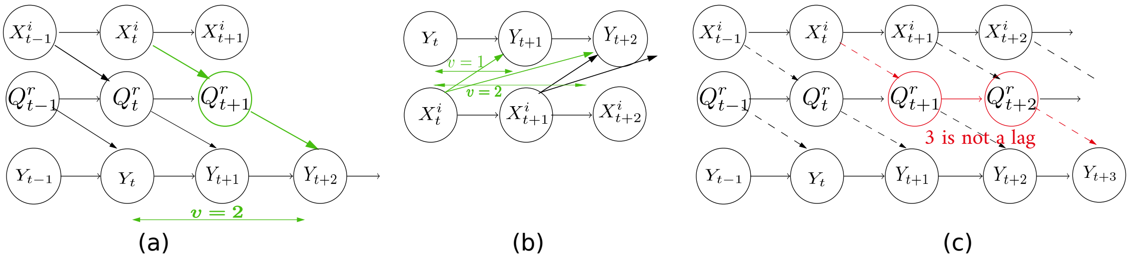
T7 “sg-unconfounded” (summary-graph-unconfounded) causal path: A causal path in the full time graph that does not appear as a confounded path in the summary graph

T8 “pb-unconfounded” (past-blocked-unconfounded) causal path: A causal path between two nodes Q_t^i and Q_s^j in the full time graph for which all confounding paths that do not contain more than one time step (nodes) from each of the $Q^k, k \neq i, j$ time series, are blocked by $Q_{t'}^i$ or $Q_{s'}^j, t' < t, s' < s$.

T9 "lag": v is a lag for the ordered pair of a time series X^i and the target Y (X^i, Y) if there exists a collider-free path X^i_t- - -Y_{t+v} that does not contain a link of this form $Q^r_{t'} \to Q^r_{t'+1}$, with t' arbitrary, for any $r \not\equiv i, j$, nor any duplicate node, and any node in this path does not belong to X^i, Y. See explanatory Figure 8.1.

T10 "single-lag dependencies": We say that a set of time series $(\mathbf{X}, Y)$ have "single-lag dependencies" if all the $X^i \in \mathbf{X}$ have only one lag v for each pair X^i, Y. Otherwise we refer to "multiple-lag dependencies".

As we define in T9, an integer v is a lag for the ordered pair of a time series X^i and the target Y (X^i, Y) if there exists a collider-free path X^i_t- - -Y_{t+v} that does not contain a link of this form $Q^r_{t'} \to Q^r_{t'+1}$ with t' arbitrary, for any $r \not\equiv i, j$, nor any duplicate node, and no other node in this path other than X^i_t, Y_{t+v} belongs in X^i, Y. Figure 8.1 shows some example graphs and the lags between the candidate and the target time series, based on the definition T9. The integers defined by the highlighted green path between X^i and Y in graphs (a) and (b) are example lags for the singla-lag (a) and multi-lag graph (b) accordingly, while the path in (c) does not define a lag because it contains a link $Q^r_{t+1} \to Q^r_{t+2}$. If the links between the time series were direct links, then the correct lag for (X^i, Y) in (c) would be 2.



(a) (b) (c)

Figure 8.1: In (a) we have a single lag depedendency graph, and the integer 2 is the lag for (X^i, Y). (b) shows a multi-lag dependency graph where both integers 1 and 2 are lags for (X^i, Y). On the contrary, the red coloured path in (c) that corresponds to the integer 3 is not a lag, because it contains the link $Q^r_{t+1} \to Q^r_{t+2}$.

Having introduced the necessary terminology, we assume that the graph satisfies the following assumptions. Note that the first four are standard assumptions of time series analysis and causal discovery, while assumptions A5 - A9 impose some kind of restrictions on the connectivity of the graph.

8.3.3 Assumptions

A1 **Causal Markov condition** [1] in the full time graph

[1] see definition 2

A2 **Causal Faithfulness** in the full time graph [2]

A3 **No backward arrows** in time $X^i_{t'} \not\to X^j_t, \forall t' > t$

A4 **Stationary** full time graph: the full time graph is invariant under a joint time shift of all variables

A5 The full time graph is **acyclic**.

A6 The **target** time series is a **sink node** in the summary graph; it does not affect any other variables in the graph.

A7 There is an arrow $X^i_{t-1} \to X^i_t, Y_{t-1} \to Y_t \forall i, t \in \mathbb{Z}$. Note that arrows $U^i_{t-1} \to U^i_t$ need not exit, we then call U memoryless.

A8 There are no arrows $Q^i_{t-s} \to Q^i_t$ for $s > 1$.

A9 Every variable U^i that affects Y **directly** (no intermediate observed nodes in the path in the summary graph) or that is connected with an observed collider in the summary graph should be memoryless ($U^i_{t-1} \not\to U^i_t$) and should have single-lag dependencies with Y in the full time graph.[3]

Below, we present three theorems for detection of causes in the full time graph. **Theorem 2a** provides **sufficient conditions for direct and indirect sg-unconfounded causes in single-lag dependency graphs**. **Theorem 2b provides sufficient conditions for direct and indirect causes in multi-lag dependency graphs. Theorem 3** provides **necessary conditions for identifying all the direct sg-unconfounded causes** of a target time series, assuming the imposed graph constraints.

8.3.4 Intuition for proposed Theorems

Intuition for proposed conditions in Theorems 2a/2b and 3: The idea is to *isolate* the path $X^i_{t-1} \to X^i_t$ - -$Q^j_{t'} \dashrightarrow Y_{t+w_i}, w_i \in Z, t' < t + w_i$ in the full time graph, in order to be able to examine similar conditions proposed in (Mastakouri *et al.*, 2019b). In this path "- -" means $\dashleftarrow$ or $\dashrightarrow$ and $Q^j_{t'}$ (if observed) in addition to any other intermediate variable in the path X^i_t - -$Q^j_{t'} \dashrightarrow Y_{t+w_i}$ must $\not\in \{X^i, Y\}$. Mastakouri *et al.* (2019b) proposed sufficient conditions for causal feature selection in a DAG with non-sequential data, where a cause of a potential cause was known or could be assumed due to time-ordered pair of variables.

Here the goal is to propose necessary and sufficient conditions that will differentiate between $Q^j_{t'}$ being a common cause or - -$Q^j_{t'} \dashrightarrow$ being a (in)direct edge to Y_{t+w_i} in the full time graph.

[2] see definition 5

[3] Note that this assumption is only required for the completeness of the algorithm against direct false negatives (Theorem 3). The violation of this assumption does not spoil Theorem 2a/2b. The existence of a **latent variable with memory** affecting the target time series Y **directly**, or of a **latent variable affecting directly the target with multiple lags** renders impossible the existence of a conditioning set that could d-separate the future of the target variable and the past of any other observed variable.

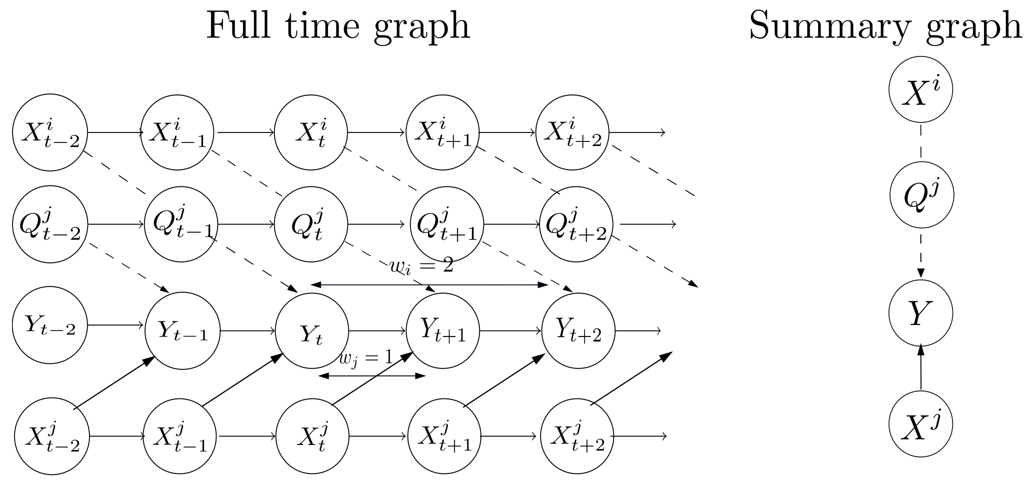


Figure 8.2: Visualization of a simple full time graph of two observed, one potentionally hidden and one target time series. The summary graph is presented to emphasize that the notions of “pb-confounded” and “sg-confounded” are different and to point out the challenge of identifying sg-unconfounded causal relations in time series, where the past of each time series introduces dependencies that are not obvious in the summary graph.

Figure 8.2 visualizes why time-series raise an additional challenge for identifying sg-unconfounded causal relations. While the influence of X^j on Y is unconfounded in the summary graph, the influence $X^j_t \to Y_{t+1}(\equiv Y_{t+w_j})$ is confounded in the full time graph due to its own past; for example X^j_t and Y_t are confounded by X^j_{t-1}. Therefore we need to condition on $Y_t(\equiv Y_{t+w_j-1})$ to remove past dependencies. If no other time series were present, that would be sufficient. However, in the presence of other time series affecting the target Y, Y_{t+w_j-1} becomes a collider that unblocks dependencies. If for example we want to examine X^i as a candidate cause, we need first to condition on $Y_{t+w_i-1} \equiv Y_{t+1}$, the past of the Y_{t+w_i}. Following, we need to condition on node from each time series $\mathbf{X} \setminus X^i$ that enter $Y_{t+w_i-1} \equiv Y_{t+1}$ (which is a collider) to avoid all the dependencies that might be created by conditioning on it. It is enough to condition only on these nodes for the following reason: If a node $X^{j \neq i}$ has a w_j lag-dependency with Y, then there is an (un)directed path from $X^j_{t+w_{ij}-1}$ to Y_{t+w_i-1}. If this path is a confounding one, then conditioning on $X^j_{t+w_{ij}-1}$ is not necessary, but also not harmful because the future of this time series in the full graph is still independent of Y_{t+w_i}. This independence is forced by the fact that the $X^j_{t+w_{ij}}$ is a collider because of the stationarity of graphs and this collider is by construction *not* in the conditioning set. If $X^j, j \neq i$ is connected with Y_{t+w_i-1} via a directed link (as in Figure 8.2), then conditioning on $X^j_{t+w_{ij}-1}$ is necessary to block the parallel path created by its future values $X^j_{t+w_{ij}-1} \to X^j_{t+w_{ij}} \dashrightarrow Y_{t+v}$. Based on this idea of isolating the path of interest, we build the conditioning set as described in Theorem

2a/2b and its converse Theorem 3, where we prove the necessity and sufficiency of our conditions.

8.3.5 Theorems

Theorem 2a. *[Sufficient conditions for a direct or indirect sg-unconfounded cause of Y in single-lag dependency graphs] Assuming A1-A5 and single-lag dependency graphs, let w_i be the minimum lag (see T9) between X^i and Y. Further, let $w_{ij} := w_i - w_j$. Then, for every time series $X^i \in \mathbf{X}$ we define a conditioning set $\mathbf{S^i} = \{X^1_{t+w_{i1}-1}, X^2_{t+w_{i2}-1}, \ldots, X^{i-1}_{t+w_{ij}-1}, X^{i+1}_{t+w_{ij}-1}, \ldots, X^n_{t+w_{in}-1}\}$.*

If

$$X^i_t \not\perp\!\!\!\perp Y_{t+w_i} \mid \{\mathbf{S^i}, Y_{t+w_i-1}\} \tag{1}$$

and

$$X^i_{t-1} \perp\!\!\!\perp Y_{t+w_i} \mid \{\mathbf{S^i}, X^i_t, Y_{t+w_i-1}\} \tag{2}$$

are true, then

$$X^i_t \dashrightarrow Y_{t+w_i}$$

and the path between the two nodes is sg-unconfounded.

Proof. **(Proof by contradiction)**
We need to show that in single-lag dependency graphs, if $X^i_t \not\dashrightarrow Y_{t+w_i}$ or if the path $X^i_t \dashrightarrow Y_{t+w_i}$ is sg-confounded then at least one of the conditions (1) and (2) is violated.

First assume that there is no directed path between X^i_t and Y_{t+w_i}: $X^i_t \not\dashrightarrow Y_{t+w_i}$. Then, there is a confounding path $X^i_t \dashleftarrow Q^j_{t'} \dashrightarrow Y_{t+w_i}, t' \leq t$ without any colliders. (Colliders cannot exist in the path by the definition of the lag T9.) In that case we will show that either condition (1) or (2) is violated. If all the existing confounding paths $X^i_t \dashleftarrow Q^j_{t'} \dashrightarrow Y_{t+w_i}, t' \leq t$ contain an observed confounder $Q^j_{t'} \equiv X^j_{t'} \in \{\mathbf{S^i}, Y_{t+w_i-1}\}$ (there can be only one confounder since in this case there are no colliders in the path), then condition (1) is violated, because we condition on $X^j_{t'}$ which d-separates X^i_t and Y_{t+w_i}. If in all the existing confounding paths the confounder node $Q^j_{t'} \notin \{\mathbf{S^i}, Y_{t+w_i-1}\}, t' \leq t$ but some observed non-collider node is in the path and this node belongs to $\{\mathbf{S^i}, Y_{t+w_i-1}\}$, then condition (1) is violated, because we condition on $\mathbf{S^i}$ which d-separates X^i_t and Y_{t+w_i}. If there is at least one confounding path and its confounder node does no belong in $\{\mathbf{S^i}, Y_{t+w_i-1}\}$ and no other observed (non-collider or descendant of collider) node which is in the path belongs in $\{\mathbf{S^i}, Y_{t+w_i-1}\}$ then condition (2) is violated for the following reasons: Let's name $p1 : X^i_t \dashleftarrow Q^j_{t'} \dashrightarrow Y_{t+w_i}, t' \leq t$. We know the existence of the path $p2 : X^i_{t-1} \rightarrow X^i_t$, due to assumption A7.

(1I) If $p1$ and $p2$ have X^i_t in common, then X^i_t is a collider. Thus, adding X^i_t in the conditioning set would unblock the path between X^i_{t-1} and Y_{t+w_i}.

(1II) If $p1$ and $p2$ have X^i_{t-1} in common, that means X^i_{t-1} lies on $p1$. Thus X^i_t is not in the path from X^i_{t-1} to Y_{t+w_i} and hence adding X^i_t to the conditioning set could not d-separate X^i_{t-1} and Y_{t+w_i}.

In both cases condition (2) is violated.
Now, assume that there is a directed path $X^i_t \dashrightarrow Y_{t+w_i}$ but it is "sg-confounded" (there exist also a parallel confounding path $p3 : X^i_t \dashleftarrow Q^j_{t'} \dashrightarrow Y_{t+w_i}, t' \leq t$. Then, if $p3$ and $p2$ have X^i_t in common, then condition (2) is violated due to (1I). If $p3$ and $p2$ have X^i_{t-1} in common, then condition (2) is violated due to (1II). In all the above cases we show that if conditions (1) and (2) hold true in single-lag dependency graphs, then X^i_t is an "sg-unconfounded" direct or indirect cause of Y_{t+w_i}. □

Theorem 2b. *[Sufficient conditions for a (possibly confounded) direct or indirect cause of Y in multi-lag dependency graphs] Assuming A1-A5, and allowing multi-lag dependency graphs, let w_i be the minimum lag (see T9) between X^i and Y. Further, let $w_{ij} := w_i - w_j$. Then, for every time series $X^i \in \mathbf{X}$ we define a conditioning set $\mathbf{S^i} = \{X^1_{t+w_{i1}-1}, X^2_{t+w_{i2}-1}, ..., X^{i-1}_{t+w_{ij}-1}, X^{i+1}_{t+w_{ij}-1}, ..., X^n_{t+w_{in}-1}\}$.*

If conditions (1) and (2) of Theorem 2a hold true for the pair X^i_t, Y_{t+w_i}, then

$$X^i_t \dashrightarrow Y_{t+w_i}$$

We can think of $\mathbf{S^i}$ as the set that contains only one node from each time series X^j and this node is the one that enters the node Y_{t+w_i-1} due to a directed or confounded path (if w_j exists then the node is the one at $t + w_{ij} - 1$).

Proof. **(Proof by contradiction)**
We need to show that in multi-lag dependency graphs, if $X^i_t \not\dashrightarrow Y_{t+w_i}$ then at least one of the conditions 1 and 2 is violated.

First assume that there is no directed path between X^i_t and Y_{t+w_i}: $X^i_t \not\dashrightarrow Y_{t+w_i}$. Then, there is a confounding path $X^i_t \dashleftarrow Q^j_{t'} \dashrightarrow Y_{t+w_i}, t' \leq t$ without any colliders. (Colliders cannot exist in the path by the definition of the lag T8.)In that case we will show that either condition 1 or 2 is violated. If all the existing confounding paths $X^i_t \dashleftarrow Q^j_{t'} \dashrightarrow Y_{t+w_i}, t' \leq t$ contain an observed confounder $Q^j_{t'} \equiv X^j_{t'} \in \{\mathbf{S^i}, Y_{t+w_i-1}\}$ (there can be only one confounder since in this case there are no colliders in the path), then condition 1 is violated, because we condition on $X^j_{t'}$ which d-separates X^i_t and Y_{t+w_i}. If in all the existing confounding paths the confounder node $Q^j_{t'} \notin \{\mathbf{S^i}, Y_{t+w_i-1}\}, t' \leq t$ but some observed non-collider node is in the path and this node belongs to $\{\mathbf{S^i}, Y_{t+w_i-1}\}$, then condition 1 is violated, because we condition on $\mathbf{S^i}$ which d-separates X^i_t and Y_{t+w_i}. If there is at least one confounding path and its confounder node does no belong in $\{\mathbf{S^i}, Y_{t+w_i-1}\}$ and no other observed (non-collider or descendant of collider) node which is in the path belongs in $\{\mathbf{S^i}, Y_{t+w_i-1}\}$ then condition 2 is violated for the following reasons: Let's name $p1 : X_t^i \dashleftarrow Q_{t'}^j \dashrightarrow Y_{t+w_i}, t' \leq t$. We know the existence of the path $p2 : X_{t-1}^i \rightarrow X_t^i$, due to assumption A7.

(1I) If $p1$ and $p2$ have X_t^i in common, then X_t^i is a collider. Therefore, adding X_t^i in the conditioning set would unblock the path between X_{t-1}^i and Y_{t+w_i}.

(1II) If $p1$ and $p2$ have X_{t-1}^i in common, that means X_{t-1}^i lies on $p1$. In this case X_t^i is not in the path from X_{t-1}^i to Y_{t+w_i} and hence adding X_t^i to the conditioning set could not d-separate X_{t-1}^i and Y_{t+w_i}.

In both cases condition 2 is violated. □

Remark 1. *Theorem 2b conditions hold for any lag as defined in T9; not only for the minimum lag. The reason why we refer to the minimum lag in 2b is to have conditions closer to its converse Theorem 3.*

Theorem 3. *[Necessary conditions for a direct sg-unconfounded cause of Y in single-lag dependency graphs]*

Let the assumptions and the definitions of Theorem 2a hold, in addition to Assumptions A6-A9.

If X_t^i is a direct, "sg-unconfounded" cause of Y_{t+w_i} ($X_t^i \rightarrow Y_{t+w_i}$), then conditions (1) and (2) of Theorem 2a hold.

Proof. **(Proof by contradiction)**
Assume that the direct path $X_t^i \rightarrow Y_{t+w_i}$ exists and it is unconfounded. Then, condition (1) is true. Now assume that condition (2) does not hold. This would mean that the set $\{\mathbf{S^i}, X_t^i, Y_{t+w_i-1}\}$ does not d-separate X_{t-1}^i and Y_{t+w_i}. Note that a path p is said to be *d-separated* by a set of nodes in Z if and only if p contains a chain or a fork such that the middle node is in Z, or if p contains a collider such that neither the middle node nor any of its descendants are in the Z. Hence, a violation of condition (2) would imply that (a) there is some middle node or descendant of a collider in $\{\mathbf{S^i}, X_t^i, Y_{t+w_i-1}\}$ and no non-collider node in this path belongs to this set, or (b) that there is a collider-free path between X_{t-1}^i and Y_{t+w_i} that does not contain any node in $\{\mathbf{S^i}, X_t^i, Y_{t+w_i-1}\}$.

(a) *There is some middle node or descendant of a collider in $\{\mathbf{S^i}, X_t^i, Y_{t+w_i-1}\}$ and no non-collider node in this path belongs to this set:*
(a1:) If there is at least one path $p1 : X_{t-1}^i \dashrightarrow Y_{t+w_i-1} \dashleftarrow Y_{t+w_i}$ where Y_{t+w_i-1} is a middle node of a collider and none of the non-collider nodes in the path belongs to $\{\mathbf{S^i}, X_t^i\}$: Such a path could be formed only if in addition to X^i some $Q_{t'}^j$ directly caused Y. Then $p1 : X_{t-1} \dashrightarrow Y_{t+w_i-1} \dashleftarrow Q_{t'}^j \rightarrow Y_{t+w_i}, t' \leq t+w_i$. (Due to our assumption for single-lag dependencies (see T10) a path of the form $X_{t-1} \dashrightarrow Y_{t+w_i-1} \dashleftarrow X_s^i - -Y_{t+w_i}$ could not exist). Then, due to stationarity of graphs the node $Q_{t'-1}^j$ will enter Y_{t+w_i-1}. If this $Q_{t'}^j$ is hidden ($Q_{t'}^j \equiv U_{t'}^j$), then due to assumption A9 this time series will be memoryless ($U_{t'-1}^j \not\rightarrow U_{t'}^j$). Therefore, the

collider Y_{t+w_i-1} in the conditioning set will not unblock any path between X^i_{t-1} and Y_{t+w_i} that could contain $U^j_s, s > t'$. If $Q^j_{t'}$ is observed ($Q^j_{t'} \equiv X^j, j \neq i$) then due to assumption A7 the path $p1$ will be $X^i_{t-1} \dashrightarrow Y_{t+w_i-1} \dashleftarrow X^j_{t+w_{ij}-1} \rightarrow X^j_{t+w_{ij}} \dashrightarrow Y_{t+w_i}$. However, this path is always blocked by $X^j_{t+w_{ij}-1} \in \mathbf{S^i}$ due to the rule we use to construct $\mathbf{S^i}$. That means a non-collider node in the conditioning set will necessarily be in the path $p1$, which contradicts the original statement.

(a2:) If there is at least one path $p2 : X^i_{t-1} \dashrightarrow X^i_t \dashleftarrow Y_{t+w_i}$ *where* X^i_t *is a middle node of a collider and none of the non-collider nodes in the path belongs to* $\{\mathbf{S^i}, Y_{t+w_i-1}\}$: This could only mean that there is a confounder between the target Y_{t+w_i} and X^i_t. However this contradicts that $X^i_t \rightarrow Y_{t+w_i}$ is "sg-unconfounded".

(a3:) If there is at least one path $p3 : X^i_{t-1} \dashrightarrow X^j_{t'} \dashleftarrow Y_{t+w_i}$ *where* $X^j_{t'} \in \mathbf{S^i}$ *with* $t' \leq t + w_i - 1$ *is a middle node of a collider and no non-collider node in the path belongs to* $\{\mathbf{S^i} \setminus X^j_{t'}, X^i_t, Y_{t+w_i-1}\}$: In this case, $t' \equiv t + w_{ij} - 1$ because $X^j_{t'} \in \mathbf{S^i}$. By construction of $\mathbf{S^i}$ all the observed nodes in $\mathbf{X} \setminus X^i$ that enter the node Y_{t+w_i-1} belong in $\mathbf{S^i}$. That means that $X^j_{t'}$ enters the node Y_{t+w_i-1}. Hence, in the path $p3$ Y_{t+w_i-1} will necessarily be a non-collider node which belongs to the conditioning set. This contradicts the original statement "and no non-collider node in the path belongs to $\{\mathbf{S^i} \setminus X^j_{t'}, X^i_t, Y_{t+w_i-1}\}$".

(a4:) If a descendent D of a collider G in the path $p4 : X^i_{t-1} \dashrightarrow G \dashleftarrow C \dashrightarrow Y_{t+w_i}$ *belongs to the conditioning set* $\{\mathbf{S^i}, X^i_t, Y_{t+w_i-1}\}$ *and no non-collider node in the path belongs to it*: Due to the single-lag dependencies assumption, $w_C \equiv w_i$ otherwise there are multiple-lag effects from C to Y. That means that, independent of C being hidden or not, the C in the collider path will enter the node Y_{t+w_i-1}. If $C \in \mathbf{X}$ then because C enter the node Y_{t+w_i-1}, $C \in \{\mathbf{S^i}, X^i_t, Y_{t+w_i-1}\}$. In the first case Y_{t+w_i-1} only and in the latter case also C are a non-collider variable in the path $p4$ that belongs to the conditioning set, which contradicts the statement of (a4). If the collider $G \in \mathbf{X}$, as explained in (a3) at least one non-collider variable in the path will belong in the conditioning set, which contradicts the statement (a4). Finally, if G and C are hidden, if $w_D \equiv w_C$ then the node Y_{t+w_i-1} is necessarily in the path as a pass-through node, which contradicts the statement (a4). If $w_D \not\equiv w_C$ then the single-lag assumption is violated.

(b) *There is a collider-free path between* X^i_{t-1} *and* Y_{t+w_i} *that does not contain any node in* $\{\mathbf{S^i}, X^i_t, Y_{t+w_i-1}\}$:
Such a path would imply the existence of a hidden confounder between X^i_{t-1} and Y_{t+w_i} or the existence of a direct edge from X_{t-1} to Y_{t+w_i}. The former cannot exist because we know that X_t is an sg-unconfounded direct cause of Y_{t+w_i}. The latter would imply that there are multiple lags of direct dependency between X_t and Y_{t+w_i} which contradicts the assumption of single-lag dependencies.

Therefore we showed that whenever $X_t^i \to Y_{t+w_i}$ is an sg-unconfounded causal path, conditions (1) and (2) are necessary. □

Since it is unclear how to identify the lag in T9, we introduce the following lemmas for the detection of the minimum lag that we require in the theorems. We provide the proofs of the lemmas in Appendix B Section B.1).

Lemma 1. *If the paths between X^j and Y are directed then the minimum lag w_j as defined in T9 coincides with the minimum non-negative integer w'_j for which $X_t^j \not\perp\!\!\!\perp Y_{t+w'_j} \mid X^j_{past(t)}$. The only case where $w'_j \not\equiv w_j$ is when there is a confounding path between X^j and Y that contains a node from a third time series with memory. In this case $w'_j = 0$.*

Lemma 2. *Theorems 2a/2b and 3 are valid if the minimum lag w_j as defined in T9 is replaced with w'_j obtained in lemma 1.*

At this point, we relax the assumption A6 with the following Theorem 4, by stating that the above two theorems still apply even if the target time series is not a sink node, but instead none of its descendants belongs in its candidate causes.

Theorem 4 (Theorems 2a and 3 still apply)**.** *Given the target time series Y and the candidate causes $\mathbf{X}$, assuming A1 - A9, if the target Y is not a sink node, but, instead, none of each direct or indirect descendants belongs in $\mathbf{X}$: $\mathbf{DE}_Y^{\mathcal{G}} \notin \mathbf{X}$, then Theorem 2a and 3 still apply. That means the conditions of Theorem 2a are still sufficient for identifying direct and indirect causes, and conditions of Theorem 3 are still necessary for identifying all the direct unconfounded causes in single-lag dependencies.*

Proof. The proof of Theorem 2a applies without changes. Regarding Theorem 3 Assume that the direct path $X_t^i \to Y_{t+w_i}$ exists and it is unconfounded. Then, condition 1 of Theorem 3 is true. Now assume that condition 2 of Theorem 3 does not hold. This would mean that the set $\{\mathbf{S^i}, X_t^i, Y_{t+w_i-1}\}$ does not d-separate X_{t-1}^i and Y_{t+w_i}. (Recall that a path p is said to be *d-separated* by a set of nodes in Z if and only if p contains a chain or a fork such that the middle node is in Z, or if p contains a collider such that neither the middle node nor any of its descendants are in the Z.) Hence, a violation of condition 2 would imply that (a) there is some middle node or descendant of a collider in $\{\mathbf{S^i}, X_t^i, Y_{t+w_i-1}\}$ and no non-collider node in this path belongs to this set, or (b) that there is a collider-free path between X_{t-1}^i and Y_{t+w_i} that does not contain any node in $\{\mathbf{S^i}, X_t^i, Y_{t+w_i-1}\}$.

(A) *There is some middle node or descendant of a collider in $\{\mathbf{S^i}, X_t^i, Y_{t+w_i-1}\}$ and no non-collider node in this path belongs to this set* $\Rightarrow$ the proof given in 8.3.5 remains unaffected if all $\mathbf{DE}_Y^{\mathcal{G}} \notin \mathbf{X}$, because any collider D or descendent of collider between some X_t^j and Y_{t+w_i} will be unobserved, therefore will not be possible to belong in the conditioning set $\{\mathbf{S^i}, X_t^i, Y_{t+w_i-1}\}$.

(B) *There is a collider-free path between X^i_{t-1} and Y_{t+w_i} that does not contain any node in* $\{\mathbf{S^i}, X^i_t, Y_{t+w_i-1}\}$ ⇒ the proof given in 8.3.5 remains unaffected.

□

Using the condition in Lemma 1 via LASSO regression and the two conditions in Theorems 2a/2b and 3 we build an algorithm to identify direct and indirect causes on time series. The input is a 2D array $\mathbf{X}$ (candidate time series) and a vector Y (target), and the output a set with indices of the time series that were identified as causes. The complexity of our algorithm is $\mathcal{O}(n)$ for n candidate time series, assuming constant execution time for the conditional independence test.

Algorithm 2: *SyPI* Algorithm for Theorems 2a/2b and 3.

```
Input: $\mathbf{X}, Y$.
Output: causes_of_R
$n_{\text{vars}}$ = shape($\mathbf{X}$, 1); causes_of_R= []
$w = min\_lags(\mathbf{X}, Y)$
for $i = 1$ to $n_{vars}$ do
    $\mathbf{S_i} = \bigcup_{j=1, j\neq i}^{n_{\text{vars}}} \{X^j_{t+w[i]-w[j]-1}\}$
    pvalue1 $= cond\_ind\_test(X^i_t, Y_{t+w[i]}, [\mathbf{S_i}, Y_{t+w[i]-1}])$
    if $pvalue1 < threshold1$ then
        pvalue2 $= cond\_ind\_test(X^i_{t-1}, Y_{t+w[i]}, [\mathbf{S_i}, X^i_t, Y_{t+w[i]-1}])$
        if $pvalue2 > threshold2$ then
            causes_of_R = [causes_of_R, $X^i_t$]
        end
    end
end
```

8.4 Experiments

8.4.1 Simulated experiments: time series construction

To test our method, we build simulated full time graphs with a varying number of hidden variables, respecting the aforementioned assumptions. We sample 100 random graphs for the following tuples of hyperparameters: (# samples, # hidden variables, # observed variables, density of edges between candidate time series, density of edges between time series and target series, noise variance). Then we report the false positive (FPR) and false negative rates (FNR) for the 100 graphs. The values that are tested for each hyperparameter in the tuple are the following: # samples ∈ $(500, 1000, 2000, 3000)$, # hidden variables $\in (0,1,2)$, # observed variables $\in (1,2,3,4,5,6,7,8)$, Bernoulli(p) existence of edge between candidate time series $\in (0.1, 0.15, 0.2, 0.25)$, Bernoulli($p$) existence of edge between candidate time series and target series $\in (0.1, 0.2, 0.3)$ and noise variance $\in (10\%, 20\%, 30\%)$. To construct the time series, every time step is calculated as the weighted sum of the previous step of all the incoming time series with the previous step of the current time series. The weights between two time series are either set to zero or they are drawn from a uniform distribution in the range $[0.7, 0.95]$ (this way we prevent too weak edges, which would result in almost non-faithfulness distributions that render the problem of detecting causes impossible).

The two conditional independence tests are calculated with partial correlation, since our simulations are linear, but there is no restriction for non-linear systems (see extension in 8.6.2). For the "lag" calculation step of our method, we use LASSO in a bivariate form between each node in $\mathbf{X}$ in the summary graph and Y (for the non-linear case LASSO could be replaced with a non-linear regressor). After some exploratory search across different values for the regularization parameter and the threshold on the coefficients of this step, we conclude that for regularization $\lambda = 0.001$ and any threshold in the region 0.1 to 0.15 for the returned coefficients of LASSO, the results are mostly stable. Thus, we fixed these two parameters once before running the experiments, without re-adjusting them for the different types of graphs.

All the aforementioned experiments were implemented with unique direct lag of 1. Although our theory is complete against false negatives only for single-lag dependencies, we wanted to test the performance of SyPI in the graphs with multiple lags. Therefore we examined the performance for four and five observed, one additional hidden and one target time series, for two, three and four coexisting lag direct effects. We decide for the existence of a lag sampling from a Bernoulli distribution with $p = 0.5$.

SyPI vs LASSO-Granger

For the final part of our simulated experiments, we compare our algorithm to LASSO-Granger (Arnold *et al.*, 2007) for two hidden and three, four and five observed time series. SyPI operates with two thresholds for the p values of the two tests, one (*threshold1*) for rejecting independence in the first condition, and a second (*threshold2*) for accepting dependency in the second condition. LASSO-Granger (Arnold *et al.*, 2007) operates with one hyper-parameter: the regularization parameter λ. To ensure a fair comparison of the two methods, we tuned the λ regularizer for LASSO-Granger (not our method) to allow for at least the same FNR as our method, for the same type of graphs. For all the aforementioned experiments apart from the comparison of the two methods, we used *threshold1*= 0.01 and *threshold2*= 0.2. Finally, we produced ROC curves for the two methods as follows: for LASSO-Granger, we varied the λ parameter across $\{0.00001, 0.0001, 0.001, 0.0025, 0.005, 0.01, 0.025, 0.05, 0.1, 0.5, 0.6, 0.7, 0.8, 0.9\}$. For our method (SyPI), we varied only *threshold1* and *threshold2*, keeping their ratio equal to 1, using values in $\{0.01, 0.02, \ldots, 0.12\}$. Note that in our simulations we allow for cycles among the candidate time series, as long as they do not include the target node.

SyPI vs seqICP and PCMCI

We performed ten experiments with twenty random graphs each, with two to six observed and one to two hidden series, for sample size 2000 and medium density. For our method we kept the same thresholds, as we defined above: *threshold1*$=0.01$ and *threshold2*$=0.2$.

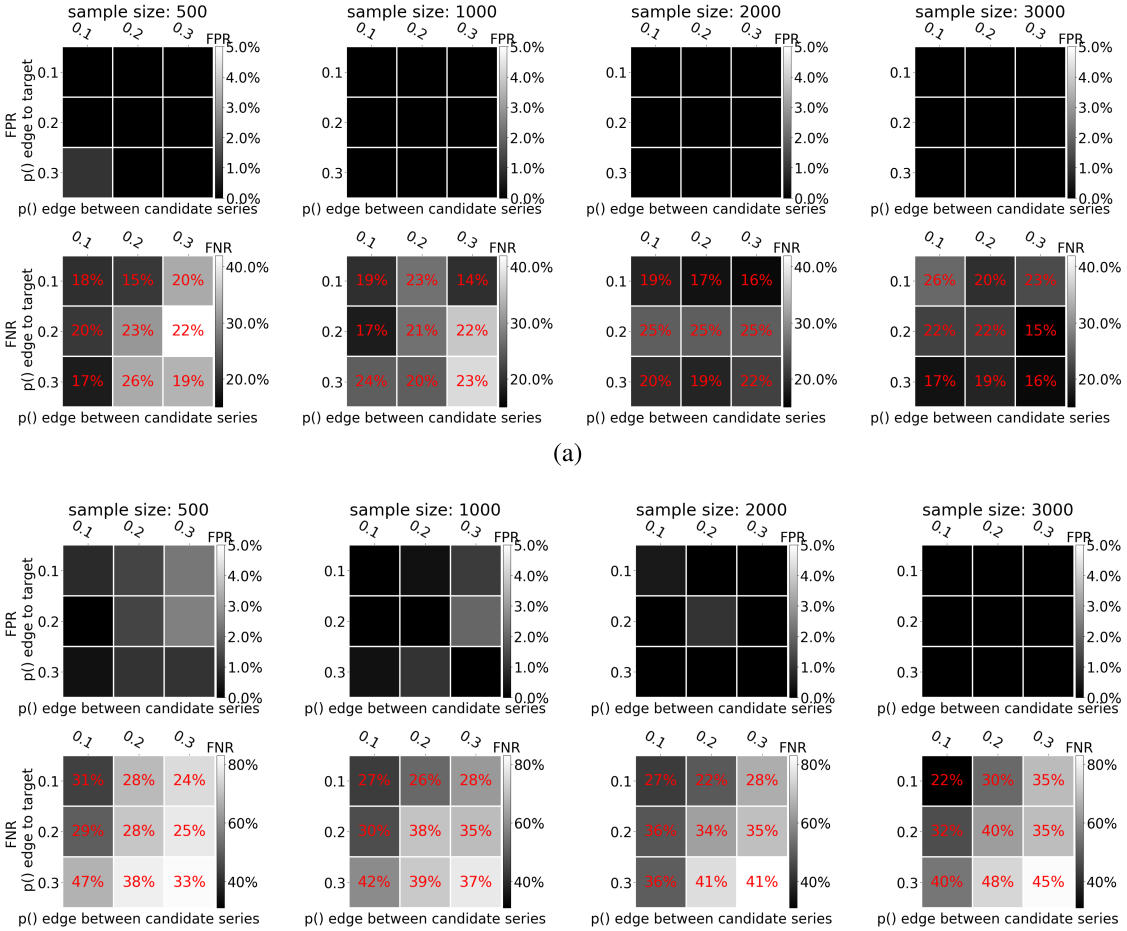
8.4.2 Experiments on real-data

Finally, we examine the performance of SyPI on real data, where there is no guarantee that the required assumptions hold true. The official records of dairy product prices in Europe are used (EU, 2019) (data provided in Appendix B). In this datasets, the product 'Butter' was assigned as the target variable. According to the manufacturing process of dairy products, as described in Soliman and Mashhour (2011), the source material for butter production is 'Raw Milk'. Moreover, according to the same source, butter is not used as an ingredient for the other dairy products in the list. Thus, we can hypothesize that the direct cause of the Butter prices is the price of Raw Milk, and that the remaining nodes in the graph (other cheese, WMP, SMP, Whey Powder) do not cause butter's price. We examine three countries, two of which provide data for the "Raw Milk" (Germany 'DE' (8 time series) and Ireland 'IE' (6 time series)) and one for which the "Raw Milk" prices are not provided (United Kingdom 'UK' (4 time series)). This last dataset was selected on purpose, as this could represent a realistic scenario of a hidden confounder. In this latter case, our method must not identify any cause in the dataset. Due to the extremely low sample sizes (< 180) that are provided, the identification of dependencies is particularly hard. For this reason, we adjust the threshold on the lag detector at zero and the *threshold1* at 0.05 for accepting dependence in the first condition. We leave anything else unchanged as in the simulation experiments.

8.5 Results

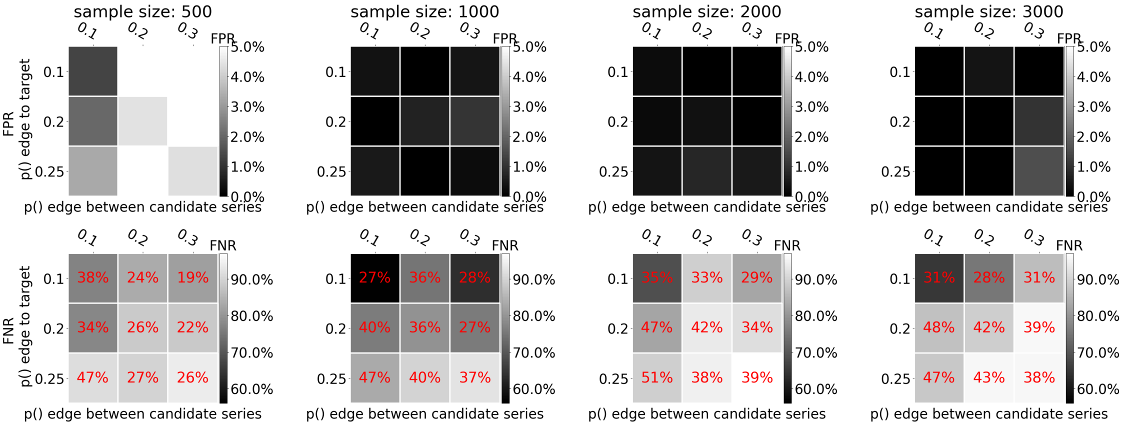
8.5.1 FPR and FNR for various densities and graph size

First, we wanted to examine the performance of our method for various density of edges among the candidate series, and between the candidates and the target time series (see 8.4.1). In Figure 8.3a-8.2c we present results for a medium noise level (20%) and for sample sizes 500, 1000, 2000 and 3000. Above sample size 500 the results are similar for larger or smaller noise levels (see Section B.2 in Appendix B). Lacking space, we present results for one, four and eight observed time series, one additional hidden and one target, to show how the graph size affects the rates. With red colour in each cell we present the percentage of the FNR that corresponds to the direct causes that were missed since our method is complete for direct only. Since our claims refer to complete conditions for unconfounded direct causes, we also encounter as false positives the confounded direct causes. Overall, we see that our algorithm performs with almost zero FPR independent of the noise, the density or the size of the graphs. FNR are low for the direct causes starting from 16% for small and sparse graphs and not exceeding 45% for very large and dense graphs.



(a)

(b)



(c)

Figure 8.2: FPR and FNR for varying numbers of observed, 1 additional hidden and 1 target series, for different sample size (columns) and sparsity of edges among the candidate causes (x-axis) and between the candidate causes and the target (y-axis). The total FNR (for indirect and direct causes) is depicted by the gray scale, where black means 0% and white means 100%. The FNR that refers to the direct causes (for which our method is proven to be complete) is written in red in the middle of each cell. **(a)** 1 observed time series. The FPR are practically zero, and the total FNR 20% for dense graphs. Notice that the FNR of the direct causes is always low, starting from just 16% for dense up to 26% for sparse graphs. **(b)** 4 observed time series. As we can see, for sample sizes > 500 the FPR remain practically zero, and the FNR for direct causes 22% for sparse and 45% for dense graphs. **(c)** 8 observed time series. For sample sizes > 500 the FPR still remain practically zero. The FNR of the direct causes is just 31% for sparse graphs and up to 38% for large and very dense ones.

The results for different number of observed time series and noises are presented in the supplementary B.2.

8.5.2 FPR and FNR for varying # of hidden nodes

Figure 8.3 presents the behaviour of our algorithm in moderately dense graphs, for 2000 sample size, 20% noise variance and a varying number of hidden variables. As it can be seen, the false positive rate is close to zero, independent of the number of hidden variables. Although the false negative rate that refers to both direct and indirect causes increases with the number of time series, the percentage that corresponds to direct causes ranges just from 30 to 40%. Results are similar for different densities of edges (see subsection B.2.2 in Appendix B).

Figure 8.3: FPR and FNR for different number of coexisitng lags. Notice that the FPR is very low as expected by Theorem 2a/2b. Since our method is complete only for single-lag dependencies, we notice that the FNR both for direct causes (dashed lines) for which our method is complete, and for indirect causes increases.

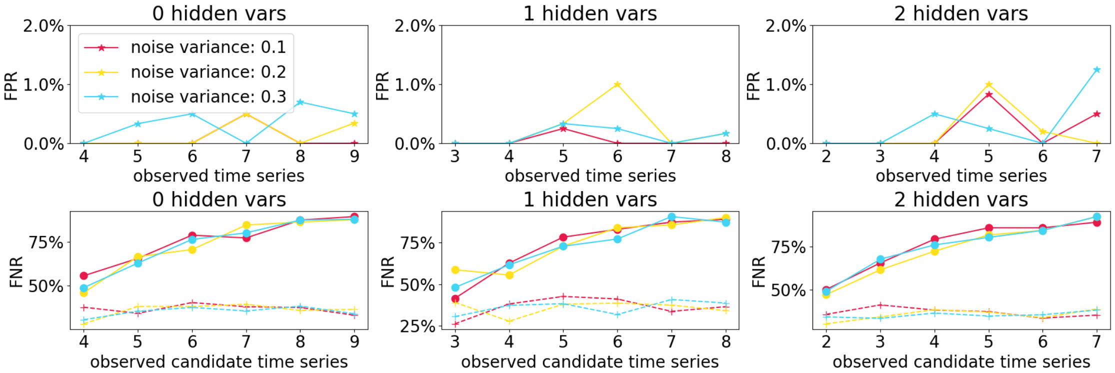
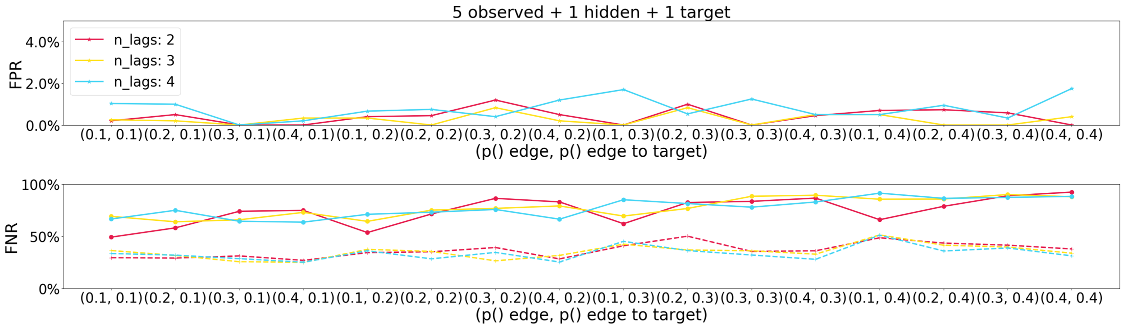


Figure 8.3: FPR and FNR for varying numbers of hidden and observed series, noise variance and sample size 2000, for moderate edge density. FPR is very low (max 1.2% for high noise) for any number of hidden series. Notice that although the total FNR increases with the graph size, the FNR for the direct causes (dashed lines), for which our method is complete, does not exceed 40%.

8.5.3 FPR and FNR for “multiple-lag dependencies”



We examine the false positive and false negative rates for varying number of *lags* between pairs of time series. To test that, we vary the Bernoulli probability that defines the existence of an edge between the time series ($p1 = \{0.1, 0.2, 0.3, 0.4\}$), and between the time series and the target ($p2 = \{0.1, 0.2, 0.3, 0.4\}$). We fix the sample size at 2000, the noise variance at 20%, for two, three and four lags. We examine the above combination for a moderate density of graphs with one hidden, one target and five observed time series. As depicted in Figure 8.3, our method seems to perform very well in terms of FPR, independent of the number of coexisting lags among the time series. As the proposed method is complete only for single-lag dependencies, a larger number of missed targets is expected in multiple-lag settings. However, we see that the FNR that refer to direct causes only does not exceed 40%.

8.5.4 Comparison against LASSO-Granger causality

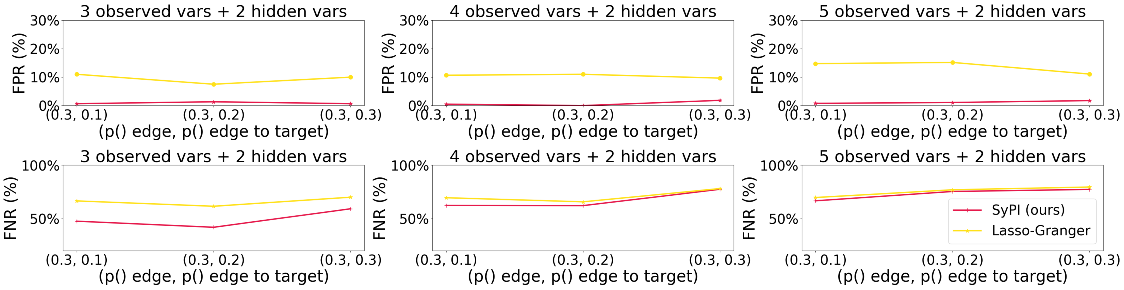


Figure 8.4: Comparison of our method against LASSO-Granger, for sample size 2000, 2 hidden variables, 20% noise variance, for varying number of observed time series and sparsity of edges. As we can see, we tuned the regulariser for the LASSO-Granger to achive similar FNR for similar graphs as *SyPI*. Nevertheless, *SyPI* still performs with lower or equal FNR and with a stable almost zero FPR. In contrast, LASSO-Granger reaches up to 16% FPR. Not tuning λ for LASSO-Granger led to even larger FPR.

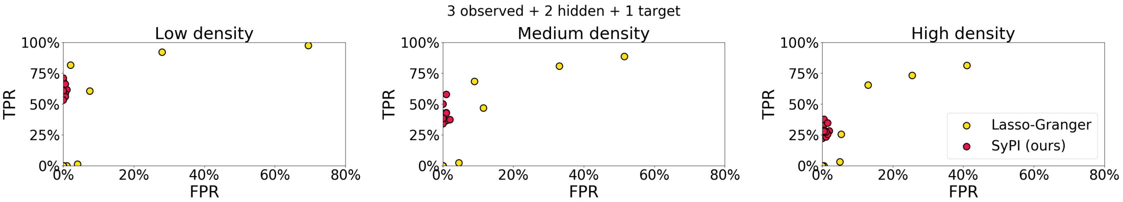


Figure 8.5: Yellow: ROC curve of LASSO-Granger for different values of the λ parameter. Red: ROC curve of our method for different values of *threshold1* and *threshold2* with fixed ratio of 1. The ROC curves were calculated over 100 random graphs, for different density of edges (three columns) and a moderate number of observed series with additional two hidden ones. Our method's ROC curve is always above the Granger's ROC.

As a final step, we compare the performance of SyPI against the widely used LASSO-Granger algorithm. We test the performance of the two methods for relatively dense graphs, for two hidden, one target and three, four and five observed time series. Figure 8.4 shows that even in such confounded graphs, SyPI always performs with almost zero FPR, while LASSO-Granger yields up to 16%, for similar or even larger FNR. Finally, Figure 8.5 depicts the ROC curve for the performance of SyPI and LASSO-Granger for the same graphs. Since SyPI consists of two conditions and, as such, two p-values, we did not manage to find logical pairs of thresholds that increase further the FPR. As it can be seen, SyPI outperforms LASSO-Granger at all operating points.

8.5.5 Comparison against seqICP and PCMCI

As it can be seen in figure 8.6, our method SyPI outperforms both methods for all type of full time graphs, yielding FPR $<$ 1.5% and FNR between 20% and 40%. SeqICP yielded up to 12% FPR and around 95% FNR for almost all the graphs. This result is not surprising, as with hidden confounders seqICP will detect only a subset of the ancestors AN(Y)), and in addition, it assumes that interventions exist in the input dataset. PCMCI yielded up to 25% FPR and oscillated around 25% FNR. In terms of performance times, SyPI was the fastest, followed by PCMCI; seqICP was rather slow for more than 5 time series.

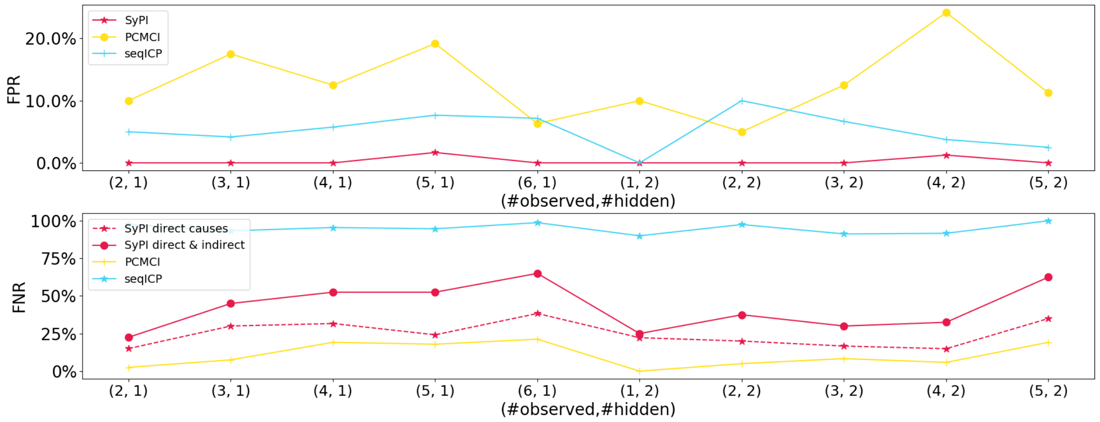
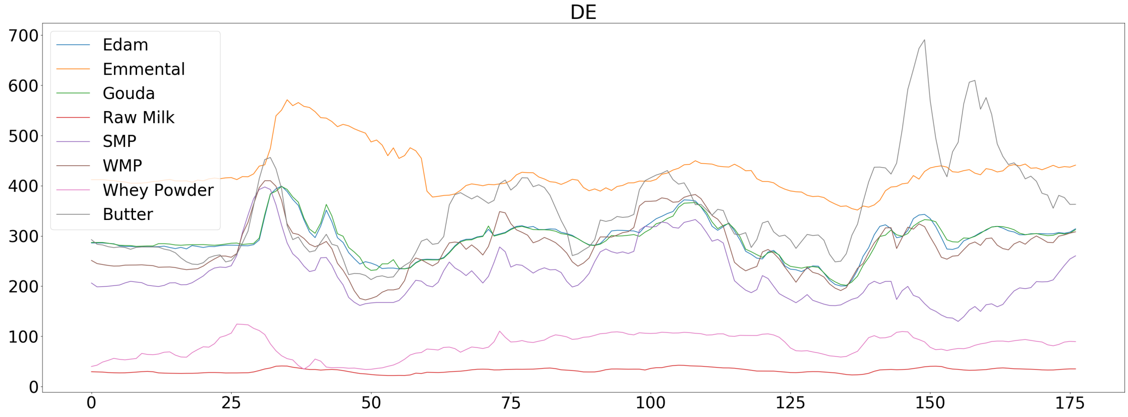


Figure 8.6: Comparison of SyPI against seqICP and PCMCI, for the same full time graphs. False positive and false negative rates are reported over 20 random graphs of similar type (# observed, # hidden time series) for each of the 10 types. Our method SyPI outperforms both methods, with FPR $<$ 1.5% and FNR $20-40\%$. SeqICP yielded 12% FPR and 95% FNR (this is not surprising, as with hidden confounders seqICP will detect only a subset of the ancestors AN(Y)). PCMCI yielded 25%4 FPR with 25% FNR.

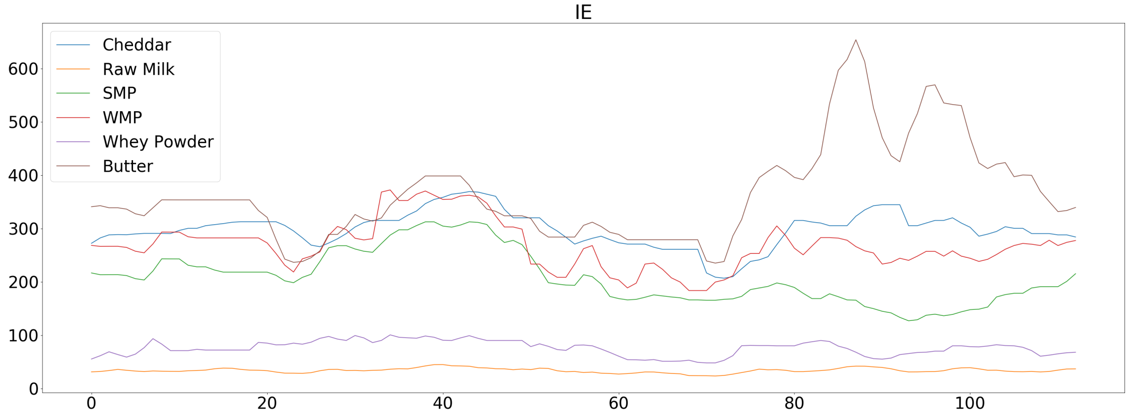
8.5.6 Experiments on real data: Dairy product prices

We applied our algorithm on datasets with the prices of the dairy products in Europe. More specifically, we used the datasets for ’DE’ (8 time series), ’IE’ (6 time series) and ’UK’ (6 time series). The sample sizes were 178, 113 and 168 accordingly. Data are depicted in figure 8.6.

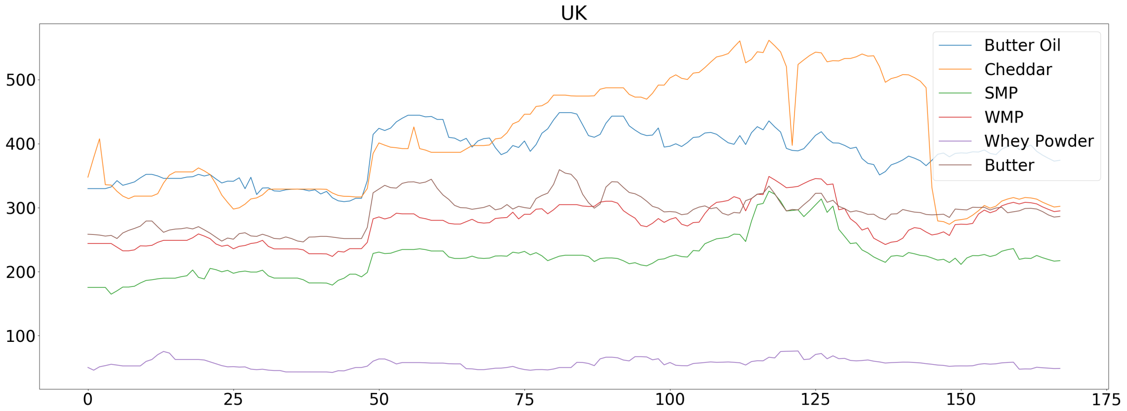
SyPI successfully identified ’Raw Milk’ as the direct cause of ’Butter’ in the ’IE’ dataset and correctly rejected all the remaining 4 nodes (100 % TPR, 100% TNR). In the ’DE’ dataset, ’Raw Milk’ was also correctly identified and there was only one false positive (’Edam’); all the remaining 6 nodes were successfully rejected (100 % TPR and 84% TNR). Notably, in the ’UK’ dataset where no measurements for ’Raw Milk’ prices were provided (hidden confounder), SyPI did not identify any cause (100% TNR).



(a) Dairy product prices for Germany. Raw Milk prices provided in the dataset.



(b) Dairy product prices for Ireland. Raw Milk prices provided in the dataset.



(c) Dairy product prices for the United Kingdom. Notice that Raw Milk prices are no provided in the dataset. This dataset was on purpose selected, as it would represent a realistic case of a hidden confounder between butter and the rest dairy products.

Figure 8.6: Dairy product prices provided by the EU for a span of 8.5-14.5 years (one price recording per month). Here we provide the ground truth time-series from the dairy product prices for each of the three countries 'DE', 'IE' and 'UK'. For UK Raw Milk prices are not provided.

8.6 Discussion

8.6.1 Efficient conditioning set in terms of minimum asymptotic variance

In contrast to other approaches, our method does not search over a large set of possible combinations to identify the right conditioning sets. Instead, for each potential cause X^i it directly constructs its 'separating set' for the nodes X^i_{t-1} and Y_{t+w_i} (condition 2), from a pre-processing step that identifies ($\mathbf{S^i}$) the nodes of the time series that enter Y_{t+w_i-1}. Therefore, the resulting set $\{\mathbf{S^i}, Y_{t+w_i-1}, X^i_t\}$ contains covariates that enter the outcome node Y_{t+w_i}, and not the potential cause X^i_{t-1}. Adjustment sets that include parents of the potential cause node are considered inefficient in terms of asymptotic variance of the causal effect estimate (Henckel *et al.*, 2019), as they can reduce the variance of the *cause* if they are strongly correlated with it, and thus reduce the signal. On the other hand, adding nodes that explain variance in the *outcome* node can contribute to a better signal to noise ratio for the dependences under consideration. According to Theorem 3.1. of (Henckel *et al.*, 2019) our conditioning set yields a more accurate causal effect estimate, compared to a set that would include incoming nodes to X^i_{t-1} or X^i_t. Therefore, our set could strengthen the statistical outcome of the conditional independence test.

8.6.2 Linear and non-linear systems

The proposed method can be used for both linear and non-linear relationships among the time series. For the linear case a partial correlation test is sufficient to examine the conditional dependencies. The algorithm can easily be adapted for the non-linear case, with the KCI (Zhang *et al.*, 2012), KCIPT (Doran *et al.*, 2014) or FCIT (Chalupka *et al.*, 2018) test for the conditional independence testing.

8.6.3 Comparison with related work

Pfister *et al.* (2019) also aim at causal feature selection. However, their method (seqICP) requires sufficient interventions in the dataset, which are required to affect only the input and not the target. In the presence of hidden confounders, seqICP will detect only a subset of the ancestors of target Y, and only if the dataset contains sufficient interventions on the predictors. Given the assumptions presented in Section 8.3.3, we proved that our method will detect all the parents of Y (not only a subset), even in the presence of latent confounders, without requiring interventions in the dataset. Our method's complexity ($\mathcal{O}(n)$) is also smaller than seqICP ($\mathcal{O}(n \log n)$). A method with a greater goal - that of full graph causal discovery - which however could easily be adjusted for our narrower goal is PCMCI by Runge *et al.* (2019a). Nevertheless, PCMCI relies on causal sufficiency, which is often violated in real datasets. As shown in Figure 8.6, our method outperforms both seqICP and PCMCI. Finally, a method that focuses on the full graph causal discovery on time series is SVAR-FCI by Malinsky and Spirtes (2018b). Although our method focuses on the narrower goal of causal feature selection, and there is no direct way of comparison between the two methods, it is still worth mentioning some techincal differences on a high level. SVAR-FCI is computationally intensive because it performs exhaustive conditional independence tests for all lags and conditioning sets. On the contrary, SyPI calculates in advance both the lag and the conditioning set for each conditional independence. Therefore SyPI significantly reduces the required statistical testing.

8.6.4 Technical assumptions of SyPI

Although our technical assumptions are many, we do not consider them extreme, given the hardness of the problem of hidden confounding. (Entner and Hoyer, 2010b) or (Malinsky and Spirtes, 2018b) do not need these assumptions, as they exhaustively perform CI tests for all lags and time series. Assumptions A7 and A8 assure that X are time series with dependency from their previous time step. By assuming memoryless hidden confounders (A9), we avoid the problem that auto-lag hidden confounders create by inducing infinite-lag associations; a case in which also (Malinsky and Spirtes, 2018b) do not find causal relationships as stated there.

8.6.5 Multiple-lag effects

Although our algorithm performs equally well in terms of false positive rate in simulations with “multiple-lag dependencies”, our theory is necessary only for “single-lag dependencies” (see T10). We could allow for “multiple-lag dependencies” if we were willing to condition on larger sets of nodes. However, we do not find this acceptable for statistical reasons. For the conditioning sets of the proposed conditions, we require one node the most, from each observed time series. In a naive extension for multiple lags, n coexisting time lags would require n nodes from each time series to be added in the conditioning set, but the theory is getting cumbersome. As a future work, in Section 9.2.2 in Chapter 9 we show how only in the multi-lag bivariate case (one candidate, one target), with memoryless hidden confounders, it is still possible to have sufficient and necessary conditions, subject to some extensions in the theory.

8.7 Conclusion

In this chapter, we presented necessary and sufficient conditions for a time series to causally influence a target one, even when causal sufficiency is violated, subject to some connectivity assumptions that we impose on the full time graph, which seemed hard to avoid. Other methods that are based on similar ideas as the FCI algorithm are often unreliable, because they require multiple testing and large conditioning sets. On the contrary, the proposed method restrict the problem to the narrower task of the causal feature selection, with the advantage that it requires only two conditional independence tests per candidate cause, with a relatively small conditioning set. SyPI is the first complete and sound algorithm (subject to appropriate graphical assumptions) for direct causal feature selection in time series, that does not require causal sufficiency. Therefore, it overcomes the shortcomings of the widely used Granger Causality. Our simulated experiments demonstrated that for varying graph size and densities, SyPI outperforms LASSO-Granger, seqICP and PCMCI. Finally, in experiments on real data SyPI yielded almost 100% TPR and TNR, despite the potential violation of our assumptions and the low sample size.

Chapter 9

Conclusions & Future work

9.1 Conclusions

Trying to infer causal relationships among real observations has always been a characteristic of human thinking and behaviour. Since the science of Causal Inference was founded, 50 years ago, this procedure of construction of causal statements based on observational data has been formalised with theorems and methods that impose conditions, in order to infer the existence and the direction of causal relationships. Such a task is especially hard, when based on observations alone, due to the hidden confounding factor, which is particularly prominent in real datasets. Many methods try to discover the underlying causal graph, which can be a computationally intensive task. Furthermore, these methods cannot differentiate among Markov equivalent classes, if additional assumptions are not imposed. This thesis focuses on the narrower, yet challenging and non-trivial sub-problem of causal feature selection: the detection of direct and indirect causes of a given target. Our motivation behind this sub-problem arose by the need for detection of causal brain features for the human upper limb movements, and from the gap that has been observed in the literature regarding techniques for causal feature selection in the presence of latent variables. Furthermore, the observed heterogeneity across subjects' behavioural response to brain stimulation protocols revealed a need for personalisation of brain stimulation targets and a lack of a systematic causal detection method, that could causally relate the activity of the human motor cortex to the observed motor performance.

This dissertation, through electrophysiological and non-invasive brain stimulation experiments on humans, as well as through novel theoretical methods for causal feature selection, contributes in a twofold manner in the scientific community: First, via the incremental neuroscientific findings that help in the better understanding of the functionality of human motor cortex, and, second, via two novel causal feature selection methods from observational data, for independent and sequential data, which tackle problems that occur in causal inference on real datasets.

Overall, the work presented in this thesis aims to emphasize the caution that someone should practise when making causal claims and proposing causal methods for complex systems, like the human brain.

9.1.1 Neuroscientific findings

At first, a transfer-learning-based pipeline is proposed in order to build "personalised" regression models that relate the global configuration of EEG rhythms to motor performance, using only a few trials. These single-trial predictive models showed initial evidence for considerable heterogeneity in the activity of motor cortex across different subjects, during similar 3D-reaching movements. This heterogeneity was found to be in accordance with the observed variability in response to non-invasive brain stimulation, over the contralateral motor cortex. Our findings further support this line of argument by evidencing a substantial heterogeneity amongst subjects: Features in the alpha, beta, and gamma range, used by the predictors, turn out to be sometimes negatively and sometimes positively correlated with motor performance. In line with previous findings, our models reveal the alpha, beta and (high) gamma frequency ranges as decisive for motor performance. These initial results are an indication that a one-for-all stimulation approach is unlikely to consistently improve motor performance. Of course, decoding models as the ones trained in this initial study *do not* immediately reflect causal relationships, and as such, they *do not* allow to directly read off optimal stimulation parameters for each subject (Haufe *et al.*, 2014; Weichwald *et al.*, 2015). Nevertheless, they do allow to reject non-relevant and hence non-causal features. Thus, a decoding model, which is able to predict well single-trial motor performance, is a necessary prerequisite for personalised stimulation protocols.

Being motivated and inspired by this initially observed heterogeneity in motor cortex activity across subjects, we proceeded with a crossover non-invasive brain stimulation study on 20 healthy participants. We applied real 70 Hz and sham transcranial alternating current stimulation over the contralateral motor cortex, during a visuo-motor 3D reaching task, in a randomized order, in parallel with EEG recordings. Our goal was to examine potential causes of the observed variability in the behavioural response to the stimulation. Our findings supported a potential role of β-power as a mediator of γ-tACS on motor performance. In particular, our empirical results were in favour of a causal model in which β-power may mediate the effect of γ-tACS on motor performance. It is important to stress, however, that up to this point, inference methods as applied here cannot prove causal relationships due to the hidden confounding factor that cannot be excluded. Nevertheless, in the context of neurophysiological procedures underlying the effect of γ-stimulation on β-power and on the observed motor behaviour, a possible explanatory factor that supports the proposed causal model could be the modulation of γ-aminobutyric acid (GABA) concentration: Firstly, β-oscillations have been shown to be the summed output of principal cells temporally aligned by GABAergic interneuron rhythmicity (Yamawaki *et al.*, 2008). GABA levels have been found to strongly correlate with β-power and to exhibit elevated values in bradykinesia and in Parkinson's disease (McAllister *et al.*, 2013). In addition, high-γ deep brain stimulation in motor cortex has been reported to significantly decrease β-power (Gulberti *et al.*, 2015). This argument supports our finding of the inhibitory effect of γ-stimulation on the ongoing

β-oscillations. Combining these two points, the behavioural response to γ-tACS may be explained by a decrease of β-power and hence of GABA levels, modulated by the stimulation. Therefore, it is plausible that whenever γ-tACS leads to the inhibition of human movements, this may be caused by an increase in GABAergic drive, which hinders the decrease of β-power.

The aforementioned encountered heterogeneity in response to tACS on this crossover study, validated related sparse findings from other stimulation studies, pointing out one more time the need for personalisation of brain stimulation parameters. A first step towards this personalisation is the differentiation of responders and non-responders prior to the application of stimulation treatment. Such an early screening of the non-responders could help avoid unnecessary or even harmful stimulation treatments. To that end, twenty-two more healthy participants were recruited, to whom the same visuo-motor EEG/tACS experiment was performed. Resting-state high-gamma power prior to stimulation was found to enable the differentiation of a newly coming subject between a responder and a non-responder. Specifically, we demonstrated in the first experimental group that subjects' resting-state EEG predicts their motor response (arm speed) to gamma (70 Hz) tACS over the contralateral motor cortex. We then validated in a prospective stimulation study with the twenty-two new subjects that the proposed screening pipeline achieves a reliable stratification of subjects into a responder and a non-responder group. Strong resting-state γ-power over contralateral motor cortex was found to be indicative of a positive stimulation response to tACS (in terms of arm speed). The finding about the predictive role of high-gamma power is in line with our current understanding of the neurophysiological effects of γ-tACS and the role of γ-power in fronto-parietal networks for motor performance (Gonzalez Andino *et al.*, 2005). The explanation is the following: Resting-state γ-power in primary motor cortex positively correlates with γ-aminobutyric acid (GABA) levels (Chen *et al.*, 2014; Muthukumaraswamy *et al.*, 2009; Bartos *et al.*, 2007; Wang and Buzsáki, 1996; Brunel and Wang, 2003). Because γ-tACS over motor cortex decreases GABA levels (Nowak *et al.*, 2017), and decreases in motor cortex GABA levels correlate with increased motor performance (Stagg *et al.*, 2011), high resting-state γ-power may signal a brain state in which motor performance can be improved through tACS-induced reduction of GABA levels. Low resting-state γ-power, in contrast, would signal a brain state in which GABA levels are already low, thus limiting the extent of potential further reduction by γ-tACS. We note that this explanation is also in line with our finding that stimulation response is contingent on the current brain state.

Finally, applying the novel causal inference method presented in Chapter 7 on EEG resting state periods from the aforementioned experiments resulted in findings very much in line with established neuroscientific conclusions. This was the first time that such conclusions were also found through a purely causal method. More specifically, the application of our proposed method on our EEG data gave performance-specific causes across subjects, which are consistent with the known roles of physiological α, β and γ brain rhythms in upper-limb movements. In particular, channels in the β power were found to be causal by our CFS method for subjects that did not improve their motor performance. This is in line with established neuroscientific conclusions that have reported β activity to be significantly elevated in patients with motor disorders (tremors, slowed movements) such as Parkinson's disease (McAllister *et al.*, 2013; Brown, 2007; Khanna and Carmena, 2017). Furthermore, in healthy subjects, elevated β-power has been found to play an anti-kinetic role (Khanna and Carmena, 2017). On the other hand, our method detected causal motor channels in the γ band, in subjects who managed to reduce their reaching times and improved their motor performance. This appears in accordance with the literature, as increased γ activity over the motor cortices has been suggested to be prokinetic and has been associated with large ballistic movements (Muthukumaraswamy, 2010; Nowak *et al.*, 2018). Finally, our detected causal channels in the ipsilateral hemisphere at α-band are consistent with neurophysiological studies that report increased α-power over ipsilateral sensorimotor cortex during selection of movement (Brinkman *et al.*, 2014). Although there is no ground truth for our neurophysiological results, the findings appear plausible and meaningful, given the current understanding of the aforementioned physiological brain rhythms in movement. Therefore, our method contributes to the more precise localisation of causal cortical electrode-areas.

It is crucial to point out that there is not a one-to-one mapping between the causal brain features detected here and the stimulation targets, as it is still unknown how the stimulation current in a specific frequency entrains the ongoing brain oscillations. For example, as it is shown and discussed in detail in Chapter 5, β-rhythms may play a mediating role between γ stimulation and motor performance. Someone can consider the problem of selecting personalised stimulation targets and frequencies as a two-step procedure: first, understand the effect of stimulation on brain activity, and second, detect the link between brain activity and motor response. In the graphical chain *stimulation parameters → brain activity → behavioural response*, the proposed causal method contributes to the second link. Thus, it narrows the original problem of personalised stimulation to the new question: *stimulation parameters → detected causal brain activity*. Hence, the search for personalised stimulation parameters can now be reduced to the detection of those that up- or down-modulate accordingly, the causal brain features which our algorithm identifies.

9.1.2 Theoretical contributions & Methods

The theoretical contributions of this thesis focus on the field of causal feature selection, both from independent random variable settings and from sequential data.

In Chapter 7 a novel causal feature selection method for independent random variables was presented. Given a pool of features and a target random variable, we proved that assuming a cause of each feature is known, a single conditional independence test with a single conditioning variable is sufficient to identify the target's direct and indirect causes, even in the presence of latent confounders. This assumption can naturally be met in set-ups where two nodes constitute consecutive timestamps of a variable's state in a system, and an edge from the previous to the present state can be assumed. With one targeted conditional independence test per variable and only one conditioning variable, the complexity of the algorithm scales linearly with the number of features, substantially strengthening the statistical power of our tests, and allowing us for a weaker assumption of faithfulness. Excluding the assumption of causal sufficiency and speeding up the process of causal feature selection are two key points that facilitate the application of causal inference on real datasets.

Causal feature selection in time series data is a fundamental problem in several fields, when the causes of a time series of interest (i.e. revenue, temperature) need to be identified, while latent variables cannot be excluded. In Chapter 8 a novel causal feature selection method for sequential data was presented. Two theorems were proposed and it was proved that their conditions are necessary for direct causes in single-lag dependency graphs, even in the presence of latent variables, and sufficient for direct and indirect causes in multi-lag dependency graphs. To the best of our knowledge, this novel causal method (SyPI) is the first complete and sound algorithm (subject to appropriate graphical assumptions) for direct causal feature selection in time series that does not assume causal sufficiency, thus overcoming the shortcomings of Granger Causality and the state of the art methods. In contrast to approaches inspired by conditional independence based algorithms for causal discovery, SyPI directly constructs the 'adjustment set' for each potential cause X^i, from a pre-processing step that identifies the nodes of the time series that enter the previous node of the target Y_{t+w_i}. Therefore, the resulting conditioning set contains covariates that enter the outcome node Y_{t+w_i-1}, and not the potential cause X^i_{t-1}. According to Henckel *et al.* (2019) the proposed conditioning set has a smaller asymptotic variance compared to a set that would include incoming nodes to the candidate causes X^i_{t-1} or X^i_t. Therefore, this choice also contributes to a reasonable signal to noise ratio for the dependences under consideration. This could strengthen the statistical outcome of the conditional independence test.

9.1.3 Experimental datasets

During the experimental work presented in Chapters 5 and 6, we recorded an extended dataset with EEG recordings from 41 healthy participants, during a visuomotor reaching task in a crossover study, with alternating blocks of real and sham tACS applied over the contralater motor cortex of the subjects. In parallel, the coordinates of the arm trajectory were also recorded. This dataset will become public to the scientific community, respecting all the personal data according to the DSGVO regulations, aiming at contributing in shortening the lack of open real datasets.

9.2 Future work

The methods and neuroscientific contributions presented in this dissertation can become the basis for improvements in future work that can bring personalised brain stimulation a step closer to its realization. We split the possible directions of future work into (a) new stimulation experiments, and (b) methodological, technical extensions.

9.2.1 New stimulation experiments

More specifically, the findings in Chapter 7 that were derived by the proposed causal method, can be validated in a prospective stimulation study, under the guidance of a trained doctor that can exclude potentially harmful stimulation targets, which due to ethical reasons should not be tested on humans. Such experiments would be priceless for the validation of our methods, but at the same time could become dangerous for humans, as they include an exploratory procedure on the human brain. In general, ground truth in the human brain is a particularly hard thing to derive, and the closest possible approach is through brain stimulation. Furthermore, the second causal method for time series that is presented in Chapter 8 can be applied in neuroscientific data, preferably from intracranial recordings for less noise and better focality, to examine directly the causal role of brain time series in different frequencies and areas in continuous response signals.

9.2.2 Methodological technical extensions

Here we propose some possible extensions on the novel causal feature selection method SyPI proposed in Chapter 8.

Expansion of SyPI for multiple lags

The conditions of the proposed theorems in Section 8.3 are necessary only for "single-lag dependencies" (see T10). We could allow for "multiple-lag dependencies" if we were willing to condition on larger sets of nodes, which we do not find acceptable for statistical reasons. Right now we require one node the most from each observed time series for the conditioning set. In a naive approach, n coexisting time lags would require n nodes from each time series in the conditioning set, but the theory is getting cumbersome.

The reason why our conditions are not necessary for "multiple-lag dependencies" is the difficulty in identifying just one lag from each time series to look at and to add in the conditioning set. If we did not put a lot of weight on keeping the conditioning set to a minimum size for assuring a decent statistical strength, we could still construct a conditioning set with as many nodes per time series as the multiple lags and have necessary conditions. In single-lag effects we describe why a single node from each time-series is necessary and sufficient and we show why this single lag can be the minimum lag as defined in 1. Without making any strong claims about the multi-lag case as it is out of the scope of this thesis, the following point could be used as a basis for extension: If we use the following condition, instead of the one defined in lemma 1, as $\max(v) \neq \inf$ s.t. $A_t \not\perp\!\!\!\perp B_{t+v} \mid \{A_{\text{past}(t)}, A_{\text{future}(t)}, B_{\text{past}(t+v)}, B_{\text{future}(t+v)}\}$, then in the bivariate case described below it is enough to use the *maximum* integer v as the w_i in the theorems and still have necessary and sufficient conditions. Only in a bivariate (2 observed series) full-time graph with one candidate time series and one target time series, where hidden confounders are memoryless and with unique lag, given the above condition, max v as defined could be enough for differentiating between the time series causing the target with multiple lags and the time-series being confounded. If a node X_t^i has a direct edge both to Y_{t+1} and Y_{t+2}, then the "maximum" v would be equal to 2. If we used this as w_i in our conditions then, $Y_{t+w_i} \equiv Y_{t+2}$. Then conditioning on X_t^i, which is w_i steps back, and on $Y_{t+1} \equiv Y_{t+w_i-1}$, would render X_{t-1}^i and Y_{t+w_i} independent, so the two conditions of our theorems would hold. Of course this extension does not hold beyond the bivariate case.

Therefore, future work could potentially expand the proposed conditions of Theorems 2a/2b and 3 or modify the way the conditioning set is built, in order to account for multiple-lag dependencies in multi-variate settings.

Expansion of SyPI for non-linear relationships

A straightforward extension that should be made as a future work, is the identification of the lag for non-linear causal mechanisms. In Chapter 8 we take into account linear relationships and we use Lasso-Granger to identify the lag between pairs of time series. An easy and important extension is to use a non-linear version of Granger Causality for the lag-identification step.

Less assumptions for the target variable

Finally, a direction that could become a topic of future work is the modification of the necessary conditions of Theorem 3 so that they do not require any connectivity restriction regarding the target, such as A6, or such as the relaxed one we propose in Theorem 4. In Theorem 4 we relaxed the strict assumption A6 that required the target not to have any descendants of its own, by proving that Theorems 2a/2b and 3 still hold true, if none of the targets' descendants belongs to its candidate causes. Although this already relaxes the connectivity restrictions, further work could be done in the direction of allowing the target node to have similar properties as the candidates. Assuming that the target does not cause any of its candidate causes, basically, reassures that there are no cycles in the summary graph. Our method is tolerant in cycles among the candidate causes alone, which means that even so, it will still detect all and only all the direct causes. However, it will no longer be complete if there are cycles that include nodes of the target time series, which is the reason for the aforementioned assumptions. Getting rid of this restriction and still maintaining necessary conditions is still an open question, which will definitely require at least modifications in the way that the conditioning set is constructed.

Appendix A

Supplementary material for chapter 7

A.1 Sufficient but not necessary

We present an example where both direct causes are rejected. In this example although both M^1 and M^2 nodes are causes of R, both are rejected. P^1 and R are not d-separated by M^1 due to the path including P^2 and M^2 or because M^1 acts as a collider. On the other hand, P^2 and R are not d-separated by M^2 due to the path including P^1 and M^1.

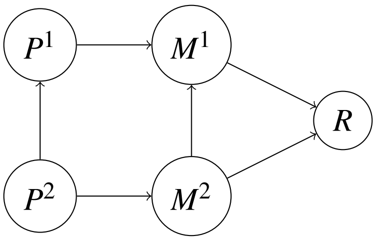
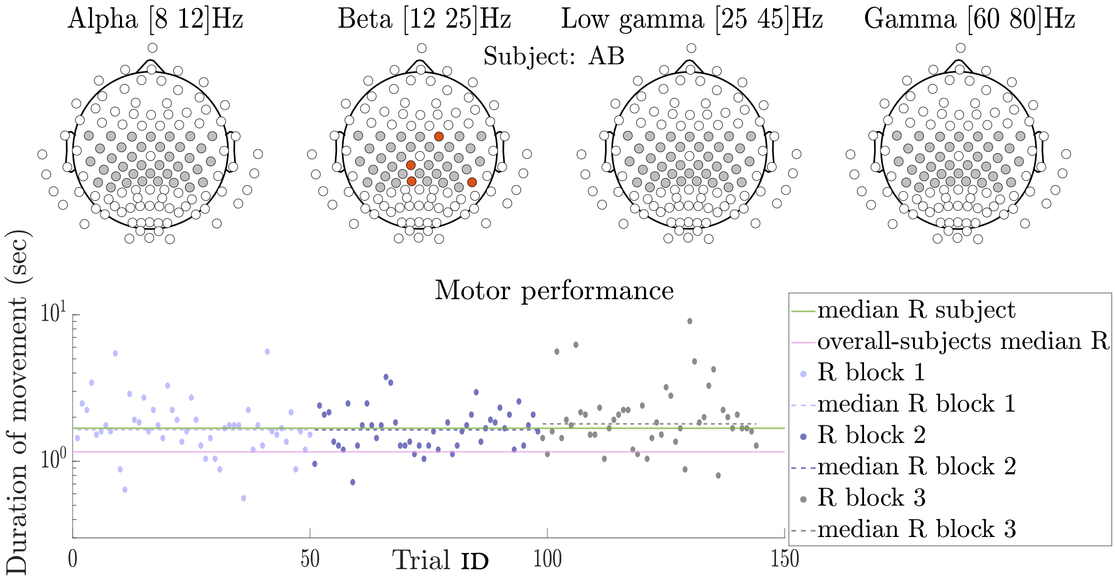


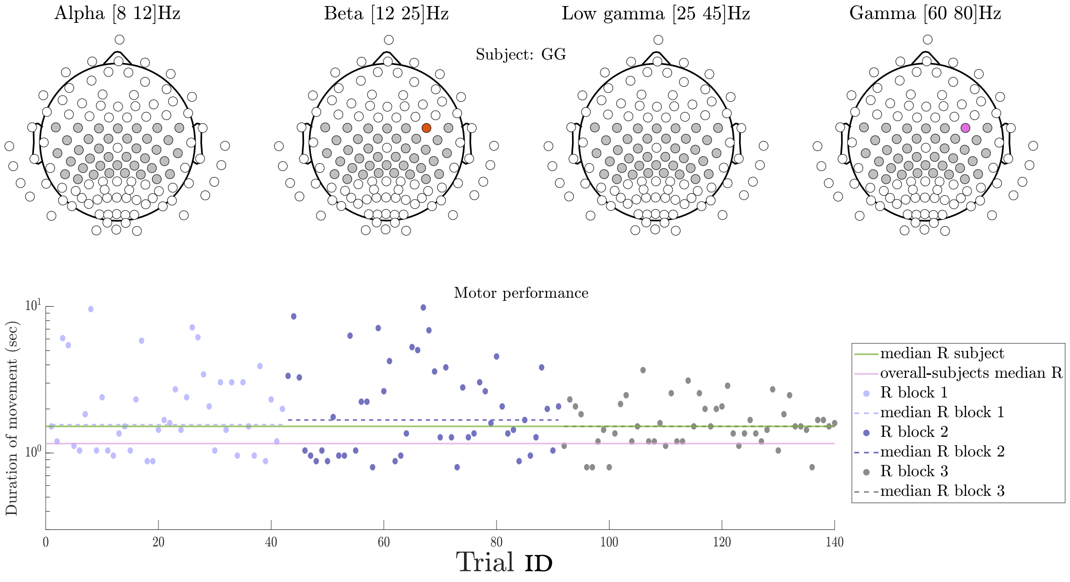
Figure A.1: Example of DAG where causes are rejected because our theorem is sufficient but not necessary. Here, if all direct edges are equally strong, then both M^1 and M^2 are not identified by our theorem, due to the confounding path formed by the P variables.

A.2 Detected causal features for all subjects - Grouping based on motor performance

Here we present the detected causes for all subjects we processed with our causal method. In total, our algorithm detected causal brain features in seventeen out of twenty-one subjects. Our findings group subjects in three main categories that couple detected causes with subject's performance: those that γ power is detected when subjects improve their performance (Figure A.3), those that β power is detected when subjects worsen or do not improve their performance (Figure A.2), and finally those that α power is detected in the ipsilateral hemisphere (Figure A.4).

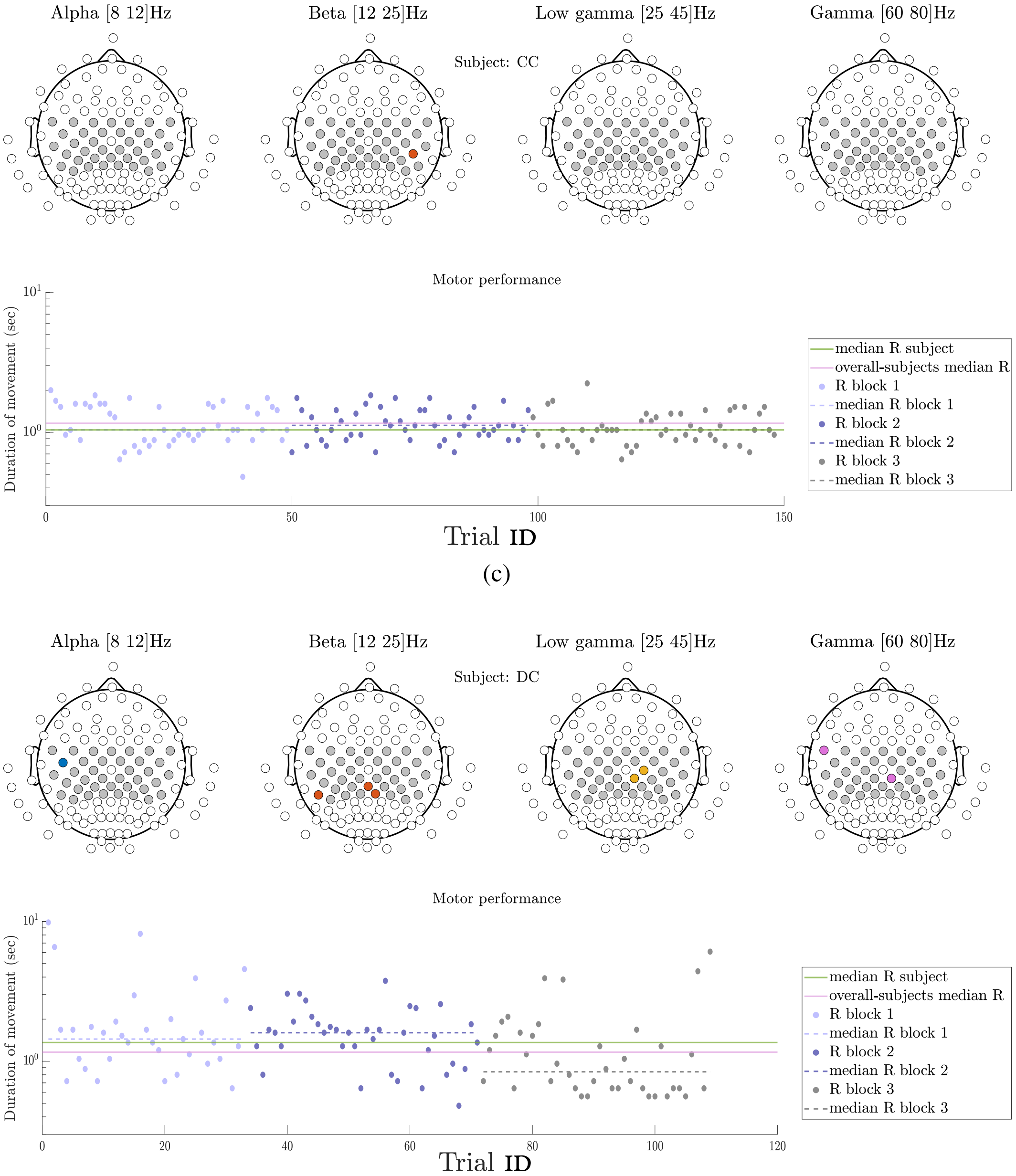

Alpha [8 12]Hz
Beta [12 25]Hz
Low gamma [25 45]Hz
Gamma [60 80]Hz
Subject: AB
Motor performance
Duration of movement (sec)
10^1
10^0
0
50
100
150
Trial ID
median R subject
overall-subjects median R
R block 1
median R block 1
R block 2
median R block 2
R block 3
median R block 3

(a)


Alpha [8 12]Hz
Beta [12 25]Hz
Low gamma [25 45]Hz
Gamma [60 80]Hz
Subject: GG
Motor performance
Duration of movement (sec)
10^1
10^0
0
20
40
60
80
100
120
140
Trial ID
median R subject
overall-subjects median R
R block 1
median R block 1
R block 2
median R block 2
R block 3
median R block 3

(b)

Figure A.2

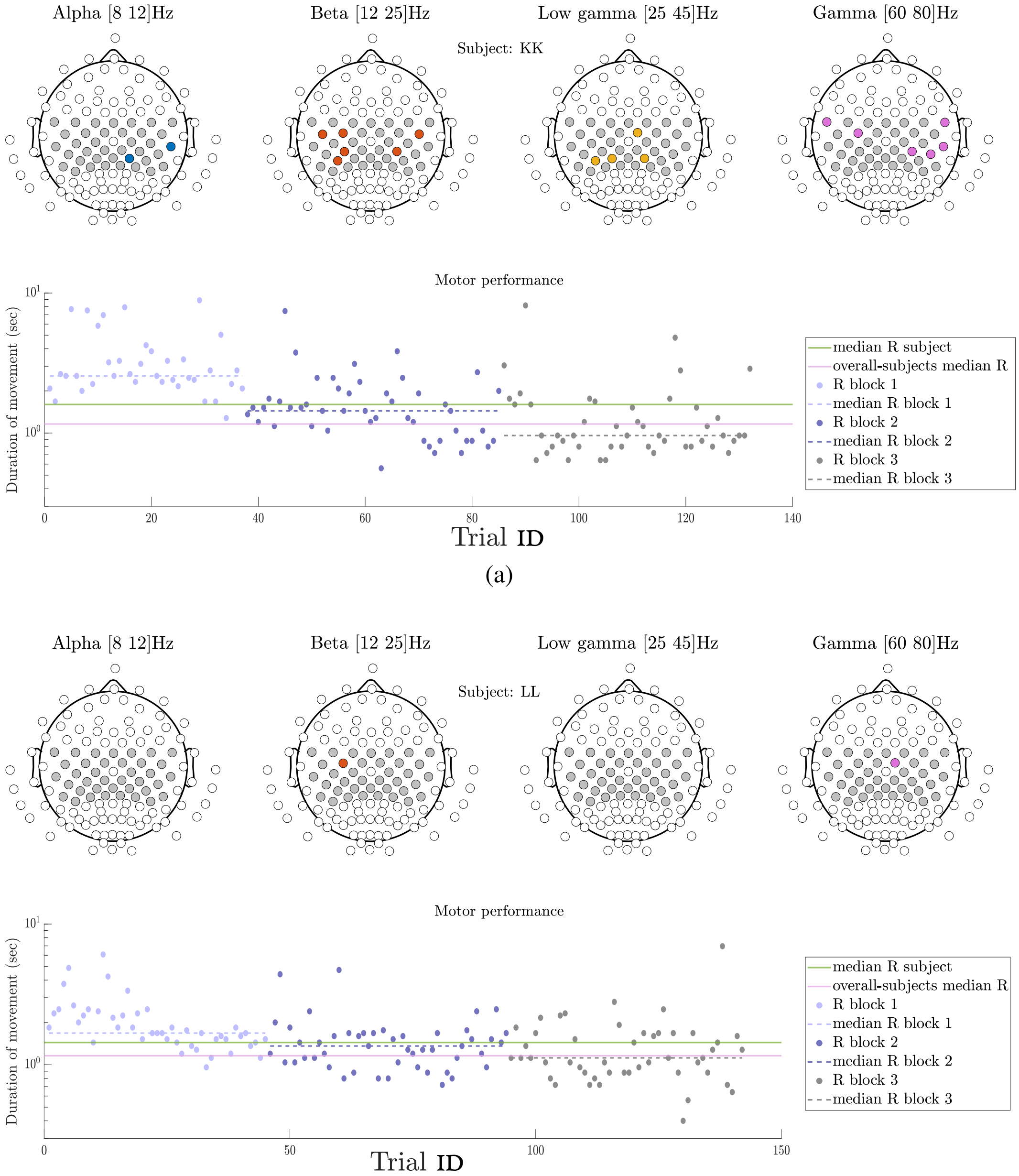


(c)

(d)

Figure A.2: Electrodes over contralateral motor cortex in the beta power at subjects that remain stable or worsen during the reaching trial, are detected as causal features from our algorithm. Y-axis is in logarithmic scale.

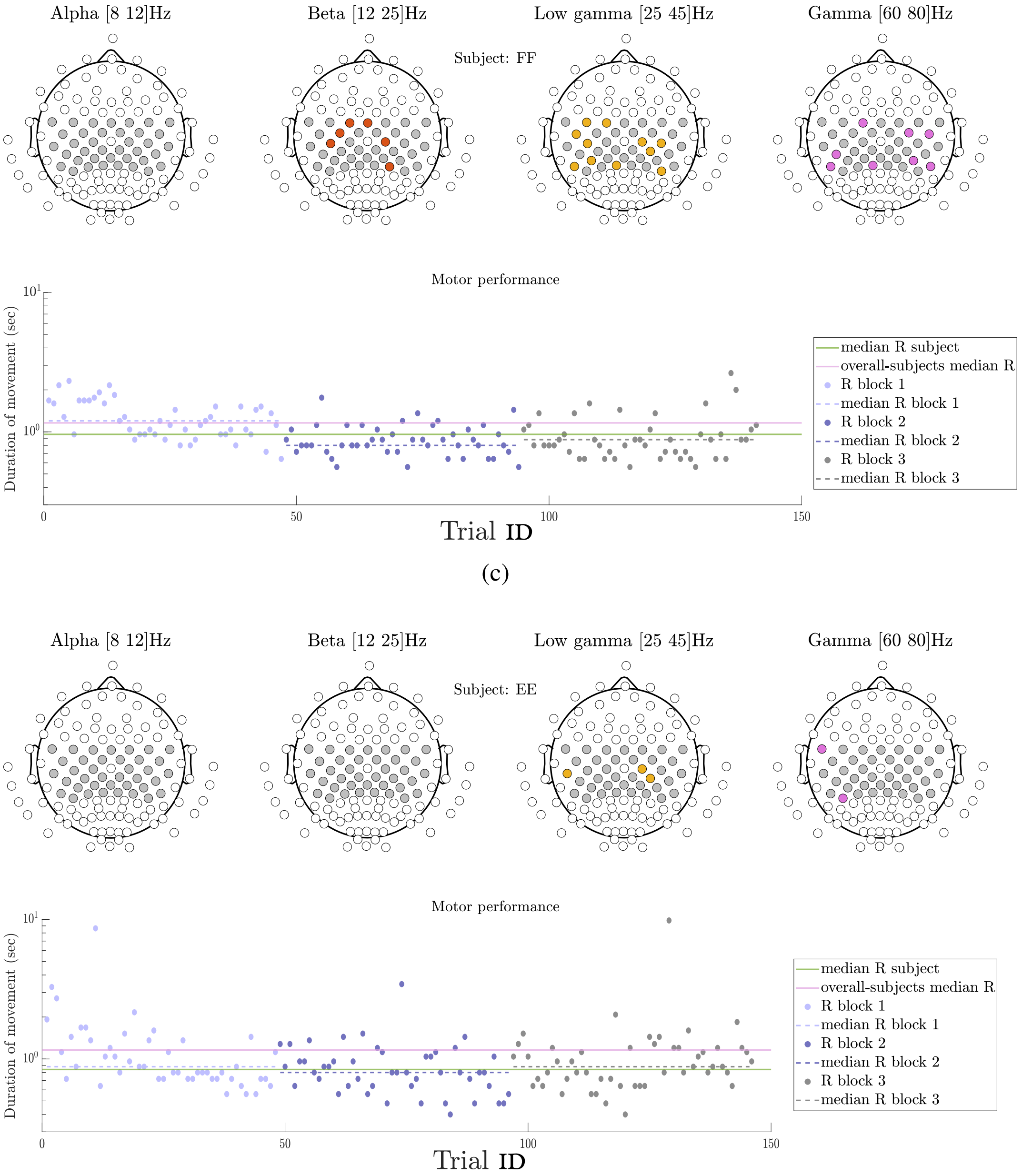
Figure A.2 depicts the subjects that did not improve their movement duration throughout the sequence of reaching trials or who got worse (larger durations for completing the trial). We observe that our algorithm detects causes over motor channels in the beta range (second headplot), which is consistent with the literature findings about the predominant role of beta power in slow or unstable movements. In addition, we observe that among these subjects, for those who in general had performances better than the average, our algorithm detects also some electrodes in the gamma range (fourth headplot), which complies with the facilitatory role of gamma from the literature.



(a)

(b)

Figure A.3

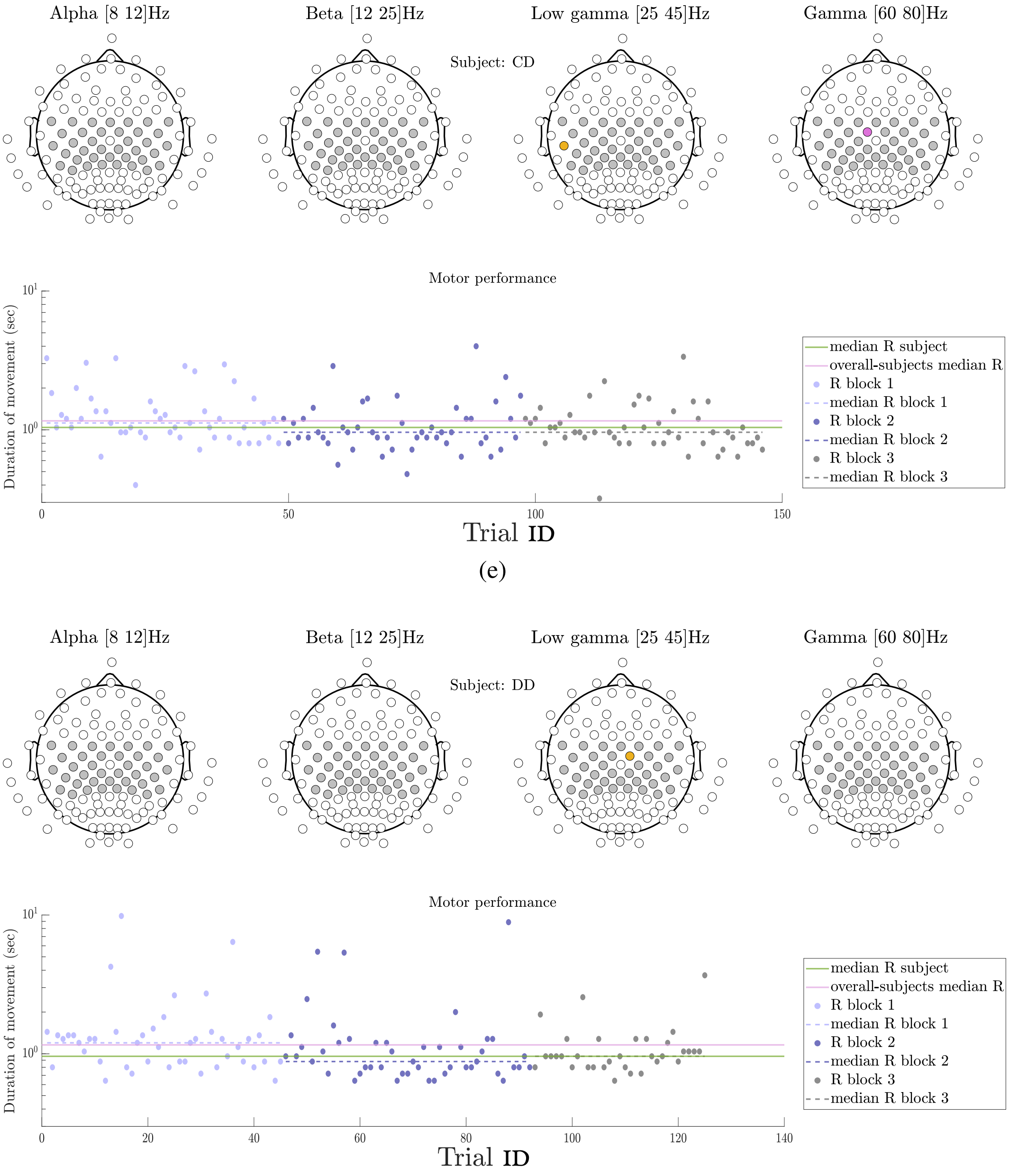

Alpha [8 12]Hz
Beta [12 25]Hz
Low gamma [25 45]Hz
Gamma [60 80]Hz
Subject: FF
Motor performance
10^1
10^0
Duration of movement (sec)
0
50
100
150
Trial ID
median R subject
overall-subjects median R
R block 1
median R block 1
R block 2
median R block 2
R block 3
median R block 3

(c)

Alpha [8 12]Hz
Beta [12 25]Hz
Low gamma [25 45]Hz
Gamma [60 80]Hz
Subject: EE
Motor performance
10^1
10^0
Duration of movement (sec)
0
50
100
150
Trial ID
median R subject
overall-subjects median R
R block 1
median R block 1
R block 2
median R block 2
R block 3
median R block 3

(d)

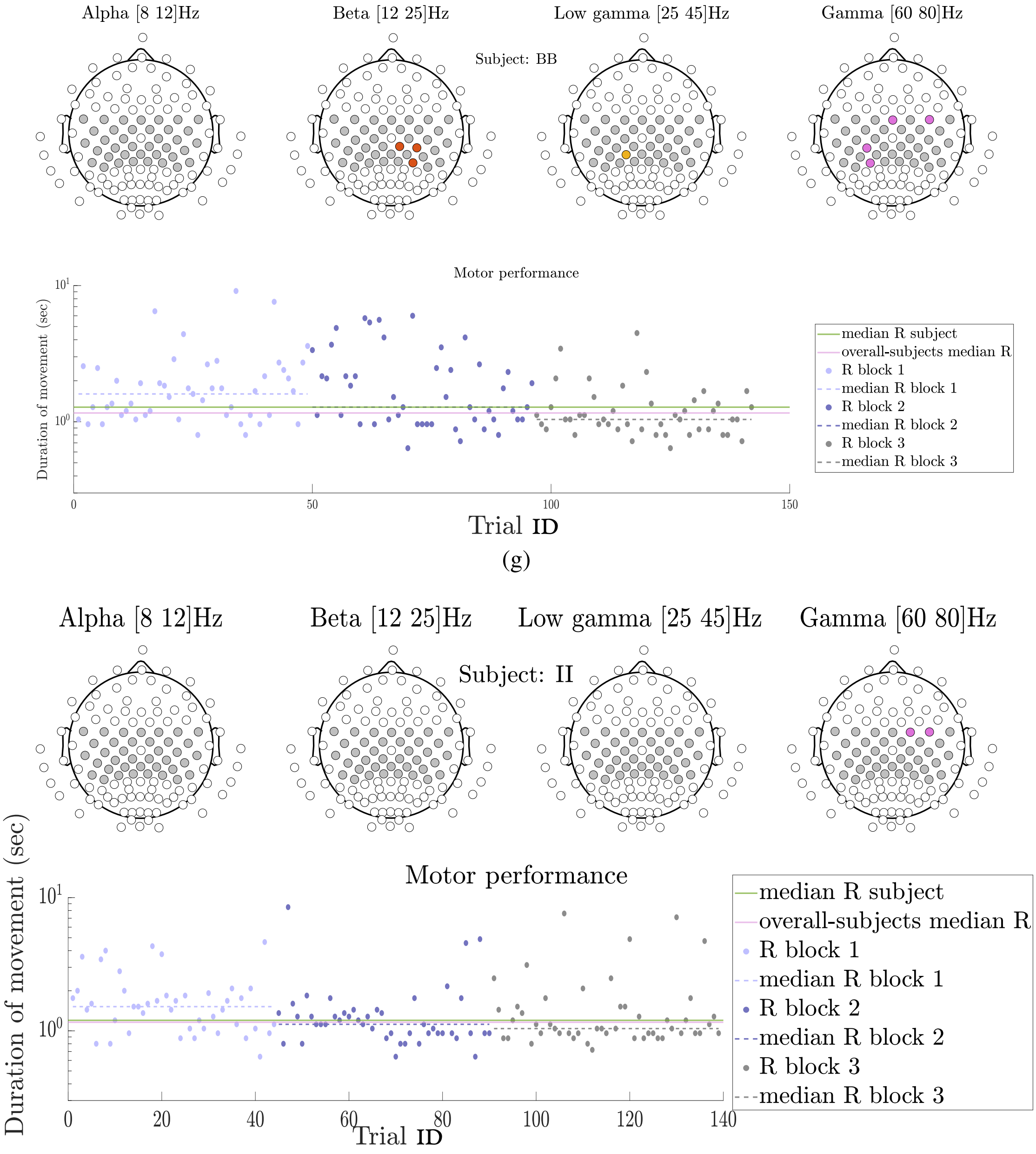
Figure A.3


Alpha [8 12]Hz
Beta [12 25]Hz
Low gamma [25 45]Hz
Gamma [60 80]Hz
Subject: CD
Motor performance
10^1
10^0
Duration of movement (sec)
0
50
100
150
Trial ID
median R subject
overall-subjects median R
R block 1
median R block 1
R block 2
median R block 2
R block 3
median R block 3
Alpha [8 12]Hz
Beta [12 25]Hz
Low gamma [25 45]Hz
Gamma [60 80]Hz
Subject: DD
Motor performance
10^1
10^0
Duration of movement (sec)
0
20
40
60
80
100
120
140
Trial ID
median R subject
overall-subjects median R
R block 1
median R block 1
R block 2
median R block 2
R block 3
median R block 3

(e)

(f)

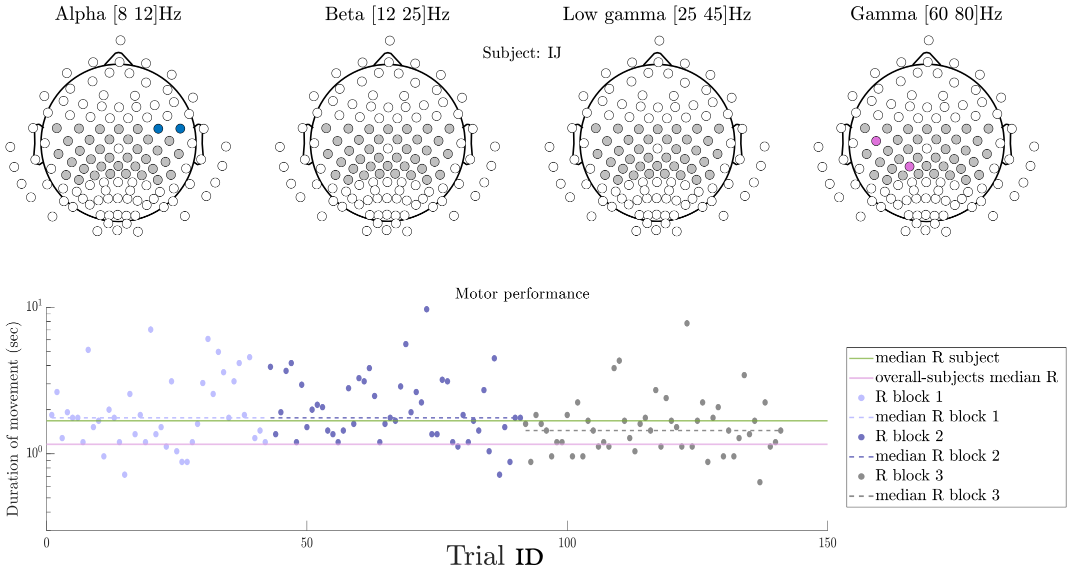
Figure A.3


Alpha [8 12]Hz
Beta [12 25]Hz
Low gamma [25 45]Hz
Gamma [60 80]Hz
Subject: BB
Motor performance
Duration of movement (sec)
10^1
10^0
0
50
100
150
Trial ID
median R subject
overall-subjects median R
R block 1
median R block 1
R block 2
median R block 2
R block 3
median R block 3
Alpha [8 12]Hz
Beta [12 25]Hz
Low gamma [25 45]Hz
Gamma [60 80]Hz
Subject: II
Motor performance
Duration of movement (sec)
10^1
10^0
0
20
40
60
80
100
120
140
Trial ID
median R subject
overall-subjects median R
R block 1
median R block 1
R block 2
median R block 2
R block 3
median R block 3

(g)

(h)

Figure A.3

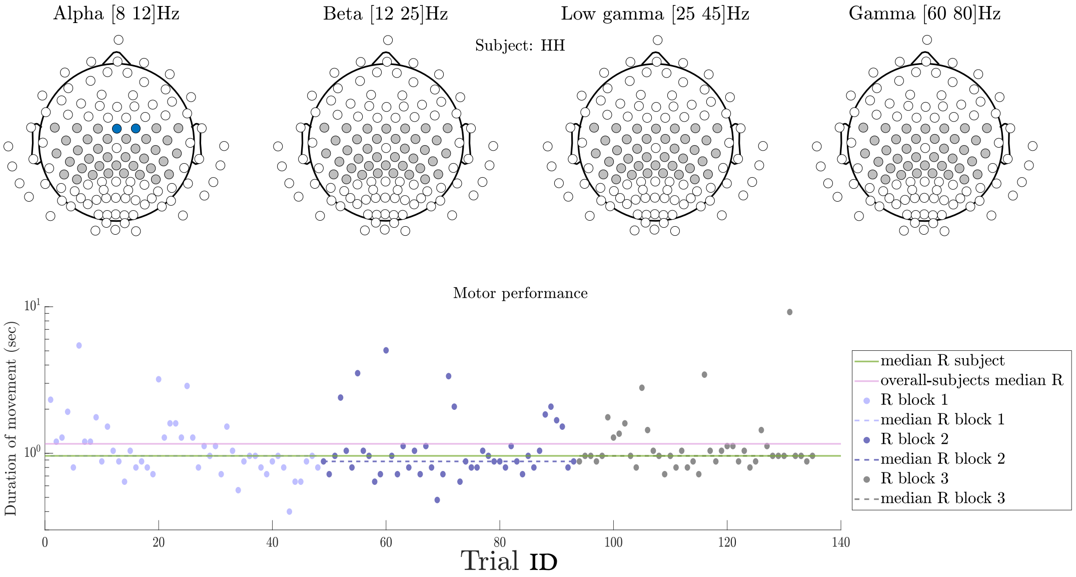


(i)

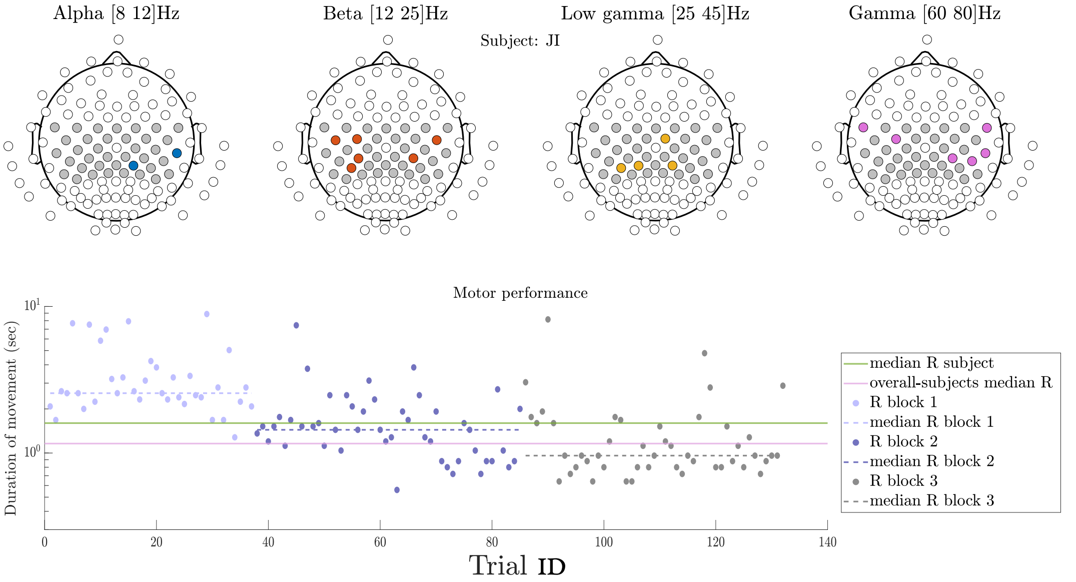
Figure A.3: Electrodes over contralateral motor cortex in the low and high gamma power at subjects that improve their reaching movement duration over the trials, are detected as causal features from our algorithm. Y-axis is in logarithmic scale.

Figure A.3 depicts the subjects that improved their movement, decreasing the duration of their reaching movements throughout the sequence of the trials. We observe that our algorithm detects causes over motor channels in the gamma range (3rd and 4th headplot), which is in accordance with the facilitatory role of cortical gamma power in motor performance. For subjects whose average performance is far below the median performance despite their improvement, also some electrodes in beta range arise.

Figure A.4 depicts the subjects for who our algorithm detected causes over ipsilateral motor channels in the alpha range (1st headplot). For subject JJ, who slightly improves her duration times towards the end, gamma power also arises as a causal feature on the contralateral motor cortex. Finally on subject HG, channels both on contralateral and ipsilateral cortex were detected as causal. Our causal findings in the ipsilateral motor cortex at α-band are consistent with neurophysiological studies that report evidence of increased α-band power over ipsilateral sensorimotor cortex during selection of movement Brinkman *et al.* (2014). Yet, no association of alpha power and arm speed has been reported.

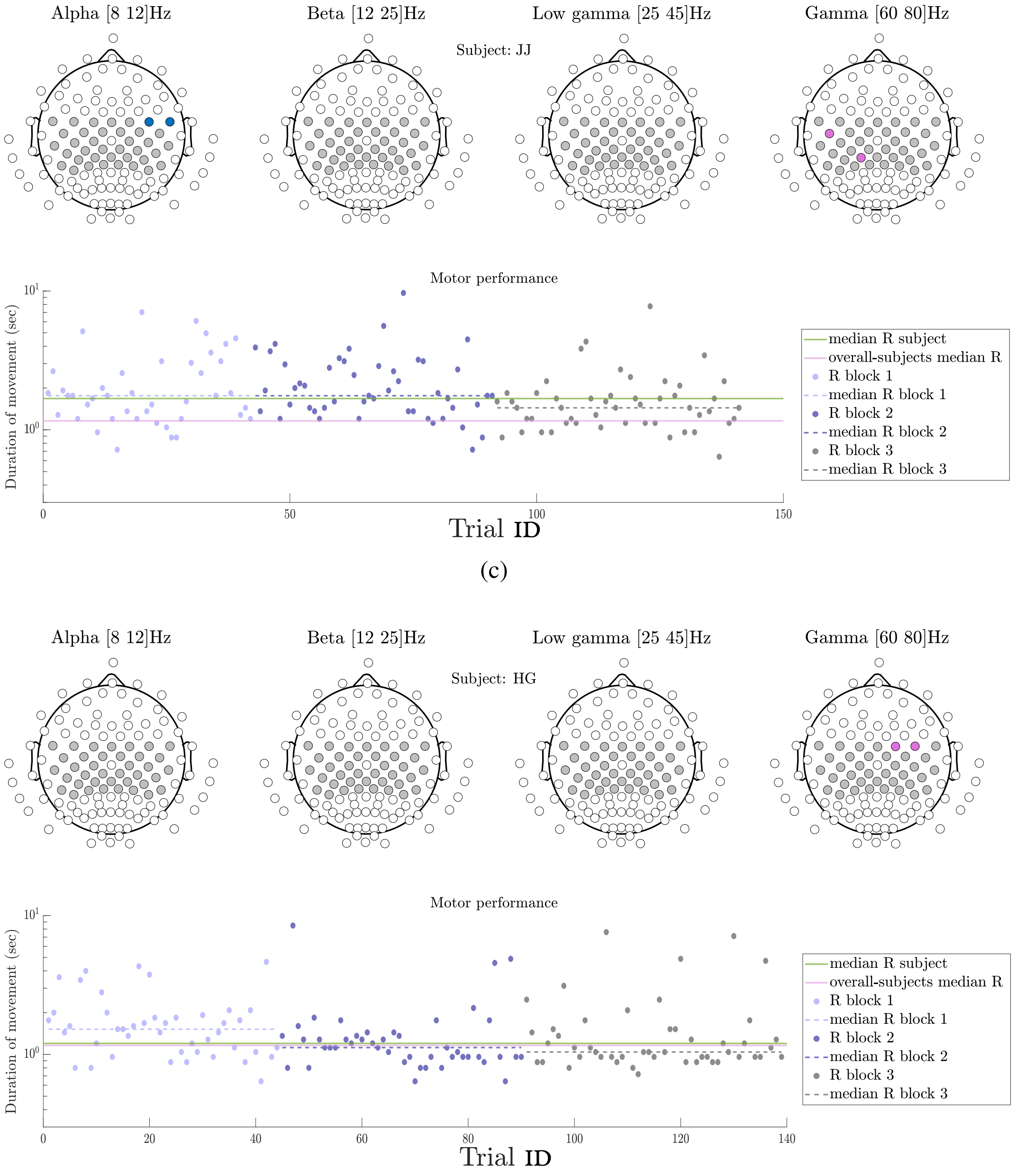

Alpha [8 12]Hz
Beta [12 25]Hz
Low gamma [25 45]Hz
Gamma [60 80]Hz
Subject: HH
Motor performance
Duration of movement (sec)
10^1
10^0
0
20
40
60
80
100
120
140
Trial ID
median R subject
overall-subjects median R
R block 1
median R block 1
R block 2
median R block 2
R block 3
median R block 3

(a)


Alpha [8 12]Hz
Beta [12 25]Hz
Low gamma [25 45]Hz
Gamma [60 80]Hz
Subject: JI
Motor performance
Duration of movement (sec)
10^1
10^0
0
20
40
60
80
100
120
140
Trial ID
median R subject
overall-subjects median R
R block 1
median R block 1
R block 2
median R block 2
R block 3
median R block 3

(b)

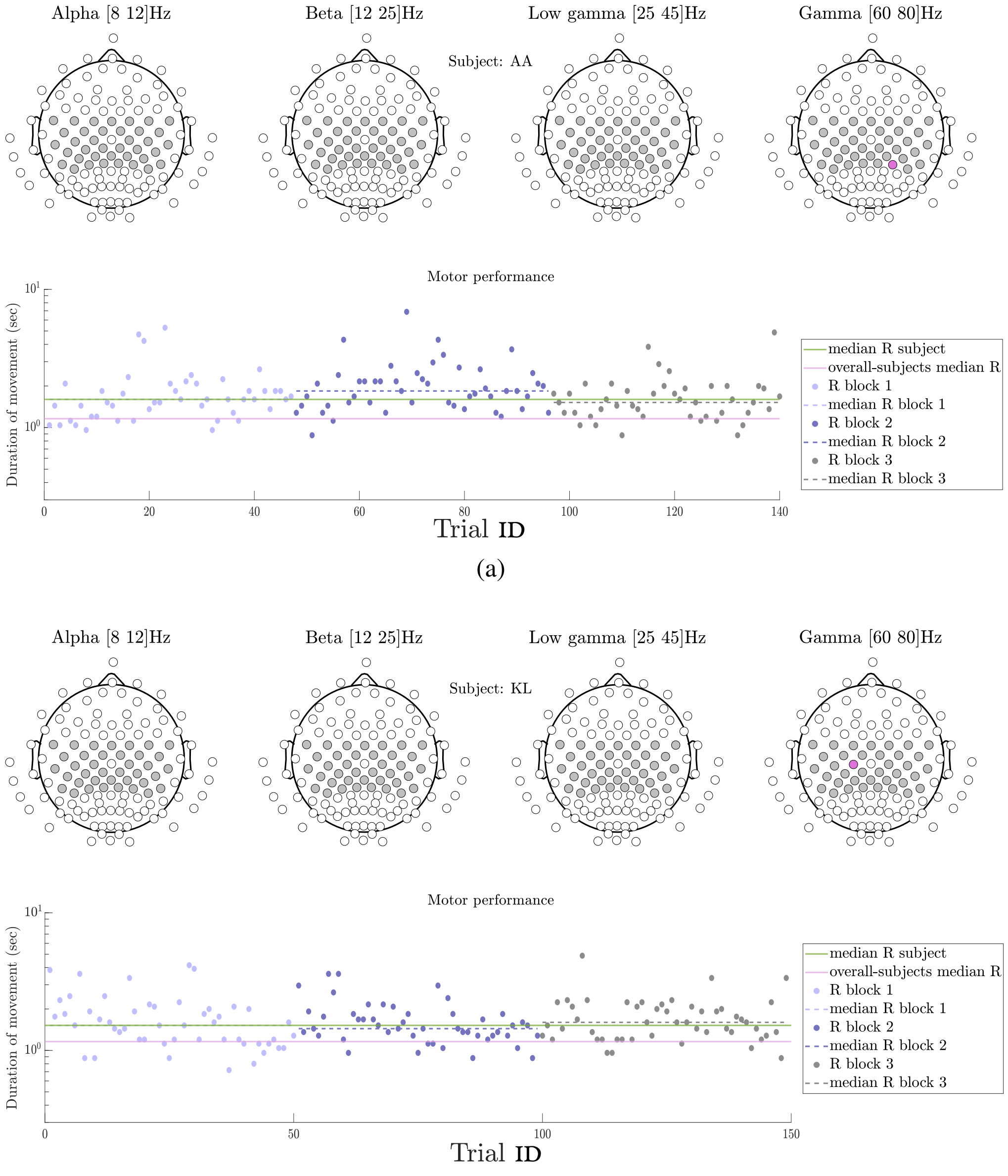
Figure A.4



(c)

(d)

Figure A.4: Electrodes mainly over ipsilateral motor cortex in the alpha power, are detected as causal features from our algorithm in some subjects. Y-axis is in logarithmic scale.



(a)

(b)

Figure A.5: For subjects AA and KL with very slight improvement of reaching movements either in the middle or in the end of the experiment, our algorithm detected one electrode at the gamma range. Y-axis is in logarithmic scale.

Finally in Figure A.5, for subjects AA and KL with very slight improvement of reaching times, which they don't manage to maintain, our algorithm detected one electrode at the gamma range, which may imply again the facilitatory role of gamma, which however is not strong enough to result in improved times.

Subject	Alpha	Beta	Low Gamma	Gamma	Above Group Average	Performance
AA	-	-	-	CPP4h	False	Inhibited but then improved
AB	-	FC2, CCP1h, CPP1h, CP6	-	-	False	Full inhibition
BA	-	-	-	-	True	Full improvement
BB	-	CCP2h, CCP4h, CPP4h	CP1	FCz, FC4, CCP3h, CPP3h	False	Full improvement
CC	-	CCP6h	-	-	True	Inhibited but then improved
CD	-	-	C5	FCC1h	True	Full improvement
DC	FCC5h	CPP2h, CP5, CPz	C2, CCP2h	FC5, CCP2h	False	Inhibited but then improved
DD	-	-	FCC2h	-	True	Improvement but then inhibited
EE	-	-	C2, C5, CCP4h	FC5, CPP5h	True	Improvement but then inhibited
FF	-	FC1, FCz, C2, C3, CPP4h, FCC3h	FC3, FC1, C2, C4, C3, CCP4h, CPP6h, CPP1h, FCC5h, CP5, CP3	FC1, CCP5h, CPP1h, FCC4h, FCC6h, CP5, CP4, CP6	True	Improvement but then inhibited
GG	-	FC4	-	FC4	False	Inhibited but then improved
GH	-	-	-	-	True	Full improvement
HG	C2, C1, CCP3h, CPP5h	CP1	-	-	True	Full improvement
HH	FC2, FCz	-	-	-	True	Improvement but then inhibited
II	-	-	-	FC2, FC4	False	Full improvement
IJ	CP5	-	FC3, FC6, C4	FC5, FC3, FC4, FC6, C2, C4, C6, CCP2h, CCP4h, CPP6h, FCC5h, FCC2h, FCC6h, CP3	False	Full improvement
JI	C2	-	-	-	False	Full improvement
JJ	FC4, FC6	-	-	FCC5h, CP1	False	Full improvement
KK	C6, CP2	CCP4h, CCP3h, FCC3h, FCC5h, FCC6h, CP3	FCC2h, CP3, CP1, CP2	FC5, FC6, C6, CCP4h, CCP6h, FCC3h	False	Full improvement
KL	-	-	-	C1	False	Improvement but then inhibited
LL	-	FCC3h	-	FCC2h	False	Full improvement

Table A.1: Detected causal features for all twenty-one subjects.

Appendix B

Supplementary material for chapter 8

B.1 Proof of lemmas 1, 2

Lemma 1. *If the paths between X^j and Y are directed then the minimum lag w_j as defined in T9 coincides with the minimum non-negative integer w'_j for which $X^j_t \not\perp\!\!\!\perp Y_{t+w'_j} \mid X^j_{past(t)}$. The only case where $w'_j \not\equiv w_j$ is when there is a confounding path between X^j and Y that contains a node from a third time series with memory. In this case $w'_j = 0$.*

Proof. This is obvious by the fact that in the first two cases and when a memoryless confounder exists in the path X^j_t- - -$Y_{t+w'_j}$, the path does not contain horizontal arrows of the type $Q^r_s \to Q^r_{s+1}$. □

Lemma 2. *Theorems 2a/2b and 3 are valid if the minimum lag w_j as defined in T9 is replaced with w'_j obtained in lemma 1.*

Proof. Claims of theorem 2a/2b remain unaffected because the conditions of theorem 2a/2b hold for any lag according to remark 1. According to lemma 1 the only occasion that the minimum non-negative integer w'_j identified by its simple condition, does not coincide with the minimum lag w_j of the definition in T9 is when there exist confounding paths X^j_t- - -Y_{t+w_j} in which the confounder or any intermediate node in the path has memory. In this case w'_j will always be 0. If the confounder in the paths is hidden, then, due to assumption A9 it will be memoryless. In this case the w'_j will coincide with the minimum lag and therefore according to the proof of theorem 3 the appropriate node of X^j will be in the conditioning set and no cause will be rejected. Therefore, it is enough to show that theorem 3 is valid using $w'_j = 0$ when there is an observed confounder in the path.

Assume that condition (2) is violated. Then this will mean that the set $\{X^i_t, S^i, Y_{t+w_i-1}\}$ does not d-separate X^i_{t-1} and Y_{t+w_i}. This would mean that there is a path X^i_{t-1} - - - Y_{t+w_i} in which one of the elements of this set is a collider or descendent of collider and there is no non-collider node in the conditioning set. The proof for the cases (a1), (a2), (a4) and (b) remain the same for the proof of theorem 3. Assume that $X^j_{t+w_{ij'}-1}$ is a collider and no non-collider node in the path belong to the conditioning set. However the observed common causes of $X^j_{t+w_{ij'}-1}$ and Y_{t+w_i-1} are always in the path. Because all these observed common causes are connected via a directed path with Y_{t+w_i-1}, their minimum lag will be correctly identified and so by construction they will be added in the conditioning set. This contradicts the statement "and there is no non-collider node in the path that belongs in the conditioning set". Therefore, we showed that condition (2), thus theorem 3 is not violated. □

B.2 Additional results for simulations with varying noise levels, # observed and hidden time-series

For completeness, here we provide results for all the different number of observed time series that were tested during the simulations.

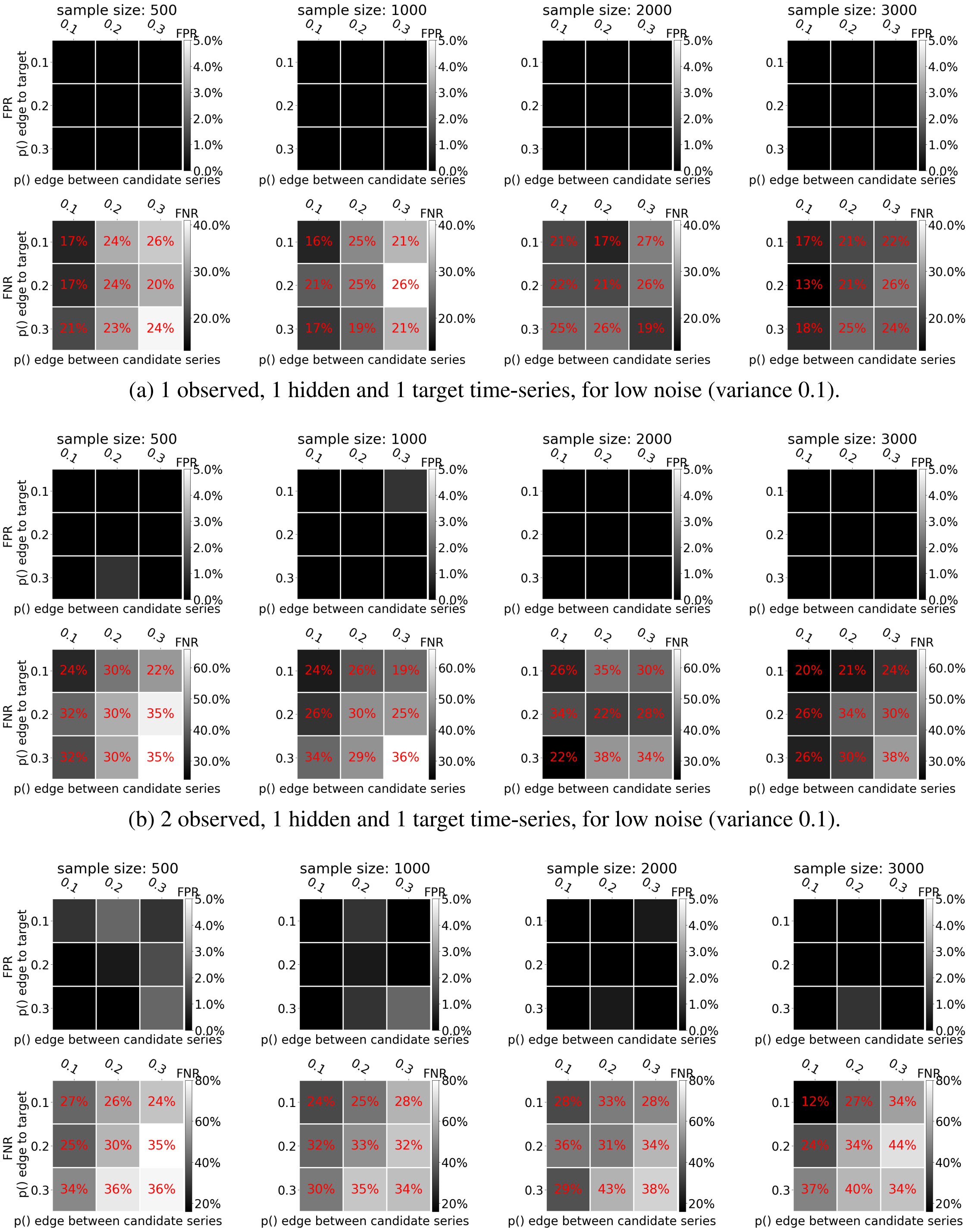
In practice for our simulations where our models are linear with weights < 1 we assume that a shorter indirect edge will have a stronger indirect effect compared to a longer indirect edge. Therefore, we assume that the minimum integer that corresponds to the shortest lag between X^i and Y will also correspond to the maximum coefficient given by the LASSO regression.

B.2.1 FPR and FNR for various densities

Here we provide additional heatmaps for all the noise variance levels and for all various number of observed time series that were simulated, for one hidden time series. The false positive and false negative rates are calculated over 100 random graphs created for each combination tested here.

Overall, the noise in the data does not seem to affect the results for sample sizes ≥ 1000. The false positive rate (FPR) is constantly close to zero for sample size > 500, and is not affected by the density and the size of the graph. The total false negative rate that refers to both direct and indirect missed causes (FNR) seem to gradually increase with the size and the density of the graph. On the other hand, the FNR that refers to the direct causes, for which we proved that our method is complete and sound, does not increase above 50% in very dense and large simulated graphs.

Results for low noise (0.1 noise variance):

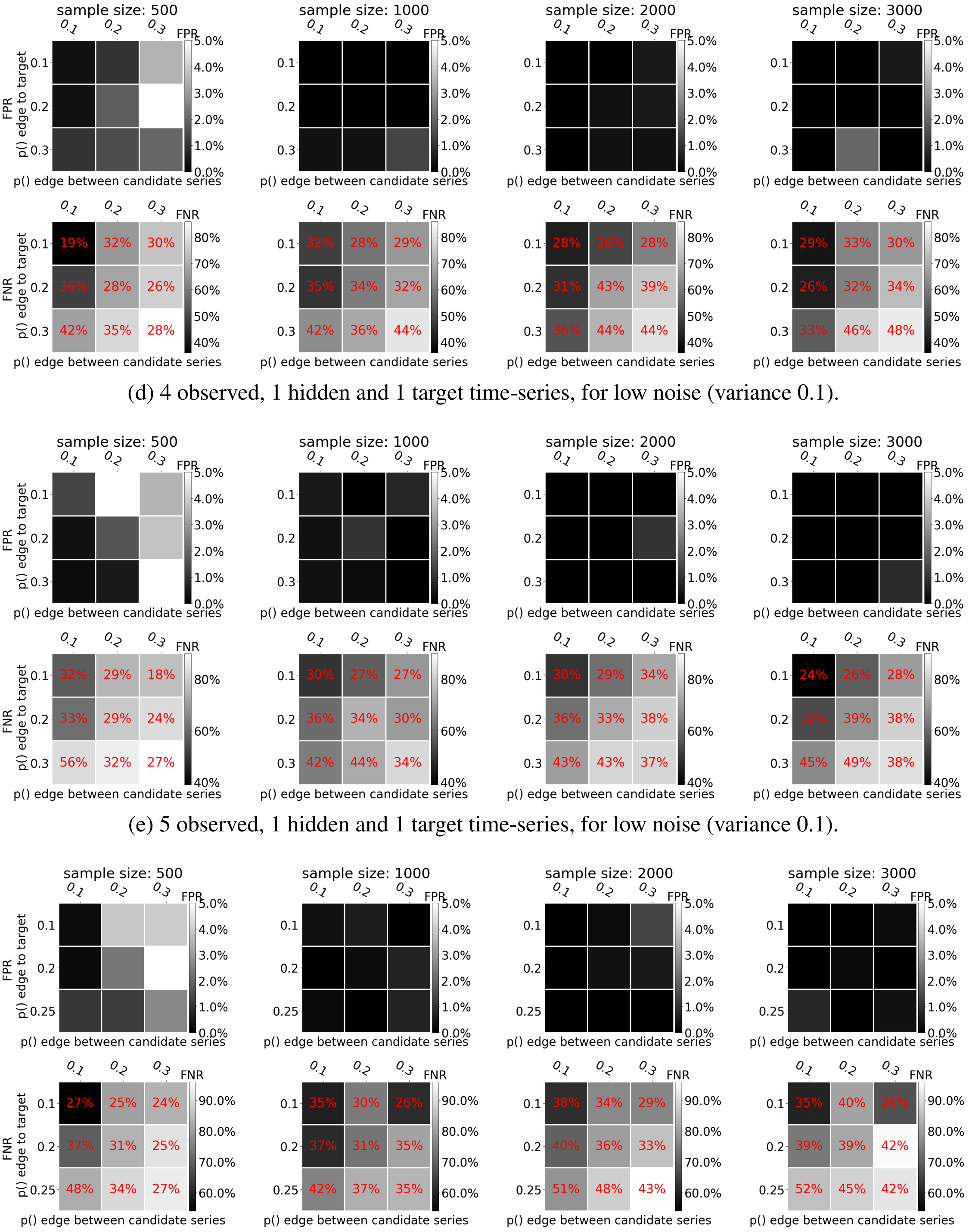


(a) 1 observed, 1 hidden and 1 target time-series, for low noise (variance 0.1).

(b) 2 observed, 1 hidden and 1 target time-series, for low noise (variance 0.1).

(c) 3 observed, 1 hidden and 1 target time-series, for low noise (variance 0.1).

Figure B.1

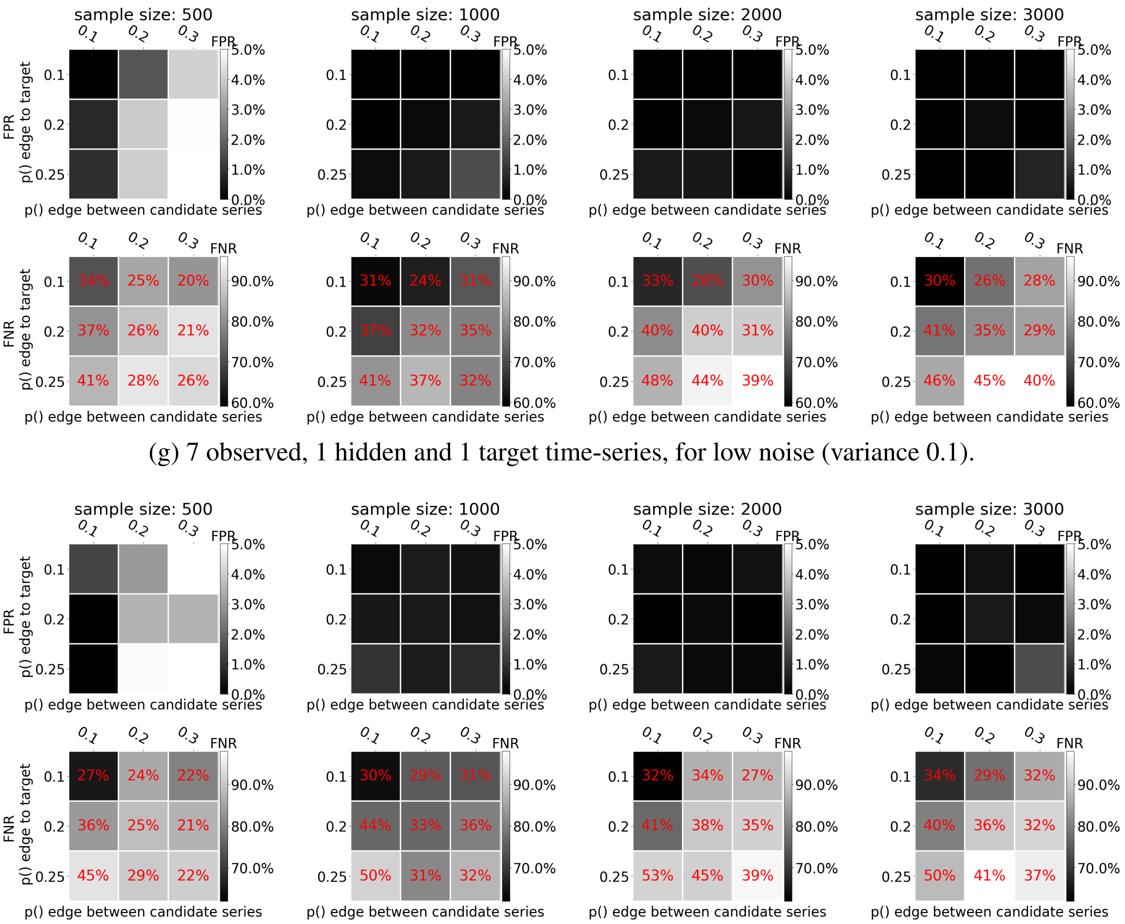


(d) 4 observed, 1 hidden and 1 target time-series, for low noise (variance 0.1).

(e) 5 observed, 1 hidden and 1 target time-series, for low noise (variance 0.1).

(f) 6 observed, 1 hidden and 1 target time-series, for low noise (variance 0.1).

Figure B.1

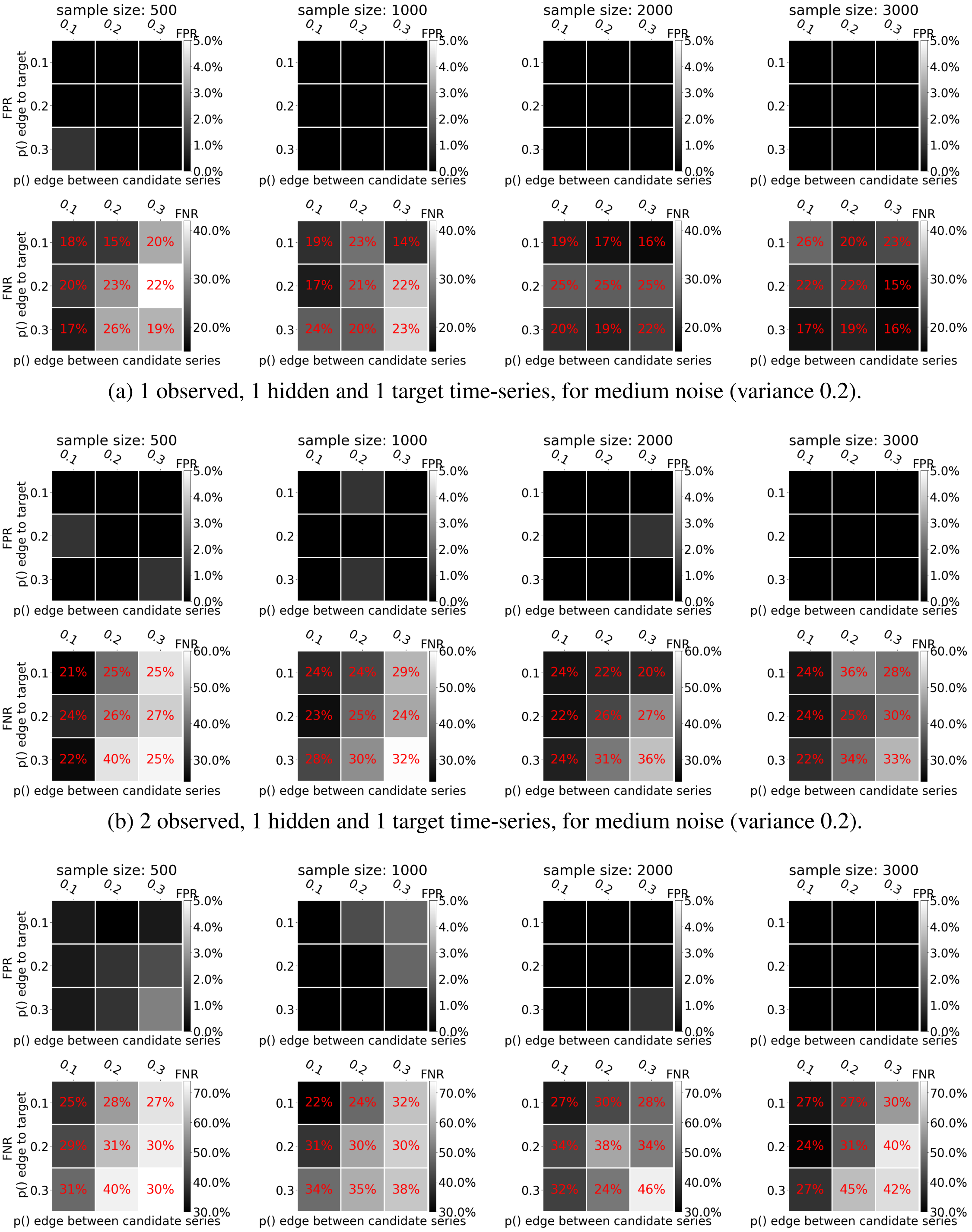


(g) 7 observed, 1 hidden and 1 target time-series, for low noise (variance 0.1).

(h) 8 observed, 1 hidden and 1 target time-series, for low noise (variance 0.1).

Figure B.1: FPR and FNR for low noise, various observed, 1 additional hidden and 1 additional target time-series, for different sample size (columns) and sparsity of edges among the candidate causes (x-axis) and between the candidate causes and the target (y-axis). The total FNR (for indirect and direct causes) is depicted by the heatmap color. The FNR that refers to the direct causes (for which our method is proved to be complete) is depicted with red in the middle of each cell. Overall we see tat for sample size above 500 the false positives are very low and they keep decreasing as the number of examples increase. False negatives for both direct and indirect causes increase with the number of nodes and the density of the graph, however the FNR that refers only to the direct causes for which our method provides necessary conditions (red coloured numbers) ranges just from 12% up to 52% for dense large graphs.

Results for medium noise (0.2 noise variance):

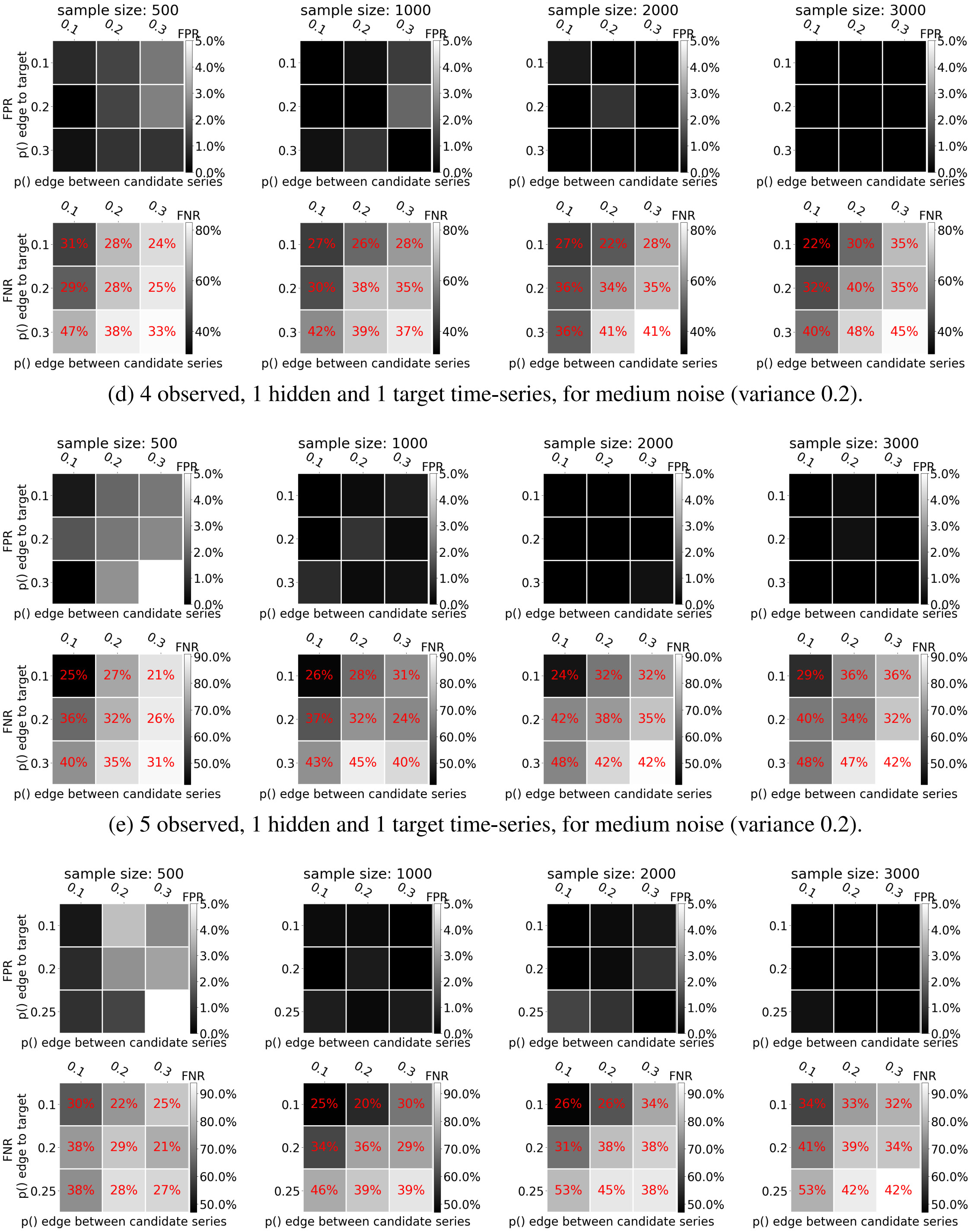


(a) 1 observed, 1 hidden and 1 target time-series, for medium noise (variance 0.2).

(b) 2 observed, 1 hidden and 1 target time-series, for medium noise (variance 0.2).

(c) 3 observed, 1 hidden and 1 target time-series, for medium noise (variance 0.2).

Figure B.2

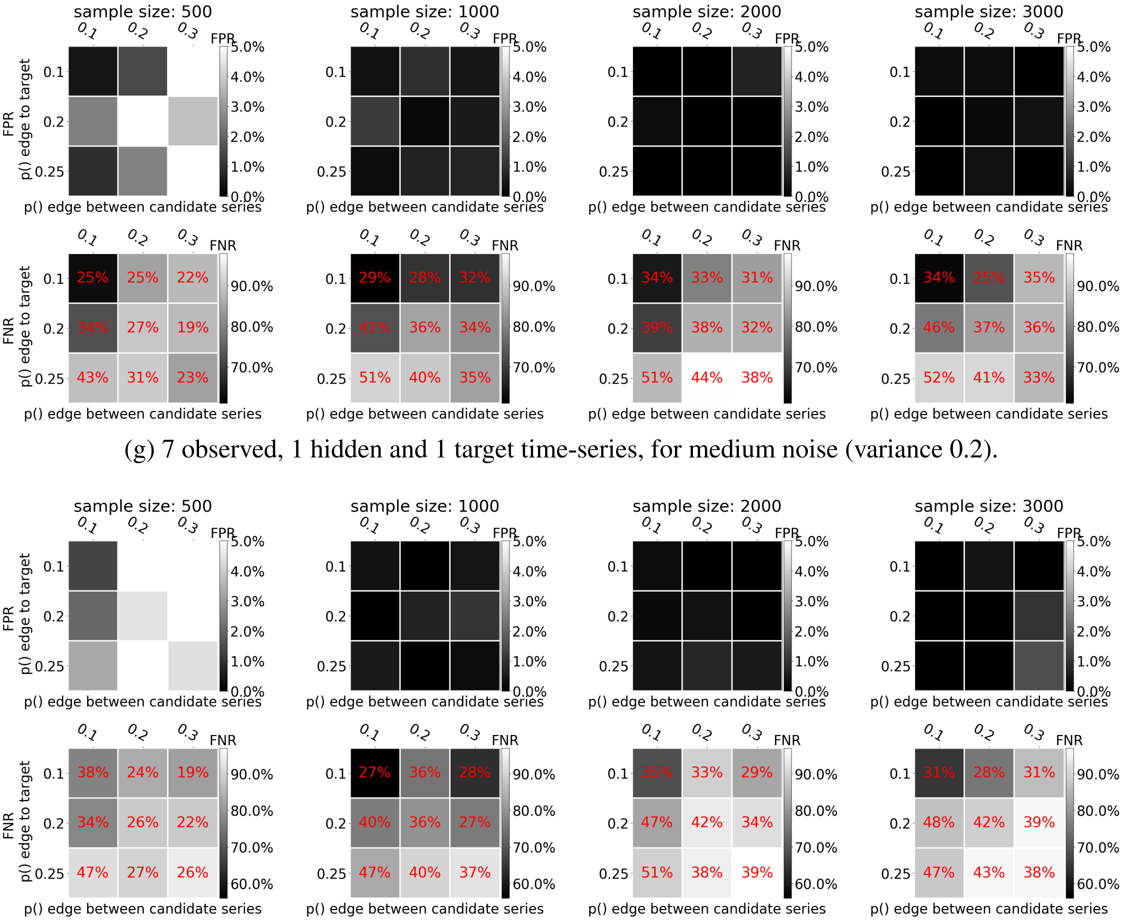


(d) 4 observed, 1 hidden and 1 target time-series, for medium noise (variance 0.2).

(e) 5 observed, 1 hidden and 1 target time-series, for medium noise (variance 0.2).

(f) 6 observed, 1 hidden and 1 target time-series, for medium noise (variance 0.2).

Figure B.2

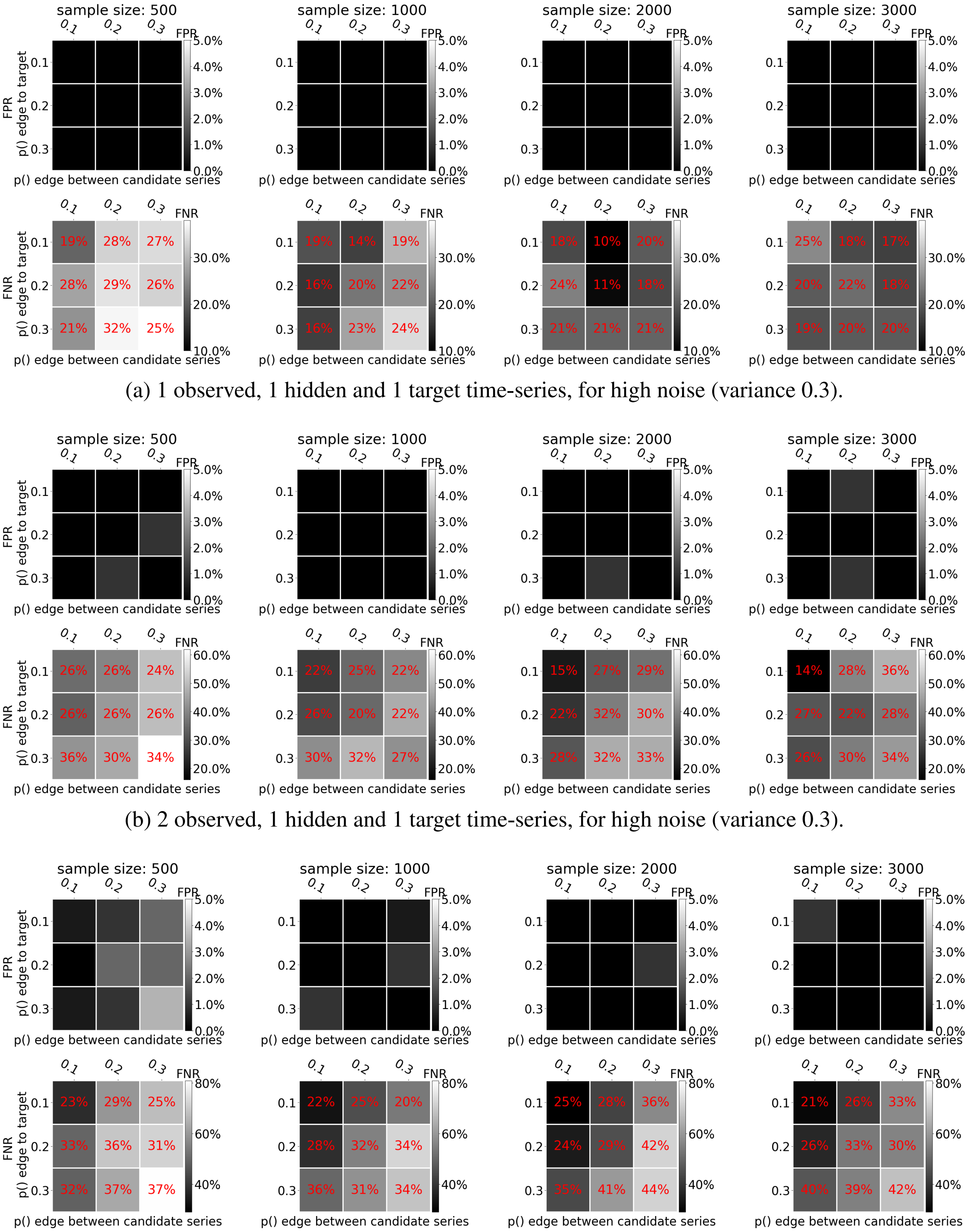


(g) 7 observed, 1 hidden and 1 target time-series, for medium noise (variance 0.2).

(h) 8 observed, 1 hidden and 1 target time-series, for medium noise (variance 0.2).

Figure B.2: FPR and FNR for medium noise, various observed, 1 additional hidden and 1 additional target time-series, for different sample size (columns) and sparsity of edges among the candidate causes (x-axis) and between the candidate causes and the target (y-axis). Similar to the rest of the noise levels, the total FNR (for indirect and direct causes) is depicted by the heatmap color. The FNR that refers to the direct causes (for which our method is proved to be complete) is depicted with red in the middle of each cell. Overall we see tat for sample size above 500 the false positives are very low and they keep decreasing as the number of examples increase. False negatives for both direct and indirect causes increases with the number of nodes and the density of the graph, however the FNR that refers only to the direct causes for which our method provides necessary conditions (red coloured numbers) ranges just from 15% up to 52% for dense large graphs.

Results for high noise (0.3 noise variance):

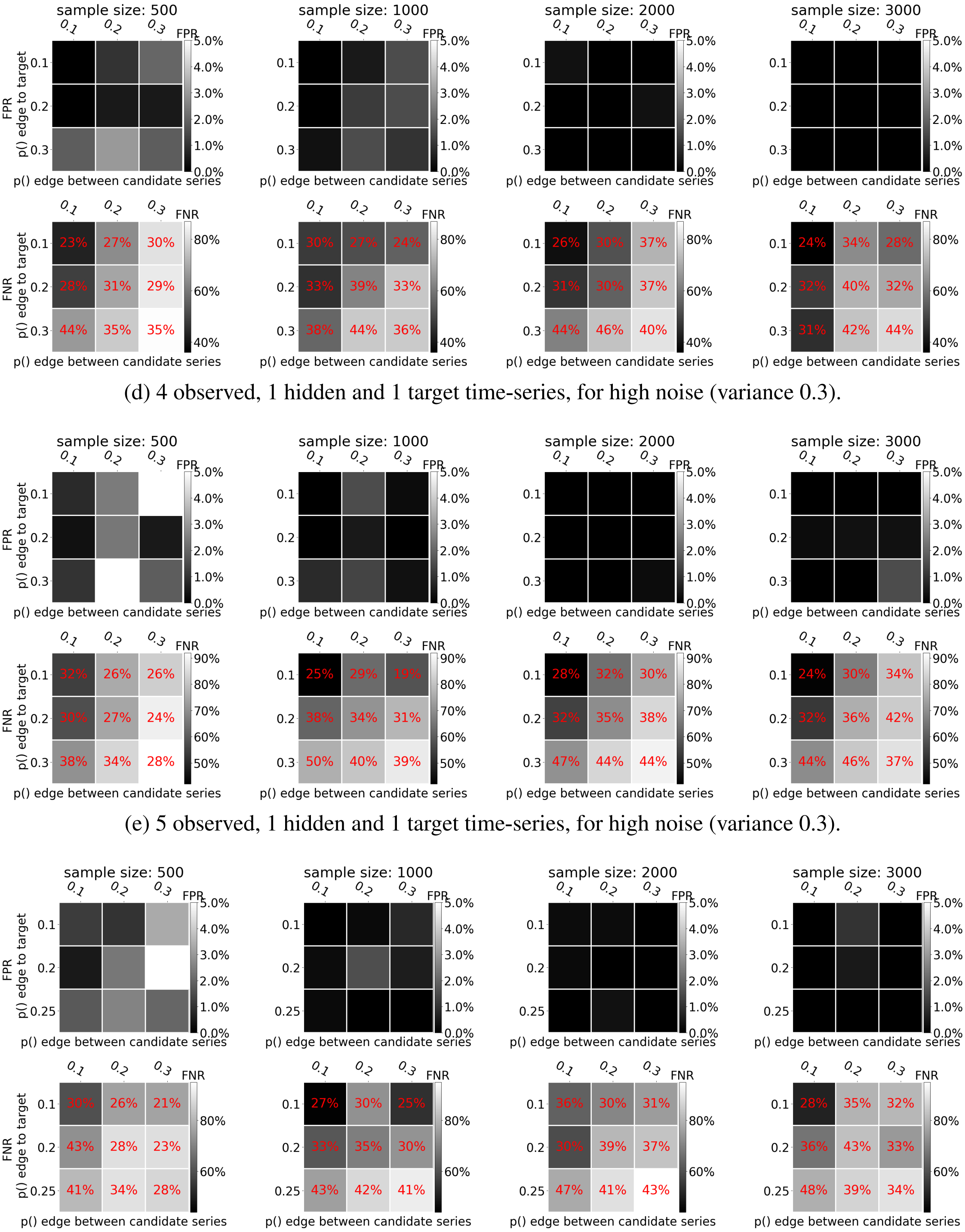


(a) 1 observed, 1 hidden and 1 target time-series, for high noise (variance 0.3).

(b) 2 observed, 1 hidden and 1 target time-series, for high noise (variance 0.3).

(c) 3 observed, 1 hidden and 1 target time-series, for high noise (variance 0.3).

Figure B.3

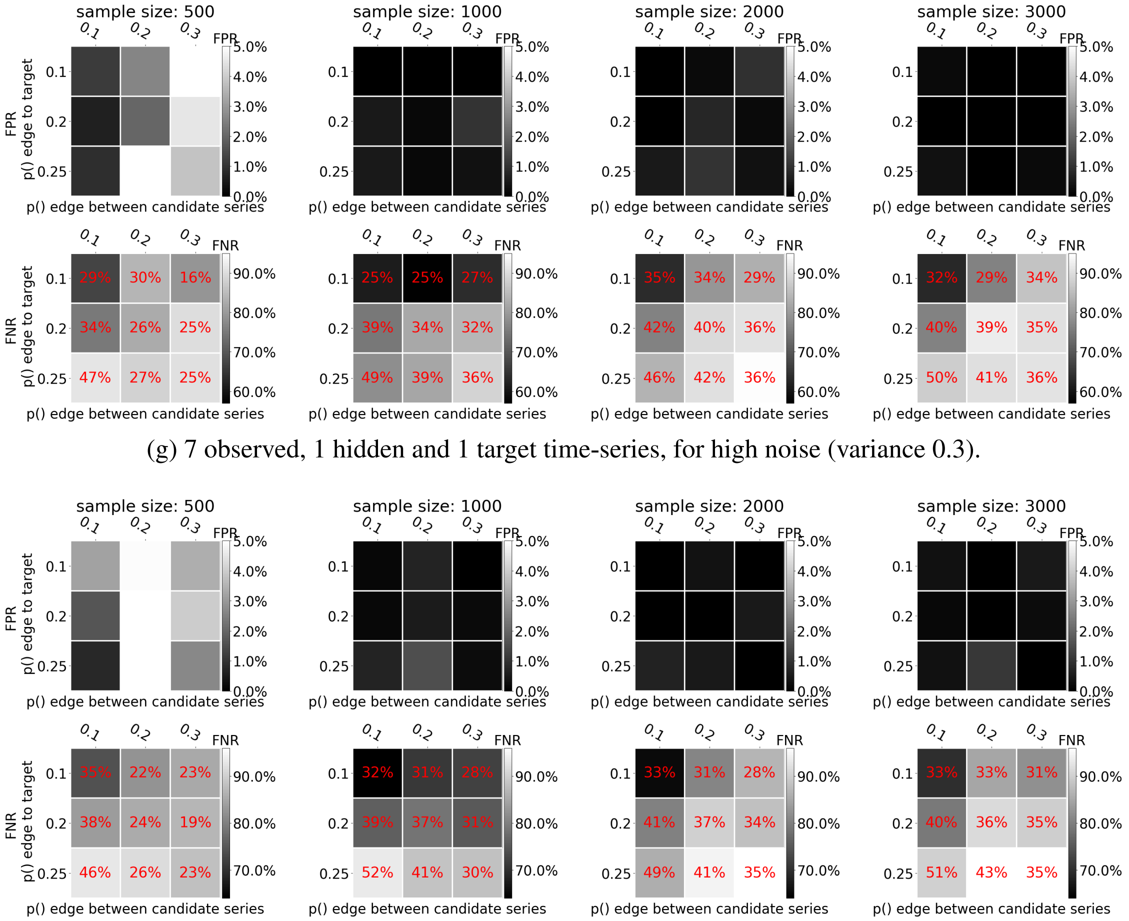


(d) 4 observed, 1 hidden and 1 target time-series, for high noise (variance 0.3).

(e) 5 observed, 1 hidden and 1 target time-series, for high noise (variance 0.3).

(f) 6 observed, 1 hidden and 1 target time-series, for high noise (variance 0.3).

Figure B.3



(g) 7 observed, 1 hidden and 1 target time-series, for high noise (variance 0.3).

(h) 8 observed, 1 hidden and 1 target time-series, for high noise (variance 0.3).

Figure B.3: FPR and FNR for high noise, various observed, 1 additional hidden and 1 additional target time-series, for different sample size (columns) and sparsity of edges among the candidate causes (x-axis) and between the candidate causes and the target (y-axis). Similar to the rest of the noise levels, the total FNR (for indirect and direct causes) is depicted by the heatmap color. The FNR that refers to the direct causes (for which our method is proved to be complete) is depicted with red in the middle of each cell. Overall we see tat for sample size above 500 the false positives are very low and they keep decreasing as the number of examples increase. False negatives for both direct and indirect causes increases with the number of nodes and the density of the graph, however the FNR that refers only to the direct causes for which our method provides necessary conditions (red coloured numbers) ranges just from 11% up to 51% for dense large graphs.

B.2.2 FPR and FNR with varing number of hidden variables and various densities

In the presence of zero hidden variables our method has practically 0 false positives, which reaches up to 0.7% for large noise, which again is practically zero.

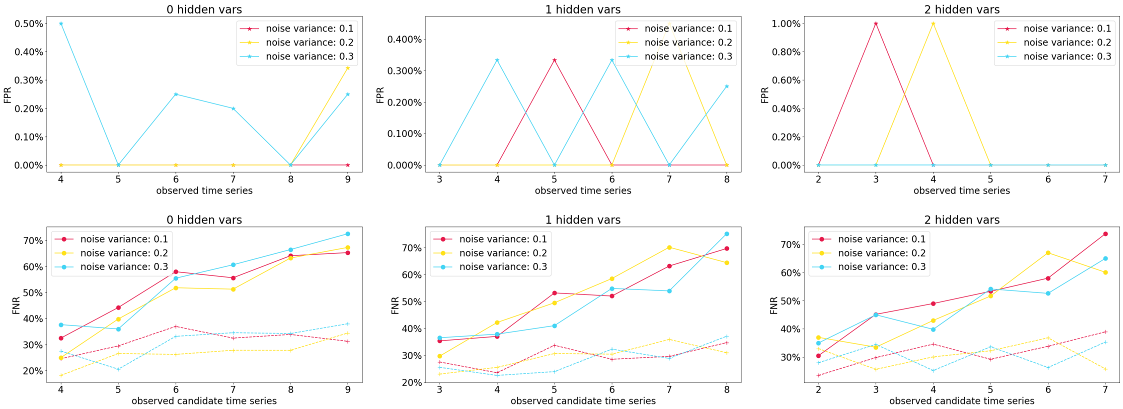


Figure B.4: FPR and FNR for various number of hidden and observed series, noise variance and sample size 2000, for sparse edges among the $\mathbf{X}$ and Y (0.1, 0.1). As we can see, FPR is very low (max 1%) for any number of hidden series. Although the total FNR is gradually increasing with the graph size, notice that the FNR that corresponds to direct causes (dashed lines, for which our method is complete) does not exceed 35%.

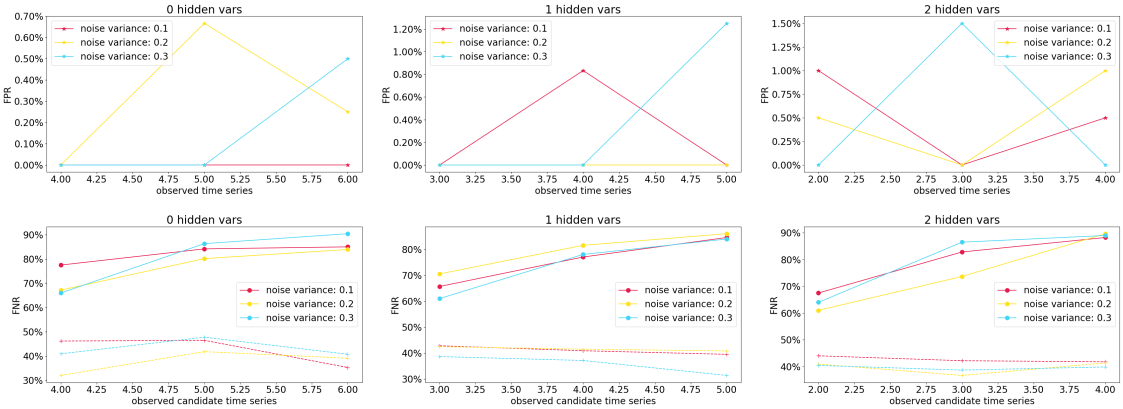


Figure B.5: FPR and FNR for various number of hidden and observed series, noise variance and sample size 2000, for dense edges among the $\mathbf{X}$ and Y (0.3, 0.3). As we can see, FPR remains very low (max 1.5% for high noise) for any number of hidden series. Although the total FNR is gradually increasing with the graph size, notice that the FNR that corresponds to direct causes (dashed lines, for which our method is complete) does not exceed 45%.

Bibliography

Aliferis, C. F., Tsamardinos, I., and Statnikov, A. (2003). Hiton: a novel markov blanket algorithm for optimal variable selection. In *AMIA annual symposium proceedings*, volume 2003, page 21. American Medical Informatics Association.

Anand, S. and Hotson, J. (2002). Transcranial magnetic stimulation: neurophysiological applications and safety. *Brain and cognition*, **50**(3), 366–386.

Ang, K. K., Guan, C., Phua, K. S., Wang, C., Zhou, L., Tang, K. Y., Joseph, E., Gopal, J., Kuah, C. W. K., and Chua, K. S. G. (2014). Brain-computer interface-based robotic end effector system for wrist and hand rehabilitation: results of a three-armed randomized controlled trial for chronic stroke. *Frontiers in neuroengineering*, **7**, 30.

Ang, K. K., Chua, K. S. G., Phua, K. S., Wang, C., Chin, Z. Y., Kuah, C. W. K., Low, W., and Guan, C. (2015). A randomized controlled trial of eeg-based motor imagery brain-computer interface robotic rehabilitation for stroke. *Clinical EEG and neuroscience*, **46**(4), 310–320.

Antal, A. and Herrmann, C. S. (2016). Transcranial alternating current and random noise stimulation: possible mechanisms. *Neural plasticity*, **2016**.

Antal, A., Boros, K., Poreisz, C., Chaieb, L., Terney, D., and Paulus, W. (2008). Comparatively weak after-effects of transcranial alternating current stimulation (tacs) on cortical excitability in humans. *Brain stimulation*, **1**(2), 97–105.

Arnold, A., Liu, Y., and Abe, N. (2007). Temporal causal modeling with graphical granger methods. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 66–75.

Axmacher, N., Mormann, F., Fernández, G., Elger, C. E., and Fell, J. (2006). Memory formation by neuronal synchronization. *Brain research reviews*, **52**(1), 170–182.

Bales, J. W., Bonow, R. H., and Ellenbogen, R. G. (2018). Closed head injury. In *Principles of Neurological Surgery*, pages 366–389. Elsevier.

Bartos, M., Vida, I., and Jonas, P. (2007). Synaptic mechanisms of synchronized gamma oscillations in inhibitory interneuron networks. *Nature reviews neuroscience*, **8**(1), 45–56.

Belouchrani, A., Abed-Meraim, K., Cardoso, J., and Moulines, E. (1993). Second-order blind separation of temporally correlated sources. In *Proc. Int. Conf. Digital Signal Processing*, pages 346–351. Citeseer.

Bendat, J. S. and Piersol, A. G. (2011). *Random data: analysis and measurement procedures*, volume 729. John Wiley & Sons.

Benington, J. H. and Frank, M. G. (2003). Cellular and molecular connections between sleep and synaptic plasticity. *Progress in neurobiology*, **69**(2), 71–101.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, **57**(1), 289–300.

Bestmann, S. and Walsh, V. (2017). Transcranial electrical stimulation. *Current Biology*, **27**(23), R1258–R1262.

Brinkman, L., Stolk, A., Dijkerman, H. C., de Lange, F. P., and Toni, I. (2014). Distinct roles for alpha-and beta-band oscillations during mental simulation of goal-directed actions. *Journal of Neuroscience*, **34**(44), 14783–14792.

Brown, P. (2003). Oscillatory nature of human basal ganglia activity: relationship to the pathophysiology of parkinson's disease. *Movement disorders: official journal of the Movement Disorder Society*, **18**(4), 357–363.

Brown, P. (2007). Abnormal oscillatory synchronisation in the motor system leads to impaired movement. *Current opinion in neurobiology*, **17**(6), 656–664.

Brunel, N. and Wang, X.-J. (2003). What determines the frequency of fast network oscillations with irregular neural discharges? i. synaptic dynamics and excitation-inhibition balance. *Journal of neurophysiology*, **90**(1), 415–430.

Buch, E. R., Santarnecchi, E., Antal, A., Born, J., Celnik, P. A., Classen, J., Gerloff, C., Hallett, M., Hummel, F. C., Nitsche, M. A., *et al.* (2017). Effects of tdcs on motor learning and memory formation: a consensus and critical position paper. *Clinical Neurophysiology*, **128**(4), 589–603.

Butler, A. J., Shuster, M., O'Hara, E., Hurley, K., Middlebrooks, D., and Guilkey, K. (2013). A meta-analysis of the efficacy of anodal transcranial direct current stimulation for upper limb motor recovery in stroke survivors. *Journal of Hand Therapy*, **26**(2), 162–171.

Butler, R., Bernier, P.-M., Mierzwinski, G. W., Descoteaux, M., Gilbert, G., and Whittingstall, K. (2019). Cortical distance, not cancellation, dominates inter-subject eeg gamma rhythm amplitude. *NeuroImage*, **192**, 156–165.

Cabel, D. W., Cisek, P., and Scott, S. H. (2001). Neural activity in primary motor cortex related to mechanical loads applied to the shoulder and elbow during a postural task. *Journal of neurophysiology*, **86**(4), 2102–2108.

Campbell, A. W. (1905). *Histological studies on the localisation of cerebral function*. University Press.

Cartwright, N. (2010). What are randomised controlled trials good for? *Philosophical studies*, **147**(1), 59.

Cecere, R., Rees, G., and Romei, V. (2015). Individual differences in alpha frequency drive crossmodal illusory perception. *Current Biology*, **25**(2), 231–235.

Chalupka, K., Perona, P., and Eberhardt, F. (2018). Fast conditional independence test for vector variables with large sample sizes. *arXiv preprint arXiv:1804.02747*.

Chen, C.-M. A., Stanford, A. D., Mao, X., Abi-Dargham, A., Shungu, D. C., Lisanby, S. H., Schroeder, C. E., and Kegeles, L. S. (2014). Gaba level, gamma oscillation, and working memory performance in schizophrenia. *NeuroImage: Clinical*, **4**, 531–539.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., Robins, J., *et al.* (2017). Double/debiased machine learning for treatment and causal parameters. Technical report.

Chickering, D. M. (2002). Optimal structure identification with greedy search. *Journal of machine learning research*, **3**(Nov), 507–554.

Cirillo, G., Di Pino, G., Capone, F., Ranieri, F., Florio, L., Todisco, V., Tedeschi, G., Funke, K., and Di Lazzaro, V. (2017). Neurobiological after-effects of non-invasive brain stimulation. *Brain stimulation*, **10**(1), 1–18.

Cocchi, L. and Zalesky, A. (2018). Personalized transcranial magnetic stimulation in psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, **3**(9), 731–741.

Colombo, D., Maathuis, M. H., Kalisch, M., and Richardson, T. S. (2012). Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics*, pages 294–321.

Cozens, J. A. and Bhakta, B. B. (2003). Measuring movement irregularity in the upper motor neurone syndrome using normalised average rectified jerk. *Journal of Electromyography and Kinesiology*, **13**(1), 73–81.

Daniusis, P., Janzing, D., Mooij, J., Zscheischler, J., Steudel, B., Zhang, K., and Schölkopf, B. (2012). Inferring deterministic causal relations. *arXiv preprint arXiv:1203.3475*.

Darvas, F., Pantazis, D., Kucukaltun-Yildirim, E., and Leahy, R. (2004). Mapping human brain function with meg and eeg: methods and validation. *NeuroImage*, **23**, S289–S299.

Datta, A. (2012). Inter-individual variation during transcranial direct current stimulation and normalization of dose using mri-derived computational models. *Frontiers in psychiatry*, **3**, 91.

Davis, N. J. and Koningsbruggen, M. V. (2013). non-invasive brain stimulation is not non-invasive. *Frontiers in systems neuroscience*, **7**, 76.

Davis, N. J., Tomlinson, S. P., and Morgan, H. M. (2012). The role of beta-frequency neural oscillations in motor control. *Journal of Neuroscience*, **32**(2), 403–404.

Dayan, E., Censor, N., Buch, E. R., Sandrini, M., and Cohen, L. G. (2013). Noninvasive brain stimulation: from physiology to network dynamics and back. *Nature neuroscience*, **16**(7), 838–844.

Destexhe, A., Hughes, S. W., Rudolph, M., and Crunelli, V. (2007). Are corticothalamic upstates fragments of wakefulness? *Trends in neurosciences*, **30**(7), 334–342.

Di Lazzaro, V., Oliviero, A., Pilato, F., Saturno, E., Dileone, M., Mazzone, P., Insola, A., Tonali, P., and Rothwell, J. (2004). The physiological basis of transcranial motor cortex stimulation in conscious humans. *Clinical neurophysiology*, **115**(2), 255–266.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental brain research*, **91**(1), 176–180.

Di Pino, G., Pellegrino, G., Assenza, G., Capone, F., Ferreri, F., Formica, D., Ranieri, F., Tombini, M., Ziemann, U., Rothwell, J. C., *et al.* (2014). Modulation of brain plasticity in stroke: a novel model for neurorehabilitation. *Nature Reviews Neurology*, **10**(10), 597–608.

Dickson, C. T., Magistretti, J., Shalinsky, M., Hamam, B., and Alonso, A. (2000). Oscillatory activity in entorhinal neurons and circuits: Mechanisms and function. *Annals of the New York Academy of Sciences*, **911**(1), 127–150.

Dmochowski, J. P., Datta, A., Bikson, M., Su, Y., and Parra, L. C. (2011). Optimized multi-electrode stimulation increases focality and intensity at target. *Journal of neural engineering*, **8**(4), 046011.

Doran, G., Muandet, K., Zhang, K., and Schölkopf, B. (2014). A permutation-based kernel conditional independence test. In *UAI*, pages 132–141.

Ebbesen, C. L. and Brecht, M. (2017). Motor cortexto act or not to act? *Nature Reviews Neuroscience*, **18**(11), 694.

Eichler, M. (2007). Causal inference from time series: What can be learned from Granger causality. In *Proceedings of the 13th International Congress of Logic, Methodology and Philosophy of Science*, pages 1–12. King's College Publications London.

Engel, A. K. and Fries, P. (2010). Beta-band oscillationssignalling the status quo? *Current opinion in neurobiology*, **20**(2), 156–165.

Entner, D. and Hoyer, P. O. (2010a). On causal discovery from time series data using fci. *Probabilistic graphical models*, pages 121–128.

Entner, D. and Hoyer, P. O. (2010b). On causal discovery from time series data using FCI. *Probabilistic graphical models*, pages 121–128.

Espenhahn, S. (2018). *The relationship between cortical beta oscillations and motor learning*. Ph.D. thesis, UCL (University College London).

EU (2019). European union prices of dairy products. `https://ec.europa.eu/info/food-farming-fisheries/farming/facts-and-figures/markets/prices/price-monitoring-sector/`.

Ferrier, D. (1875). Experiments on the brain of monkeys.no. i. *Proceedings of the Royal Society of London*, **23**(156-163), 409–430.

Filmer, H. L., Dux, P. E., and Mattingley, J. B. (2014). Applications of transcranial direct current stimulation for understanding brain function. *Trends in neurosciences*, **37**(12), 742–753.

Frank, M. G. (2009). *Brain Rhythms*, pages 482–483. Springer Berlin Heidelberg, Berlin, Heidelberg.

Fregni, F., Simon, D., Wu, A., and Pascual-Leone, A. (2005). Non-invasive brain stimulation for parkinsons disease: a systematic review and meta-analysis of the literature. *Journal of Neurology, Neurosurgery & Psychiatry*, **76**(12), 1614–1623.

Fritsch, G. (1870). Uber die elektrische erregbarkeit des grosshirns. *Arch, anat. Physiol. Wiss. Med.*, **37**, 300–332.

Frølich, L. and Dowding, I. (2018). Removal of muscular artifacts in eeg signals: a comparison of linear decomposition methods. *Brain informatics*, **5**(1), 13–22.

Fu, Q.-G., Suarez, J. I., and Ebner, T. J. (1993). Neuronal specification of direction and distance during reaching movements in the superior precentral premotor area and primary motor cortex of monkeys. *Journal of neurophysiology*, **70**(5), 2097–2116.

Fu, S. and Desmarais, M. C. (2010). Markov blanket based feature selection: a review of past decade. In *Proceedings of the world congress on engineering*, volume 1, pages 321–328. Newswood Ltd.

Fujiyama, H., Hyde, J., Hinder, M. R., Kim, S.-J., McCormack, G. H., Vickers, J. C., and Summers, J. J. (2014). Delayed plastic responses to anodal tdcs in older adults. *Frontiers in aging neuroscience*, **6**, 115.

Fukumizu, K., Gretton, A., Sun, X., and Schölkopf, B. (2008). Kernel measures of conditional dependence. In *Advances in neural information processing systems*, pages 489–496.

Gaetz, W., Liu, C., Zhu, H., Bloy, L., and Roberts, T. P. (2013). Evidence for a motor gamma-band network governing response interference. *Neuroimage*, **74**, 245–253.

Georgopoulos, A., Caminiti, R., and Kalaska, J. (1984). Static spatial effects in motor cortex and area 5: quantitative relations in a two-dimensional space. *Experimental Brain Research*, **54**(3), 446–454.

Georgopoulos, A. P., Kalaska, J. F., Caminiti, R., and Massey, J. T. (1982). On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *Journal of Neuroscience*, **2**(11), 1527–1537.

Glymour, C., Zhang, K., and Spirtes, P. (2019). Review of causal discovery methods based on graphical models. *Frontiers in Genetics*, **10**.

Gomez-Rodriguez, M., Peters, J., Hill, J., Schölkopf, B., Gharabaghi, A., and Grosse-Wentrup, M. (2011). Closing the sensorimotor loop: haptic feedback facilitates decoding of motor imagery. *Journal of neural engineering*, **8**(3), 036005.

Gonzalez Andino, S. L., Michel, C. M., Thut, G., Landis, T., and Grave de Peralta, R. (2005). Prediction of response speed by anticipatory high-frequency (gamma band) oscillations in the human brain. *Human brain mapping*, **24**(1), 50–58.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438.

Granger, C. W. (1980). Testing for causality: a personal viewpoint. *Journal of Economic Dynamics and control*, **2**, 329–352.

Green, J., Forster, A., Bogle, S., and Young, J. (2002). Physiotherapy for patients with mobility problems more than 1 year after stroke: a randomised controlled trial. *The Lancet*, **359**(9302), 199–203.

Greenland, S. (2000). An introduction to instrumental variables for epidemiologists. *International journal of epidemiology*, **29**(4), 722–729.

Grefkes, C., Nowak, D. A., Eickhoff, S. B., Dafotakis, M., Küst, J., Karbe, H., and Fink, G. R. (2008). Cortical connectivity after subcortical stroke assessed with functional magnetic resonance imaging. *Annals of neurology*, **63**(2), 236–246.

Gretton, A., Herbrich, R., Smola, A., Bousquet, O., and Schölkopf, B. (2005a). Kernel methods for measuring independence. *Journal of Machine Learning Research*, **6**(Dec), 2075–2129.

Gretton, A., Bousquet, O., Smola, A., and Schölkopf, B. (2005b). Measuring statistical dependence with hilbert-schmidt norms. In *International conference on algorithmic learning theory*, pages 63–77. Springer.

Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Schölkopf, B., and Smola, A. J. (2008). A kernel statistical test of independence. In *Advances in neural information processing systems*, pages 585–592.

Grosse-Wentrup, M. and Schölkopf, B. (2012). High gamma-power predicts performance in sensorimotor-rhythm brain–computer interfaces. *Journal of Neural Engineering*, **9**(4), 046001.

Grosse-Wentrup, M., Mattia, D., and Oweiss, K. (2011). Using brain–computer interfaces to induce neural plasticity and restore function. *Journal of neural engineering*, **8**(2), 025004.

Grosse-Wentrup, M., Janzing, D., Siegel, M., and Schölkopf, B. (2016). Identification of causal relations in neuroimaging data with latent confounders: An instrumental variable approach. *NeuroImage*, **125**, 825–833.

Grundey, J., Thirugnanasambandam, N., Kaminsky, K., Drees, A., Skwirba, A., Lang, N., Paulus, W., and Nitsche, M. A. (2012). Rapid effect of nicotine intake on neuroplasticity in non-smoking humans. *Frontiers in pharmacology*, **3**, 186.

Gulberti, A., Moll, C. K. E., Hamel, W., Buhmann, C., Koeppen, J., Boelmans, K., Zittel, S., Gerloff, C., Westphal, M., Schneider, T., *et al.* (2015). Predictive timing functions of cortical beta oscillations are impaired in parkinson's disease and influenced by l-dopa and deep brain stimulation of the subthalamic nucleus. *NeuroImage: Clinical*, **9**, 436–449.

Guo, S., Seth, A. K., Kendrick, K. M., Zhou, C., and Feng, J. (2008). Partial granger causalityeliminating exogenous inputs and latent variables. *Journal of neuroscience methods*, **172**(1), 79–93.

Guyon, I., Statnikov, A., and Batu, B. B. (2019). *Cause Effect Pairs in Machine Learning*. Springer.

Hall, S. D., Stanford, I. M., Yamawaki, N., McAllister, C., Rönnqvist, K., Woodhall, G. L., and Furlong, P. L. (2011). The role of gabaergic modulation in motor function related neuronal network activity. *Neuroimage*, **56**(3), 1506–1510.

Hampson, M. and Hoffman, R. E. (2010). Transcranial magnetic stimulation and connectivity mapping: tools for studying the neural bases of brain disorders. *Frontiers in systems neuroscience*, **4**, 40.

Hashemirad, F., Zoghi, M., Fitzgerald, P. B., and Jaberzadeh, S. (2016). The effect of anodal transcranial direct current stimulation on motor sequence learning in healthy individuals: a systematic review and meta-analysis. *Brain and cognition*, **102**, 1–12.

Hatsopoulos, N. G. and Suminski, A. J. (2011). Sensing with the motor cortex. *Neuron*, **72**(3), 477–487.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., and Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage*, **87**, 96–110.

Heckerman, D., Geiger, D., and Chickering, D. M. (1995). Learning bayesian networks: The combination of knowledge and statistical data. *Machine learning*, **20**(3), 197–243.

Helfrich, R. F., Herrmann, C. S., Engel, A. K., and Schneider, T. R. (2016). Different coupling modes mediate cortical cross-frequency interactions. *Neuroimage*, **140**, 76–82.

Henckel, L., Perković, E., and Maathuis, M. H. (2019). Graphical criteria for efficient total effect estimation via adjustment in causal linear models. *arXiv preprint arXiv:1907.02435*.

Herrmann, C. S., Rach, S., Neuling, T., and Strüber, D. (2013). Transcranial alternating current stimulation: a review of the underlying mechanisms and modulation of cognitive processes. *Frontiers in human neuroscience*, **7**, 279.

Hoyer, P. O., Janzing, D., Mooij, J. M., Peters, J., and Schölkopf, B. (2009). Nonlinear causal discovery with additive noise models. In *Advances in neural information processing systems*, pages 689–696.

Huang, B., Zhang, K., Zhang, J., Sanchez-Romero, R., Glymour, C., and Schölkopf, B. (2017). Behind distribution shift: Mining driving forces of changes and causal arrows. In *2017 IEEE International Conference on Data Mining (ICDM)*, pages 913–918. IEEE.

Hummel, F., Celnik, P., Giraux, P., Floel, A., Wu, W.-H., Gerloff, C., and Cohen, L. G. (2005). Effects of non-invasive cortical stimulation on skilled motor function in chronic stroke. *Brain*, **128**(3), 490–499.

Hung, Y.-C., Tseng, N.-F., Balakrishnan, N., *et al.* (2014). Trimmed granger causality between two groups of time series. *Electronic Journal of Statistics*, **8**(2), 1940–1972.

Hyvärinen, A. and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural networks*, **13**(4-5), 411–430.

Hyvärinen, A. and Pajunen, P. (1999). Nonlinear independent component analysis: Existence and uniqueness results. *Neural Networks*, **12**(3), 429–439.

Hyvärinen, A. and Smith, S. M. (2013). Pairwise likelihood ratios for estimation of non-gaussian structural equation models. *Journal of Machine Learning Research*, **14**(Jan), 111–152.

Iezzi, E., Conte, A., Suppa, A., Agostino, R., Dinapoli, L., Scontrini, A., and Berardelli, A. (2008). Phasic voluntary movements reverse the aftereffects of subsequent theta-burst stimulation in humans. *Journal of neurophysiology*, **100**(4), 2070–2076.

Jayaram, V., Alamgir, M., Altun, Y., Scholkopf, B., and Grosse-Wentrup, M. (2016). Transfer learning in brain-computer interfaces. *IEEE Computational Intelligence Magazine*, **11**(1), 20–31.

Jensen, O., Goel, P., Kopell, N., Pohja, M., Hari, R., and Ermentrout, B. (2005). On the human sensorimotor-cortex beta rhythm: sources and modeling. *Neuroimage*, **26**(2), 347–355.

Johnson, L., Alekseichuk, I., Krieg, J., Doyle, A., Yu, Y., Vitek, J., Johnson, M., and Opitz, A. (2019). Dose-dependent effects of transcranial alternating current stimulation on spike timing in awake nonhuman primates. *BioRxiv*, page 696344.

Joundi, R. A., Jenkinson, N., Brittain, J.-S., Aziz, T. Z., and Brown, P. (2012). Driving oscillatory activity in the human cortex enhances motor performance. *Current Biology*, **22**(5), 403–407.

Kadosh, R. C., Levy, N., O'Shea, J., Shea, N., and Savulescu, J. (2012). The neuroethics of non-invasive brain stimulation. *Current Biology*, **22**(4), R108–R111.

Kalaska, J. F., Cohen, D., Hyde, M. L., and Prud'Homme, M. (1989). A comparison of movement direction-related versus load direction-related activity in primate motor cortex, using a two-dimensional reaching task. *Journal of Neuroscience*, **9**(6), 2080–2102.

Kandel, E. R., Schwartz, J. H., Jessell, T. M., of Biochemistry, D., Jessell, M. B. T., Siegelbaum, S., and Hudspeth, A. (2000). *Principles of neural science*, volume 4. McGraw-hill New York.

Kasten, F. H., Duecker, K., Maack, M. C., Meiser, A., and Herrmann, C. S. (2019). Integrating electric field modeling and neuroimaging to explain inter-individual variability of tacs effects. *Nature communications*, **10**(1), 1–11.

Keeser, D., Meindl, T., Bor, J., Palm, U., Pogarell, O., Mulert, C., Brunelin, J., Möller, H.-J., Reiser, M., and Padberg, F. (2011). Prefrontal transcranial direct current stimulation changes connectivity of resting-state networks during fmri. *Journal of Neuroscience*, **31**(43), 15284–15293.

Khanna, P. and Carmena, J. M. (2017). Beta band oscillations in motor cortex reflect neural population signals that delay movement onset. *Elife*, **6**, e24573.

Kiers, L., Cros, D., Chiappa, K., and Fang, J. (1993). Variability of motor potentials evoked by transcranial magnetic stimulation. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, **89**(6), 415–423.

Koller, D. and Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.

Kong, R., Li, J., Orban, C., Sabuncu, M. R., Liu, H., Schaefer, A., Sun, N., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., *et al.* (2019). Spatial topography of individual-specific cortical networks predicts human cognition, personality, and emotion. *Cerebral cortex*, **29**(6), 2533–2551.

Korb, K. B., Hope, L. R., Nicholson, A. E., and Axnick, K. (2004). Varieties of causal intervention. In *Pacific Rim International Conference on Artificial Intelligence*, pages 322–331. Springer.

Krause, M. R., Vieira, P. G., Csorba, B. A., Pilly, P. K., and Pack, C. C. (2019). Transcranial alternating current stimulation entrains single-neuron activity in the primate brain. *Proceedings of the National Academy of Sciences*, **116**(12), 5747–5755.

Kropotov, J. D. (2010a). *Quantitative EEG, event-related potentials and neurotherapy*, pages 77–95. Academic Press.

Kropotov, J. D. (2010b). *Quantitative EEG, event-related potentials and neurotherapy*, pages 59–76. Academic Press.

Kropotov, J. D. (2016a). *Functional neuromarkers for psychiatry: Applications for diagnosis and treatment*, pages 89–105. Academic Press.

Kropotov, J. D. (2016b). *Functional neuromarkers for psychiatry: Applications for diagnosis and treatment*, pages 291–321. Academic Press.

Kryger, M. H., Roth, T., Dement, W. C., *et al.* (2017). *Principles and practice of sleep medicine*, pages 335–347. Elsevier.

Kwakkel, G., Kollen, B. J., van der Grond, J., and Prevo, A. J. (2003). Probability of regaining dexterity in the flaccid upper limb: impact of severity of paresis and time since onset in acute stroke. *Stroke*, **34**(9), 2181–2186.

Lafon, B., Henin, S., Huang, Y., Friedman, D., Melloni, L., Thesen, T., Doyle, W., Buzsáki, G., Devinsky, O., Parra, L. C., *et al.* (2017). Low frequency transcranial electrical stimulation does not entrain sleep rhythms measured by human intracranial recordings. *Nature communications*, **8**(1), 1–14.

Lauritzen, S. L. (1996). *Graphical models*, volume 17. Clarendon Press.

Lillegraven, J. A., Thompson, S. D., Mcnab, B. K., and Patton, J. L. (1987). The origin of eutherian mammals. *Biological Journal of the Linnean Society*, **32**(3), 281–336.

López-Alonso, V., Cheeran, B., Río-Rodríguez, D., and Fernández-del Olmo, M. (2014). Inter-individual variability in response to non-invasive brain stimulation paradigms. *Brain stimulation*, **7**(3), 372–380.

Lum, P. S., Burgar, C. G., Shor, P. C., Majmundar, M., and Van der Loos, M. (2002). Robot-assisted movement training compared with conventional therapy techniques for the rehabilitation of upper-limb motor function after stroke. *Archives of physical medicine and rehabilitation*, **83**(7), 952–959.

Lustenberger, C., Boyle, M. R., Foulser, A. A., Mellin, J. M., and Fröhlich, F. (2015). Functional role of frontal alpha oscillations in creativity. *Cortex*, **67**, 74–82.

Malik, A. S. and Amin, H. U. (2017). *Designing EEG experiments for studying the brain: Design code and example datasets*. Academic Press.

Malinsky, D. and Spirtes, P. (2018a). Causal structure learning from multivariate time series in settings with unmeasured confounding. In *Proceedings of 2018 ACM SIGKDD Workshop on Causal Discovery*, pages 23–47.

Malinsky, D. and Spirtes, P. (2018b). Causal structure learning from multivariate time series in settings with unmeasured confounding. In *Proceedings of 2018 ACM SIGKDD Workshop on Causal Disocvery*, volume 92 of *Proceedings of Machine Learning Research*, pages 23–47.

Mastakouri, A. and Schölkopf, B. (2020). Causal analysis of covid-19 spread in germany. *Advances in Neural Information Processing Systems*, **33**.

Mastakouri, A. A. (2020). Stratification of behavioral response to transcranial current stimulation by resting-state electrophysiology. *bioRxiv*.

Mastakouri, A. A., Weichwald, S., Özdenizci, O., Meyer, T., Schölkopf, B., and Grosse-Wentrup, M. (2017). Personalized brain-computer interface models for motor rehabilitation. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3024–3029. IEEE.

Mastakouri, A. A., Schölkopf, B., and Grosse-Wentrup, M. (2019a). Beta power may meditate the effect of gamma-tacs on motor performance. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5902–5908. IEEE.

Mastakouri, A. A., Schölkopf, B., and Janzing, D. (2019b). Selecting causal brain features with a single conditional independence test per feature. In *Advances in Neural Information Processing Systems*, pages 12532–12543.

Mastakouri, A. A., Schölkopf, B., and Janzing, D. (2020). Necessary and sufficient conditions for causal feature selection in time series with latent common causes. *arXiv preprint arXiv:2005.08543*.

Matsumoto, H. and Ugawa, Y. (2017). Adverse events of tdcs and tacs: a review. *Clinical neurophysiology practice*, **2**, 19–25.

Matyas, F., Sreenivasan, V., Marbach, F., Wacongne, C., Barsy, B., Mateo, C., Aronoff, R., and Petersen, C. C. (2010). Motor control by sensory cortex. *Science*, **330**(6008), 1240–1243.

McAllister, C. J., Rönnqvist, K. C., Stanford, I. M., Woodhall, G. L., Furlong, P. L., and Hall, S. D. (2013). Oscillatory beta activity mediates neuroplastic effects of motor cortex stimulation in humans. *Journal of Neuroscience*, **33**(18), 7919–7927.

McMenamin, B. W., Shackman, A. J., Maxwell, J. S., Bachhuber, D. R., Koppenhaver, A. M., Greischar, L. L., and Davidson, R. J. (2010). Validation of ica-based myogenic artifact correction for scalp and source-localized eeg. *Neuroimage*, **49**(3), 2416–2432.

Meinel, A., Castaño-Candamil, S., Reis, J., and Tangermann, M. (2016). Pre-trial eeg-based single-trial motor performance prediction to enhance neuroergonomics for a hand force task. *Frontiers in human neuroscience*, **10**, 170.

Meyer, T., Peters, J., Zander, T. O., Schölkopf, B., and Grosse-Wentrup, M. (2014). Predicting motor learning performance from electroencephalographic data. *Journal of neuroengineering and rehabilitation*, **11**(1), 24.

Miller, R. (2007). Theory of the normal waking eeg: from single neurones to waveforms in the alpha, beta and gamma frequency ranges. *International journal of psychophysiology*, **64**(1), 18–23.

Moisa, M., Polania, R., Grueschow, M., and Ruff, C. C. (2016). Brain network mechanisms underlying motor enhancement by transcranial entrainment of gamma oscillations. *Journal of Neuroscience*, **36**(47), 12053–12065.

Moliadze, V., Antal, A., and Paulus, W. (2010). Boosting brain excitability by transcranial high frequency stimulation in the ripple range. *The Journal of physiology*, **588**(24), 4891–4904.

Moliadze, V., Schmanke, T., Andreas, S., Lyzhko, E., Freitag, C. M., and Siniatchkin, M. (2015). Stimulation intensities of transcranial direct current stimulation have to be adjusted in children and adolescents. *Clinical Neurophysiology*, **126**(7), 1392–1399.

Mooij, J. M., Peters, J., Janzing, D., Zscheischler, J., and Schölkopf, B. (2016). Distinguishing cause from effect using observational data: methods and benchmarks. *The Journal of Machine Learning Research*, **17**(1), 1103–1204.

Moran, D. W. and Schwartz, A. B. (1999). Motor cortical representation of speed and direction during reaching. *Journal of neurophysiology*, **82**(5), 2676–2692.

Mori, F., Ribolsi, M., Kusayanagi, H., Siracusano, A., Mantovani, V., Marasco, E., Bernardi, G., and Centonze, D. (2011). Genetic variants of the nmda receptor influence cortical excitability and plasticity in humans. *Journal of neurophysiology*, **106**(4), 1637–1643.

Muthukumaraswamy, S. D. (2010). Functional properties of human primary motor cortex gamma oscillations. *Journal of neurophysiology*, **104**(5), 2873–2885.

Muthukumaraswamy, S. D., Edden, R. A., Jones, D. K., Swettenham, J. B., and Singh, K. D. (2009). Resting gaba concentration predicts peak gamma frequency and fmri amplitude in response to visual stimulation in humans. *Proceedings of the National Academy of Sciences*, **106**(20), 8356–8361.

Nakayama, H., Jørgensen, H. S., Raaschou, H. O., and Olsen, T. S. (1994). Recovery of upper extremity function in stroke patients: the copenhagen stroke study. *Archives of physical medicine and rehabilitation*, **75**(4), 394–398.

Neuling, T., Rach, S., Wagner, S., Wolters, C. H., and Herrmann, C. S. (2012). Good vibrations: oscillatory phase shapes perception. *Neuroimage*, **63**(2), 771–778.

Nitsche, M. A. and Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *The Journal of physiology*, **527**(3), 633–639.

Nitsche, M. A., Jaussi, W., Liebetanz, D., Lang, N., Tergau, F., and Paulus, W. (2004). Consolidation of human motor cortical neuroplasticity by d-cycloserine. *Neuropsychopharmacology*, **29**(8), 1573–1578.

Nowak, M., Hinson, E., van Ede, F., Pogosyan, A., Guerra, A., Quinn, A., Brown, P., and Stagg, C. J. (2017). Driving human motor cortical oscillations leads to behaviorally relevant changes in local gabaa inhibition: a tacs-tms study. *Journal of Neuroscience*, **37**(17), 4481–4492.

Nowak, M., Zich, C., and Stagg, C. J. (2018). Motor cortical gamma oscillations: What have we learnt and where are we headed? *Current behavioral neuroscience reports*, **5**(2), 136–142.

Obeso, I., Oliviero, A., and Jahanshahi, M. (2016). Non-invasive brain stimulation in neurology and psychiatry. *Frontiers in neuroscience*, **10**, 574.

Ogarrio, J. M., Spirtes, P., and Ramsey, J. (2016). A hybrid causal search algorithm for latent variable models. In *Conference on Probabilistic Graphical Models*, pages 368–379.

Okada, Y. C., Wu, J., and Kyuhou, S. (1997). Genesis of meg signals in a mammalian cns structure. *Electroencephalography and clinical neurophysiology*, **103**(4), 474–485.

Oostenveld, R. and Praamstra, P. (2001). The five percent electrode system for high-resolution eeg and erp measurements. *Clinical neurophysiology*, **112**(4), 713–719.

Organization, W. H. (2002). *The world health report 2002: reducing risks, promoting healthy life*. World Health Organization.

Palm, U., Ayache, S. S., Padberg, F., and Lefaucheur, J.-P. (2014). Non-invasive brain stimulation therapy in multiple sclerosis: a review of tdcs, rtms and ect results. *Brain stimulation*, **7**(6), 849–854.

Paninski, L., Fellows, M. R., Hatsopoulos, N. G., and Donoghue, J. P. (2004). Spatiotemporal tuning of motor cortical neurons for hand position and velocity. *Journal of neurophysiology*, **91**(1), 515–532.

Parazzini, M., Fiocchi, S., Liorni, I., and Ravazzani, P. (2015). Effect of the interindividual variability on computational modeling of transcranial direct current stimulation. *Computational intelligence and neuroscience*, **2015**.

Pearl, J. (2009). *Causality*. Cambridge university press.

Pearl, J. (2013). On the testability of causal models with latent and instrumental variables. *arXiv preprint arXiv:1302.4976*.

Pearl, J. (2014). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier.

Pearl, J., Verma, T., *et al.* (1991). A theory of inferred causation. *KR*, **91**, 441–452.

Peña, J. M., Björkegren, J., and Tegnér, J. (2005). Scalable, efficient and correct learning of markov boundaries under the faithfulness assumption. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 136–147. Springer.

Penfield, W. and Rasmussen, T. (1950). The cerebral cortex of man; a clinical study of localization of function.

Peters, J., Janzing, D., and Schölkopf, B. (2017). *Elements of causal inference: foundations and learning algorithms*. MIT press.

Pfister, N., Bühlmann, P., and Peters, J. (2019). Invariant causal prediction for sequential data. *Journal of the American Statistical Association*, **114**(527), 1264–1276.

Pitcher, J. B., Ogston, K. M., and Miles, T. S. (2003). Age and sex differences in human motor cortex input–output characteristics. *The Journal of physiology*, **546**(2), 605–613.

Pogosyan, A., Gaynor, L. D., Eusebio, A., and Brown, P. (2009). Boosting cortical activity at beta-band frequencies slows movement in humans. *Current biology*, **19**(19), 1637–1641.

Polanía, R., Nitsche, M. A., Korman, C., Batsikadze, G., and Paulus, W. (2012a). The importance of timing in segregated theta phase-coupling for cognitive performance. *Current Biology*, **22**(14), 1314–1318.

Polanía, R., Paulus, W., and Nitsche, M. A. (2012b). Modulating cortico-striatal and thalamo-cortical functional connectivity with transcranial direct current stimulation. *Human brain mapping*, **33**(10), 2499–2508.

Polania, R., Nitsche, M. A., and Ruff, C. C. (2018). Studying and modifying brain function with non-invasive brain stimulation. *Nature neuroscience*, **21**(2), 174–187.

Raffin, E. and Siebner, H. R. (2014). Transcranial brain stimulation to promote functional recovery after stroke. *Current opinion in neurology*, **27**(1), 54.

Raj, A., Bauer, S., Soleymani, A., Besserve, M., and Schölkopf, B. (2020). Causal feature selection via orthogonal search. *arXiv preprint arXiv:2007.02938*.

Ramos-Murguialday, A., Broetz, D., Rea, M., Läer, L., Yilmaz, Ö., Brasil, F. L., Liberati, G., Curado, M. R., Garcia-Cossio, E., Vyziotis, A., *et al.* (2013). Brain–machine interface in chronic stroke rehabilitation: a controlled study. *Annals of neurology*, **74**(1), 100–108.

Ramsey, J., Zhang, J., and Spirtes, P. L. (2012). Adjacency-faithfulness and conservative causal inference. *arXiv preprint arXiv:1206.6843*.

Ridding, M. and Ziemann, U. (2010). Determinants of the induction of cortical plasticity by non-invasive brain stimulation in healthy subjects. *The Journal of physiology*, **588**(13), 2291–2304.

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive brain research*, **3**(2), 131–141.

Rosenkranz, K., Kacar, A., and Rothwell, J. C. (2007). Differential modulation of motor cortical plasticity and excitability in early and late phases of human motor learning. *Journal of Neuroscience*, **27**(44), 12058–12066.

Rubenstein, P. K., Weichwald, S., Bongers, S., Mooij, J. M., Janzing, D., Grosse-Wentrup, M., and Schölkopf, B. (2017). Causal consistency of structural equation models. *arXiv preprint arXiv:1707.00819*.

Runge, J. (2018). Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **28**(7), 075310.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., and Sejdinovic, D. (2019a). Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, **5**(11), eaau4996.

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Muñoz-Marí, J., *et al.* (2019b). Inferring causation from time series in earth system sciences. *Nature communications*, **10**(1), 1–13.

Santarnecchi, E., Polizzotto, N. R., Godone, M., Giovannelli, F., Feurra, M., Matzen, L., Rossi, A., and Rossi, S. (2013). Frequency-dependent enhancement of fluid intelligence induced by transcranial oscillatory potentials. *Current Biology*, **23**(15), 1449–1453.

Santosa, F. and Symes, W. W. (1986). Linear inversion of band-limited reflection seismograms. *SIAM Journal on Scientific and Statistical Computing*, **7**(4), 1307–1330.

Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N., and Wolpaw, J. R. (2004). Bci2000: a general-purpose brain-computer interface (bci) system. *IEEE Transactions on biomedical engineering*, **51**(6), 1034–1043.

Schnitzler, A. and Gross, J. (2005). Normal and pathological oscillatory communication in the brain. *Nature reviews neuroscience*, **6**(4), 285–296.

Schulz, R., Gerloff, C., and Hummel, F. C. (2013). Non-invasive brain stimulation in neurological diseases. *Neuropharmacology*, **64**, 579–587.

Scott, D. W. (2015). *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons.

Sehm, B., Schäfer, A., Kipping, J., Margulies, D., Conde, V., Taubert, M., Villringer, A., and Ragert, P. (2012). Dynamic modulation of intrinsic functional connectivity by transcranial direct current stimulation. *Journal of neurophysiology*, **108**(12), 3253–3263.

Sela, T., Kilim, A., and Lavidor, M. (2012). Transcranial alternating current stimulation increases risk-taking behavior in the balloon analog risk task. *Frontiers in neuroscience*, **6**, 22.

Shafi, M. M., Westover, M. B., Fox, M. D., and Pascual-Leone, A. (2012). Exploration and modulation of brain network interactions with noninvasive brain stimulation in combination with neuroimaging. *European Journal of Neuroscience*, **35**(6), 805–825.

Shah, R. D. and Peters, J. (2018). The hardness of conditional independence testing and the generalised covariance measure. *arXiv preprint arXiv:1804.07203*.

Sharma, N., Baron, J.-C., and Rowe, J. B. (2009). Motor imagery after stroke: relating outcome to motor network connectivity. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, **66**(5), 604–616.

Shimizu, S., Hoyer, P. O., Hyvärinen, A., and Kerminen, A. (2006). A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, **7**(Oct), 2003–2030.

Shimizu, S., Inazumi, T., Sogawa, Y., Hyvärinen, A., Kawahara, Y., Washio, T., Hoyer, P. O., and Bollen, K. (2011). Directlingam: A direct method for learning a linear non-gaussian structural equation model. *Journal of Machine Learning Research*, **12**(Apr), 1225–1248.

Silvanto, J., Cattaneo, Z., Battelli, L., and Pascual-Leone, A. (2008). Baseline cortical excitability determines whether tms disrupts or facilitates behavior. *Journal of neurophysiology*, **99**(5), 2725–2730.

Soliman, I. and Mashhour, A. (2011). Dairy marketing system performance in egypt.

Song, L., Bedo, J., Borgwardt, K. M., Gretton, A., and Smola, A. (2007a). Gene selection via the bahsic family of algorithms. *Bioinformatics*, **23**(13), i490–i498.

Song, L., Smola, A., Gretton, A., Borgwardt, K. M., and Bedo, J. (2007b). Supervised feature selection via dependence estimation. In *Proceedings of the 24th international conference on Machine learning*, pages 823–830.

Spirtes, P., Glymour, C. N., Scheines, R., and Heckerman, D. (2000). *Causation, prediction, and search*. MIT press.

Sreeraj, V. S., Dinakaran, D., Parlikar, R., Chhabra, H., Selvaraj, S., Shivakumar, V., Bose, A., Narayanaswamy, J. C., and Venkatasubramanian, G. (2018). High-definition transcranial direct current simulation (hd-tdcs) for persistent auditory hallucinations in schizophrenia. *Asian journal of psychiatry*, **37**, 46–50.

Stagg, C. J., Bachtiar, V., and Johansen-Berg, H. (2011). The role of gaba in human motor learning. *Current Biology*, **21**(6), 480–484.

Stark, E., Drori, R., Asher, I., Ben-Shaul, Y., and Abeles, M. (2007). Distinct movement parameters are represented by different neurons in the motor cortex. *European Journal of Neuroscience*, **26**(4), 1055–1066.

Stecher, H. I., Pollok, T. M., Strüber, D., Sobotka, F., and Herrmann, C. S. (2017). Ten minutes of α-tacs and ambient illumination independently modulate eeg α-power. *Frontiers in human neuroscience*, **11**, 257.

Steriade, M. (2005). Brain electrical activity and sensory processing during waking and sleep states. In *Principles and practice of sleep medicine*, pages 101–119. Elsevier.

Strobl, E. V. and Visweswaran, S. (2019). Markov blanket ranking using kernel-based conditional dependence measures. In *Cause Effect Pairs in Machine Learning*, pages 359–372. Springer.

Strube, W., Bunse, T., Malchow, B., and Hasan, A. (2015). Efficacy and interindividual variability in motor-cortex plasticity following anodal tdcs and paired-associative stimulation. *Neural plasticity*, **2015**.

Strüber, D., Rach, S., Neuling, T., and Herrmann, C. S. (2015). On the possible role of stimulation duration for after-effects of transcranial alternating current stimulation. *Frontiers in cellular neuroscience*, **9**, 311.

Swaiman, K. F., Ashwal, S., Ferriero, D. M., Schor, N. F., Finkel, R. S., Gropman, A. L., Pearl, P. L., and Shevell, M. (2017). *Swaiman's Pediatric Neurology E-Book: Principles and Practice*, pages 87–96. Elsevier Health Sciences.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, **58**(1), 267–288.

Tortella, G., ML Selingardi, P., L Moreno, M., P Veronezi, B., and R Brunoni, A. (2014). Does non-invasive brain stimulation improve cognition in major depressive disorder? a systematic review. *CNS & Neurological Disorders-Drug Targets (Formerly Current Drug Targets-CNS & Neurological Disorders)*, **13**(10), 1759–1769.

Triccas, L. T., Burridge, J., Hughes, A., Pickering, R., Desikan, M., Rothwell, J., and Verheyden, G. (2016). Multiple sessions of transcranial direct current stimulation and upper extremity rehabilitation in stroke: a review and meta-analysis. *Clinical Neurophysiology*, **127**(1), 946–955.

Tsamardinos, I. and Aliferis, C. F. (2003). Towards principled feature selection: relevancy, filters and wrappers. In *AISTATS*.

Tsamardinos, I., Brown, L. E., and Aliferis, C. F. (2006). The max-min hill-climbing bayesian network structure learning algorithm. *Machine learning*, **65**(1), 31–78.

Tu, R., Zhang, C., Ackermann, P., Mohan, K., Glymour, C., Kjellström, H., and Zhang, K. (2018). Causal discovery in the presence of missing data. *arXiv preprint arXiv:1807.04010*.

Turlach, B. A. (1993). Bandwidth selection in kernel density estimation: A review. In *CORE and Institut de Statistique*. Citeseer.

Uhler, C., Raskutti, G., Bühlmann, P., Yu, B., *et al.* (2013). Geometry of the faithfulness assumption in causal inference. *The Annals of Statistics*, **41**(2), 436–463.

Vannorsdall, T. D., Van Steenburgh, J. J., Schretlen, D. J., Jayatillake, R., Skolasky, R. L., and Gordon, B. (2016). Reproducibility of tdcs results in a randomized trial: failure to replicate findings of tdcs-induced enhancement of verbal fluency. *Cognitive and Behavioral Neurology*, **29**(1), 11–17.

Veniero, D., Strüber, D., Thut, G., and Herrmann, C. S. (2019). Noninvasive brain stimulation techniques can modulate cognitive processing. *Organizational Research Methods*, **22**(1), 116–147.

Vosskuhl, J., Huster, R. J., and Herrmann, C. S. (2015). Increase in short-term memory capacity induced by down-regulating individual theta frequency via transcranial alternating current stimulation. *Frontiers in human neuroscience*, **9**, 257.

Vosskuhl, J., Strüber, D., and Herrmann, C. S. (2018). Non-invasive brain stimulation: a paradigm shift in understanding brain oscillations. *Frontiers in human neuroscience*, **12**, 211.

Voti, P. L., Conte, A., Suppa, A., Iezzi, E., Bologna, M., Aniello, M., Defazio, G., Rothwell, J., and Berardelli, A. (2011). Correlation between cortical plasticity, motor learning and bdnf genotype in healthy subjects. *Experimental brain research*, **212**(1), 91–99.

Wach, C., Krause, V., Moliadze, V., Paulus, W., Schnitzler, A., and Pollok, B. (2013). Effects of 10 hz and 20 hz transcranial alternating current stimulation (tacs) on motor functions and motor cortical excitability. *Behavioural brain research*, **241**, 1–6.

Wade, E., Parnandi, A. R., and Matarić, M. J. (2011). Using socially assistive robotics to augment motor task performance in individuals post-stroke. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2403–2408. IEEE.

Wagner, T., Valero-Cabre, A., and Pascual-Leone, A. (2007). Noninvasive human brain stimulation. *Annu. Rev. Biomed. Eng.*, **9**, 527–565.

Walker-Batson, D., Smith, P., Curtis, S., Unwin, H., and Greenlee, R. (1995). Amphetamine paired with physical therapy accelerates motor recovery after stroke: further evidence. *Stroke*, **26**(12), 2254–2259.

Wang, X.-J. and Buzsáki, G. (1996). Gamma oscillation by synaptic inhibition in a hippocampal interneuronal network model. *Journal of neuroscience*, **16**(20), 6402–6413.

Weichwald, S., Meyer, T., Özdenizci, O., Schölkopf, B., Ball, T., and Grosse-Wentrup, M. (2015). Causal interpretation rules for encoding and decoding models in neuroimaging. *NeuroImage*, **110**, 48–59.

Wiener, N. (1956). The theory of prediction. *Modern mathematics for engineers*.

Wiethoff, S., Hamada, M., and Rothwell, J. C. (2014). Variability in response to transcranial direct current stimulation of the motor cortex. *Brain stimulation*, **7**(3), 468–475.

Yamada, M., Jitkrittum, W., Sigal, L., Xing, E. P., and Sugiyama, M. (2014). High-dimensional feature selection by feature-wise kernelized lasso. *Neural computation*, **26**(1), 185–207.

Yamawaki, N., Stanford, I., Hall, S., and Woodhall, G. (2008). Pharmacologically induced and stimulus evoked rhythmic neuronal oscillatory activity in the primary motor cortex in vitro. *Neuroscience*, **151**(2), 386–395.

Yang, L.-Z., Zhang, W., Wang, W., Yang, Z., Wang, H., Deng, Z.-D., Li, C., Qiu, B., Zhang, D.-R., Kadosh, R. C., Li, H., and Zhang, X. (2020). Neural and psychological predictors of cognitive enhancement and impairment from neurostimulation. *Advanced Science*, **7**(4), 1902863.

Zhang, K. and Chan, L.-W. (2006). Extensions of ica for causality discovery in the hong kong stock market. In *International Conference on Neural Information Processing*, pages 400–409. Springer.

Zhang, K. and Hyvarinen, A. (2012). On the identifiability of the post-nonlinear causal model. *arXiv preprint arXiv:1205.2599*.

Zhang, K., Peters, J., Janzing, D., and Schölkopf, B. (2012). Kernel-based conditional independence test and application in causal discovery. *arXiv preprint arXiv:1202.3775*.

Zhang, K., Zhang, J., Huang, B., Schölkopf, B., and Glymour, C. (2016). On the identifiability and estimation of functional causal models in the presence of outcome-dependent selection. In *UAI*.

Zhang, K., Huang, B., Zhang, J., Glymour, C., and Schölkopf, B. (2017). Causal discovery from nonstationary/heterogeneous data: Skeleton estimation and orientation determination. In *IJCAI: Proceedings of the Conference*, volume 2017, page 1347. NIH Public Access.

Ziemann, U., Paulus, W., Nitsche, M. A., Pascual-Leone, A., Byblow, W. D., Berardelli, A., Siebner, H. R., Classen, J., Cohen, L. G., and Rothwell, J. C. (2008). Consensus: motor cortex plasticity protocols. *Brain stimulation*, **1**(3), 164–182.