Signal detection and causal inference in functional Magnetic Resonance Imaging

Natalia Z. Bielczyk
Signal detection and causal inference in functional Magnetic Resonance Imaging

Natalia Z. Bielczyk
The work leading to this thesis was supported by the following sources of funding:
(1) NWO-Vidi 864-12-003 grant personally awarded to Christian F Beckmann.
(2) Funding from the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreements no 305697 (OPTIMISTIC), no 278948 (TACTICS) and no 603016 (MATRICS), managed by Jeffrey C Glennon.
© Natalia Z. Bielczyk, 2020. No part of this thesis may be reproduced in any form without prior written permission of the author.
Signal detection and causal inference in functional Magnetic Resonance Imaging

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op donderdag 18 juni 2020
om 16:30 uur precies

door

Natalia Zofia Bielczyk
geboren op 14 januari 1986
te Katowice, Polen
Promotoren:
Prof. dr. J.K. Buitelaar
Prof. dr. C.F. Beckmann

Copromotor:
Dr. J.C. Glennon

Manuscriptcommissie:
Prof. dr. D.G. Norris
Dr. R.B. Mars
Prof. dr. M. Hanke
“We all have two lives.  
The second one starts when we realize that we only have one.”

Confucius
# Contents

1 Introduction  1
   1.1 Aims and Overview of this Thesis  1
   1.2 Methods versus approaches  1
   1.3 Methods for data analysis in fMRI  3
      1.3.1 Motion correction  3
      1.3.2 Spatial normalization  3
      1.3.3 Spatial smoothing  3
      1.3.4 Estimating Hemodynamic Response Function  4
      1.3.5 Region definition  5
      1.3.6 After preprocessing the datasets  6
   1.4 Signal detection  6
   1.5 Detection of connections  6
   1.6 Causal inference in fMRI  8
   1.7 Wider implications for psychiatric research  9

2 Methods for signal detection in fMRI  11
   2.1 Statistical Parametric Mapping (SPM)  11
      2.1.1 General Linear Model (GLM)  11
      2.1.2 Fitting a GLM model  12
      2.1.3 GLM in the context of functional Magnetic Resonance Imaging  13
      2.1.4 Finding effects in second-level analysis  13
      2.1.5 Correcting for multiple comparisons  14
         Family-Wise Error (FWE)  14
         Random Field Theory (RFT)  14
         False Discovery Rate (FDR)  16
      2.1.6 Basic findings from Statistical Parametric Mapping in fMRI  16
      2.1.7 Statistical Parametric Mapping in studies on psychiatric disorders  20
   2.2 Multivoxel Pattern Analysis (MVPA)  21
   2.3 What next?  23

3 Momentum: a robust test for normality  25
   3.1 Momentum, a robust test for normality: Introduction  26
   3.2 Momentum, a robust test for normality: Methods  27
      3.2.1 Defining fractional moments of the distribution  27
      3.2.2 Deriving confidence intervals  29
      3.2.3 Correction for multiple comparisons  33
      3.2.4 Momentum as a signal detection method  34
      3.2.5 The full Momentum procedure  34
      3.2.6 Comparing Momentum with other tests for normality  35
      3.2.7 Synthetic datasets for comparison  37
8 Mapping large scale networks to diagnostic categories  163
  8.1 Mapping large scale networks to diagnostic categories: Introduction . 164
    8.1.1 Major Depressive Disorder 164
    8.1.2 Causality 164
  8.2 Etiology of MDD 165
    8.2.1 Constructs in MDD 165
    8.2.2 Cognitive constructs 167
    8.2.3 Physiological constructs 168
    8.2.4 Neuronal constructs 168
  8.3 Treatment 169
  8.4 Circuit for MDD 170
  8.5 Modeling MDD 171
    8.5.1 Neural mass models and attractor landscapes 171
    8.5.2 Treatments in the context of dynamical systems 174
  8.6 Circuit to Construct Mapping 175
    8.6.1 What is CCM 175
    8.6.2 Execution of CCM 176
    8.6.3 Conditions to perform CCM 178
  8.7 Limitations of the CCM approach 180
    8.7.1 Plasticity and neurodegeneration 180
    8.7.2 Heterogeneity 180
    8.7.3 Application of treatments to the CCM 182
    8.7.4 Temporal dynamics in the resting state 182
    8.7.5 Effective connectivity in EEG/EMG and fMRI research 183
  8.8 Concluding remarks 183

9 General discussion 185
  9.1 Summary 185
  9.2 Conclusions: what information can we extract from fMRI data? 186
    9.2.1 Signal detection 186
    9.2.2 Connection detection 187
    9.2.3 Causal inference in fMRI 188
  9.3 How to use concepts developed in this work in a broader context? 188
    9.3.1 Signal detection 188
    9.3.2 Connection detection 189
    9.3.3 Causal inference in fMRI 190
List of Figures

2.1 The General Linear Model ........................................... 15
2.2 Statistical Parametric Mapping applied to Human Connectome Project datasets ........................................ 18
2.3 Strategies for mapping cognition onto the brain ............... 20
2.4 A comparison of a several classifiers in scikit-learn on synthetic datasets ............................................. 22

3.1 The concept behind Momentum: Momentum curves .......... 27
3.2 The null distributions against the distribution of signal mixed with the noise, for one fractional moment .......... 28
3.3 The histogram of the distances to the analytic solution for the null, and for the noise mixed with signal, for one fractional moment ............................................. 29
3.4 Signals used in the simulation ...................................... 29
3.5 Moments between 0.0 and 5.0 for sinusoidal signals .......... 30
3.6 Moments between 0.0 and 5.0 for binary trains, p(on)=p(off) .......... 30
3.7 Moments between 0.0 and 5.0 for binary trains, p(on)=2*p(off) ......... 31
3.8 Moments between 0.0 and 5.0 for binary trains, p(on)=0.5*p(off) .......... 31
3.9 Moments between 0.0 and 5.0 for rare events ................... 32
3.10 Momentum as a signal detection technique .................... 33
3.11 Normal distribution and other popular non-Gaussian distributions ............................................. 37
3.12 Exemplary samples created from binary signals and sigmoidal signals mixed with white noise ......................... 38
3.13 P-values obtained by applying the standard tests for normality and Momentum ............................................. 38
3.14 Effect sizes obtained by applying of the popular test for normality and Momentum ............................................. 39
3.15 P-values and associated effect sizes for "biological signals" of varying SNRs ............................................. 40
3.16 Second level analysis moment by moment ....................... 43

4.1 Mixture modeling .................................................... 51
4.2 Validation on synthetic benchmark datasets ..................... 57
4.3 Comparison of BIC for all the 8 mixture modeling methods on the synthetic dataset ............................................. 58
4.4 The evaluation on synthetic datasets, for network sizes between 5 and 50 nodes ............................................. 59
4.5 Comparison of BIC for all the 8 mixture modeling methods on the resting state HCP dataset ............................................. 60
4.6 Comparison of BIC for all the 8 mixture modeling methods on the HCP datasets from the n-back Working Memory task ............................................. 61
4.7 The model fit with the lowest BIC: Gauss-Gamma fit based on the inverse of the LW-regularized covariance ........... 62
4.8 The "group sparse connectome" for the conventional methods with canonical thresholds .......................... 62
4.9 The "group sparse connectome" for permutation testing and for the mixture modeling methods ............... 63
4.10 The "group sparse connectome" for proportional thresholding methods ................................................. 64
4.11 The group difference between resting state and n-back WM task for six methods including the mixture modeling .... 65
4.12 Results as in Fig. 4.11, corrected for multiple comparisons ................................................................. 66
4.13 Comparison between results obtained on the subject- and cohort level, for mixture modeling and permutation testing ... 67

5.1 Causal research in fMRI .......................... 81
5.2 The full pipeline for the DCM generative model ................ 87
5.3 Linear Non-Gaussian Acyclic Model (LiNGAM) ............. 93
5.4 Bayesian nets ........................................ 96

6.1 The test network and its two perturbations ................. 116
6.2 Exemplary dynamics for a DAG of 4 nodes ................. 116
6.3 Residuals between signals generated from the original network and from the perturbed networks .......... 117
6.4 Defining $\Delta$ for the two-node noisy system .............. 119
6.5 The relationship between $\Delta$ and GC ...................... 120
6.6 Results of the classification study with the use of the Linear Discriminant Analysis ............................. 121
6.7 Impact of the relative hemodynamic lags computed as the difference between time to peak of the hemodynamic response in the upstream and downstream region on $\Delta$ ............ 123
6.8 Impact of TR on $\Delta$ ................................... 124

7.1 A two node network with one directed connection ........ 136
7.2 Cumulants derived from the DCM generative model ....... 138
7.3 Discriminative power for all cumulants ...................... 139
7.4 Cumulants in polar coordinates ............................ 140
7.5 Success rate for all the individual cumulants, averaged over 28 simulations from the synthetic benchmark datasets ...... 141
7.6 Success rate for all the individual cumulants, for all simulations in benchmark datasets ......................... 143
7.7 Dependence of the grand mean performance on the choice of cumulants ............................................. 145
7.8 Comparison between the classifier based on the fractional cumulants and several other methods, on the benchmark simulation no 2 .................................................. 148
7.9 Overall success rate of Fractional cumulants against other methods ......................................................... 150
7.10 Robustness of the methods against the background scale-free neuronal noise ................................. 151
7.11 Robustness of different methods to varying SNRs ........ 152

8.1 Diagnosis, treatment, and brain dynamics in MDD ........ 166
8.2 Wilson-Cowan model and a "ball" metaphor. 173
8.3 Circuit to Construct Mapping (CCM). 175
8.4 Conditions to perform CCM. 178
8.5 Attention as an example of a construct with multiple neural mechanisms underneath. 181

9.1 Contribution to the standard pipeline for pairwise inference in fMRI presented in this Thesis. 186
## List of Tables

4.1 The mean density of connections per subject, in the resting state and in the n-back WM task, at the canonical thresholds 68
4.2 Index of Overlap between the two scanning sessions 68
4.3 ICC scores in the resting state and in the n-back WM task across subjects 68
4.4 The percentage of non-zero links in Figs 4.11 and 4.12 68

5.1 Summary of all the methods for effective connectivity in fMRI 104
Chapter 1

Introduction

1.1 Aims and Overview of this Thesis

In this Thesis, I present the state of the art in the field of methods for data analysis in functional Magnetic Resonance Imaging (fMRI), and introduce my original contributions to the field.

First, I discuss the topic of signal detection in fMRI. In Chapter 2, I introduce signal detection in task- and in resting state fMRI. In Chapter 3, I introduce my input to how methods for signal detection can be improved with the use of more sensitive tests for normality of the distribution by incorporating fractional moments of the distribution. In Section 3.5, I propose some possible applications of such a normality testing to the fMRI data analysis protocols.

Second, I discuss the methods for thresholding functional connectomes. In Chapter 4, I propose a new method for thresholding connectomes with the use of mixture modeling.

Third, I discuss the topic of causal inference in fMRI studies. In Chapter 5, I review the field of methods for effective connectivity in fMRI. In Chapter 6, I analyze the dos and don’ts in causal inference in fMRI using synthetic datasets generated with the use of the Dynamic Causal Modeling generative model. In Chapter 7, I propose an advance to methods for effective connectivity in fMRI with the use of fractional cumulants of the BOLD fMRI distribution.

Fourth, in Chapter 8, I propose some potential applications of the methods proposed in this Thesis. I use an example of Major Depression (MDD) to explain the complexity and challenges associated with modeling network-derived biomarkers of psychopathology.

Lastly, in Chapter 9, I summarize the novel contributions to the field proposed in this Thesis. I also discuss the intrinsic limitations of data analysis in fMRI and I propose a list of potential applications where methods derived in this Thesis might be reapplied or adapted.

1.2 Methods versus approaches

In cognitive neuroimaging, there is a wide variety of data analytic techniques. Even within the subfield of fMRI research, this is a very broad category. At the end of 2018, the search term:

"new method" AND ("functional Magnetic Resonance Imaging" OR fMRI)

returns 2,129 records at PubMed (www.ncbi.nlm.nih.gov/pubmed), while
the analogous search term

"new approach" AND ("functional Magnetic Resonance Imaging" OR fMRI)
returns 1,151 records.

This thesis is dedicated to methods in signal detection and causal inference in fMRI. In fMRI experiments, only the echo of the underlying biological processes can be observed. Analytic and statistical techniques are applied to render or approximate underlying biological processes.

There are certain conceptual differences between a method, and an approach. An approach is, in general, a way of looking at the data. E.g., let us assume that one is interested in the influence of behavioral factor (e.g., language comprehension) on the time-varying functional structure of the brain networks. He proposes to operationalize this influence by performing a task involving an experimental stimulus and a block design (e.g., blocks when prerecorded excerpts from Aesop’s fables are presented to the participants, followed by control questions related to the content of these fables). Then, the experimental hypothesis is tested. The hypothesis states that the cognitive state influenced by experimental manipulation is associated with a significant change in modularity (Newman, 2006; Bullmore and Sporns, 2009) in cortical networks activated during language comprehension. Many other ways of quantifying global differences between functional connectomes (e.g., small worldness or network centrality, Bullmore and Sporns, 2009) could have been chosen. However, modularity within brain network is of interest here as this study is a part of a larger study in schizophrenia. In schizophrenia, network modularity is altered on many levels of brain organization (Alexander-Bloch et al., 2010; Cinelli et al., 2018) - therefore, looking from that clinical perspective, network modularity is an important biomarker to take into account. For these reasons, in this study, modularity was chosen as a variable representing the global functional architecture in the brain. On the contrary, the following question could be asked: does language comprehension influence the activity of neuronal populations in cortical and subcortical structures? Does language comprehension influence presence or absence of significant pairwise correlations between different neuronal populations? Does language comprehension influence the way that the neuronal populations communicate? Or, in other words, the strength of the causal influences between them? (this question assumes that the activity of a given population is no different than its firing rate: the mean value of the action potentials fired per neuron per second Kandel, 2012; Gerstner et al., 1997). Pairwise correlation between populations A and B is usually computed using partial correlation (Baba, Shibata, and Sibuya, 2004) between firing rates in these populations. The causal influence of population A on population B, is the net gain in firing rate in population B caused by the direct influence from population A. There is certain underlying biological truth to this definition, and there is no room to choose a definition with respect to what neuronal activity or causality mean. In terms of definition, causality can be quite controversial (see: Chapters 1.6 and 5) but in general, it reflects a physical process of A driving activity in B.
1.3 Methods for data analysis in functional Magnetic Resonance Imaging

FMRI datasets are complex from neurophysiological point of view. Therefore, a range of methods in this field were developed with a goal to regress confounding variables from the fMRI datasets before testing any hypotheses (Caballero-Gaudes and Reynolds, 2017).

1.3.1 Motion correction

fMRI datasets are burdened with multiple confounds. One such confound are motion artifacts that occur as a result of a combination of multiple factors: an uncontrolled body movement in the scanner as well as systematic, periodic movements reflecting the breathing pattern (Zaitsev, Maclaren, and Herbst, 2016). To tackle this issue, multiple strategies have been developed (Goto et al., 2016). The simplest strategy to mitigate the effect of motion artifacts is often referred to as scrubbing (Power et al., 2012). This is a volume censoring technique in which volumes that clearly reflect micromovements at this particular time point (e.g., by a rapid shift in the mean volume intensity throughout the brain) are excluded from the dataset. However, in most instances, scrubbing itself is not sufficient to eliminate motion artifacts from the data. Therefore, volumes are additionally aligned (typically via affine spatial registration, Jenkinson and Smith, 2001).

More sophisticated strategies involve denoising with the use of Independent Component Analysis (ICA-AROMA, Pruim et al., 2015), estimating coherent noise components using Principal Component Analysis (CompCor method, Behzadi et al., 2007), or modeling with the use of wavelets (Patel et al., 2014).

1.3.2 Spatial normalization

Furthermore, every brain has its own, individual shape. Therefore, before doing the group analysis, one should perform spatial normalization of the individual data. During this procedure, brain scans are warped to a common template. The warp-field might be parametrized by nonlinear basis functions (Ashburner and Friston, 1999) such as cosines and polynomials. The most popular templates for spatial normalization, are the Montreal Neurological Institute (MNI), and Talairach space (Handbook of Functional MRI Data Analysis). The choice of the registration template obviously highly influences the results of any further analysis (Calhoun et al., 2017).

1.3.3 Spatial smoothing

Even after the motion artifact removal, the signal-to-noise ratio (SNR) is still low in fMRI. One possible way of increasing the SNR during the preprocessing is spatial smoothing (Triantafyllou, Hoge, and Wald, 2006). The purpose of spatial smoothing is to cope with functional anatomical variability not compensated by spatial normalization. In that sense, spatial smoothing increases the statistical power. The idea behind the smoothing is simple. In fMRI, the BOLD fMRI response on the top of single, underlying neuronal events is a convolution with a spatio-temporal point-spread function
that makes the response smooth in space and time (Shmuel et al., 2007). The main component of the spatial noise is Gaussian; it is essentially random, independent among voxels and roughly centered around zero. Then, if we average intensity across several neighboring voxels, the noise should average out to zero, whereas the underlying signal should average to some non-zero value. Therefore, spatial smoothing can decrease the noise while not decreasing the mean signal, and the associated SNR grows. Smoothing is beneficial when:

1. Fine-grained resolution of the data is not essential, and expected clusters of active voxels are bigger than just a few voxels. This can be determined, e.g., by reviewing the available literature on the topic,

2. The results are averaged over a group of subjects in a brain region where functional anatomy and organization are not known in fine detail,

3. There is a need to improve on the SNR, e.g., when the data undergoes Statistical Parametric Mapping (see: Chapter 2).

On the contrary, spatial smoothing should be avoided when:

1. Keeping voxel resolution in the datasets is important for testing the research hypothesis. An example of such a scenario is when we expect significant statistical effects in little structures such as small nuclei of the thalamus,

2. The subject of the study is an ROI developed with high resolution and/or in single subjects,

3. One can be confident that the experimental task will generate large SNRs.

1.3.4 Estimating Hemodynamic Response Function

Lastly, the recording captured in the experiment is effectively not a neuronal activity per se but rather, a convolution between neuronal activity and the local *hemodynamic* response: a rapid increase in neural activity evokes a cascade of physiological changes in the local blood vessels including changes in the volume of cerebral blood per unit of brain tissue (CBV), increase in the rate of cerebral blood flow, and in the concentration of oxyhemoglobin and deoxyhemoglobin (Logothetis et al., 2001; Goense and Logothetis, 2008). Of course, since hemodynamics acts as a low pass filter (Sauvage et al., 2017), it results in a dropout of information from the BOLD fMRI signal (for more details, see: Chapters 5 and 6). The hemodynamic response on the BOLD fMRI recording can be modeled (Buxton et al., 2004; Lindquist et al., 2009). Therefore to some extent, detrimental influence of the slow hemodynamics on the results of the data analysis can be controlled. There is a rapid progress in the domain of methods for estimating the local hemodynamic response functions (HRFs) based on resting state recordings. Deconvolution methods are a subfield of methods for data analysis in fMRI.
Firstly, cubature Kalman filtering was used to estimate hemodynamic responses from the data (Havlicek et al., 2011). Another method, by Sreenivasan, Havlicek, and Deshpande, 2015, uses homomorphic filtering to obtain underlying neuronal responses from the fMRI signal. One recently developed method for blind deconvolution for resting state fMRI by Wu et al., 2013, treats resting state fMRI as a spontaneous event-related time series. In this framework, BOLD fMRI fluctuations are driven by discrete events that can be detected at the local peaks of the BOLD fMRI response. Subsequently, local HRF responses are fitted to best predict the BOLD fMRI dynamics as a convolution between the series of point events and the HRF response.

However, all these methods assume the shape of the hemodynamic response is constant over time - which is, in general, not true (Handwerker, Ollinger, and D’Esposito, 2004). Therefore, even if modeled and regressed out from the data, hemodynamics will remain a confounding factor for any methods for data analysis in fMRI.

### 1.3.5 Region definition

In research studies where voxel resolution of the data is not necessary to test the experimental hypothesis, Regions Of Interest (ROIs) are often determined before testing research hypotheses. The choice of ROIs in important and highly influences the results of connectomic studies (Marrelec and Fransson, 2011).

Usually, region definition (or, brain parcellation) is achieved with the use of a brain atlas. There is a broad selection of methods available for brain parcellation. Multiple methods use biology as the starting point to section the brain: anatomy (Wake Forest University pickatlas, WFU, Maldjian et al., 2003 or Automatic Anatomical Labeling, AAL, Tzourio-Mazoyer et al., 2002), histology (Glasser and Essen, 2011; Eickhoff et al., 2005; Amunts et al., 2013), or white-matter fibre structure (Mars et al., 2011; Mars et al., 2012; Sallet et al., 2013; Neubert et al., 2014; Neubert et al., 2015). However, such a biology-informed parcellation can bring confounds to the BOLD fMRI signal: while averaging the time series over ROIs, we might mix signals involved in various neuronal processes with each other. From the perspective of network analysis, mixing signals is equivalent to inducing a pink noise in the underlying neuronal dynamics. Therefore, one should pursue efforts aiming at best functional segregation into ROIs instead. The optimal brain parcellation is a widely discussed topic in the field of fMRI (Stanley et al., 2013). There are a few classes of available functional parcellations: Ward algorithm (Ward, 1963; Thirion et al., 2014), Bayesian methods (Janssen et al., 2015; Janssen, Jylänki, and Gerven, 2016), mixture modeling (Golland, Goldstein, and Malach, 2007; Tucholka et al., 2008; Lashkari et al., 2010; Lashkari et al., 2012), k-means algorithm (Flandin et al., 2002; Yeo et al., 2011; Kahnt et al., 2012), hierarchical clustering (Eickhoff et al., 2011; Michel et al., 2012; Orban et al., 2014; Heuvel, Mandl, and Pol, 2008; Bellec et al., 2006; Bellec et al., 2010; Blumensath et al., 2013), spectral clustering (Thirion et al., 2006; Chen et al., 2012; Craddock et al., 2012), the new, semi-automated classification technique by Glasser et al., 2016, and hierarchical ICA by Oort et al., 2017.
Chapter 1. Introduction

1.3.6 After preprocessing the datasets...

After denoising the fMRI datasets from the confounding variables, one can eventually test the hypotheses poised in the study. The subject of this thesis, are methods to study brain activity and dynamics while treating the brain as an isolated system interacting with the environment. Since in practice, there are no efficient ways to randomize brain states and test how they influence behaviors, we perform the experiments the other way around: we randomize experimental stimuli, and record the outcome brain states.

The methods characterized and proposed in this thesis cover three aspects of the investigated brain dynamics: (1) finding activation patterns (signal versus noise), (2) determining significance of functional links between distinct areas, and (3) finding the direct causal influences between distinct areas.

1.4 Signal detection

In Chapter 2, the topic of signal detection in task- and in resting state fMRI is introduced. I define signal detection as finding non-random patterns of activation during cognitive tasks, or during rest. Signal detection is important for testing research hypotheses associated with a set of predefined ROIs as well as for finding new ROIs in a functional way as mentioned above, in Section 1.3.5.

Historically, the first method to discriminate between signal and noise adapted to fMRI, was the General Linear Model (GLM, Friston et al., 2007). In simplest terms, GLM compares the means of the signal during activation and rest, and as such, it finds voxels activated during cognition.

However, can we also talk about signals in the BOLD fMRI time series in the resting state? Since fMRI has poor time resolution, we do not expect any strong oscillatory activity. How to detect a signal in such conditions? There is no a priori hypothesis about the form of the signals in the resting brain but in general, looking for non-Gaussianity in the distributions of BOLD fMRI samples can be a proxy for signal detection. Independent Component Analysis (Hyvärinen and Oja, 2000) in application to fMRI is a technique based on the observation that the distribution of BOLD fMRI samples is non-Gaussian, therefore voxels can be classified based on the type of non-Gaussianity that they represent. Non-Gaussianity of the BOLD fMRI sample distribution is, therefore, interpreted as some form of a signal.

In Chapter 3, I introduce my input to how methods for signal detection can be improved using more sensitive tests for normality of the distribution incorporating fractional moments of the distribution.

1.5 Detection of connections

The brain is a densely connected system of interacting nodes. Assuming that the network was successfully created using either an anatomical atlas or a functional parcellation, we can further study these interactions. Functional connectivity (FC) is a very broad term to describe statistical associations between variables, e.g., between neuronal activity in distinct brain regions. In fMRI, this concept is used in many contexts (Smith et al., 2013).
1.5. Detection of connections

FC serves to study the (co)activity in the neuronal networks, and to investigate links between neuronal activity and cognitive skills (Smith et al., 2015; Finn et al., 2015; Tavor et al., 2016; Smith, 2016; Chauvin et al., 2017) or clinical-behavioural covariates (Lynall et al., 2010; Garrity et al., 2007; Greicius et al., 2007; Harrison et al., 2009; Rausch et al., 2016; Oldehinkel et al., 2016; Mulders et al., 2015). It is also used to gain insights into hierarchical structures in the brain in rest and cognition (Smith et al., 2015; Bola and Borchardt, 2016), e.g., the hierarchical structure of sensory systems (Arcaro et al., 2015; Merhar et al., 2016).

Functional connectivity is typically operationalized either as Pearson correlation (Pearson and Hartley, 1972) or partial correlation (Baba, Shibata, and Sibuya, 2004; Marrelec et al., 2006). In partial correlation, the association between two random variables is corrected for the linear influence of all the other variables in the network. Other alternative ways of operationalizing functional connectivity include, e.g., mutual information (Press et al., 2007) or coherence (Everitt, 2002). The latter is less popular in analyzing fMRI datasets given its poor temporal dynamics, yet it is still used in fMRI research at times (Sun, Miller, and D’Esposito, 2004).

Regardless of which method was chosen to operationalize FC, for any given pair of random time series, the computed functional connectivity will be non-zero - even if the data represents pure noise. The reason is because every two finite, unrelated time series will be weakly (positively or negatively) correlated by chance. Therefore, the question: how to filter out the spurious connections induced by the background noise, and select the significant functional connections? In Chapter 4, I review the most popular approaches to address this issue, e.g., proportional thresholding (Achard and Bullmore, 2007; Bassett and Bullmore, 2009; Heuvel et al., 2008) and permutation testing (Welch, 1990). Furthermore, I propose mixture modeling as an advance to the current methods (Bielczyk et al., 2018), using the example of thresholding partial correlation matrices computed for the human visual system in resting state and under visual stimulation, based on the Human Connectome Project (Essen et al., 2013) datasets. One advance that mixture modeling has over its predecessors, is the fact that when using mixture modeling, every connectome can be treated individually: the thresholding is done based on the distribution of all connections within the subject-specific network, and then computing a False Discovery Rate for every threshold using a mixture of two distributions (one of them representing the pseudo-null distribution, and the other representing the distribution of signals). This allows for choosing a threshold at a pre-specified confidence level operationalized by FDR, for a single subject. In contrast to mixture modeling, in permutation testing a reference population of subjects is necessary for computing confidence levels (by repeatedly shuffling the node labels between subjects and building a separate null distribution for every connection in the network). This is an advantage as mixture modeling allows for case studies as well as for analyzing data sets with only a few subjects involved (which is often the case, e.g., in translational psychiatry studies).
1.6 Causal inference in fMRI

Finally, if we are interested not only in the (undirected) statistical associations between activity in distinct brain regions, but also in the direct mutual influences between them, we need to search for causal dependencies. Causality is a complex research problem. In fMRI specifically, rendering causal interactions from the data is a multifaceted problem because of the multiple shortcomings of the fMRI data (Smith et al., 2011; Friston, 2011; Bielczyk et al., 2017b; Bielczyk et al., 2019). I review these shortcomings in detail and discuss three main classes of methods developed in this field in Chapter 5. In short, the methods can be grouped into classes according to various criteria, but I concentrated on the division into three groups based on the assumptions made on the connectivity structure within the investigated network:

1. Multivariate methods that search for directed graphs without imposing any particular structure on the graph (further referred to as the network-wise models),

2. Multivariate methods (further referred to as hierarchical network-wise models) including an additional assumption of acyclicity, i.e., an assumption that the information travels through the brain by feed-forward projections only. As a result, the network can always be represented in a form of a Directed Acyclic Graph (DAG, Thulasiraman and Swamy, 1992),

3. Pairwise methods that involve a two-stage procedure. Firstly, a map of nondirectional functional connections is rendered. Secondly, the directionality of the causal link is assessed within each connection. Since these methods focus on pairwise connections rather than the network architecture, they do not impose the acyclicity assumption.

It should be highlighted that thresholding connectomes is the first step of the two-step pairwise inference procedure. Therefore, an advance to the algorithms for thresholding functional connectomes with the use of mixture modeling proposed in this thesis (Bielczyk et al., 2018, Section 1.5), also provides an advance to the methods for causal inference in fMRI.

Furthermore, I perform a theoretical study with the use of Dynamic Causal Modeling generative model Friston, Harrison, and Penny, 2003 to determine which preprocessing strategies can be beneficial or detrimental to the causal inference in fMRI (Bielczyk et al., 2017b). The study was described in Chapter 6. In the study, I reported two major findings. Firstly, I found that the scale-free background noise coming from suboptimal brain parcellation is more detrimental to the causal inference than the hemodynamics itself. Therefore, the functional parcellation into areas is highly advisable, as apposite to anatomical parcellations of the brain. Secondly, I found out that lagged methods such as Granger Causality (Granger, 1969; Seth, Barrett, and Barnett, 2015) might not be reliable as even a small local variability in the hemodynamic lags, can cause a flip in the directionality of the causal link. I conclude that a primary feature of a method for causal inference in fMRI should be resilience to the background noise within the network. Furthermore, I do not use any methods that involve lags in the communication between ROIs, and are solely based on information hidden in
1.7 Wider implications for psychiatric research

Modeling causal interactions in brain networks can be the basis to further network modeling, e.g., modeling brain networks with the use of neural mass models (Wilson and Cowan, 1972; Deco, Jirsa, and McIntosh, 2013). In this class of models, the activity of neuronal populations (or, brain regions) is represented by dynamical variables. These models have interesting dynamical properties: the dynamics in the network can settle to a stable pattern, a so-called attractor state. Analysis of the distribution of attractor states and their basins of attraction can give insights into the differences between the neurotypical and clinical subjects.

Even though computational psychiatry is a developed field in general (Huys, Maia, and Frank, 2016; Montague et al., 2012; Wang and Krystal, 2014; Deco and Kringelbach, 2014; Friston et al., 2014), the (mal)-functioning of the large scale networks in psychiatric disorders remains poorly understood. In Chapter 8, I use the example of Major Depression (MDD) to introduce the most urgent questions in modeling of the large-scale brain networks in disease. I also propose ways to characterize the global dynamics of the brain in disease.
Chapter 2


Characterizing the dynamics of the BOLD time series and finding activation patterns based on this dynamics, are the fundamentals of cognitive research in fMRI. The BOLD response is (most likely) stochastic, and the experimental datasets have a low number of datapoints compared to time series collected in other domains, e.g., electroencephalography. In this Chapter, I review and compare the most popular approaches to finding activation patterns in resting state and under cognitive stimulation. Furthermore, I discuss their pros, cons, and limitations.

Keywords: signal detection, activation patterns, cognition, General Linear Model, Independent Component Analysis

2.1 Statistical Parametric Mapping (SPM)

2.1.1 General Linear Model (GLM)

General Linear Model (GLM) is the most popular technique for finding brain activations from fMRI datasets. GLM was first formulated by John Nelder and Robert Wedderburn in 1970s (Nelder and Wedderburn, 1972), and adapted to fMRI research by Karl Friston and colleagues (Friston et al., 1995).

GLM is a study of a relationship between a set of dependent variables (in this case, representing components of behavior) and one or more independent variables, also known as predictors or explanatory variables, with the use of either a linear or a nonlinear model. In epidemiology and clinical psychiatry, regression models are widely used both for prediction, and for causal discovery. In fMRI, GLM modeling is univariate (although in general, it can be multivariate as well).

In univariate linear regression, the set of dependent variables is modeled as a linear combination of predictor variables. In this model, regressor coefficients fitted to the model are the parameters associated with causal links between independent and dependent variables. The solution to the inference procedure is a set of regression coefficients that allows for fitting the dependent variables as a linear combination of predictors, with the minimum inaccuracy (expressed as a noise term in the model).
In the most general, multivariate form, GLM model involves 4 matrices:

\[ Y = XB + \Sigma \]  \hspace{1cm} (2.1)

where \( Y \) is a matrix containing a series of multivariate measurements, \( X \) is a matrix of observations for predictor variables often referred to as a design matrix, \( B \) - the matrix that should be estimated, and \( \Sigma \) - the matrix of stochastic noise terms. \( \Sigma \in N(0, V) \), where \( V \) - the covariance matrix for the multivariate noise. In this model, \( X \) and \( Y \) are given while \( B \) and \( \Sigma \) should be estimated.

### 2.1.2 Fitting a GLM model

Optimal fitting strategy for this model depends on the structure of the noise covariance matrix \( V \). There are two cases:

1. If \( \Sigma \) is iid (covariance matrix \( V \) is trivial: \( \Sigma = \sigma^2 I \)), then the Ordinary Least Square estimate is optimal (OLS, Edwards, 1976; Hayashi, 2000). We aim to find the solution to the GLM by find value of \( B \) that minimizes \( (Y - XB)'(Y - XB) \). Taking the derivative with respect to \( B \), we get \( X'XB = X'Y \). The solution is:

\[ \hat{B} = (X'X)^{-1}X'Y \]  \hspace{1cm} (2.2)

Then, \( \sigma^2 \) can be estimated as

\[ r = Y - \hat{Y} = (I - X(X'V^{-1}X)^{-1}X'V^{-1})Y = RY \]  \hspace{1cm} (2.3)

where \( R \) - the residual inducing matrix. Then

\[ \hat{\sigma}^2 = \frac{r'Tr}{tr(RV)} \]  \hspace{1cm} (2.4)

2. If \( \Sigma \) is not iid (covariance matrix \( V \) is not trivial: \( \Sigma = \sigma^2 V \)), then Generalized Least Square estimate is optimal (GLS, Aitken, 1934):

\[ \hat{B} = (X'V^{-1}X)^{-1}X'V^{-1}Y \]  \hspace{1cm} (2.5)

Estimating \( \sigma^2 \) is more complex in this case (Madsen, 2010). To briefly summarize this algorithm: estimating \( B \) depends on the noise variance \( V \) and vice versa. Therefore, we need to employ an iterative procedure as follows:

(a) Assume \( V = I \) and estimate the OLS solution,
(b) Estimate the parameters of \( \hat{V} \) using residuals,
(c) Reestimate \( B \) using the estimated covariance matrix \( \hat{V} \) from previous step,
(d) Iterate until convergence.

For more details on this estimation, please check Bishop, 2006.
2.1.3 GLM in the context of functional Magnetic Resonance Imaging

In fMRI, GLM is often referred to as a massive univariate approach, Fig. 2.1A. This is because the GLM model is estimated for every voxel separately. Of course, since fMRI data is typically spatially smoothed as described in Section 1.3.3, in fact BOLD signals in neighboring voxels are correlated.

In fMRI, there are two types of regressors. Firstly, regressors of interest represent the intentional design of the experiment (i.e., events such as experimental stimuli and cues). To test research hypotheses, e.g., to compare two particular conditions with each other, contrasts between conditions can be applied.

Secondly, regressors representing other factors that can influence the results of the experiment, should also be appended to the model. These regressors are usually referred to as regressors of no interest or nuisance regressors. Possible nuisance regressors include:

1. Slow drift (modeled as a cosine basis set),
2. Physiological artifacts like heartbeat and respiration,
3. Head motion.

In fMRI, GLM also involves Linear Time Invariant systems (LTI, Hespanha, 2009), i.e., convolutions with a hemodynamic response of a known shape. This convolution is added to the model to characterize the influence of non-linear hemodynamic response on the outcome BOLD fMRI responses (Fig. 2.1). In both standard packages for data analysis in fMRI - Statistical Parametric Mapping package (Friston et al., 2007) and FSL (Jenkinson et al., 2012), hemodynamic response function can be modeled in a few different ways. However, since the hemodynamic response differs across the brain, typically, temporal basis functions are used since a linear combination of functions can account for delay and dispersion in the hemodynamic response. Regressors are convolved with each of the basis functions to give the final design matrix.

Now, let us assume that the model 2.1 has been fitted, so that $\mathbf{B}$ and $\mathbf{V}$ are estimated. Can we find activation in a given voxel given the computed effect magnitudes $B_1, \ldots, B_n$? If $\mathbf{B}$ is a vector of regressor coefficients and $c$ is a contrast vector, then the t-statistic yields:

$$T = \frac{c' \hat{\mathbf{B}}}{\sqrt{\text{var}(c' \mathbf{B})}}$$

with the null hypothesis of no effect, and the number of degrees of freedom equal to $f = \frac{\ell^2 \text{tr} \left( \mathbf{R} \right)}{\text{tr} \left( \mathbf{R}^2 \right)}$. A few contrasts can also be tested at once by introducing a contrast matrix.

2.1.4 Finding effects in second-level analysis

Eq. 2.6 refers to finding voxel-wise activations in a single subject. However, in most studies a cohort of subjects participates in the experiment to leverage
the power of the study. This allows for a second-level analysis: finding statistically significant differences between two groups of subjects (e.g., an experimental group that receives a particular cognitive stimulation versus a control group that goes through an analogous resting state study in the same conditions). There can be a large inter-subject variability in measured responses to experimental manipulation that one needs to take into account. There are two possible strategies to model brain activations in group studies:

1. Fixed effect analysis: assumes that effect size is the same in all subjects and the only source of variation is the measurement error,

2. Mixed effect analysis: assumes that there are two sources of variation: the differences in effect size between subjects and the measurement error.

In any of these two scenarios, second order analysis links single subject \( B \) values to the population distribution of \( B \) values.

### 2.1.5 Correcting for multiple comparisons

**Family-Wise Error (FWE)**

A Family-Wise Error (FWE) is the probability of making one or more false discoveries, or type I errors, among all the hypotheses when performing multiple hypotheses tests. A simple example of FWE is Bonferroni correction, which treats all tests as independent from each other, and requires dividing the desired confidence level by the number of tests (Miller, 1966). Another approach to FWE can be fixing the threshold p-value to one (uncorrected) level that is more stringent than p-value accepted for a single test (typically, \( p < 0.05 \)), e.g., \( p < 0.0001 \).

**Random Field Theory (RFT)**

Unlike the Bonferroni correction for FWE that controls for the expected number of false positives among single voxels, Random Field Theory (RFT, Worsley, 1994; Sigmund and Worsley, 1995; Worsley, 1996; Cao and Worsley, 1999; Worsley, 1995) corrects for the expected number of false positives among regions defined as clusters of co-activated voxels (Brett, Penny, and Kiebel, 2003). In this case, the correction should not depend on the number of voxels as they are more or less arbitrary units of volume, whereas a region is a topological feature (Fig. 2.1B).

Random Field Theory treats statistical maps obtained from the GLM analysis as a discretization of an underlying continuous field with associated topological features such as activity amplitude, cluster size, number of clusters, etc. Therefore, RFT employs topological inference: inferring peak height, cluster extent, number of clusters, etc. In RFT, the basic volume unit is a resel rather than a voxel. Resel was a term coined by Worsley et al., 1992, and is a measure of the number of resolution elements in the statistical map defined simply as a block of voxels of the same size as the full width at half maximum (FWHM) of the smoothing kernel (see: Section 1.3.3).

The algorithm for applying RFT to statistical maps obtained from GLM modeling is as follows:
2.1. Statistical Parametric Mapping (SPM)

The General Linear Model (GLM) is a statistical model that is used to analyze fMRI data. It is based on the assumption that the BOLD fMRI signal can be modeled as a linear combination of multiple regressors, each representing a different component of the signal, such as the effect of a specific experimental condition or the baseline signal. The subject of GLM modeling is finding parameters $\beta_i$ that give the best fit to the experimental data. Since the modeling is performed for each voxel separately, we obtain a 3D map of activation by mapping parameters $\beta_i$ back onto the brain. 

Random Field Theory (RFT) treats statistical maps obtained from the GLM analysis as a discretization of the underlying continuous field with associated topological features such as activity amplitude, cluster size, number of clusters, etc. We can expect activation with varying probabilities depending on the threshold $t$.

1. Counting number of resels $R$ in the dataset (usually computed as the number of voxels divided by 3D FWHM),

2. Calculating the Euler characteristics. The higher the threshold $t$, the less probable it becomes to find a blob of activity in an image after thresholding. Euler characteristics is an expression which corresponds to an approximate for the probability of finding an above threshold blob in the statistical image (Eq. 2.7).

$$EC = R \frac{Mn(2)^{3/2}}{2\pi^2} \exp\left(-\frac{t^2}{2}\right)(t^2 - 1)$$

Using this formula is very simple. E.g., to set a probability of obtaining a false positive blob at $p = 0.05$, one should pick a threshold $t$ at which $EC = 0.05$.

Today, RFT is the first line of choice for correcting for multiple comparisons in fMRI GLM studies (Worsley et al., 1996). Indeed, RFT is typically a better choice than FWE for the smoothed imaging pictures as it is hard to determine how many independent observations there are in the smoothed image. As RFT relates to testing significance for blobs of activity rather than single voxels, it is often colloquially related to as “blobology” (Poldrack, 2018).
Applying GLM together with RFT in one pipeline is often referred to as Statistical Parametric Mapping (SPM, Friston et al., 1995).

**False Discovery Rate (FDR)**

FWER is usually too stringent while RFT is highly parametric. One alternative to both these approaches, is using the False Discovery Rate (FDR). Rather than controlling the FWE (one or more false detections amongst all the tests), FDR controls for the *expected number of false positives amongst all the positives.*

The most popular approach to FDR is the Benjamini-Hochberg method (Benjamini and Hochberg, 1995) which controls for the false discovery rate for all discoveries (at level $\alpha$). The algorithm is very simplistic: just order all $p$-values representing single tests in ascending order $p_1, \ldots, p_m$, and find the largest $k$ such as $p_k \leq \frac{k}{m} \alpha$. Then, reject all the null hypotheses for $i = 1, \ldots, k$.

Using FDR to threshold activation patterns found in the GLM analysis has a long tradition in fMRI studies (Genovese, Lazar, and Nichols, 2002).

**2.1.6 Basic findings from Statistical Parametric Mapping in fMRI**

Statistical Parametric Mapping is the default pipeline in studies localizing cognitive processes in the brain. One of the most extensive studies in this domain was performed by Barch et al., 2013 based on the high quality Human Connectome Project datasets (HCP, Essen et al., 2013). HCP is an extensive study in which, next to two resting state scanning sessions, a battery of seven cognitive tasks were performed on a large cohort of healthy individuals:

1. N-back working memory task (Drobeshevsky, Baumann, and Schneider, 2006): presenting blocks of trials consisting of pictures representing four different classes: faces, places, tools, and body parts. Within each run, the 4 different stimulus types are presented in blocks, and the subject is asked to recognize the class of an object that is currently presented ("0-back" condition) or was presented two trials before ("2-back" condition). In the n-back task, it was found that a broad range of regions are active, e.g., bilateral dorsal and ventral prefrontal cortex, dorsal parietal cortex and dorsal anterior cingulate. These brain regions are believed to be responsible for cognitive control. At the same time, the default mode network (DMN) was deactivated which is consistent with prior findings that DMN and executive networks anticorrelate (Fig. 2.2A),

2. Motor task (Buckner et al., 2011): subjects are presented with visual commands asking them to perform one of the following five actions: (1) tap left pointing finger, (2) tap right pointing finger, (3) squeeze left toes, (4) squeeze right toes, (5) move their tongue. In the motor task, the SNR in the motor cortex was very high, and the activation patterns were dependent on the class of presented objects as expected (Fig. 2.2B). The comparison between classes of objects, demonstrating the division of motor cortex into subregions mapping separate limbs, is presented in Fig. 2.2F,
3. Gambling task (a.k.a. incentive processing task, Delgado et al., 2000): subjects play a card guessing game. They are asked to guess the number on a card marked with a question mark to win or lose money. The subjects are also told that potential card numbers range from 1 to 9. To indicate if they believe the mystery card number is more or less than 5, they need to press one of two buttons on the response box. The subjects receive the card number as a feedback; it is generated by the associated software as a function of whether the trial was a "reward", a "loss", or a "neutral" trial. In the reward trials, the feedback is presented as a green arrow pointing upwards with a dollar sign. In the loss trials, the feedback is presented as a red arrow pointing downwards next to 0.50USD. In the neutral trials, number 5 and a gray double headed arrow are presented. In the gambling task, multiple brain regions involved in reward processing - including bilateral striatum and bilateral insula - were activated. Fewer regions are active in the group map comparing reward with punishment, although there is some differential activation in the visual cortex and the striatum (Fig. 2.2C).

4. Relational processing task (Smith, Keramatian, and Christoff, 2007): the stimuli consist of 6 different shapes filled with one of 6 different textures. In the relational processing condition, subjects are presented with two pairs of objects: one pair at the top of the screen and the other pair at the bottom of the screen. The subjects are first asked which dimension differs across the top pair of objects (either shape or texture) and then whether the bottom pair of objects also differs along that same dimension (e.g., "if the top pair differs in texture, then does the bottom pair also differ in texture?"). The relational processing task activated bilateral anterior prefrontal cortex (Fig. 2.2D).

5. Social cognition task (a.k.a. Theory of Mind task, Castelli et al., 2000; Castelli, Happe, and Frith, 2002): subjects are presented with short video clips of geometrical objects (squares, circles, triangles) either moving randomly across the screen or interacting in some way. After presentation, the subjects are asked whether the objects had a social interaction, i.e., an interaction that suggests that objects respect each other’s feelings or thoughts). In the social cognition task, temporal parietal junction and superior temporal cortex regions were activated on the group level (Fig. 2.2E).

6. Emotional processing task (Hariri et al., 2002): subjects are presented with blocks of trials representing the emotion processing condition and the control condition, respectively. In the emotion processing condition, the subjects are asked which of two faces presented on the bottom of the screen matches the face presented at the top of the screen. The faces have either angry or fearful expressions. In the control condition, the subjects are asked which of two shapes presented at the bottom of the screen matches the shape presented at the top of the screen. In the emotional processing task, robust bilateral activation of the amygdala extending into the hippocampus and bilateral activation in the medial and the lateral orbital frontal cortices were found.
Furthermore, some activation was found in the visual regions, including the fusiform face area responsible for face perception - which was expected given that there were faces presented as the task stimuli (Fig. 2.2G).

Figure 2.2: Statistical Parametric Mapping applied to seven cognitive tasks from the Human Connectome Project (Essen et al., 2013) datasets. Reprinted from Barch et al., 2013, copyright (2018), with permission from Elsevier. A: N-back working memory task. In this task, a broad range of regions were found active, e.g., bilateral dorsal and ventral prefrontal cortex, dorsal parietal cortex and dorsal anterior cingulate - regions believed to be responsible for cognitive control. At the same time, the default mode network (DMN) was deactivated. This result is consistent with prior findings that DMN and executive networks anticorrelate. B: Motor task. In this task, different parts of motor system were activated for different classes of presented objects. C: Gambling task. In this task, multiple regions typically involved in reward processing - including the bilateral striatum and the bilateral insula - were activated. Fewer regions were active in the group map comparing reward with punishment, although there was some differential activation in the visual cortex and the striatum. D: Relational processing task. In this task, the bilateral anterior prefrontal cortex was activated. E: Social cognition task. In this task, temporal parietal junction and superior temporal cortex regions were found active. F: Summary of the results for the motor task. G: emotional processing task. In this task, robust bilateral activation of the amygdala extending into the hippocampus and bilateral activation in the medial and lateral orbital frontal cortices were found. Additionally, some activation was found in the visual regions, including the fusiform face area. H: Language processing task. In this task, activation was found in ventral lateral prefrontal cortex and in both superior and inferior temporal cortices, including the anterior temporal poles. Activation was left-lateralized.

7. Language processing task (Binder et al., 2011): consisted of two runs of 4 blocks of a story task interleaving with 4 blocks of a math task. The story blocks were given in a form of brief auditory stories of 5-9 sentence long exempts from Aesop’s fables. The stories are followed
Chapter 2

2.1. Statistical Parametric Mapping (SPM)

by a 2-alternative choice question asked to monitor attention of the participant. An example from Binder et al., 2011: "For example, after a story about an eagle that saves a man who had done him a favor, subjects were asked, 'That was about revenge or reciprocity?'" In the language task, a group activation was found in the ventral lateral prefrontal cortex and in both the superior and the inferior temporal cortices, including the anterior temporal poles. Activation was left-lateralized (Fig. 2.2H).

The results of the analysis by Barch et al., 2013 are highly dependent on the details of experimental paradigms used in the study. In fact, there is a broad selection of possible tasks that might represent different aspects of virtually any cognitive process. Therefore, to develop more robust mapping of cognition onto the brain, more sophisticated tools are necessary. Namely, pooling information over large databases containing group results coming from a variety of tasks and using meta-analytic tools such as Neurosynth (Yarkoni et al., 2011), is recommended (Varoquaux et al., 2018). Combining multiple experimental tasks into studying cognition is a complex research problem, also because it involves the so-called inverse inference problem (Poldrack, 2011). Namely, what we should aim for, in fact, is the forward inference (Henson, 2005): manipulating a specific psychological function and identifying effects of this experimental intervention on the activity of the brain. However, we often infer the other way around: different tasks are believed to induce similar psychological functions because they activate similar regions of the brain (Poldrack, 2011). In fact, there are multiple ways of approaching this inference as described by Varoquaux et al., 2018. Ontology defines cognitive functions and the relations between them. We have as follows:

1. Forward term mapping: maps differences between brain responses for a given term and its neighbors within ontology,

2. Forward inference with ontology contrasts (the standard analysis): GLM models the brain responses as linear combinations of multiple effects. This analysis allows for creating contrasts between conditions. Ontology contrasts are contrasts comparing responses to the same category of cognitive stimuli (e.g., presenting faces) but across different studies,

3. Reverse inference with logistic regression: large-scale decoding by fitting a predictive model (Poldrack, 2012). The method used in the study is based on stacked regressions (Breiman, 1996): two layers of logistic regressions discriminating between cognitive terms. The first layer discriminates between ontology-related terms while the second layer predicts which specific term is most relevant,

4. NeuroSynth (Yarkoni et al., 2011) reverse inference: meta-analysis with the use of Neurosynth,

5. Mapping using both decoding and ontology: leveraging ontologies of cognitive concepts and multi-label brain decoding to map the neural mechanisms behind these concepts (Varoquaux et al., 2018).
The authors demonstrated and compared these approaches by building an atlas of cognition based on 30 fMRI studies including a total of 196 experimental conditions. The exemplary differential results from this analysis are presented in Fig. 2.3.

**Figure 2.3:** Strategies for mapping cognition onto the brain, and differential results when applied to studies over visual and auditory processing. Reprinted from Varoquaux et al., 2018, Copyright (2018) by PLoS Computational Biology, at the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. Top panel: visual regions in the left hemisphere. Bottom panel: auditory regions in the left hemisphere. A: Forward term mapping. B: Forward inference with ontology contrasts (standard analysis). C: Reverse inference with logistic regression. D: NeuroSynth (Yarkoni et al., 2011) reverse inference. E: A new approach by Varoquaux et al., 2018: mapping with the use of decoding and ontology.

### 2.1.7 Statistical Parametric Mapping in studies on psychiatric disorders

In research studies on cognitive pathologies such as Attention Deficit Hyperactivity Disorder or Autism Spectrum Disorder (Kupfer and Regier, 2011), the SNR ratio necessary to discriminate between an experimental condition and resting state is not sufficient as one needs to additionally detect differences in activation between the experimental groups. Therefore, in general, these studies require more careful choice of experimental paradigms
as well as higher effect sizes. There are two main types of such studies. On one hand, one can perform a case-control study, in which group effects are compared between neurotypical controls and the diagnostic group. On the other hand, one can treat diagnostic traits as continuous variables and correlate the performance along the cognitive dimensions represented by these variables with the activation patterns across the brain.

2.2 Multivoxel Pattern Analysis (MVPA)

Statistical Parametric Mapping is a standard approach to analyze activation patterns in the brain. However, as a massive univariate approach, it assumes that voxels are separate units of brain tissue. Unlike SPM, Multivoxel Pattern Analysis (MVPA, Davatzikos et al., 2005; Norman et al., 2006; Mahmoudi et al., 2012) involves searching for reproducible *spatial patterns* of BOLD fMRI activity that differ between experimental conditions. In other words, the desired spatial patterns should allow to best *classify* between experimental conditions.

Unlike the GLM, MVPA is not one algorithm but the whole *class* of algorithms: the patterns can be rendered from experimental datasets with the use of multiple supervised classification methods. The brief overview of the available types of MVPA classifiers is given below:

1. Linear Support Vector Machines (LSVMs, Cortes and Vapnik, 1995): a non-probabilistic binary linear classifier that divides objects into classes separated by parallel lines, as distant from each other as possible. Simple versions of SVMs are only applicable to two-class classification problems. In multi-class classification problems, one needs to split the research question into multiple binary classification problems. SVMs are most often used in MVPA in cognitive neuroimaging,

2. K-Nearest Neighbors algorithm (kNN, Altman, 1992): a nonparametric, iterative algorithm assigning each data point to the same class as their closest neighbors in (typically) Euclidean coordinates for as long as it takes to reach equilibrium with no further reassignments,

3. Gaussian Process classifier (Seeger, 2004; Rasmussen and Williams, 2006): a classifier fitting multivariate Gaussian distributions to model different classes in the multivariate feature space. Since Gaussian distribution has indefinite domain, unlike in SVM or kNN, distributions modeling different classes can be strongly overlapping,

4. Decision Tree classifier (Rokach and Maimon, 2008): a hierarchical classifier in which features are not independent. In a Decision Tree, leaves represent classes while branches represent conjunctions between features that lead to those class labels,

5. Random Forest classifier (Ho, 1995): an ensemble learning method in which multiple decision trees are built for the different subsets of the data. The prediction accuracy is then computed as a mean accuracy over all instantiations of the tree. This method allows for solving the problem of low robustness in the classic decision trees (small change in the training dataset can lead to a large change in the outcome tree architecture, and to a major change in the final predictions),
6. Neural Nets (NNs, Bishop, 2006): a broad class of hierarchical networks that imitate neuronal networks. There are multiple possible architectures of NNs, including feed forward networks and recurrent networks. In NNs, the most often implemented learning mechanisms involve Hebbian learning (Hebb, 1949; Farley and Clark, 1954) and backpropagation (Goodfellow, Bengio, and Courville, 2016).


There is no good answer to the question, "Which classifier is best?": the choice of the classification method for the given study should depend on the properties of the particular dataset (Fig. 2.4, Varoquaux and Müller, 2011). More insights about the nature of these classifiers can be found in Bishop, 2006.

![Figure 2.4: A comparison of several classifiers in scikit-learn on synthetic datasets.](image)

Feature selection and dimensionality reduction are two important aspects of MVPA analysis (Mahmoudi et al., 2012). Features for MVPA may be created from the statistical maps estimated using GLM. However, if one analyzes patterns of activity across all single voxels in the brain, MVPA becomes computationally expensive. An alternative scenario can be choosing a smaller subset of voxels for the analysis. This method is often referred to as the searchlight analysis (Kriegeskorte, Goebel, and Bandettini, 2006).

MVPA should be used with caution. As studies on synthetic data demonstrated, MVPA can be sensitive to the magnitude of voxel-level variability in the effect sizes within single subjects (Davis et al., 2014). Furthermore, as studies in macaques have shown, decoding presented visual stimuli from the visual cortex with the use of MVPA is hard (Berker and Tsao, 2015).
2.3 What next?

SPM and MVPA are standard approaches to find activation patterns from experimental fMRI datasets. All these methods explore very basic features of the data. The core of Statistical Parametric Mapping is GLM modeling which effectively compares the mean of the BOLD fMRI signal between conditions. In MVPA, features are also typically derived from GLM modeling, and as such, based on similarly basic characteristics. The question would be: are there any features other than the mean signal, which would allow for discriminating between different conditions?
Chapter 3

Momentum: a robust test for normality based on fractional moments of distribution

Normality testing is necessary in multiple domains of natural sciences: from epidemiological studies, through genome-wide association studies to regression models in biology, to economy and other fields.

In this work, I propose Momentum: a new, robust test for normality, based on a novel concept of using fractional moments of the sample distribution to fully characterize the shape of the distribution. Unlike some other methods for quantifying the sample deviation from a normal distribution such as Kullback-Leibler divergence, Momentum is well defined for a sample of any length. Moreover, as it integrates information about the shape of the distribution over all its moments, it is more sensitive than other methods for testing normality based on a few moments such as Shapiro-Wolf test (based on a combination of the first and the second moment) or Jarque-Bera test (based on a combination of the third and the fourth moment). In this work, I define the normality test and the rules for deriving the confidence intervals. Furthermore, I compare the performance of Momentum with a few other tests for normality using synthetic datasets.

I also propose a number of applications of Momentum as the method for normality testing and signal detection in fMRI. The method can be applied in two ways. Firstly, as a holistic signal detection method, Momentum can be used to create 3D maps of effect sizes across the brain (either voxel-wise or ROI-wise). This can help, e.g., in analyzing resting state datasets and deriving new, spatially acute biomarkers of cognition and psychopathology (even at a single voxel resolution) using second level analysis in clinical cohorts. Secondly, as a method allowing for a thorough description of the distribution, sets of fractional moments can serve to better describe differences in mean BOLD distributions between groups, moment by moment.

Overall, there is a huge potential for using Momentum to reveal new biomarkers of cognition and psychopathology from fMRI datasets (and multiple applications in other branches of science).

Keywords: normality, signal detection, stochastic systems, distribution, cohort studies, task-fMRI, resting state fMRI

This Chapter is in preparation for submission as:

Chapter 3. Momentum: a robust test for normality

3.1 Introduction

Normality of a distribution is defined by equivalence to a Gaussian distribution (usually, "normal" also implies that the distribution is scaled to the mean of 0 and the variance of 1). According to the studies by Curran-Everett (Curran-Everett and Benos, 2004), about 50% of the published articles contain at least one statistical error. Therefore, it is crucial to check the normality assumption for all the statistical procedures that require this assumption (Ghasemi and Zahediasl, 2012; Curran-Everett, 2017). The most popular statistical methods requiring normality of the sample, are: Pearson correlation (Pearson and Hartley, 1972), T-test, F-test, Bartlett’s test (Snedecor and Cochran, 1989), ANOVA, and regression models (Dixon, 1983).

Secondly, normality tests are used whenever normality of the sample distribution is expected in the data analysis. When this assumption is not fulfilled, it is a warning that the data is contaminated, or that there are some (possibly, latent) confounding variables in the study. As an example, in the genome-wise association studies, the distribution of effect sizes for the single-nucleotide polymorphisms (SNPs) across the genome is compared to a normal distribution to determine whether there is a signal in the data or rather, if all the detected effects can be explained as random. Another example can be psychometric studies: psychological tests and questionnaires are standardized under the assumption that the distribution of behavioral phenotypes in a cohort is normal. In case the sample was not normal, it can be a sign that the test/questionnaire is measuring a combination of two or more independent traits and the psychometric tool should either be simplified or factorized into multiple dimensions.

According to Central Limit Theorem, a mixture of multiple independent random variables always tends towards a normal distribution (Laplace, 1812; Billingsley, 1995). Therefore, the presence of non-normality in the data can also be associated with presence of a particular signal. This is the reason why in certain circumstances, normality tests can also be recommended as a method of choice for signal detection. This can be advised especially when analyzing datasets that are very noisy and of poor time resolution. E.g., in fMRI, Independent Component Analysis (ICA) is an algorithm used to separate a multivariate time series into a linear combination of sources with a non-Gaussian distribution (Hyvärinen, Karhunen, and Oja, 2001; Beckmann, 2012). ICA uses measures of non-normality to compare between voxels, and to group proximal voxels of similar source characteristics together. One popular method for signal analysis in such conditions, is also the maximum entropy principle (Jaynes, 1957a; Jaynes, 1957b). In this approach, it is assumed that the probability distribution that best represents the data is the one with the highest entropy.

In this work, I introduce Momentum: a new test for normality based on a (predefined) set of fractional moments of the distribution. Using a range of moments of fractional orders allows for creating a thorough characteristics of a distribution. As such, it is valuable especially when there is no presumption on the particular properties of the distribution. In this Chapter, I compare the performance of Momentum with other, popular tests for normality using synthetic datasets. Furthermore, I discuss the potential applications in fMRI research and beyond.
3.2 Materials and methods

3.2.1 Defining fractional moments of the distribution

In this work, I propose to re-parametrize the difference between a given distribution and a normal distribution, from integration over the original domain (as implemented, e.g., in Kullback-Leibler divergence) to integration over the space spanned by the moments of the distribution. Momentum is based on the continuous, central fractional moments of the distribution. Let us imagine a sample \( X = (x_1, ..., x_N) \), with mean \( \mu = 0 \). The (central) moments of the distribution of this sample, are defined as follows:

\[
M_k = \frac{1}{N} \sum_{i=1}^{N} x_i^k
\]  

(3.1)

where \( k \geq 0 \). Note that \( k \in \mathbb{N} \rightarrow M_k \in \mathbb{R} \), and \( k \in \mathbb{R} \rightarrow M_k \in \mathbb{C} \). The mean of the distribution is equal to zero which implies that some elements of the sample have negative values. For \( k = 0.5 \), \(-1^k = i\), therefore, for the left part of the distribution, the fractional moments become complex. Then, since the RHS of the Eq. 3.1 is continuous with respect to the order \( k \), fractional moments will form a curve in the complex plane.

**Figure 3.1:** Momentum curves. **A:** The null distribution for the fractional moments between 0.0 and 5.0, \( T = 1,000 \) independent samples taken from the normal distribution, 100 instances. Different colors stand for different moment orders. The shaded errorbar denotes standard deviations of the sample of 100 fractional moments of a given order. **B:** An exemplary signal: trains of on- and off-states mixed with a pure Gaussian noise and normalized, \( N = 1,000 \) independent samples. **C:** Fractional moments for 100 realizations of such a sample (red), \( SNR = 1.0 \), compared to 100 realizations of white noise (blue). The distribution of fractional moments for the signal is shifted with respect to the null along the moment order dimension.
In Fig. 3.1A, I present 100 instances of the whole set of fractional moments between 0.0 and 5.0 for \( N = 1,000 \) independent samples from normal distribution with the mean of 0.0 and variance of 1.0. Different colors stand for different moment orders. The shaded errorbar denotes standard deviations in the sample of 100 fractional moments of a given order. The Momentum curve starts from \((1, 0)\) for \( k = 0.0 \), traverses the upper half-plane and arrives at \((0, 0)\) for \( k = 1.0 \). It travels back through the lower half-plane towards \((1, 0)\) for \( k = 2.0 \) as the variance was fixed to 1.0 through the normalization. Every time \( k \) becomes an integer, the curve crosses the real axis. It is easy to notice that (1) for integer moments, the scatter goes along with the x-axis (as all the moments become real numbers); (2) for fractional moments, the scatter becomes more complex. For fractional moments lower than \( k = 3.0 \), the scatter of the data points is almost one-dimensional, and aligned along an axis which is not parallel towards neither real nor imaginary axis. The angle of this axis changes smoothly between fractional moments. For higher moments (\( k > 3.0 \)), the scatter becomes two dimensional.

In Fig. 3.1B, I present a mixture of a signal with trains of binary states \((\pm s)\), with switches between high and low states governed by Poissonian process, with white noise (since only distributions are analyzed, the spectral characteristics of the noise do not matter, hence the white noise). In Fig. 3.1C, I compare 100 instances of white noise (blue) with 100 instances of such a mixed signal (red). The Gaussian white noise has unique characteristic curve (green) that can be derived analytically, therefore, the deviation from this normal characteristic can serve as a measure for normality (red).

**Figure 3.2:** The null distributions (blue scatter) against the distribution of signal mixed with the noise (red scatter), for one fractional moment \( k = 4.5 \). In this case, the signal takes a form of trains of on- and off- states occurring with equal probabilities. Green dot - the analytic solution. Red dot - the mean of the distribution of the data-points for the signal mixed with the noise. \( S \) - SNR. The distributions of moments for different samples drawn from the Gaussian distribution (blue scatter) are two-dimensional Gaussians while distributions of moments for different samples drawn from the Gaussian mixed with signal (red scatter), are highly non-Gaussian and yield a crescent shape.
3.2. Momentum, a robust test for normality: Methods

A zoom into onto one particular fractional moment in a function of SNR is presented in Fig. 3.2 and 3.3.

3.2.2 Deriving confidence intervals

Particular fractional moments will differentiate the signal from the noise better of worse, depending on the type of the signal. This can easily be demonstrated. Let us sample signals of fixed length (of \( N = 10,000 \) samples) of the following classes:

1. A slow sine wave with an amplitude \( s \) (Fig. 3.4A),
2. Trains of binary states \((\pm s)\), with switches between high and low states governed by a Poissonian process (Fig. 3.4B,C,D),
3. Rare events with an amplitude \( s \) (Fig. 3.4E).

Each signal is demeaned before adding the white noise. After adding the signal to the noise, I further normalize the time series.

**Figure 3.3:** The histogram of distances to the analytic solution, for one fractional moment \( k = 4.5 \). Blue: sample derived from white noise (the null). Red: sample derived from white noise mixed with the signal. \( S \cdot SNR \).

**Figure 3.4:** Signals used in the simulation. A: A slow sine wave. B: On- and off-states, \( p(on) = p(off) \). C: On- and off-states, \( p(on) = 2 \cdot p(off) \). D: On- and off-states, \( p(on) = 0.5 \cdot p(off) \). E: Rare events.
In Fig. 3.5, 3.6, 3.7, 3.8 and 3.9, I present mean percentiles over 1,000 instances of the mixture of signal with noise, for fractional moments between 0.1 and 5.0. If the fractional moment does not carry information, its mean percentile is the same as in pure noise (and equal to 50%). The higher the mean percentile (the red line), the more information the given moment carries for this particular type of signal. E.g., the moment \( k = 4.0 \) can be very to little informative depending on the type of the signal in the sample, e.g., it is the most informative for sinus (Fig. 3.5), and the least informative for binary trains and rare events.

**Figure 3.5:** Moments between 0.0 and 5.0 for sinusoidal signals, Fig. 3.4A. For the sinusoidal signals, moments around \( k = 3.0 \) and \( k = 5.0 \) are the least informative as they are the least discriminative between the samples derived from white noise, and mixtures of white noise with sinusoidal signals.

**Figure 3.6:** Moments between 0.0 and 5.0 for binary trains, \( p(\text{on})=p(\text{off}) \), Fig. 3.4B. Similarly as for the sinusoidal signals, moments around \( k = 3.0 \) and \( k = 5.0 \) are the least discriminative as they are the least informative between the samples derived from white noise, and mixtures of white noise with binary signals. In general, the discriminative power is also lower than for the oscillatory signals of the same SNR (Fig. 3.5).
3.2. Momentum, a robust test for normality: Methods

Figure 3.7: Moments between 0.0 and 5.0 for binary trains, p(on)=2*p(off), Fig. 3.4C. In general, discriminability is better than for the binary trains, p(on)=p(off) (Fig. 3.6).

Figure 3.8: Moments between 0.0 and 5.0 for binary trains, p(on)=0.5*p(off), Fig. 3.4D. In general, discriminability is better than for the binary trains, p(on)=p(off) (Fig. 3.6).

For this reason, the confidence intervals should be derived for each fractional moment separately. As mentioned before, the variability in fractional moments is two-dimensional. Therefore, to derive the confidence intervals, I consider the distribution of modules of moments of order $k$ for a time series of length $N$ representing the white noise (thus, derived from a normal distribution):

$$M_k = \left| \frac{1}{N} \sum_{i=1}^{N} x_i^k \right| = \frac{1}{N} \sum_{i=1}^{N} |x_i^k|$$

(3.2)
Chapter 3. Momentum: a robust test for normality

Figure 3.9: Moments between 0.0 and 5.0 for rare events, Fig. 3.4E. In general, discriminability is weaker than for the other types of signals.

Note that when \( x_i \) is negative, \( M_k \) becomes a complex number. Let us first consider a random variable \( x \) representing a single sample. Since this variable is sampled from a normal distribution,

\[
\rho_0(x) = \sqrt{\frac{2}{\pi}} e^{-x^2/2}
\]

where \( x \in [0, \infty] \). Then, for a variable \( z = x^k \), I have a chi-square distribution:

\[
\mu_k = \langle x^k \rangle = \frac{2^{k/2}}{\sqrt{\pi}} \Gamma\left(\frac{1 + k}{2}\right)
\]

\[
\sigma_k^2 = \langle x^{2k} \rangle - \langle x^k \rangle^2 = \frac{2^k}{\sqrt{\pi}} \Gamma\left(\frac{1 + 2k}{2}\right) - \frac{2^k}{\sqrt{\pi}} \Gamma\left(\frac{1 + k}{2}\right)^2
\]

where \( \Gamma \) - the Euler function. If \( x \) is scaled to \( y = sx \), then \( \mu_s = s\mu_k \) and \( \sigma_s^2 = s^2\sigma_k^2 \). Now, if have a series of \( N \) independent samples taken from a normal distribution \( \rho_0(x), z = x_1 + x_2 + ... + x_N \), then \( \mu_z = \mu_1 + ... + \mu_N \) and \( \sigma_z^2 = \sigma_1^2 + ... + \sigma_N^2 \), as the sum, the mean and the variance are additive. Therefore, finally, \( \mu_N = N\mu_s = Ns\mu_k, \sigma_N^2 = Ns^2\sigma_k^2 \). Let us define a variable \( \xi = z/N \). Then, the result is as follows:

\[
\mu_\xi = s\mu_k
\]

\[
\sigma_\xi^2 = \frac{s^2\sigma_k^2}{N}
\]

These two quantities, give us the expected mean and variance of the modules of order \( k \) in a Gaussian sample with the mean of 0, variance of \( s \), and length of \( N \). Knowing these values, we can test the hypothesis that a given sample comes from a Gaussian distribution based on the value of its moment \( k \), by demeaning the sample, computing its length and variance, and computing the p-value for this moment under the assumption that the sample comes from a Gaussian distribution.
In fact, the aforementioned solution is not exact. However, from the Central Limit Theorem, the sum of samples from any distribution approaches the Gaussian distribution. Therefore for any distributions of $N > 10$ samples, this approximation should be satisfactory.

### 3.2.3 Correction for multiple comparisons

![Diagram](image)

**Figure 3.10:** Momentum as a signal detection technique. 

A: A stochastic time series is compressed into a distribution of its values. 

B: Since the spectral properties of the signal are neglected, samples of the same distribution are lumped into the same classes, e.g., two series of on- and off-states with an addition of noise will have the same distribution regardless of the total number of on- and off-states. 

C: Complex moments form a curve in the complex plane. The normal distribution has one characteristic curve (green). The stochastic time series deviates from the normality (red). The deviation from normality can serve as a measure of the amount of a signal in the time series (arrows).

In Momentum, multiple moments of the distribution are computed at a time. However, moments of the distribution are *not* independent from each other as they always form a smooth curve in the complex plane. As such, they are positively correlated. Therefore, there is certain redundancy between fractional moments, which raises the question of how to optimally correct for multiple comparisons. The state-of-the-art method for correcting for multiple comparisons for positively correlated tests, is the False Discovery Rate
Chapter 3. Momentum: a robust test for normality

(FDR) correction with the use of the Benjamini-Hochberg procedure Benjamini and Hochberg, 1995).

3.2.4 Momentum as a signal detection method

As mentioned in the Introduction, in datasets with poor temporal characteristics and low SNR, normality tests can serve as a method for approximating SNR in the data 3.10. In such cases, the time series is first normalized and lumped into a distribution (Fig. 3.10A). In a way, it is form of a data compression as this operation breaks the temporal order of the samples and causes that the temporal dynamics is lost. Any signals that can be represented by the same distributions, are lumped into the same categories, e.g., the two series of on- and off-states with an addition of noise will have the same distribution, regardless of the frequency of the switches between on- and off-states (Fig. 3.10B). The deviation from normality can then serve as a measure of the amount of the signal in the time series (Fig. 3.10C).

When using Momentum as a tool for signal detection, it is convenient to additionally convert the outcome p-values into effect sizes \( d \) with the use of the following transform:

\[
d = -\log_{10}(p)
\]

3.2.5 The full Momentum procedure

The summary of how to apply Momentum to any dataset, can be found in the following box:

Momentum procedure step by step:

1. Demean the sample and compute its length \( N \) and variance \( s \),

2. Specify the desired set of fractional moment orders to describe the distribution in the sample (e.g., from 0.1 to 5.0, with intervals of 0.1, and with the omission of \( k = 1.0 \) and \( k = 2.0 \)). The optimal range depends on the properties of the investigated data (expected SNR, length of the sample). In the relatively short and noisy datasets, moments higher than \( k = 5.0 \) will have a very high variance (see: Fig. 3.1A). Therefore moments lower than \( k = 5.0 \) are advised in most applications (and, in certain cases, even moments lower than \( k = 3.0 \)). Also, as mentioned before, two moment orders should be skipped from this set: the mean \( (k = 1.0) \), since the mean in the sample was set to 0 in the previous step, moment \( k = 1.0 \) is not informative and should be skipped) and the variance \( (k = 2.0), \) since the variance of the sample is further used to derive norms for other moments),

3. Compute the set of fractional moments in your sample (Eq. 3.1),

4. For every moment order in the set separately, compute the norms, i.e., the mean and the standard deviation of the distribution of fractional moments for that order, under the assumption that the
sample of length \( N \) is derived from a Gaussian distribution with mean 0, and variance \( s \). Use Eq. 3.6 and 3.7 for this purpose,

5. Given the norms for every fractional moment derived in the previous step, compute a p-value for every fractional moment computed from the sample you are now testing for normality,

6. Perform FDR correction for multiple comparisons (e.g., Benjamini-Hochberg) on the set of obtained p-values, and find the minimal value in the set of corrected p-values,

7. Optionally, you can also convert the p-value to an effect size (Eq. 3.8).

3.2.6 Comparing Momentum with other tests for normality

Previous studies comparing the power between standard tests for normality clearly demonstrate that some tests have higher power than others. In particular, the study by Yap and Sim, 2010 demonstrated that in typical situations, the following methods have the highest power among 8 tested methods:

1. Shapiro-Wilk test (SW, Shapiro and Wilk, 1965; Pearson and Hartley, 1972; Ghasemi and Zahediasl, 2012), which uses the following statistic:

   \[
   SW = \frac{\left(\sum_{i=1}^{N} \alpha_i \tilde{x}_i \right)^2}{\sum_{i=1}^{N} (x_i - \mu)^2} \tag{3.9}
   \]

   where \( N \) - length of the sample, \( \mu \) - the sample mean, \( \tilde{x}_i \) - i-th lowest value in the sample, and

   \[
   (\alpha_1, \ldots, \alpha_N) = \frac{m^T C^{-1}}{(m^T C^{-1} C^{-1} m)^{1/2}} \tag{3.10}
   \]

   where \( m = [m_1, \ldots, m_N] \), and \( m_i \) are expected values of independent and identically distributed random variables sampled from the normal distribution, \( C \) - covariance matrix of this order statistic.

   SW is a very popular method for testing normality. Although SW has a reputation of the most sensitive available test, it has its limitations. Most importantly, this test is biased by the sample size: the larger the sample, the more likely the result will be statistically significant. Also, SW does not work well in samples that contain multiple identical values.

   In this study, I used the open source scipy Python implementation of the SW test, scipy.stats.shapiro.

2. Jarque-Bera test (JB, Jarque and Bera, 1980; Jarque and Bera, 1987; Ghasemi and Zahediasl, 2012) based on a statistic built on skewness and kurtosis of the distribution. JB uses the following statistic:

   \[
   JB = \frac{N}{6} \left( S^2 + \frac{(K - 3)^2}{4} \right) \tag{3.11}
   \]

   where \( S \) - sample skewness, \( K \) - sample kurtosis.
Chapter 3. Momentum: a robust test for normality

JB test has lower power for small samples as compared with other tests, and gives a poor performance when applied to distributions with short tails, especially if the shape of the distribution is bimodal (Thadewald and Buning, 2004). This test was reported to perform worse than SW in a study by Razali and Wah, 2011.

In this study, I used the open source scipy Python implementation of the JB test, scipy.stats.jarque_bera.

3. d’Agostino & Pearson’s test, also referred to as the "Omnibus" test (AP, D’Agostino, 1970; D’Agostino, 1971; D’Agostino and Pearson, 1973; D’Agostino, Belanger, and Jr, 1990). AP performs two tests for normality: "skewtest" for deviation of skewness from normality ($Z_1(S)$, measured in Z-scores) and "kurtosistest" for deviation of kurtosis from normality ($Z_2(K)$, measured in Z-scores). Then, AP creates the joint statistic:

$$AP = Z_1(S)^2 + Z_2(K)^2$$ (3.12)

where $S$ - sample skewness, $K$ - sample kurtosis. Expressions for $Z_1(S)$ and $Z_2(K)$ are complex and can be found in (D’Agostino, 1970).

In this study, I used the open source scipy Python implementation of the AP test, scipy.stats.normaltest.

On the other hand, I did not use the following tests for comparison with Momentum:

1. Kolmogorov-Smirnov test (Smirnov, 1948): it used to be a very popular test in the natural sciences over the past fifty years. However, its power is lower as compared to other available tests (Stephens, 1974),

2. Kullback-Leibler divergence (KL, Kullback and Leibler, 1951): as testing for normality is equivalent to collecting evidence that the given sample comes from a non-normal distribution, it involves computing certain measure of the difference between the given distribution and a normal distribution. One alternative approach to this, is converting the distributions to probability density functions, and using KL to quantify the distance between the two distributions. KL has numerous limitations, e.g., it gives asymmetric measure of the distance (i.e., KL divergence between distributions A and B is not equal to the KL divergence between distributions B and A). When applied numerically to discrete datasets, it also has the binning problem: KL divergence is effectively computed based on the histogram of the data, and it can give different results depending on the choice of the binning strategy,

3. Graphical methods (e.g., quantile-quantile plots, QQ, Ehret, 2011). The graphical methods typically have a narrow range of applications (primarily, genome-wide studies in case of QQ),

4. Anderson-Darling test (AD, Anderson and Darling, 1952; Anderson and Darling, 1954) which uses the following statistic:

$$AD = -N - \sum_{i=1}^{N} \frac{2i - 1}{N} [\ln F(x_i) + \ln(1 - F(x_{N+1-i}))]$$ (3.13)
3.2. Momentum, a robust test for normality: Methods

where $N$ - length of the sample, $x_i$ - samples in ascending order, $F(x)$ - cumulative distribution function of the normal distribution (this test can also work for comparing the sample distributions other than normal). There is a mixed evidence with respect to the power of the AD test. Just as JB, AD test was reported to perform worse than SW in the study by Razali and Wah, 2011. However, Stephens, 1974 found AD to be one of the best empirical distribution function statistics for detecting deviations from normality. In this study, I did not include Anderson-Darling test for a practical reason: the only available Python implementation of this test (scipy.stats.anderson) does not return the p-values (but rather, it returns a list of threshold statistic values for a number of five arbitrary confidence levels).

3.2.7 Synthetic datasets for comparison

I compare Momentum with other methods for assessing normality of the distribution using two classes of synthetic datasets:

1. Samples drawn from non-Gaussian distributions: uniform, chi-square (50 degrees of freedom), Gamma (scale 2.0, shape 2.0), geometric, Laplace (location=0.0, scale=1.0), Poisson ($\lambda = 5.0$), and power distribution (shape 5.0, Fig. 3.11), for $N = 10,000$ samples,

![Fig. 3.11: Normal distribution and other popular non-Gaussian distributions. All samples of length $N = 10,000$.](image)

2. Samples representing "biological signals," created by adding signals presented in Fig. 3.4 with varying magnitudes, to the white noise (Fig. 3.12) with mean of 0.0 and variance of 1.0.

The full pipeline has been posted in open access on GitHub, under the link https://github.com/cryptofan/Momentum. The pipeline is available in two versions: compatible with Python 2.7 and Python 3.6.
Chapter 3. Momentum: a robust test for normality

Figure 3.12: Exemplary samples created from binary signals and sigmoidal signals mixed with white noise \((N = 10,000)\). Even though the data contains a clear signal in both cases, histograms of these samples resemble normal distributions to a much higher extent than histograms of samples of the same length drawn from other classic distributions (Fig. 3.11).

3.3 Results

Figure 3.13: P-values obtained by applying the standard tests for normality and Momentum to samples drawn from classic distributions, with 100 iterations for every simulation. The results are presented together with standard deviations computed from 100 iterations. Momentum is the only test that perfectly handles small data samples for the full set of classic distributions.

1. Comparison between methods on samples drawn from classic non-Gaussian distributions (Fig. 3.11), for various sample sizes \(N\): Fig. 3.13 and 3.14. Unlike other methods, Momentum can perfectly detect non-normality in every type of classic distribution for samples of \(N = 500\) and larger. For samples of size \(N = 100\) and less (represented by \(N = 50\)), samples drawn from Laplace distribution are the hardest to distinguish from the normal distribution with the use of Momentum - which is an intuitive result given that the Laplace distribution is closest in shape to the Gaussian distribution. Other methods for testing normality are less successful: they fail to detect non-normality for short samples drawn from any classic distribution, and fail at recognizing non-normality in samples drawn from certain distributions even for large sample sizes of \(N = 2,000\). This effect is even more
pronounced when looking at the effect sizes (Fig. 3.14) than when looking at p-values. For SW, recognizing non-normality in samples drawn from any distribution yields some issues. This is a surprising result given that, as mentioned before, SW is believed to be the most sensitive test for normality. For JB, chi-Square and Poisson distributions are particularly problematic. For AP, non-normality in all distributions but the Uniform and the Geometric distribution is difficult to detect.

Note: unlike other methods, Momentum does not return a mean value of \( p = 0.5 \) for samples of low SNRs. This is caused by the fact, that for a sample drawn from a pure normal distribution, single fractional moments return corrected p-values distributed around 0.5, and Momentum by default picks the lowest value in the set of p-values (corrected for multiple comparisons). Therefore, the expected \( p \) value at \( SNR = 0.0 \) is actually lower than 0.5. However, these p-values are still substantially higher than \( p > 0.05 \). Therefore, we do not observe any false positives in this study.

2. Method comparison for samples representing "biological signals," created by adding signals presented in Fig. 3.4 to white noise (Fig. 3.12) with varying SNRs: Fig. 3.15. Momentum is a clear leader in this comparison: it is 3 times more sensitive than the standard tests for normality. Among other methods, JB seems to perform best which is, again, surprising given that currently, SW is acknowledged as the most sensitive available test.
Figure 3.15: P-values and associated effect sizes for "biological signals" of varying SNRs, for $N = 10,000$ samples and 100 iterations for every simulation. The results are presented together with standard deviations computed from 100 iterations of the simulation. Momentum is able to detect non-normality for signals of circa 3 times lower SNRs than other tests for normality. Note: unlike other methods, Momentum does not return a mean value of $p = 0.5$ for samples of low SNRs. This is caused by the fact, that for a sample drawn from a pure normal distribution, single fractional moments return corrected p-values distributed around 0.5, and Momentum by default picks the lowest value in the set. Therefore, the expected $p$ value at $SNR = 0$ is lower than 0.5. However, these p-values are still substantially higher than $p > 0.05$. 
3.4 Discussion

3.4.1 Potential applications for fractional moments of the distribution

Mathematically, fractional moments of the distribution are well defined (Dremlin, 1994; Matsui and Pawlas, 2014). However, this concept was not applied in the natural sciences to date. One of the reasons is that for the normalized time series, the fractional moments become complex numbers. For this reason, the features characterized by the fractional moments cannot be visualized and conceptualized as easily as the features characterized by the integer moments (e.g., skewness can be interpreted as a measure of "asymmetry" of the distribution whereas kurtosis can be interpreted as its "flatness"). However, the range of potential applications of fractional moments is much broader than just assessing the normality of a distribution. A dense set of fractional moments can provide more comprehensive description of the distribution than a sparse set of central moments (i.e. mean, variance, skewness, and kurtosis). Such a dense set could serve, e.g., as an input feature set for Independent Component Analysis.

3.4.2 Normality tests for signal detection

To characterize signal in a stochastic time series, normality of distributions can be tested. Signal detection in stochastic time series is a complex research problem. In many applications in natural sciences, signal detection boils down to the analysis of spectral properties of the time series, e.g., through the Fourier analysis. However, this approach is often suboptimal, especially when it comes to signals of very poor SNRs and temporal dynamics such as in the resting state fMRI datasets (Ogawa et al., 1993; Boxerman et al., 1995; Friston et al., 1995; Friston et al., 2007) or in gene expression datasets (Maze et al., 2014). In this case, compressing a signal to a distribution and characterizing the nature of non-normality in this distribution, can serve as a method to detect signal in a noisy time series.

One particularly useful feature of normality tests for signal detection is that, since compressing a signal into a distribution breaks the temporal dynamics in the sample, normality tests are blind to the spectral characteristics (or, the color) of the background noise, i.e., will also detect signal in a presence of autocorrelated, pink noise. In certain applications, this can come across useful. E.g., it is debateable whether the background noise in the brain has white, pink, or red spectrum (He, 2014; Bédard, Kröger, and Destexhe, 2006; Dehghani et al., 2010). Characterizing the brain signals with the use of normality testing should give the same output irrespective of the answer to this question.

3.4.3 Momentum versus other methods for assessing normality of the distribution

In this work, I demonstrated that, provided a sample of mean equal to zero and a given length $N$, deriving expected value for a moment of certain order (and the confidence intervals around the expected value) can lead to a reliable estimate for whether the data is sampled from a normal distribution. When it comes to samples derived from distributions very far from
Gaussian (e.g., Poisson distribution), popular tests for normality such as SW or JB can properly detect non-normality only when the sample is long enough while Momentum is sensitive to non-normality even in samples as short of $N = 100$. Furthermore, in presence of biologically relevant signals created as a mixture of certain type of signal (e.g., a sinusoidal wave) with Gaussian noise, Momentum is far more sensitive in detecting non-Gaussianity than the available methods.

One remark is that, conceptually, Momentum has the affinity to d’Agostino & Pearson test (D’Agostino, 1970; D’Agostino, 1971; D’Agostino and Pearson, 1973; D’Agostino, Belanger, and Jr, 1990) as AP is based on comparing skewness and kurtosis of the distribution with the values expected under the assumption that the distribution was normal, and computing the weighted statistics from both these deviations (i.e., for skewness and kurtosis). The differences are two fold. Firstly, Momentum considers any arbitrary moment orders rather than solely the third and the fourth moment. Secondly, d’Agostino & Pearson test uses some parametric approximations to the variance of expected skewness and kurtosis (for more details, see: D’Agostino, 1970) while Momentum uses a very simple, nonparametric derivation based on the observation that for independent, standard normal random variables $x_i$, the sum $\frac{1}{N} \sum_{i=1}^{N} |x_i|^t$ tends to a chi-square distribution.

Momentum also solves the binning problem encountered while using some other methods for testing normality such as the Kullback-Leibler divergence (KL, Penny, 2001): since moment of any order is well defined for any length of the sample $N$ (as in Eq. 7.1), there is no need to arbitrarily bin the data to compute the test statistic.

### 3.5 Exemplary applications of Momentum in functional Magnetic Resonance Imaging

In this Chapter, I proposed Momentum as a method normality testing and signal detection. This method is sensitive, which is important especially when analyzing short samples (Fig. 3.14). Also, as discussed in Chapter 3, testing properties of the sample distributions is a good solution for analyzing datasets with poor temporal characteristics. Typically, fMRI data fulfill these conditions. Thus, Momentum has a potential as a “search engine” for new biomarkers of cognition and psychopathology based on fMRI datasets.

One important thing to realize is that, the method can be applied in two distinct ways. Firstly, as a signal detection method, it can be used to create maps of non-normality, or effect size, across the brain (either voxel-wise or ROI-wise). This can help, e.g., in analyzing resting state data and deriving new, interesting, spatially acute resting-state biomarkers (even at a single voxel resolution). This can also further lead to second level analysis and testing these local effects in the clinical cohorts.

A second possible way of applying Momentum is to consider sets of single fractional moments, and to analyze group differences along every moment separately (Fig. 3.16). In this case, second level analysis can also be performed (Fig. 3.14).
3.5. Exemplary applications of Momentum in fMRI

3.5.1 Resting state fMRI datasets

Resting state brain dynamics is a broadly studied subject in connectomics (Heuvel and Pol, 2010). In studies on the resting state dynamics, the focus is typically put on finding and interpreting patterns of co-activation between the nodes of the resting state networks (which is often referred to as functional connectivity, Greicius et al., 2003; Hansen et al., 2015; Chang and Glover, 2010). The consensus in the field is that resting state fluctuations are random or pseudo-random as they reflect cognitive processes that cannot be controlled or explained (Biswal, 2012). How to quantify intra subject differences in the local resting state dynamics though? Typically, resting state dynamics within one node is characterized in terms of entropy (Wang and Krystal, 2014) or its spectral properties (Mantini et al., 2007). Estimating
effect sizes with the use of Momentum can be an alternative to these approaches as it provides a measure of non-Gaussianity that is qualitatively different from the state-of-the-art methods. It is also an interesting research question whether or not the effect size profile across the brain in the resting state fluctuates in time. It might be an idea to create a voxel-wise map of effect size, and window the resting state data (e.g., the data from the Human Connectome Project as they are long as for the standards in the field, i.e., 2 x 2,400 samples) to research this intra-subject variability.

3.5.2 Task fMRI datasets

As mentioned before, fractional moments are hard to interpret as they describe properties of the shape of the distribution that - unlike mean, variance, skewness, or kurtosis - cannot easily be imagined. However, this does not mean that fractional moments of the distribution cannot be valid biomarkers of behavior.

The most popular approach to analyzing task-fMRI is the GLM (as described in Chapter 2) which boils down to comparing the mean BOLD signal (i.e., $k = 1.0$) between task and rest. There might be a lot of information also hidden in other moments of the BOLD distribution. Therefore, exploring what happens at single fractional moments across a wide scale of moment orders (e.g., between $k = 0.1$ and $k = 5.0$) might substantially increase the power of GLM. Furthermore, using Momentum as a holistic measure of non-normality and computing the map of effect sizes can also reveal maps of activation in the resting state which is impossible to achieve using the GLM only.

3.5.3 Longitudinal studies

Effect sizes from normality testing, as well as single fractional moments, can be used as features for almost any classification paradigm. Specifically, they can be used to measure long-term effects of a clinical therapy, or ageing effects on the brain. It is known that functional networks in the brain form during early development (Keunen, Counsell, and Benders, 2017), but how this compares to the profile of resting state activity across the brain?

3.5.4 Improving Independent Component Analysis

Lastly, as mentioned in Chapter 3, Independent Component Analysis (ICA) is an algorithm used to decompose a multivariate time series into a linear combination of sources with a non-Gaussian distribution (Hyvärinen, Karhunen, and Oja, 2001; Beckmann, 2012). ICA uses measures of non-normality to compare the activation patterns within voxels, and to group proximal voxels of similar distributions together. So far, most often, the third and the fourth moment are used to characterize the BOLD fMRI distributions in ICA. Enriching the ICA procedures by using a wider range of fractional moments allows for a richer characteristics of the local BOLD fMRI distributions, and lead to better performance of ICA in fMRI datasets.
Chapter 4

Thresholding functional connectomes by means of mixture modeling

Functional connectivity is one of the state-of-the-art concepts in studies on the large-scale functional architecture of the human brain. In network research in fMRI, functional connectivity is defined as a set of pairwise interactions between the nodes of the network. These interactions are typically operationalized through the full or partial correlations between all pairs of regional BOLD fMRI time series. Even for a pair of unrelated time series, correlations will always be nonzero, hence the question: which functional connections are significant? This question remains an open research problem. Typically, this thresholding problem is approached by proportional thresholding, or by means of parametric or non-parametric permutation testing across a cohort of subjects, for each functional connection separately. As an alternative, I propose a data-driven thresholding approach for network matrices based on mixture modeling. This approach allows for creating subject-specific sparse connectomes by modeling the full set of partial correlations as a mixture of weak connections associated with noise and weak (or, unreliable) connections, and a sparse set of reliable connections. Consequently, I propose to use alternative thresholding strategy based on the model fit using pseudo-False Discovery Rates based on the empirical null estimated as part of the mixture distribution. I evaluate the method on synthetic benchmark fMRI datasets where the underlying network structure is known, and how that it improves performance with respect to the alternative methods for thresholding connectomes (at the canonical thresholding levels). I also demonstrate that mixture modeling yields highly reproducible results when applied to the functional connectomes of the visual system derived from the n-back Working Memory task from the Human Connectome Project. The resulting sparse connectomes are further discussed in the light of the previous knowledge of the functional architecture of the visual system in humans. I also demonstrate that with the use of this method, I am able to extract similar information on the group level as can be achieved with permutation testing even though these two methods are not equivalent. I demonstrate that with both these methods, decoupling between the two hemispheres in the higher order areas of the visual cortex during visual stimulation as compared to the resting state is visible. This result is in line with previous studies suggesting lateralization in the visual processing. However, as opposed to permutation testing, this approach does not require inference at the cohort level and can be used for creating sparse connectomes at the single subject level. This is relevant for studies that involve small experimental groups.

Keywords: functional connectivity, partial correlation, sparse connectome, mixture modeling, False Discovery Rate
4.1 Introduction

Functional connectivity (FC) characterizes temporal correlations between signals in the nodes or regions-of-interest (ROIs) in the neuronal network. In fMRI, this concept is used in many contexts (Smith et al., 2013). FC serves to study the (co)activity in the neuronal networks, and to investigate links between activity in neuronal networks and cognitive abilities (Smith et al., 2015; Finn et al., 2015; Tavor et al., 2016; Smith, 2016; Chauvin et al., 2017; Chauvin et al., 2019b) or clinical-behavioural covariates (Lynall et al., 2010; Garrity et al., 2007; Greicius et al., 2007; Harrison et al., 2009; Rausch et al., 2016; Oldehinkel et al., 2016; Mulders et al., 2015). It is also used to gain insights into hierarchical structure in the brain in rest and cognition (Smith et al., 2015; Bola and Borchardt, 2016), e.g., hierarchical structures in the sensory systems (Arcaro et al., 2015; Merhar et al., 2016).

In fMRI research, functional connectivity is typically operationalized by means of partial correlation (Marrelec et al., 2006). Since any two processes - even in the absence of the underlying direct connection - will almost surely retrieve a non-zero partial correlation by chance, the partial correlation matrices should be constrained to remove unreliable connections \(^1\). As indicated in recent studies by Heuvel et al., 2017; Zalesky et al., 2016 and Ginestet et al., 2011, the choice of thresholding method can influence the amount of weak connections present in the connectomes, which, in turn, yields an effect on the structure and global properties of the sparsified networks. For this reason, the choice of thresholding method can highly influence the results (and its interpretation) in any functional connectivity study in fMRI.

There are a few leading approaches to the problem of sparsifying functional connectomes in the field. Firstly, a popular approach is proportional thresholding (Achard and Bullmore, 2007; Bassett and Bullmore, 2009; Heuvel et al., 2008). In this thresholding scheme, a predefined top percentage of all functional connections in a subject-specific connectome is selected. The main aim of this approach is to keep the number of connections fixed for all the individuals to eliminate the impact of network density on the comparison of graph metrics across groups. This method for sparsifying functional connectomes is currently the most popular approach in the field, which might be due to its simplicity.

Secondly, sparse functional connectivity can be estimated, e.g., with the use of regularizers (Bishop, 2006). Regularization techniques impose sparsity on the network by using a loss function that penalizes for the number of non-zero entries in the connectivity matrix so that the weak connections

---

\(^1\)Neurobiologically speaking, one may argue that the brain is highly integrative and therefore, the connections represented by low values of a partial correlation should arguably be referred to as weak - in the sense of indistinguishable from the noise.
are shrunk to zero. The shrinkage approach is used to drive weak connections to zero and then to accept every connection that has not been set to zero as a real connection. Originally, the goal of the regularizing techniques wasn’t thresholding connectomes specifically, however, this became one of the practical applications.

Lastly, thresholding can be performed based on connection-specific significance levels obtained from permutation testing (Welch, 1990). Each of the functional connections can then be thresholded at the edge-specific threshold, according to the edge-specific null distribution. Technically, computing significance levels through permutation testing can be done both at the population- or the single subject level. On the single subject level, thresholds are estimated from the null-distribution of connections generated by breaking correlations between time series. On the other hand, to create the null distribution at the population level, region-specific time series are permuted between subjects, and functional connectivity is computed for a number of surrogate networks obtained from permutations across subjects. However, in datasets in which signals are autocorrelated - such as fMRI data where the slow hemodynamics induces autocorrelations in the signal - building the null by shuffling the labels between subjects is preferred over permuting samples on the single subject level, as it keeps the autocorrelations intact\(^2\). For this reason, the population level approach was also used in the seminal work by Smith et al., 2011. As a result, in practice, building significance intervals in permutation testing is only possible on the population (and not on the single subject) level in the fMRI functional connectivity research.

In this work, I propose an alternative to the aforementioned methods: thresholding subject-specific connectomes by means of mixture modeling. Only one team has previously applied this method to model sparse resting-state functional connections (Tyszka et al., 2014). Here, I provide a thorough investigation of this technique.

Mixture modeling differs from permutation testing as the (pseudo)-null is built subject-wise across all possible connections in the connectome as opposed to estimating null distributions connection-wise via permutation testing. The underlying assumptions here, are that (i) the evidence for a reliable connection is unrelated to the spatial location of the nodes, (ii) reliable connections are sparse, (iii) there is a sufficient number of nodes so that the set of values for non-existing edges in the network can be used to estimate the subject-specific null distribution of functional connections.

In this approach, mixture modeling is used to separate strong connections in the connectome from a \textit{pseudo-null}, which is a mixture of noise with weak (or, unreliable) functional connections. I talk about the pseudo-null as in fact, the functional connectomes in the brain most likely have a scale-free distribution of connections rather than being sparse (Eguiluz et al., 2005; Heuvel et al., 2008). Therefore, I can only talk about a \textquote{pseudo-null} which consists of the \textquote{true null} distribution of functionally disconnected pairs of nodes, and, in addition to that, a part of the scale-free distribution that involves connections too weak to be discerned from the noise with any statistical inference methods. As mainly the strong connections are of interest

\(^2\)I also demonstrate the effect of breaking autocorrelations on the results of permutation testing in Fig. 4.13.
in the functional connectivity studies, this model choice is a justifiable
simplification.

Mixture modeling is a valuable alternative as, on the contrary to permu-
tation testing, it allows for creating connectomes both at the single subject
and at the group level. Furthermore, mixture modeling solves the problem
related to subject-specific proportional thresholding. Proportional threshol-
ding allows for weak connections to pass the thresholding in some subjects
(e.g., in subjects whose individual connectomes have low number of strong
connections compared to other subjects). This, in turn, changes the global
properties of the sparse networks. In mixture modeling, this is not the case
as the total number of connections in the sparsified network is not fixed per
subject: strong and weak connections are determined based on the shape of
the subject-specific distribution of connections; mixture modeling provides
with a natural, data-driven separation into the two classes.

The mixture modeling approach used here is popular in other contexts
in fMRI research, especially as the basis for thresholding Independent Com-
ponent Analysis-derived maps (Beckmann and Smith, 2004; Beckmann et
al., 2005). It is also used in other applications such as Genome-wide associ-
ation studies (GWAS) in polygenic disorders (Thompson et al., 2015) where
a mixture model is fitted to the distributions of effect sizes for all SNPs.

In this study, I validate mixture-model-based thresholding on the bench-
mark synthetic fMRI datasets (Smith et al., 2011) derived from the Dynamic Causal Modeling generative model (Friston, Harrison, and Penny,
2003; Smith et al., 2011). Furthermore, I apply the new thresholding ap-
proach to experimental fMRI datasets from the Human Connectome Project
(HCP, Essen et al., 2013) by creating a sparse connectome of the human
visual system at rest and under visual stimulation (Barch et al., 2013). I
chose to study the human visual system because this network incorporates
one of best known functional architectures in the human brain, therefore,
it also allows for the qualitative comparison between the methods. I used
the n-back working memory (WM) task data from the Human Connectome
Project (Barch et al., 2013), because this task involves ongoing visual stim-
ulation by presenting objects in the visual field of the participants. Since
during the task, objects of a few categories were presented to the subjects,
my hypotheses concentrated on the connectivity between the areas respon-
sible for object recognition such as the two areas of the lateral occipital cor-
tex: LO1 and LO2 (Silson et al., 2013; Amedi et al., 2001; James et al., 2002).
Multiple studies have revealed that these areas respond to objects defined
by luminance, texture or motion, but do not respond when subjects only
view backgrounds of different textures or coherently moving dots (Grill-
Spector et al., 1998). Furthermore, object recognition in the visual system
is known as a laterialized process (Warrington and Taylor, 1978): the right
hemisphere is responsible for perceptual-, while the left hemisphere is re-
sponsible for the semantic categorization. Therefore, I test for asymmetry
in the responses to the visual stimuli between the left and the right hemi-
sphere, as compared to the resting state connectivity.

In Section 4.2.1, I introduce the operationalization of functional connect-
tivity chosen in this study. In Section 4.2.2, I introduce the mixture mod-
eling procedure. In Section 4.2.3, I list the steps undertaken to validate
the method on the synthetic benchmark datasets, and in Section 4.3.1, I
present quantitative results. In Section 4.2.3, I describe datasets and the preprocessing pipeline for computing thresholded functional connectomes, both in rest and under visual stimulation, and in Section 4.3.2, I present the results. Finally, in Section 7.4, I discuss the results in the light of the literature, and I propose other applications of the method.

4.2 Materials and methods

4.2.1 Operationalization of functional connectivity

Functional connectivity between two nodes $X, Y$ in the network is usually operationalized as either Pearson’s or partial correlation (Marrelec et al., 2006) between the time series representing activity in the nodes $X(t)$, $Y(t)$. In the fMRI functional connectivity research, partial correlation analysis is preferred as it reflects direct rather than both direct and indirect functional connections between the nodes (Smith et al., 2011). For each pair of nodes in the network, limiting influence of indirect connections is achieved by regressing out the activity from all the other nodes before computing Pearson’s correlations. For small networks, the regression step can trivially be performed by means of Ordinary Least Squares regression (OLS, Vittinghoff et al., 2005). For large networks, however, this may be problematic if the number of possible pairs in the network approaches the length of the BOLD time series, given that partialing out secondary time series involves loss in the temporal degrees of freedom. Therefore, in this work, I choose estimating partial correlation between two nodes in the network by means of the inverse covariance (a.k.a. precision matrix) normalized by the temporal autocorrelation in the two nodes (Christensen, 2011).

4.2.2 Mixture modeling

In this work, I model the distribution of pairwise partial correlation scores (transformed into pseudo-Z statistics using the Fisher $r$-to-$Z$ transform) as a mixture between the distribution of noise, and weak (or, unreliable) connections and the distribution of reliable connections. I exclude the main diagonal entries from the partial correlation matrices as they represent self-connections and further do not take inhibition-induced anti-correlations into account.

I model the pseudo-null either using Gaussian or Laplace distribution. In principle, as partial correlation relates to the second order statistic, both the null and the mixture representing reliable connections can be characterized by the Fisher-Snedecor $F$ distribution (in case of the null, it is an $F$-distribution mirrored around zero, Box, 1953). Ideally, it should be modeled as such. However, as this distribution does not belong to the exponential family, its parameters are hard to fit with the Expectation Maximization algorithm (EM, Bishop, 2006; Do and Batzoglou, 2008). For this reason, I chose to simplify to Gaussian or Laplace.

To model the distribution of reliable connections, I chose either Gamma or Inverse Gamma distribution. Again, this is a pragmatic choice as they both represent a distribution from the exponential family. Unlike simple
Gamma distribution, the Inverse Gamma distribution cannot deflate towards a flat distribution in the estimation process. In the rest of the Chapter, I use the following abbreviations: GG for the Gauss-Gamma mixture, GIG for Gauss-Inverse Gamma, LG for Laplace-Gamma mixture and LIG for Laplace-Inverse Gamma mixture. The details of the probability density and parameters fitted for each of the aforementioned distributions used in the study, are provided below:

1. Gaussian:
   \[
   p(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (4.1)
   \]
   where \( \mu \) - the mean of the distribution, \( \sigma \) - the standard deviation of the distribution.

2. Laplace:
   \[
   p(x|\mu, \theta) = \frac{1}{2\theta} e^{-\frac{|x-\mu|}{\theta}} \quad (4.2)
   \]
   where \( \mu \) - the mean of the distribution, \( \theta \) - the scale parameter.

3. Gamma:
   \[
   p(x|k, \theta) = \frac{\theta^{-k}}{\Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}} \quad (4.3)
   \]
   where \( k \) - the shape parameter, \( \theta \) - the scale parameter.

4. Inverse Gamma:
   \[
   p(x|k, \theta) = \frac{\theta^k}{\Gamma(k)} x^{-k-1} e^{-\frac{\theta}{x}} \quad (4.4)
   \]
   where \( k \) - the shape parameter, \( \theta \) - the scale parameter.

For all the competing models, I infer the relevant mixture parameters using the Expectation Maximization algorithm (Bishop, 2006; Do and Batzoglou, 2008; Llera et al., 2016) with the cut-off threshold set to 0.001. The initialization of the EM algorithm is performed in the following way. Firstly, the data is standardized before mixture model fitting. Consequently, the null distribution is initialized with mean \( \mu = 0.0 \) and standard deviation \( STD = 1.0 \). The activation distribution is initialized with mean \( \mu = 3.0 \) and \( STD = 1.0 \). Such an initialization ensures that the initial activation distribution models a negligible percentage of samples in the presence of no connection (Gaussian white noise) and an increasing percentage of samples with increased connectivity. Furthermore, the mixing proportions are initialised using a flat prior. Such an initialization provides a near optimal solution that allows me to avoid re-initialization.

After the model fitting, I propose to threshold the connectome matrices based on the pseudo-False Discovery Rate (pFDR, Efron, 2007). The False Positive Rate (FPR) is the number of falsely rejected null hypotheses among all elements belonging to the null distribution. In contrast to the FPR, FDR takes the overall estimated power of the signal into account, in addition to the estimated number of type 1 errors. By thresholding connectomes based on FDR, the focus shifts towards controlling the relative amount of positive identifications. This makes this estimate ideal for conservative estimations concerning the overall architecture of a network - since small influences in the network are more prone to be neglected in favor of strong connections.
In this work, FDR is used to determine the threshold, which is then used to sparsify the functional connectome. FDR is computed as the number of false positives with respect to the number of positively identified partial correlation values. FPR is estimated based on a model fit which represents the null distribution:

\[
FDR(x) = \frac{\int_{x}^{\infty} p_0 f_0(t) dt}{|\rho|} \text{ for } t \geq x
\]

(4.5)

where \( f_0 \) - the probability density function of the null distribution, \( p_0 \) - the size of the null distribution in relation to the whole mixture density, \(|\rho|\) - the overall number of partial correlation values and \(|\rho(t) \text{ for } t \geq x|\) - the number of partial correlation values higher than \( x \).

Visualization of this concept is presented in Fig. 4.1.

This FDR value can be referred to as a pseudo-False Discovery Rate, since it depends on the quality of the model fit. This is because FDR can return values higher than 1.0 in areas where the model overestimates the complete mixture distribution. This also means that the FDR function is not necessarily bijective in such cases, resulting in multiple thresholds with the same FDR value. Because of this effect, I choose the highest value passing the FDR cut-off as the threshold for sparsifying the functional connectivity matrix.

The thresholding is then based on the probability of assignment to the signal component from mixture modeling, as pFDR refers to the fitted null distribution. Mind that in this approach, the goodness of the fit to the null is what influences the estimate of the threshold at a given value of pFDR. Therefore, in this approach, the main purpose of fitting the mixture with a signal component is to improve the estimate of the null.
4.2.3 Validation

Synthetic benchmark datasets

I first evaluate my approach on synthetic connectome data generated as in Smith et al., 2011 using the standard Dynamic Causal Modeling (Friston, Harrison, and Penny, 2003) generative model for the BOLD fMRI responses. The details of the generative model are given in Chapter 6.2.1. These simulations allow for emulating network dynamics while controlling a variety of experimental conditions, such as the number of nodes in the network, the session duration, the time resolution of the data, the amount of thermal noise added to the BOLD response, or the variability in the hemodynamic lags. I chose to validate my approach by reimplementing simulation no 4 from (Smith et al., 2011) which represents the largest network in the benchmark datasets \((N = 50)\) nodes. I created a large population of 500 synthetic "subjects," and performed 10-min long simulation of the BOLD dynamics, with \(TR = 3.0[\text{s}]\), with addition of 1% thermal noise and with the variability in the hemodynamic lags of \(STD = 0.5[\text{s}]\).

Following the generation of the simulated BOLD datasets, I calculated the partial correlation network matrices for the synthetic subjects. On these matrices, I evaluated different mixture models, using either Gaussian or Laplacian distributions for the null and either Gamma or Inverse Gamma distributions for the non-null part of the distribution of partial correlation scores. The covariance matrix can be computed with or without regularization, although in case of regularization, the F-distribution is lost. This set of choices (Gaussian vs Laplacian for the pseudo-null/Gamma vs Inverse Gamma for the strong connections/regularization versus no regularization for the covariance matrix), gives 8 possible mixture models. I compared the 8 possible choices by means of Bayesian Information Criterion (BIC, Schwarz, 1978) to determine which mixture best fits the synthetic datasets. Since all the compared versions of mixture modeling use both the same number of data points and free parameters, this ranking would be identical using Akaike’s Information Criterion and other methods for evaluating a model fit (AIC, Akaike, 1998).

Furthermore, I compare the results between mixture modeling and the following methods for sparsifying functional connectomes:

1. Empirical precision (EP) with hard thresholding: the sample covariance matrix, inverted, normalized with autocorrelation to obtain partial correlation, and then sparsified. Since there is no optimization technique available for the sparsification step, I apply a hard threshold. What I mean by this, is setting a fixed value for partial correlation (of 0.0 in this case) at which the connectomes obtained through empirical precision are thresholded,

2. Ledoit-Wolf regularization (LW, Ledoit and Wolf, 2003) using hard thresholding: the sample covariance matrix is shrunk towards a fixed target matrix before inverting. Among the available options for the target matrices, I chose the constant correlation model, in which the shrinkage target is constructed from the average of all the sample correlations together with the vector of sample variances. This choice for
the target matrix is advisable for uniform family of variables\textsuperscript{3}, which is the case here as every time series in the data represents activity in a single ROI. In the implementation of LW regularization used in this work, there are no free parameters with respect to shrinkage (shrinkage parameter is automatically optimized due to a closed form approximation given in Ledoit and Wolf, 2003). Then, the sparse covariance matrix needs to be inverted and normalized with the use of autocorrelation. After this operation, the functional connectome is no longer sparse. Therefore, I apply a hard threshold of 0.0 to these matrices to achieve the sparsity. In this way, sparsification is performed twice: before and after inverting the covariance matrix.

3. Permutation testing: creating a sparse estimate of the partial correlation matrix by constructing a null distribution of connections estimated from time series with shuffled subject labels (confidence levels computed on the population level). The partial correlation matrix is then thresholded at a chosen significance level in the light of this null distribution, for each connection independently (so that FPR is controlled independently for each functional connection in the network),

4. Proportional thresholding: thresholding partial correlation subject-wise by selecting a top percentage of all the values in the partial correlation matrix. As mentioned in Introduction, this is a commonly used approach to thresholding functional connectomes in fMRI (Achard and Bullmore, 2007; Rubinov and Sporns, 2010; Bassett and Bullmore, 2009; Heuvel et al., 2008), recently broadly discussed in Heuvel et al., 2017. In my study, I chose two levels for proportional thresholding: top 5\% and top 10\%. In the synthetic networks used in my study, the true underlying density of connections equals 6.25\%. I decided not to fine-tune the chosen proportions to this particular value, because in real-world applications, this fine-tuning is also not possible, and choosing an arbitrary threshold is necessary.

The aforementioned methods belong to two separate categories. Mixture modeling and permutation testing are procedures fully based on estimating the null from the data, and therefore on the significance estimation. On the other hand, sparsifying connectomes through hard thresholding (obtained with or without regularization, as described in points 1 and 2), or through proportional thresholding as described in point 4, do not allow for controlling the significance. However, I decided to include all these methods in the study to create a comprehensive comparison between different thresholding approaches present in the fMRI literature.

In the analysis, I only include the upper-diagonal values in the connectomes not to duplicate the partial correlation values (which would lead to under-estimation of within-component variance). The goodness of the aforementioned methods is assessed by computing and comparing their mean performances. The mean performance is given by the percentage of correctly inferred entries of the binarized precision matrix as compared to the ground-truth binary graph adjacency matrix, averaged over all simulated subjects.

\textsuperscript{3}The original version of the Ledoit-Wolf regularization was designed for the stock market analysis, where different data coming from different instruments, such as stocks and bonds, is often fed into the same model (Ledoit and Wolf, 2003).
The respective mean performances of the considered methods are reported for the following parametrizations: hard-thresholding at the value of $0.0$, proportional thresholding through selecting the top $5\%$ and $10\%$ of all the connections for EP and LW, and at $p = 0.05$ for permutation testing and pFDR. To make the results from the mixture modeling procedure comparable with permutation testing, the significance levels for both these methods were derived on the group level.

Furthermore, I evaluate the methods on subnetworks of the original $50$-node network connectivity pattern. This is because while some of the methods - such as inverse covariance with a hard thresholding at $0.0$ - are not sensitive to the network size, other methods can be sensitive, and the performance can drop off along with the decreasing number of nodes in the network. Namely, in mixture modeling, network size can influence the estimation of significance as it affects the quality of the pseudo-null and signal component estimates, whereas in proportional thresholding setting a low value of proportion (such as top $5\%$) can result in empty connectivity matrices if the network is small. Therefore, I created smaller networks of sizes between $N = 5$ and $N = 49$ nodes iteratively, by removing one node at a time. In each subnetwork, I kept at least one pair of nodes that were originally connected in the initial $N = 50$, so that there is always at least one true connection in the network. In the process of shrinking the size of the network, the density of connection stays roughly on the same level.

**The human visual system**

For validation on fMRI datasets, I use the data from 207 unrelated subjects from the HCP500 cohort (Essen et al., 2013). The unrelated subjects were chosen to prevent bias in the estimates of the network matrix due to familiarity status of the participants.

The following two datasets are used in the analysis:

1. Resting state with eyes open. In each subject, the resting state HCP data involves $4 \times 15[\text{min}]$ runs: two scanning sessions on the two consecutive days when two sessions per day were recorded using different right-to-left (RL), and left-to-right (LR) phase-encoding directions,

2. N-back Working Memory (WM) task (Drobyshesky, Baumann, and Schneider, 2006; Barch et al., 2013). In this task, subjects’ working memory and cognitive control were tested in the n-back task paradigm in which different types of visual stimuli (faces, places, tools and body parts) were presented to the subjects in separate blocks. Each run contains 8 task blocks (of 10 trials, $2.5[\text{s}]$ each) and 4 fixation blocks ($15[\text{s}]$ each). Half of the blocks involved a 2-back task (“respond when the stimulus is the same as the one two trials before”), and the other half involved the 0-back task (a cue stimulus is presented at the start of each block, and the subject must respond to the presentation of this stimulus during the block). A $2.5[\text{s}]$-long indicates the block type (either 2-back or 0-back, plus the target cue in the latter case) at the start of the block. In each trial, the stimulus is presented for $2[\text{s}]$ and followed by $0.5[\text{s}]$ inter-stimulus interval (ITI). The n-back task was performed once per subject, and lasted for a total of $5[\text{min}]$,
which amounts to a total of 810 3D volumes (405 volumes per each phase-encoding direction).

The n-back task was chosen for comparison against resting state because this task requires sustained attention and provides an almost permanent stream of the visual stimuli (presented in blocks with ITIs of 0.5 [s], which are substantially shorter than the repetition time, TR, of 0.72 [s]). Since ITIs are very short in this particular experiment, I used the whole time series from the n-back task to derive the functional connectomes.

All the subjects were healthy individuals scanned on a 3-T Siemens Skyra scanner with 100 mTm$^{-1}$ gradient strength, $TR = 0.72$ [s], $TE = 33.1$ [ms], flip angle = 52°, $BW = 2290$ Hz/px, in-plane $FOV = 208 \times 180$ [mm], 72 slices and spatial resolution of 2 [mm], isotropic, in a multiband setup with an acceleration factor of 8, using a 32-channel head coil. Please see Ugurbil et al., 2013 for the further acquisition details. The preprocessing was performed using the HCP workbench (Marcus et al., 2013), FSL (Jenkinson et al., 2012), and Freesurfer (Fischl, Sereno, and Dale, 1999). The structural artifacts were removed using Independent Component Analysis and ICA-based X-noisefier (Salimi-Khorshidi et al., 2014; Griffanti et al., 2014). Then, spatial smoothing was applied using an unconstrained 3D Gaussian kernel of $FWHM = 3$ [mm].

The data was then parcellated into ROIs with the use of the volume-based probabilistic atlas of visual topography by Wang et al., 2015. I chose for using an atlas rather than for defining ROIs by a seed-based analysis (or, by any other functional methods) because the visual system is fine-grained and the estimation of ROIs obtained this way could be noisy. The atlas was created by employing retinotopic mapping experiments which resulted in 50 ROIs (25 per hemisphere) in total: 8 ventral-temporal (V1v, V2v, V3v, hV4, VO1, VO2, PHC1, and PCH2), 9 dorsal-lateral (V1d, V2d, V3d, V3A, V3B, LO1, LO2, TO1, and TO2), 7 parietal (IPS0, IPS1, IPS2, IPS3, IPS4, IPS5, and SPL1), and one frontal (hFEF) region. I mapped the probabilistic assignments into ROIs on one 3D map of maximum probability assignments for all the voxels.

Since the original atlas has twice higher spatial resolution than the HCP data, this 3D map was subsampled to $91 \times 109 \times 91$ voxels. As known from previous computational studies, mixing signals within ROIs is very detrimental to the connectivity research in fMRI (Smith et al., 2011; Bielczyk et al., 2017b). Therefore in case voxels within a given block of $2 \times 2 \times 2$ voxels belonged to two or more separate ROIs from Wang’s atlas, I did not assign any label to that block in the downsampled atlas.

To prevent mixing signals between ROIs, I classified each set of $2 \times 2 \times 2$ voxels into a new ROI, but only when all the 8 voxels belonged to the same ROI in the original Wang’s atlas. After this scaling, certain voxels at the boundaries between ROIs did not receive new labels and therefore, a few small regions disappeared from further analysis. The following regions were thus discarded: MST(lH, rH), hMT(lH, rH), IPS3(lH, rH), IPS4(lH, rH), IPS5(lH, rH), SPL1(lH, rH), FEF(lH, rH). The remaining 36 ROIs were included in the analysis.

Then, I prepared the data for the analysis in two different ways, for two different purposes:
1. For comparison between sparse connectomes in the resting state and under visual stimulation: I removed the initial 5 frames from the data for each version of the encoding to prevent scanning-related artifacts. Then, for the n-back WM task datasets, the BOLD time series from each version of the encoding was normalized, and the two time series were merged into a vector of 800 samples per subject. For the resting state datasets the procedure was repeated, except that the BOLD time series was additionally shortened so that it matched the length of the task datasets (800 samples per subject in total), so that the differences in the distribution of partial correlation values between task and rest are not caused by the differences in precision of the estimated partial correlations (which depends on the length of the time series).

2. For the test-retest comparison, the data was rearranged: the LR encoding from the first day was concatenated with RL encoding from the second day, and vice versa.

Subsequently, functional connectivity was estimated, and mixture modeling was performed in the same fashion as in the synthetic datasets (Section 4.2.2): by fitting 8 possible versions of the mixture model, and comparing against each other by means of BIC.

Finally, for the test-retest comparison, I used the Index of Overlap to compare the results from day 1 against day 2, both in the resting state and in the n-back WM task. Furthermore, to establish the level of inter-subject variability between sparse connectomes obtained from each method, I computed the intraclass correlation coefficient (ICC, Shrout and Fleiss, 1979; Herting et al., 2017; Vetter et al., 2017) as a standard tool in the functional connectivity research (Fiecas et al., 2013) for the resting state and for the task datasets from day 1.

### 4.3 Results

#### 4.3.1 Synthetic benchmark datasets

Among 8 possible versions of the mixture modeling, Gauss-Gamma mixture based on the inverse covariance computed with LW regularization, which I further refer to as MM(LW,GG), achieved the lowest BIC score (BIC=3282.2). The fit of this winning mixture to the synthetic datasets is presented in Fig. 4.2A. According to the BIC scores, the runner-up model is the Gauss-Inverse Gamma mixture based on the inverse covariance computed with LW regularization (hereon called MM(LW,IG), BIC=3308.6). This means that the LW regularization is consistently the preferred shrinkage method (as opposed to Lasso).

I present the detailed results of this comparison in Fig. 4.3. In Fig. 4.3, I compare the fits on the synthetic datasets between the new method based on mixture modeling and other, established methods introduced in Section 4.2.3.
4.3. Thresholding functional connectomes: Results

Figure 4.2: Validation on synthetic benchmark datasets. A: The best mixture modeling fit: Gauss-Gamma mixture computed on partial correlation estimated with LW regularization. Red: the pseudo-null. Green: the component representing strong connections. B: The mean performance across 500 synthetic subjects, for the canonical parameters (thresholding at 0 for empirical precision and LW, $p = 0.05$ for permutation testing, $FDR = 0.05$ for mixture modeling). To make the results from mixture modeling comparable with permutation testing, the significance levels for both these methods were derived on the group level. Mean performance: mean ratio of correctly identified links (connection/lack of connection). TPR - true positive rate. FPR - false positive rate. Mixture modeling gives the best trade-off between TPR and FPR, therefore the overall performance is the highest. C: The estimated FDR in a function of the true FDR. The FDR estimation with the use of mixture modeling is conservative. D: The ROC curve for the synthetic dataset, comparing mixture modeling MM(LW,GG) with all other methods. The FPR range was clipped to $[0, 0.2]$ because the values outside this range are, typically, not of interest in neuroimaging studies. Circles - the canonical values of the thresholding parameter ($p = 0.05$ for permutation testing, $FDR = 0.05$ for mixture modeling, 5% and 10% cut-off for proportional thresholding). Mixture modeling gives the highest AUC, and hard-thresholding partial correlation obtained from LW-regularized covariance matrix is the runner-up.

1. Methods based on empirical precision matrices with LW-regularization perform better than methods based on empirical precision matrices without LW-regularization (mean $BIC = 3338.5/3451.1$, respectively). The difference is significant ($p < \epsilon$ in the ranksum test where $\epsilon$ - machine precision),

2. Methods fitting Gaussian to pseudo-null perform better than methods fitting Laplace distribution (mean $BIC = 3352.5/3436.5$, respectively). The difference is significant ($p < \epsilon$ in the ranksum test),

3. Methods fitting Gamma distribution to model the distribution of connections perform better than methods fitting Inverse Gamma distribution to model the distribution of connections (mean $BIC = 3387.3/$
3396.1, respectively). The difference is significant ($p = 2.4 \times 10^{-6}$ in the ranksum test).

As a result, MM(LW,GG) is the winner mixture with the lowest BIC score in the synthetic dataset.

![Comparison of BIC for all the 8 mixture modeling methods on the synthetic dataset. A: All 8 versions. B: Methods based on LW-regularized precision matrices perform better than mixture modeling methods based on empirical precision matrices (BIC=3338.5, BIC=3451.1 respectively). C: Methods fitting a Gaussian to the pseudo-null perform better than methods fitting Laplace distribution (BIC=3352.5, BIC=3436.5 respectively). D: Methods fitting Gamma distribution to model the distribution of true connections perform better than mixture modeling methods fitting Inverse Gamma distribution (BIC=3387.3, BIC=3396.1 respectively).](image)

In Fig. 4.2B, I present the comparison between mixture modeling and other methods for thresholding connectomes given the canonical thresholds (0 for empirical precision and for LW, $p = 0.05$ for permutation testing, $p_{FDR} = 0.05$ for mixture modeling, 5% and 10% cut-off for proportional thresholding). Mixture modeling gives the best trade-off between True- and False Positive Rate, therefore the overall performance is the highest.

In Fig. 4.2C, I present the estimated FDR in a function of the true FDR. As I can observe, the FDR estimation with the use of mixture modeling is conservative. In Fig. 4.2D, I compare the ROC curve between mixture modeling and other methods. In mixture modeling, I vary the value of thresholding value of pseudo-FDR. In permutation testing, I vary the thresholding p-value. In empirical precision and LW-regularized empirical precision, I vary the value of the hard threshold. In proportional thresholding, I
4.3. Thresholding functional connectomes: Results

vary the proportion of filtered connections per subject. Among all the compared methods, AUC for the mixture modeling is the highest. The results for the canonical values of the thresholding parameter ($p = 0.05$ for permutation testing, $FDR = 0.05$ for mixture modeling, 5% and 10% cut-off for proportional thresholding) are presented in circles.

In Fig. 4.4, I present the analysis of the influence of the network size on the performance, for network sizes varying between 5 and 50 nodes. The results follow the expectations: for mixture modeling and proportional thresholding at low percentage of 5%, the results fall off along with decreasing network sizes. The performance of Mixture Modeling is stable across networks bigger than $N = 15$ nodes, and slowly decreases for network sizes between $N = 10$ and $N = 15$ nodes. Its overall performance is also minimally higher than all the other methods, including proportional thresholding partial correlation at 5%, which is close to the true network density fluctuating around 6.25%. LW and empirical precision thresholded at zero give significantly worse results than the rest of the methods, for similar reasons as when evaluated on the original network (Fig. 4.2B): they are not conservative enough and therefore, give a very high False Positive Rate.

Figure 4.4: The evaluation on synthetic datasets, for network sizes between 5 and 50 nodes. The performance of Mixture Modeling is stable across networks bigger than $N = 15$ nodes, and decreases for network sizes between $N = 10$ and $N = 15$ nodes. Its overall performance for large networks also minimally higher than all the other methods, including proportional thresholding at 5%, which is close to the true network density (6.25%).

4.3.2 The human visual system

In both task and rest, the mixture model with the lowest BIC was Gauss-Gamma mixture based on the inverse covariance computed with LW regularization, MM(LW,GG) (BIC=1515.3, BIC=1663.1 for rest and task, respectively). In the resting state datasets, the runner-up method was Gauss-Inverse Gamma mixture based on the inverse covariance computed with LW regularization, MM(LW,IG) for the resting state data with (BIC=1536.2). In the task data, the runner up was Gauss-Gamma mixture based on unregularized inverse covariance (BIC=1668.9). In Fig. 4.5, I compare the fits via BIC on the resting state HCP datasets. I obtain the following set of results:

1. Methods based on empirical precision matrices with LW-regularization perform better than methods based on empirical precision matrices
without LW-regularization (mean $BIC = 1538.0/1545.6$, respectively). The difference is not significant ($p = 0.0649$ in the ranksum test).

2. Methods fitting Gaussian to pseudo-null perform better than methods fitting a Laplace distribution (mean $BIC = 1529.9/1553.7$, respectively). The difference is significant ($p = 1.5 \times 10^{-8}$ in the ranksum test).

3. Methods fitting Gamma distribution to model the distribution of connections perform better than methods fitting Inverse Gamma distribution to model the distribution of connections (mean $BIC = 1533.8/1549.8$, respectively). The difference is significant ($p = 1.2 \times 10^{-4}$ in the ranksum test).

![Figure 4.5: Comparison of BIC for all the 8 mixture modeling methods on the resting state HCP dataset. A: All 8 versions. B: Methods based on LW-regularized precision matrices perform better than mixture modeling methods based on empirical precision matrices (BIC=1529.9, BIC=1553.7 respectively). C: Methods fitting a Gaussian to the pseudo-null perform better than methods fitting a Laplace distribution (BIC=1529.9, BIC=1553.7 respectively). D: Methods fitting Gamma distribution to model the distribution of true connections perform better than mixture modeling methods fitting Inverse Gamma distribution (BIC=1533.8, BIC=1549.8 respectively).]

As a result, MM(LW,GG) mixture is also the mixture with the lowest BIC score in the resting state HCP dataset. In Fig. 4.6, I compare the fits via BIC on the HCP datasets from the n-back WM task. I obtain the following set of results:
4.3. Thresholding functional connectomes: Results

Figure 4.6: Comparison of BIC for all the 8 mixture modeling methods in the HCP datasets from the n-back Working Memory task. A: All 8 versions. B: Methods based on LW-regularized precision matrices perform better than mixture modeling methods based on empirical precision matrices (BIC=1669.9, BIC=1676.1 respectively). C: Methods fitting a Gaussian to the pseudo-null perform better than methods fitting a Laplace distribution (BIC=1660.0, BIC=1686.2 respectively). D: Methods fitting Gamma distribution to model the distribution of true connections perform better than mixture modeling methods fitting Inverse Gamma distribution (BIC=1667.7, BIC=1678.3 respectively).

1. Methods based on empirical precision matrices with LW-regularization perform better than methods based on empirical precision matrices without LW-regularization (mean $BIC = 1669.9/1676.1$, respectively). The difference is not significant ($p = 0.0985$ in the ranksum test).

2. Methods fitting Gaussian to pseudo-null perform better than methods fitting a Laplace distribution (mean $BIC = 1660.0/1686.2$, respectively). The difference is significant ($p = 8.1 \times 10^{-11}$ in the ranksum test).

3. Methods fitting Gamma distribution to model the distribution of connections perform better than methods fitting Inverse Gamma distribution to model the distribution of connections (mean $BIC = 1667.7/1678.3$, respectively). The difference is significant ($p = 0.0059$ in the ranksum test).
Chapter 4. Thresholding functional connectomes

**Figure 4.7:** The best model fit: Gauss-Gamma based on the inverse of the LW-regularized covariance, MM(LW,GG). Red: the pseudo-null. Green: the component representing connections. Tails of the distribution were truncated to $[-4.0, +6.0]$.

**Figure 4.8:** The "group sparse connectome" for the conventional methods with canonical thresholds: a summary of the number of subjects for whom the given connection survived the thresholding. **A:** The resting state. **B:** The n-back WM task.
As a result, MM(LW,GG) mixture is also the mixture with the lowest BIC score in the n-back WM HCP dataset. Fig. 4.7 presents the best mixture fit to the normalized values of the partial correlation values from 207 unrelated subjects from the HCP datasets (Essen et al., 2013), in the resting state and under the visual stimulation.

The mean numbers of connections found per subject, in both resting state and in the n-back WM task, are presented in Table 4.1. As mixture modeling and permutation testing thresholded at the canonical thresholds are more conservative than other methods, the group connectome for these two methods have less overall number of connections per subject than in case of the other two methods. Moreover, interestingly, all the methods report a decrease in the mean number of connections per subject in the task versus rest.

**Figure 4.9:** The "group sparse connectome" for permutation testing and for the mixture modeling methods: a summary of the number of subjects for whom the given connection survived the thresholding. **A:** The resting state. **B:** The n-back WM task. Both permutation testing and mixture modeling are more conservative and give sparser group connectome than empirical precision and LW thresholded at 0 (Fig. 4.8).
In Table 4.2, the Index of Overlap in the resting state and in the n-back WM task (day 1 versus day 2) are presented. Mixture modeling is the most conservative method, therefore the associated Index of Overlap scores are the highest (close to 100). In Table 4.3, the ICC scores for similarity between subject-specific sparse connectomes (computed for scanning on day 1 only) in the resting state and in the n-back WM task are presented. Again, the most conservative methods give the highest ICC scores. Moreover, ICC scores over 0.50 are at the upper end of results reported in the literature (Shehzad et al., 2009).

Figs 4.8, 4.9 and 4.10 present group statistic for the WM task and the resting state (session 1, time series shortened to 800 samples in each subject as described in Section 4.2.3).

**Figure 4.10:** The "group sparse connectome" for proportional thresholding methods: a summary of the number of subjects for whom the given connection survived the thresholding. A: The resting state. B: The n-back WM task. The group connectomes are sparse as in permutation testing and in mixture modeling (Fig. 4.9).

Fig. 4.11 and 4.12 present the difference between the group statistic for the task and rest, uncorrected and Bonferroni-corrected, respectively. In these
4.3. Thresholding functional connectomes: Results

Plots, I provided a group statistic based on the set of subject-specific, individual sparse connectomes: for each connection, I calculated the number of subjects in which this connection was found under the visual stimulation, and subtracted the number of subjects in which this connection was found in the resting state. The intensity of connections on the circular plots refers to this relative difference: red color means that a given connection appeared as significant in more subjects during the WM task as compared to resting state, whereas intense blue relates to connections which appear in fewer subjects under visual stimulation rather than during rest. In the circular plots, the homologous regions are placed against each other (on the left, versus on the right side of the plots). This plotting routine highlights the homologous connections (which is of primary importance in the results obtained in this study).

**Figure 4.11:** The group difference between resting state and n-back WM task, for six methods including the mixture modeling, permutation testing, and proportional thresholding. To make the results from mixture modeling comparable with permutation testing, the pFDR was derived on the group level. The group statistic is provided based on the set of subject-specific, individual sparse connectomes obtained for the canonical parameters (hard thresholding at 0 for empirical precision with and without regularization, $p = 0.05$ for permutation testing, $FDR = 0.05$ for mixture modeling, 5% and 10% cut-off for proportional thresholding). Then, for each connection, I calculated the number of subjects in which this connection was found under the visual stimulation, and subtracted the number of subjects in which this connection was found in the resting state. The intensity of connections on the circular plots refers to this relative difference: red color means that a given connection appeared as significant in more subjects during the WM task as compared to resting state, whereas intense blue relates to connections which appear in fewer subjects under visual stimulation rather than during rest. The homologous regions are placed against each other (on the left, versus on the right side of the plots). The figures present the raw results.
In Figs 4.11 and 4.12, I compare the six methods previously evaluated in the synthetic datasets, for the canonical value of the thresholding parameter.

What can be observed from nonparametric testing is that, the systematic differences between functional connectivity at rest and under visual stimulation can be observed with the use of all the investigated methods (at the very conservative significance level $p = 0.01$ with a Bonferroni correction). For both permutation testing and mixture modeling, results are similar: there is a dominant effect of decoupling between homotopic areas, especially high in the visual hierarchy, including LO1 and LO2 areas. In particular, a functional decoupling between higher order visual areas in parietal cortex (IPS0, IPS1, IPS2) is revealed along two dimensions: between the homologous counterparts and between the nearby regions within each ipsilateral part of the parietal cortex. On the other hand, proportional thresholding at the 5% threshold does not reveal such an effect, but instead, it suggests unilateral functional coupling between left PHC1 and PHC2 during visual stimulation.

Mixture modeling and permutation testing give similar results when applied to experimental fMRI datasets. However, as these two methods threshold the connectome along different dimensions (across subjects in permutation testing and across connections in mixture modeling), the p-values for these two methods are not directly comparable.

![Figure 4.12: Results as in Fig. 4.11, corrected for multiple comparisons with non-parametric testing (Mann-Whitney U test at the significance level $p = 0.01$ with a Bonferroni correction).](image)

Additionally, in Fig. 4.13, I present a comparison between the results obtained for permutation testing and the null, obtained in two ways: through shuffling labels between subjects versus permuting the time series within each subject’s BOLD time series. By this comparison, I demonstrate that...
the method to create the null in permutation testing (namely, subject- versus population-level) has a major influence on the group results in the study. I confront these results with mixture modeling to demonstrate that unlike in permutation testing, results obtained with the use of mixture modeling based on confidence intervals derived on subject- and on population level, are similar.

In mixture modeling, thresholding at the $p_{FDR}$ derived for each subject separately (for the subject-specific distribution of partial correlation values), and at the cohort level (for the joint distribution of partial correlation values in the whole population), gives similar results. On the contrary, in permutation testing, thresholding at the $p$ level derived for each subject separately (by permuting samples in the time series), and at the cohort level (by shuffling the time series between the subjects), gives different results.

In Table 4.4, I present the percentages of significant connections up- and down-regulated on the group level (Figs 4.11 and 4.12).
## Table 4.1: The mean density of connections per subject, in the resting state and in the n-back WM task, at the canonical thresholds. The mixture modeling gives significantly sparser connectomes at the canonical value of the threshold than the other methods.

<table>
<thead>
<tr>
<th>no</th>
<th>method</th>
<th>Rest</th>
<th>Task</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Empirical precision</td>
<td>0.574 (± 0.017)</td>
<td>0.558 (± 0.013)</td>
<td>-0.016</td>
</tr>
<tr>
<td>2</td>
<td>Ledoit-Wolf</td>
<td>0.584 (± 0.018)</td>
<td>0.568 (± 0.017)</td>
<td>-0.016</td>
</tr>
<tr>
<td>3</td>
<td>Permutation testing</td>
<td>0.115 (± 0.013)</td>
<td>0.092 (± 0.013)</td>
<td>-0.022</td>
</tr>
<tr>
<td>4</td>
<td>MM(LW,GG)</td>
<td>0.089 (± 0.009)</td>
<td>0.068 (± 0.012)</td>
<td>-0.021</td>
</tr>
<tr>
<td>5</td>
<td>Proportional thresholding, top 5%</td>
<td>0.048 (± 0.0)</td>
<td>0.048 (± 0.0)</td>
<td>-0.0</td>
</tr>
<tr>
<td>6</td>
<td>Proportional thresholding, top 10%</td>
<td>0.098 (± 0.0)</td>
<td>0.098 (± 0.0)</td>
<td>-0.0</td>
</tr>
</tbody>
</table>

## Table 4.2: Index of Overlap between the two scanning sessions: mean per subject + standard deviation in the brackets. Naturally, methods giving the sparsest connectomes also result in the highest Index of Overlap. The group difference in Index of Overlap between task and rest is insignificant for all the methods.

<table>
<thead>
<tr>
<th>no</th>
<th>method</th>
<th>Rest</th>
<th>Task</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Empirical precision</td>
<td>60.15 (± 2.25)</td>
<td>66.09 (± 2.33)</td>
<td>+5.94</td>
</tr>
<tr>
<td>2</td>
<td>Ledoit-Wolf</td>
<td>61.53 (± 2.24)</td>
<td>67.75 (± 2.45)</td>
<td>+6.22</td>
</tr>
<tr>
<td>3</td>
<td>Permutation testing</td>
<td>93.79 (± 1.08)</td>
<td>94.15 (± 1.66)</td>
<td>+0.36</td>
</tr>
<tr>
<td>5</td>
<td>MM(LW,GG)</td>
<td>96.03 (± 0.89)</td>
<td>97.53 (± 4.87)</td>
<td>+0.38</td>
</tr>
<tr>
<td>6</td>
<td>Proportional thresholding, top 5%</td>
<td>97.66 (± 0.54)</td>
<td>97.53 (± 0.65)</td>
<td>-0.13</td>
</tr>
<tr>
<td>7</td>
<td>Proportional thresholding, top 10%</td>
<td>93.47 (± 0.84)</td>
<td>93.21 (± 0.97)</td>
<td>-0.26</td>
</tr>
</tbody>
</table>

## Table 4.3: ICC scores in the resting state and in the n-back WM task across subjects (day 1 only). The most conservative methods give the highest ICC scores, exceeding 0.5. The difference between rest and task is not pronounced for any of the methods.

<table>
<thead>
<tr>
<th>no</th>
<th>method</th>
<th>Rest</th>
<th>Task</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Empirical precision</td>
<td>0.0945</td>
<td>0.0976</td>
<td>+0.0031</td>
</tr>
<tr>
<td>2</td>
<td>Ledoit-Wolf</td>
<td>0.1012</td>
<td>0.1055</td>
<td>+0.0043</td>
</tr>
<tr>
<td>3</td>
<td>Permutation testing</td>
<td>0.5181</td>
<td>0.4506</td>
<td>-0.0675</td>
</tr>
<tr>
<td>5</td>
<td>MM(LW,GG)</td>
<td>0.6192</td>
<td>0.5795</td>
<td>-0.0397</td>
</tr>
<tr>
<td>6</td>
<td>Proportional thresholding, top 5%</td>
<td>0.6218</td>
<td>0.5831</td>
<td>-0.0387</td>
</tr>
<tr>
<td>7</td>
<td>Proportional thresholding, top 10%</td>
<td>0.5377</td>
<td>0.4828</td>
<td>-0.0549</td>
</tr>
</tbody>
</table>

## Table 4.4: The percentage of non-zero links in Figs 4.11 and 4.12.

<table>
<thead>
<tr>
<th>no</th>
<th>method</th>
<th>Fig. 4.11</th>
<th>Fig. 4.12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Empirical Precision</td>
<td>94.92</td>
<td>3.97</td>
</tr>
<tr>
<td>2</td>
<td>Ledoit-Wolf</td>
<td>93.49</td>
<td>4.60</td>
</tr>
<tr>
<td>3</td>
<td>Permutation testing</td>
<td>94.60</td>
<td>5.56</td>
</tr>
<tr>
<td>4</td>
<td>MM(LW,GG)</td>
<td>71.90</td>
<td>6.03</td>
</tr>
<tr>
<td>5</td>
<td>Proportional thresholding, top 5%</td>
<td>38.89</td>
<td>0.95</td>
</tr>
<tr>
<td>6</td>
<td>Proportional thresholding, top 10%</td>
<td>86.35</td>
<td>2.06</td>
</tr>
</tbody>
</table>
4.4 Discussion

In this work, I propose a mixture-model-based approach to estimate sparse subject-specific functional connectomes. Mixture modeling allows for setting a threshold at a user-defined level of pFDR, which is a measure for separation between strong and weak connections within every subject's individual connectome. In the synthetic datasets where the ground truth is known, my approach gives best trade-off between TPR and FPR (for the canonical parameter of $pFDR = 0.05$). Furthermore, in experimental fMRI HCP datasets, I observe increased ICC for mixture modeling as compared to other methods (this effect, however, might be due to the fact that mixture modeling gives the most conservative thresholding).

I propose this approach as an alternative to the most popular methods in the field, namely to proportional thresholding and permutation testing. Even though proportional thresholding is currently the most popular approach in the field, in this work I paid special attention to comparing my approach to permutation testing as they are both based on a concept of estimating FPR, either in a nonparametric or in a parametric way. According to my results, Mixture Modeling yields a better AUC than permutation testing when applied to synthetic datasets, and similar results when applied to the HCP datasets. However, mixture modeling and permutation testing are not equivalent as they sparsify the functional connectome along different dimensions: connection-wise and across the population in permutation testing, versus across the functional connectome (subject-wise or population-wise) in mixture modeling. Furthermore, given one value of pFDR for mixture modeling and $p$ for permutation testing, the sparse connectome is sparser for mixture modeling because the FDR is always lower or equal than FPR it is derived from (Eq. 4.5). Therefore, these two methods are not directly comparable.

I propose using mixture modeling as an alternative to permutation testing for functional connectivity research in cohorts of a few subjects, such as case studies, or invasive translational psychiatry paradigms where experimental groups typically include 4-8 subjects. Using mixture modeling is possible as long as the investigated brain network is large enough to build a distribution from the entries of the partial correlation matrix. Of course, applying this strategy is also possible for large cohorts (e.g., Chauvin et al., 2019a). According to my results in synthetic datasets, network sizes of $N = 15$ or more, are sufficient for the application of the mixture modeling approach. However, in the real world applications, this will also depend on the SNRs in the data and on the density of the underlying "true" functional connectomes.

In this work, I demonstrate one potential application of mixture modeling for thresholding connectomes using the example of the human visual system. Comparing the influence of visual stimulation on the functional connectome, mixture modeling and permutation testing reveal a very similar pattern of (de)-activation under visual stimulation on the group level. In both rest and in cognition, the group functional connectome obtained from these two methods reflects the current state of the knowledge upon the functional architecture of the human visual system. I observe a clear functional hierarchy (Haak and Beckmann, 2017), reflecting the histological (Hagmann et al., 2003) and functional (Wandell, Dumoulin, and Brewer, 2007;
Grill-Spector and Malach, 2004) findings on the architecture of the visual system in humans (and also animal studies in cats and macaques, Felleman and Essen, 1991; Tootell, Tsao, and Vanduffel, 2003; Essen and Maunsell, 1983). In my study, I observe that the primary visual cortex V1 occupies the bottom of the functional hierarchy (Felleman, Burkhalter, and Essen, 1997), and sends strong projections to the secondary visual cortex (V2). I also observe projections from V2 to V3, however projections from V2 to higher visual areas reported in the literature (Grill-Spector and Malach, 2004) are not visible.

I also observe interhemispheric symmetry in the group functional connectome, with almost identical pattern of functional connectivity in both hemispheres (Fig. 4.9). The functional coupling between the homologous areas is, however, more pronounced in the resting state than under visual stimulation (Fig. 4.12), and the difference is most pronounced in the higher levels of the visual hierarchy. This functional decoupling in the high visual areas under visual stimulation can relate to the lateralization theory of visual perception (Schotten et al., 2011), which states that visual processing in two hemispheres is different (Courtney et al., 1996; Güntürkün et al., 2000) (which is believed to have a certain adaptive value).

Lastly, I can observe a strong coupling between subsequent levels of the visual hierarchy within the dorsal and within the ventral visual stream. The ventral stream, known as the "what pathway" or "perception pathway" (Lerner et al., 2001), is thought to be responsible for object recognition, and leads from the striate cortex towards the temporal lobe (Goodale and Milner, 1992; Mishkin, Ungerleider, and Macko, 1983). The dorsal stream, a.k.a. the "where pathway" or "action pathway", encodes the spatial location of objects (Goodale and Milner, 1992; Mishkin, Ungerleider, and Macko, 1983), and leads from the striate cortex towards the parietal lobe. In my study, I can clearly delineate both streams within V1, V2 and V3. Furthermore, the functional decoupling between hemispheres under the visual stimulation is much more pronounced in the dorsal stream than in the ventral stream (Fig. 4.12). In the experimental paradigm implemented in HCP datasets (Barch et al., 2013), the visual stimuli were presented at the fixation point (in the center of the visual field), which means that the same amount of visual stimulation was delivered to both hemispheres. Therefore, the functional decoupling between the contralateral parts of the "where" pathway under the visual stimulation suggest that the lateralization in the visual processing starts at a very early stage of the visual processing, possibly even in V1. Since the visual system has the functional architecture of a convolutional network (Güclü and Gerven, 2015; Wandell, Dumoulin, and Brewer, 2007), this asymmetry amplifies in the higher levels of visual hierarchy.

4.4.1 Limitations of the method

In my approach, I do not take inhibition into account. Inhibition in the brain is primarily found on two levels of organization. Firstly, in the small scale: single inhibitory cells form local inhibitory networks, densely interconnected within single brain regions (Fino and Yuste, 2011). Secondly, in the global scale: large scale resting state networks are often anticorrelated with each other (e.g., the Default Mode Network is anticorrelated with various task-positive networks, Fox et al., 2005; Uddin et al., 2009; Chai et al., 2012). This
work is dedicated to modeling networks on a *meso-scale* level: within single Resting State Networks (Damoiseaux et al., 2006; Smith et al., 2009). On this level of description, anticorrelations between the nodes of the network are rarely encountered. In particular, in my study on the visual system, I observe no anticorrelations in the functional connectivity, neither in rest nor under visual stimulation.

For this reason, I believe that in the aforementioned applications, mixture modeling consists of just the two components - pseudo-null and one component representing connections - is valid. In case of possible extensions of this research to the full brain modeling, there might be a necessity of adding a third component, representing anticorrelations between the resting state networks. Furthermore, I choose to use a parametrized mixture mode to fit the null (using either Gaussian or Laplace distribution), but in principle, one could also use the Efron’s approach (Efron, 2007) of fitting a spline to the total distribution and then fitting an empirical null to that spline fit.

The length of the time series is also an important aspect of the mixture modeling as it affects the width of the null distribution: the shorter the time series, the broader the null and the more uncertainty in the estimation of the signal component. In this case, the estimation of partial correlation matrix was performed on a high quality datasets of 800 samples. The long time series and high data quality may explain why ICC scores are so high in this study, for all the examined methods (Table 4.2, in the range of 0.5 as opposed to 0.15 encountered in the literature, Fiecas et al., 2013). In addition, I used a single session of the n-back WM task, and sectioned the data into two time series, which does not capture the day-to-day variability in the functional connectivity during cognition.

The last issue is the intrinsic limitations of the synthetic datasets used in this study. Firstly, in the synthetic datasets, I simulated the DCM generative model highly acknowledged in the fMRI research community (Smith et al., 2011; Friston, Harrison, and Penny, 2003). Whether this model gives a satisfactory representation of the experimental fMRI datasets, remains an open question. However, out of the 8 possibilities, the same mixture achieved the lowest BIC score both when applied to the experimental HCP datasets, and to the synthetic datasets. This result suggests that the distributions of connections for the synthetic and experimental datasets are fairly comparable. Secondly, in the experimental fMRI datasets, the preprocessing pipeline can influence the (functional) connectivity studies to a large extent (Aurich et al., 2015; Marrelec and Fransson, 2011). In this study, I chose the high quality, single-site datasets from the HCP500 cohort (Essen et al., 2013), and I used the standard FIX motion regression (Salimi-Khorshidi et al., 2014; Griffanti et al., 2014). Also, I chose for a high quality parcellation of the visual cortex by Wang et al., 2015, and for the highly potent visual working memory task which is known to induce strong activation patterns across the brain, especially in the visual cortex (Barch et al., 2013). Therefore, I believe that I chose a pipeline that gave the highest chances for obtaining reliable results for the group comparison between task and rest in the experimental datasets.
4.4.2 Further research

The application of mixture modeling to threshold connectomes in the visual system, as presented in this Chapter, is a proof of concept. The visual system was chosen for an application because I was able to formulate and test hypotheses based on the prior knowledge upon the anatomical and functional architecture of the human visual system. In the future, this approach can also be applied for thresholding connectomes in other mesoscale networks.

Furthermore, in this work, I constructed and compared group connectomes in the task in during resting state. I performed this comparison connection by connection rather than by looking into higher order statistics of these connectomes, graph theoretical properties, etc. As known from the literature, thresholding procedure can highly influence these global properties in the networks (Zalesky et al., 2016), therefore a follow up study on this aspect of the thresholding should also be performed.

Lastly, as the previous research suggests that the false positives have a profound effect on the network organization, the future research might include other implementations of mixture modeling with the use of other component distributions more robust to false positives, etc.

4.4.3 Conclusions

Building functional connectomes of the large scale networks, in rest and under cognitive stimulation, is a large and still developing subfield of the fMRI research. In many contexts, it is beneficial or even mandatory to sparsify the functional connectome. E.g., in the graph theoretical studies, sparse connectomes allow for studying global properties of the functional networks in health (Bullmore and Sporns, 2009) and disease (Heuvel et al., 2010; Zhang et al., 2011b; Zhang et al., 2011a). Certain graph theoretical measures such as topological overlap between two nodes in the network or Rentian scaling (wiring efficiency) are based on the sparse connectomes only (Brain Connectivity Toolbox, Bullmore and Sporns, 2009). Next, there is a class of models characterizing the dynamics in the human connectome by means of Ising models (Stramaglia et al., 2017) which require sparsity. Furthermore, recent developments in the field of causal inference in fMRI involve a two-step inference procedure. In such a procedure, connections are spotted in the network by creating a sparse functional connectome (Patel, Bowman, and Rilling, 2006; Hyvärinen and Smith, 2013; Bielczyk et al., 2017a), and the causal discovery is performed pairwise. Sparse connectomes are also useful in multiple other disciplines that involve graph theory, i.e., in social networks (Watts and Strogatz, 1998; Mislove et al., 2007), protein-protein, protein-DNA, and gene-gene interactions (Eisen et al., 1998; Pellegrini, Haynor, and Johnson, 2004).

The selection of available methods for sparsifying functional connectomes is still growing. E.g., among the recent developments in the field, there are new, fused estimators for thresholding functional connectomes (Zille et al., 2017) dedicated to comparing two different groups of functional connectomes, i.e., connectomes derived for a group of children versus adults, or for one group of subjects undergoing two different cognitive tasks. Given the comparison between the methods on synthetic benchmark
datasets, my results agree with the previous studies suggesting that the thresholding technique has a major influence on the type of information contained in the sparse connectome, and therefore should be approached with care. E.g., in proportional thresholding, the threshold is often an arbitrary choice (Achard and Bullmore, 2007; Rubinov and Sporns, 2010; Bassett and Bullmore, 2009; Heuvel et al., 2008). As discussed in a recent theoretical study by Heuvel et al., 2017, and also in previous studies (Nichols et al., 2017; Wijk, Stam, and Daffertshofer, 2010; Alexander-Bloch et al., 2010), including weak and therefore unreliable connections (in subjects with overall weak functional connectivity) to the connectome has an effect on the organization of the functional network and graph metrics.

As a summary, since mixture modeling gives a good trade-off between a data-driven approach (as in permutation testing) and subject-specific thresholding (as in proportional thresholding), I propose using mixture modeling for thresholding connectomes as an alternative to existing approaches.
Chapter 5


In the past two decades, fMRI has been used to relate neuronal network activity to cognitive processing and behavior. Recently, this approach has been augmented by algorithms that allow us to infer causal links between component populations of neuronal networks. Multiple inference procedures have been proposed to approach this research question. So far, each method has limitations when it comes to establishing whole-brain connectivity patterns. In this Chapter, I discuss the ways to infer causality in fMRI research. I also formulate recommendations for the future.

**Keywords:** causal inference, effective connectivity research, functional Magnetic Resonance Imaging

*This Chapter was published in Network Neuroscience as:*


5.1 Introduction

5.1.1 What is causality?

Although inferring causal relations is a fundamental aspect of scientific research, the notion of causation itself is notoriously difficult to define. The basic idea is straightforward: when process $A$ is the cause of process $B$, $A$ is necessarily in the past from $B$, and without $A$, $B$ would not occur. But in practice, and in dynamic systems such as the brain in particular, the picture is far less clear. First, for any event a large number of (potential) causes can be identified. The efficacy of certain neuronal process in producing behavior is dependent on the state of many other (neuronal) processes, but also on the availability of glucose and oxygen in the brain, etc. In a neuroscientific context, we are generally not interested in most of these causes, but only in a cause that stands out in such a way that it is deemed to provide a substantial part of the explanation, e.g., causes that vary with the experimental conditions. However, the contrast between relevant and irrelevant causes (in terms of explanatory power) is arbitrary and strongly dependent on experimental setup, contextual factors, etc. E.g., respiratory movement is typically considered a confound in fMRI experiments, unless the research...
question concerns the influence of respiration speed on the dynamics of the neuronal networks.

In dynamic systems, causal processes are unlikely to be part of a unidirectional chain of events, but rather a causal web, with often mutual influences between process $A$ and $B$ (Mannino and Bressler, 2015). As a result, it is hard to maintain the temporal ordering of cause and effect and, indeed, a clear separation between them (Schurger and Uithol, 2015).

Furthermore, causation can never be observed directly, just correlation (Hume, 1772). When a correlation is highly stable, we are inclined to infer a causal link. Additional information is then needed to assess the direction of the assumed causal link, as correlation indicates for association and not for causation (Altman and Krzywiński, 2015). E.g., the motor cortex is always active when a movement is made, so we assume a causal link between the two phenomena. The anatomical and physiological properties of the motor cortex, and the timing of the two phenomena provide clues about the direction of causality (i.e., cortical activity causes the movement, and not the other way around). However, only intervention studies, such as delivering Transcranial Magnetic Stimulation (TMS, Kim, Pesiridou, and OReardon, 2009) pulses over the motor cortex or lesion studies, can confirm the causal link between the activity in the motor cortex and behavior.

Causal studies in fMRI are based on three types of correlations: correlating neuronal activity to 1) mental and behavioral phenomena; 2) physiological states (such as neurotransmitters, hormones, etc.), and 3) neuronal activity in other parts of the brain. In this review, I will focus on the last field of research: establishing causal connections between activity in two or more brain areas.

5.1.2 A note on the limitations of fMRI data

fMRI studies currently use a variety of algorithms to infer causal links (Fornito, Zalesky, and Breakspear, 2013; Smith et al., 2011). All these methods have different assumptions, advantages and disadvantages (see, e.g., Valdes-Sosa et al., 2011; Stephan and Roebroeck, 2012), and approach the problem from different angles. One important reason for this variety of approaches is the complex nature of fMRI data, which imposes severe restrictions on the possibility of finding causal relations using MRI.

Temporal resolution and hemodynamics

First, and best known, the temporal resolution of the image acquisition in MR imaging is generally restricted to a sampling rate $< 1 \text{Hz}$. Recently, multiband fMRI protocols have gained in popularity (Feinberg and Setsompop, 2013), which increased the upper limit for the scanning frequency to up to $10 \text{Hz}$, albeit at the cost of a severely decreased the SNR ratios. However, no imaging protocol (including multiband imaging) can overcome the limitation of the recorded signal itself: the lagged change in blood oxygenation, which peaks 3 to 6 seconds after neuronal firing in the adult human brain (Arichi et al., 2012). The hemodynamic response thus acts as a low-pass filter, which results in high correlations between activity in consecutive frames (Ramsey et al., 2010). Since the hemodynamic lags (understood
as the peaks of the hemodynamic response) are region- and subject-specific (Devonshire et al., 2012) and vary over time (Glomb et al., 2017), it is difficult to infer causality between two time series with potentially different hemodynamic lags (Bielczyk et al., 2017b). Computational work by Seth, Chorley, and Barnett, 2013 suggests that upsampling the signal to low TRs (< 0.1[s]) might potentially overcome this issue. Furthermore, hemodynamics typically fluctuates in time. These slow fluctuations, similarly to other low frequency artifacts such as heartbeat or body movements, should be removed from the datasets through high-pass filtering before the inference procedure (Ramsey, Sanchez-Romero, and Glymour, 2014).

**SNR**

Second, fMRI data is characterized by a relatively low SNR. In grey matter, the recorded hemodynamic response changes by 1-2% at field strengths of 1.5 – 2.0[T] (Ogawa et al., 1993; Boxerman et al., 1995), and by 5-6% at field strengths of 4.0[T]. Moreover, typical fMRI protocols generate relatively short time series. E.g., the new Human Connectome Project resting state datasets (Essen et al., 2013) do not contain more than a few hundred to maximally few thousand samples. Two most popular ways of improving on the SNR in fMRI datasets are: averaging signals over multiple voxels (Friston et al., 2007) and spatial smoothing (Triantafyllou, Hoge, and Wald, 2006).

**Caveats associated with region definition**

Third, to propose a causal model, one first needs to define the nodes of the network. A single voxel does not represent a biologically meaningful part of the brain (Stanley et al., 2013). Therefore, before attempting to establish causal connection in the network, one needs to integrate the BOLD time series over regions of interest (ROIs): groups of voxels that are assumed to share a common signal with a neuroscientific meaning. Choosing the optimal regions of interest for a study is a complex problem (Poldrack, 2007; Marrelec and Fransson, 2011; Thirion et al., 2014; Fornito, Zalesky, and Breakspear, 2013; Kelly et al., 2012). In task-based fMRI, ROIs are often chosen based on activation patterns revealed by the standard GLM analysis (Friston et al., 2007).

On the other hand, in research on resting state brain activity, the analysis is usually exploratory and the connectivity in larger, meso- and macroscale networks is typically considered. In that case, a few strategies to ROI definition are possible. First, one can define regions of interest on the basis of brain anatomy. However, a consequence of this strategy could be that BOLD activity related to the cognitive process of interest will be mixed with other, unrelated activity within the ROIs. This is particularly likely to happen given that brain structure is not exactly replicable across individuals, so that a specific area cannot be defined reliably based on location alone. As indicated in the computational study by Smith et al., 2011, and also in a recent study by Bielczyk et al., 2017b, such signal mixing is detrimental to causal inference and causes all the existing methods for causal inference in fMRI to underperform. What these studies demonstrate is that parcellating into ROIs based on anatomy rather than common activity, can induce
additional scale-free background noise in the networks. Since this noise has high power in low frequencies, the modelled BOLD response cannot effectively filter it out. As a consequence, the signatures of different connectivity patterns are getting lost.

As an alternative to anatomical parcellation, choosing ROIs can be performed in a functional, data-driven fashion. There are multiple techniques developed to reach this goal, and to list some recent examples: Instantaneous Correlations Parcellation implemented through a hierarchical Independent Component Analysis (ICP, Oort et al., 2017), probabilistic parcellation based on Chinese restaurant process (Janssen et al., 2015), graph clustering based on inter-voxel correlations (Heuvel, Mandl, and Pol, 2008), large-scale network identification through comparison between correlations among ROIs versus a model of the correlations generated by the noise (LSNI, Bellec et al., 2006), multi-level bootstrap analysis (Bellec et al., 2010), clustering voxels that reveal common causal patterns in terms of Granger Causality (DSouza et al., 2017), spatially constrained hierarchical clustering (Blumensath et al., 2013), and algorithms providing a trade-off between machine learning techniques and knowledge coming from neuroanatomy (Glasser et al., 2016).

Another possibility to reduce the effect of mixing signals is to perform Principal Component Analysis (PCA, Jolliffe, 2002; Shlens, 2014), to separate the BOLD time series within each anatomical region into a sum of orthogonal signals (eigenvariates), and to extract the signal with the highest contribution to the BOLD signal (a.k.a. the first eigenvariate, Friston, Harrison, and Penny, 2003), instead of averaging activity over full anatomical regions. Finally, one can build ROIs based on patterns of activation only (task localizers, Fedorenko et al., 2010; Heinzle, Wenzel, and Haynes, 2012). However, this approach cannot be applied to the resting state research. In this Chapter, I assume that the definition of ROIs has been performed by the researcher prior to the causal inference, and I do not discuss it any further.

5.1.3 Criteria for evaluating methods for causal inference in functional Magnetic Resonance Imaging

Given the aforementioned characteristics of fMRI data (low temporal resolution, slow hemodynamics, low SNR) and the fact that causal webs in the brain are likely dense and dynamic, is it even possible to investigate causality in the brain using MRI? Multiple distinct families of models have been developed to approach this problem over the past two decades. One can look at the methods from different angles, and classify them into different categories.

One important distinction proposed by Friston, Moran, and Seth, 2013, includes division of methods with respect to the depth of the neuroimaging measurements at which a method is defined. Most methods (such as the original formulation of Structural Equation Modeling for fMRI, McIntosh and Gonzalez-Lima, 1994, see: Section 5.2.2) operate on the experimental observables, i.e., the measured BOLD responses. These methods are referred to as directed functional connectivity measures. On the contrary, other methods (e.g., Dynamic Causal Modeling, see: Section 5.2.3) consider the underlying
neuronal processes as well. These methods are referred to as effective connectivity measures. Mind that while some methods such as Dynamic Causal Modeling are hardwired to assess effective connectivity (as they are built upon a generative model), other methods can be used both as a method to assess directed functional connectivity or effective connectivity. E.g., in Granger Causality research (see: Section 5.2.1), a blind deconvolution is often used to deconvolve the observed BOLD responses into an underlying neuronal time series (David et al., 2008; Ryali et al., 2011; Ryali et al., 2016; Hutcheson et al., 2015; Wheelock et al., 2014; Sathian, Deshpande, and Stilla, 2013; Goodyear et al., 2016), which allows for assessing effective connectivity. On the contrary, when Granger Causality is used without deconvolution (Zhao et al., 2016; Regner et al., 2016; Chen et al., 2017), it is a directed functional connectivity method. Of course, both scenarios have pros and cons, as blind deconvolution can be a very noisy operation (Bush et al., 2015). For more details, please see: Friston, Moran, and Seth, 2013.

Another important distinction was proposed by Valdes-Sosa et al., 2011. According to this point of view, methods can be divided on the basis of the approach towards temporal sequence of the samples: some of the methods are based on the temporal sequence of the signals (e.g., TE or GC), or rely on the dynamics expressed by state-space equations (so-called state-space models, e.g., DCM), while other methods do not draw information from the time-sequence, and solely focus on the statistical properties of the time series instead (so-called structural models, e.g., BNs).

In this work, I would like to propose another classification of methods for causal inference in fMRI. First, I identify nine characteristics of models used to study causality. Then, I compare and contrast the popular approaches to the causal research in fMRI according to these criteria. The list of features of causality is as follows:

1. **Sign of connections**: Can the method distinguish between excitatory and inhibitory causal relations? In this context, I do not mean synaptic effects, but rather an overall driving or attenuating impact of the activity in one brain region on the activity in another region. Certain methods only detect the existence of causal influence from the BOLD responses whereas others can distinguish between these distinct forms of influence.

2. **Strength of connections**: Can the method distinguish between weak and strong connections, apart from indicating the directionality of connections at a certain confidence level?

3. **Confidence intervals**: How are the confidence intervals for the connections determined?

4. **Bidirectionality**: Can the method pick up bidirectional connections $X \leftrightarrow Y$, or only indicate the strongest of the two connections $X \rightarrow Y$ and $Y \rightarrow X$? Some methods do not allow for bidirectional relations, since they cannot deal with cycles in the network.

5. **Immediacy**: Does the method specifically identify direct influences $X \rightarrow Y$, or does it pool across direct and indirect influences $Z_i: X \rightarrow Z_i \rightarrow Y$? I assume that $Z_i$ represent nodes in the network, and the activity in these nodes is measured (otherwise $Z_i$ become a latent confounder).
While some methods aim to make this distinction, others highlight any influence $X \rightarrow Y$, whenever it is direct or not,

6. **Resilience to confounds**: Does the method correct for possible spurious causal effects from a common source ($Z \rightarrow X, Z \rightarrow Y$, so I infer $X \rightarrow Y$ and/or $Y \rightarrow X$), or other confounders? In general, confounding variables are an issue to all the methods for causal inference, especially when a given study is non-interventional (Rohrer, 2017), however different methods can suffer from these issues to a different extent,

7. **Type of inference**: Does the method probe causality through classical hypothesis testing or through model comparison? Hypothesis-based methods will test a null hypothesis $H_0$ that there is no causal link between two variables, against a hypothesis $H_1$ that there is causal link between the two. In contrast, model-comparison-based methods do not have an explicit null hypothesis. Instead, evidence for a pre-defined set of models is computed. In particular cases, when the investigated network contains only a few nodes and the estimation procedure is computationally cheap, a search through all the connectivity patterns by means of model comparison is possible. In all the other cases, prior knowledge is necessary to select a subset of possible models for model comparison,

8. **Computational cost**: What is the computational complexity of the inference procedure? In the case of model comparison, the computational cost refers to the cost of finding the likelihood of a single model, as the range of possible models depends on the research question. This can lead to practical limitations based on computing power,

9. **Size of the network**: What range of network sizes does the method allow for? Some methods are restricted in the number of nodes that they allow for, for computational or interpretational reasons.

In certain applications, an additional criterion of *empirical accuracy in realistic simulation* could be of help to evaluate the method. Testing the method on synthetic, ground truth datasets available for the research problem at hand can give a good picture on whether or not the method gives reliable results when applied to experimental datasets. In fMRI research, multiple methods for causal inference were directly compared to each other in a seminal simulation study by Smith et al., 2011. In this study, the authors employed the Dynamic Causal Modeling generative model (Friston, Harrison, and Penny, 2003), introduced in Section 5.2.3, to create synthetic datasets with a known ground truth. Surprisingly, most of the methods struggled to perform above the chance level, even though the test networks were sparse and the noise levels introduced to the model were low compared to what one would expect in real fMRI recordings. In this Chapter, I will refer to this study throughout the text. However, I will not list empirical accuracy as a separate criterion, for two reasons. Firstly, some of the methods reviewed here, e.g., SEM, were not tested on the synthetic benchmark datasets. Secondly, the most popular method in the field, Dynamic Causal Modeling (Friston, Harrison, and Penny, 2003), builds on the same generative model that is used for comparing methods to each other in Smith’s
study. Therefore, it is hard to make a fair comparison between DCM and other methods in the field using this generative model.

In the following Chapters, the references to this "causality list" will be marked in the text with lowercase indices.

![Causal Methods](image.png)

**Figure 5.1**: Causal research in fMRI. The discussed methods can be divided into two families: Network Inference Methods, which are based on a one-step multivariate procedure, and Pairwise Inference Methods, which are based on a two-step pairwise inference procedures. As pairwise methods by definition establish causal connections on a node-by-node basis, the network as a whole cannot be guaranteed to be of any particular structure.

With respect to assumptions made on the connectivity structure, the approaches discussed here can be divided into three main groups (Fig. 5.1). The first group comprises multivariate methods that search for directed graphs without imposing any particular structure onto the graph: Granger Causality (Seth, Barrett, and Barnett, 2015), Transfer Entropy (Marrelec et al., 2006), Dynamic Causal Modeling (Friston, Harrison, and Penny, 2003), and Structural Equation Modeling (Mclntosh and Gonzalez-Lima, 1994). These methods will be referred to as network-wise models throughout the text. The second group of methods is also multivariate, but requires an additional assumption of acyclicity. Models in this group assume that information travels through the brain by feed forward projections only. As a result, the network can always be represented by a Directed Acyclic Graph (DAG, Thulasiraman and Swamy, 1992). Methods in this group include Linear Non-Gaussian Acyclic Models (LiNGAM, Shimizu et al., 2006) and Bayesian Nets (Mumford and Ramsey, 2014), and will be referred to as hierarchical network-wise models throughout the text. The last group of methods, referred to as pairwise methods, use a two-stage procedure: first, a map
of nondirectional functional connections is rendered, and second, the directionality in each connection is assessed. Since these methods focus on pairwise connections rather than complete network architectures, they by definition do not impose network assumptions like acyclicity. Patel’s tau (Patel, Bowman, and Rilling, 2006) and Pairwise Likelihood Ratios (Hyvärinen and Smith, 2013) belong to this group. In this Chapter, I do not include studying a coupling between brain region and the rest of the brain with relation to a particular cognitive task, The Psycho-Physiological Interactions (PPIs, Friston et al., 1997); I only focused on the methods for assessing causal links within brain networks, and I do not include brain-behavior causal interactions.

5.2 Network-wise methods

The first group of models involves multivariate methods: methods that simultaneously assess all causal links in the network - specifically, Granger Causality (Granger, 1969), Transfer Entropy (Schreiber, 2000), Structural Equation Modeling (Wright, 1920), and Dynamic Causal Modeling (Friston, Harrison, and Penny, 2003). These methods do not pose any constraints on the connectivity structure within the network. Granger Causality, Transfer Entropy, and Structural Equation Modeling infer causal processes through classical hypothesis testing. As there are no limits to the size of the analyzed network, these methods allow for hypothesis-free discovery. Dynamic Causal Modeling on the other hand, compares a number of predefined causal structures in networks of only a few nodes. As such, it requires a specific hypothesis based on prior knowledge.

5.2.1 Granger causality

Clive Granger introduced Granger Causality (GC) in the field of economics (Granger, 1969). GC has found its way into many other disciplines, including fMRI research (Roebroeck, Seth, and Valdes-Sosa, 2011; Bressler and Seth, 2011; Seth, Barrett, and Barnett, 2015; Solo, 2016). GC is based on prediction (Diebold, 2001): the signal in a certain region is dependent on its past values. Therefore, a time series \( Y(t) \) at time point \( t \) can be partly predicted by its past values \( Y(t-i) \). A signal in an upstream region is followed by the same signal in a downstream region with a certain temporal lag. Therefore, if prediction of \( Y(t) \) improves when past values of another signal \( X(t-i) \) are taken into account, \( X \) is said to Granger-cause \( Y \). Time series \( X(t) \) and \( Y(t) \) can be multivariate, therefore they will be further referred to as \( \vec{X}(t) \), \( \vec{Y}(t) \).

\( Y(t) \) is represented as an autoregressive process: it is predicted by a linear combination of its past states and a Gaussian noise (there is also an equivalent of GC in the frequency domain, spectral GC (Geweke, 1982; Geweke, 1984), but this method will not be covered in this review). This model is compared to a model including the past values of \( X(t) \):

\[
H_0 : \vec{Y}(t) = \sum_{i=1}^{N} B_{yi} \vec{Y}(t-i) + \vec{\sigma}(t) \quad (5.1)
\]
Chapter 5

5.2. Causal inference in fMRI, a Review: Network-wise methods

\[ H_1 : \bar{Y}(t) = \sum_{i=1}^{N} B_{yi} \bar{Y}(t-i) + \sum_{i=1}^{N} B_{xi} \bar{X}(t-i) + \sigma(t) \tag{5.2} \]

where \( \sigma(t) \) - the noise (or rather, the portion of the signal not explained by the model). Theoretically, this autoregressive (AR) model can take any order \( N \) (which can be optimized using, e.g., Bayesian Information Criterion, Schwarz, 1978) but in fMRI research it is usually set to \( N = 1 \) (Seth, Barrett, and Barnett, 2015), i.e., a lag equal to the repetition time (TR).

By fitting the parameters of the AR model, which include the influence magnitudes \( B_{yi}, B_{xi} \), the sign \( \beta \) as well as the strength of the causal direction can be readily assessed with GC. The significance of the results is evaluated by comparing variance of the noise obtained from models Eq. 5.1 and Eq. 5.2. This can be achieved either by F-tests or by permutation testing. Like all the methods in this Section, GC does not impose any constraints on the network architecture and therefore can yield bidirectional connections. As a multivariate method, GC fits the whole connectivity structure at once. Therefore, ideally, it indicates the direct causal connections only, whereas the indirect connections should be captured only through higher order paths in the graph revealed in the GC analysis. However, this is not enforced directly by the method. In fact, in the original formulation of the problem by Granger, GC between \( X \) and \( Y \) works based on the assumption that the input of all the other variables in the environment potentially influencing \( X \) and \( Y \) has been removed (Granger, 1969). In theory, this would provide resilience to confounds. However, in reality this assumption is most often not valid in fMRI (Grosse-Wentrup, 2014). As a result, direct and indirect causality between \( X \) and \( Y \) are in fact pooled. In terms of the inference type, one can look at GC in two ways. On one hand, GC is a model-comparison technique, since the inference procedure is, in principle, based on a comparison between two models expressed by Eqs. 5.1 and 5.2. On the other hand, the difference between GC and other model comparison techniques lies in the fact that GC does not optimize any cost function, but uses F-tests or permutation testing instead, and it can therefore also be interpreted as a method for classic hypothesis testing. Since the temporal resolution of fMRI is so low, typically first order AR models with a time-lag equal to 1 TR are used for the inference in fMRI. Therefore, there is no need to optimize either the temporal lag or the model order, and, as such, the computational cost of GC estimation procedure in fMRI is low. The AR model imposes a mathematical restriction on the size of the network though: the number of regions divided by the number of shifts can never exceed the number of time points (degrees of freedom).

GC is used in fMRI research in two forms: as mentioned in Section 5.1.3, GC can be either applied to the observed BOLD responses (Zhao et al., 2016; Regner et al., 2016; Chen et al., 2017), or to the BOLD responses deconvolved into neuronal time series (David et al., 2008; Ryali et al., 2011; Ryali et al., 2016; Hutcheson et al., 2015; Wheelock et al., 2014; Sathian, Deshpande, and Stilla, 2013; Goodyear et al., 2016). The purpose of the deconvolution is to model fMRI data more faithfully. However, estimating the hemodynamic response from the data - a necessity to perform this deconvolution - adds uncertainty as well.

The applicability of GC to fMRI data has been heavily debated (Stokes and Purdon, 2017). Firstly, the application of GC requires certain additional
assumptions such as signal stationarity (stationarity means that the joint probability distribution in the signal does not change over time. This also implies that mean, variance, and other moments of the distribution of the samples in the signal do not change over time), which does not always hold in fMRI data. Theoretical work by Seth, Chorley, and Barnett, 2013, and work by Roebroeck, Formisano, and Goebel, 2005, suggest that despite the limitations related to slow hemodynamics, GC is still informative about the directionality of causal links in the brain (Seth, Barrett, and Barnett, 2015). In the study by Smith et al., 2011, several versions of GC implementation were tested. However, all were characterized by a low sensitivity to link detection, low sensitivity to false positives, and low overall accuracy in the directionality estimation. The face validity of GC analysis was empirically validated using joint fMRI and MEG recordings (Mill et al., 2017), with the causal links inferred with GC matching the ground truth confirmed by MEG. On the other hand, experimental findings report that GC predominantly identifies major arteries and veins as causal hubs (Webb et al., 2013). This result can be associated with a regular pulsating behaviour with different phases in the arteries across the brain. This is a well-known effect and is even explicitly targeted with physiological noise estimates such as RETROICOR (Glover, Li, and Ress, 2000).

Another point of concern is the time lag in fMRI data, which restricts the possible scope of AR models that can be fit in the GC procedure. Successful implementations of GC in EEG/MEG research typically involve lags of less than a hundred milliseconds (Hesse et al., 2003). In contrast, for fMRI the minimal lag is one full TR, which is typically between $0.7\text{s}$ and $3.0\text{s}$ (although new acceleration protocols allow for further reduction of TR). What is more, the HRF may well vary across regions (Handwerker, Ollinger, and D’Esposito, 2004; David et al., 2008), revealing spurious causal connections: when the HRF in one region is faster than in another, the temporal precedence of the peak will easily be mistaken for causation. The estimated directionality can in the worst case, even be reversed, when the region with the slower HRF in fact causes the faster one (Bielczyk et al., 2017b). Furthermore, the BOLD signal might be non-invertible into the neuronal time series (Seth, Barrett, and Barnett, 2015), which can affect GC analysis regardless whether it is performed on the BOLD time series or the deconvolved signal.

5.2.2 Structural Equation Modeling

Structural Equation Modeling (SEM, Mcintosh and Gonzalez-Lima, 1994) is a simplified version of GC. This method was originally applied to a few disciplines: economics, psychology and genetics (Wright, 1920), and was only recently adapted for fMRI research. SEM can be considered a predecessor to Dynamic Causal Modeling (Friston, Harrison, and Penny, 2003). SEM is used to study effective connectivity in cognitive studies, e.g., on motor coordination (Kiyama et al., 2014; Zhuang et al., 2005), as well as in search for biomarkers of psychiatric disorders (Schlösser et al., 2003; Carballedo et al., 2011). It was also used for investigating heritability of large scale, resting state connectivity patterns (Carballedo et al., 2011).

The idea is to express every ROI time series in a network by a linear combination of all the other time series (with the addition of noise), which
implies no time lag in the communication. These signals are combined in a mixing matrix $B$:

$$\vec{X}(t) = B\vec{X}(t) + \vec{\sigma}(t)$$  (5.3)

where $\vec{\sigma}$ - the noise.

The assumption is that, each univariate component $X_i(t)$ is a mixture of the remaining components $X_j(t), j \neq i$. This is a simple multivariate regression equation. The most common strategy for fitting this model is a search for the regression coefficients that corresponds to the maximum likelihood (ML) solution: a set of model parameters $B$ that gives the highest probability of the observed data (McIntosh and Gonzalez-Lima, 1994; Anderson and Gerbing, 1988). Assuming that variables $X_i$ are normally distributed, the ML function can be computed and optimized. This function is dependent on the observed covariance between variables, as well as a concept of a so-called implied covariance, for the details, see: Bollen, 1989, and for a practical example of SEM inference, see: Ferron and Hess, 2007. Further, under the assumption of normality of the noise, there is a closed-form solution to this problem which gives the ML solution for parameters $B$, known as Ordinary Least Squares (OLS) approximation (Hayashi, 2000; Bentler, 1985).

In SEM applications to fMRI datasets, it is a common practice to establish the presence of connections with the use of anatomical information derived, e.g., from Diffusion Tensor Imaging (Protzner and McIntosh, 2006). In that case, SEM inference focuses on estimating the strength of causal effects and not on identifying the causal structure.

SEM does not constrain the weight of connections, therefore it can retrieve both excitatory and inhibitory connections, as well as bidirectional connections. The connection coefficients $B_{ij}$ can take any rational numbers and as such, they can reflect the strength of the connections. Since OLS gives a point estimate for $\beta$, it does not provide a measure of confidence that would determine whether the obtained $\beta$ is significantly different from zero. This issue can be overcome in multiple ways. First, one can perform parametric tests, e.g., a $t$-test. Second, one can obtain confidence intervals through nonparametric permutation testing (generate a null distribution of $B$ values by the repeated shuffling of node labels across subjects and creating surrogate subjects). Third, one can perform causal inference through model comparison: various models are fitted one by one, and the variance of the residual noise resulting from different model fits is compared, using either an $F$-test, or goodness of fit (GFI, Zhuang et al., 2005). Highly optimized software packages such as LiSREL (Joreskog and Thillo, 1972) allow for an exploratory analysis with SEM by comparing millions of models against each other (James et al., 2009). Lastly, one can fit the $B$ matrix with new methods including regularization that enforces sparsity of the solution (Jacobucci, Grimm, and McArdle, 2016), and therefore eliminates weak and noise-induced connections from the connectivity matrix.

As with GC, SEM was designed to reflect direct connections: if regions $X_i$ and $X_j$ are connected only through a polysynaptic causal web, $B_{ij}$ should come out as zero, and the polysynaptic connection should be retrievable from the path analysis. Again similar to GC, SEM is resilient to confounds only under the assumption that the model represents an isolated system, and all the relevant variables present in the environment are taken into
Moreover, to obtain the maximum likelihood solution for $B$ parameters, one needs to make a range of assumptions on the properties of the noise in the network. Typically, a Gaussian white noise is assumed, although background noise in the brain is most probably scale-free (He, 2014). Inference can be performed either through the classical hypothesis testing (as the computationally cheap version) or through model comparison (as the computationally heavier version)\(^7,8\).

In summary, SEM is a straightforward approach: it simplifies the causal inference by reducing the complex network with a low-pass filter at the output to a very simple linear system, but this simplicity comes at the cost of a number of assumptions. In the first decade of fMRI research, SEM was often a method of choice (Zhuang et al., 2008; Schlösser et al., 2008). However, recently, using Dynamic Causal Modeling has become more popular in the field. One recently proposed approach in this domain published by Schwab et al., 2018, extends linear models by introducing time-varying connectivity coefficients, which allows for tracking the dynamics of causal interactions over time. In this approach, linear regression is applied to each node in the network separately (to find causal influence of all the remaining nodes in the network on that node). The whole graph is then composed from node-specific DAGs node by node - and that compound graph can be cyclic.

### 5.2.3 Dynamic Causal Modeling

Both the aforementioned network-wise methods were developed in other disciplines, and only later applied to fMRI data. Yet, using prior knowledge about the properties of fMRI datasets can prove useful when searching for causal interactions. Dynamic Causal Modeling (DCM, Friston, Harrison, and Penny, 2003) is a hypothesis testing tool which uses state-space equations reflecting the structure of fMRI datasets. This technique was also implemented for other neural recording methods: EEG and MEG (Kiebel et al., 2008). DCM is well received within the neuroimaging community (the original article by Friston, Harrison, and Penny, 2003 had been cited over 3,800 times at the time of submitting this thesis).

In this work, I describe the original work by Friston, Moran, and Seth, 2013 because, despite multiple recent developments (Kiebel et al., 2007; Stephan et al., 2007; Marreiros, Kiebel, and Friston, 2008; Stephan et al., 2008; Li et al., 2011; Daunizeau, Stephan, and Friston, 2012; Seghier and Friston, 2013; Friston et al., 2011; Havlicek et al., 2015; Frässle et al., 2016b; Razi and Friston, 2016; Prando et al., 2017; Frässle et al., 2017), it remains the most popular version of DCM in the fMRI community. The idea of DCM is as follows. First, one needs to build a generative model (Fig. 5.2). This model has two levels of description: the neuronal level (Fig. 5.2, (iii)), and the hemodynamic level (Fig. 5.2, (v)). Both of these levels contain parameters which are not directly recorded in the experiment and need to be inferred from the data. This model reflects scientific evidence on how the BOLD response is generated from neuronal activity.

At the neuronal level of the DCM generative model, simple interactions between brain areas are posited, either bilinear (Friston, Harrison, and Penny, 2003) or nonlinear (Stephan et al., 2008). In the simplest, bilinear version of the model, the bilinear state equation reads:
5.2. Causal inference in fMRI, a Review: Network-wise methods

\[ \dot{z} = (A + \sum_j u_j B^j)z + Cu \]  

where \( z \) - the dynamics in the nodes of the network, \( u \) - the experimental inputs, \( A \) - the connectivity matrix characterizing causal interactions between the nodes of the network, \( B \) - the modulatory influence of experimental inputs on the connections within the network, \( C \) - the experimental inputs to the nodes of the network (Fig. 5.2). The hemodynamic level is more complex and follows the biologically informed Balloon-Windkessel model (Buxton, Wong, and Frank, 1998), for details please see: Chapter 6. The Balloon-Windkessel describes the BOLD signal observed in fMRI experiments as a function of neuronal activity but also region-specific and subject-specific physiological features such as the time constant of signal decay, the rate of flow-dependent elimination, and the hemodynamic transit time or resting oxygen fraction. This is a weakly non-linear model with free parameters estimated for each brain region. These parameters determine the shape of the hemodynamic response (Fig. 5.2, (iv)), which typically peaks at 4–6[s] after the neuronal activity takes place, to match the lagged oxygen consumption in the neuronal tissue mentioned in Section 5.1.2. The Balloon-Windkessel model is being iteratively updated based on new experimental findings, e.g., to mimic adaptive decreases to sustained inputs during stimulation or the post-stimulus undershoot in the hemodynamic response (Havlicek et al., 2015). I introduce the hemodynamic level of the DCM model in detail in Chapter 6.

In this Section, the deterministic, bilinear single state per region DCM will be described (Friston, Harrison, and Penny, 2003). The DCM procedure starts with defining hypotheses based on observed activations, which involves defining which regions are included in the network (usually based
Chapter 5. Causal inference in fMRI: a Review

on activations found through the General Linear Model, Friston et al., 2007) and then defining a model space based on the research hypotheses. In the latter model selection phase, a range of literature-informed connectivity patterns and inputs to the networks (referred to as "models") are posited (Fig. 5.2, (i)). The definition of a model space is the key to the DCM analysis. The models should be considered carefully in the light of the existing literature. The model space represents the formulation of a prior over models, therefore, it should always be constructed prior to the DCM analysis. Subsequently, for every model one needs to set priors on the parameters of interest: connectivity strengths and input weights in the model (Fig. 5.2, (ii)) and the hemodynamic parameters. The priors for hemodynamic parameters are experimentally informed Gaussian distributions (Friston, Harrison, and Penny, 2003). The priors for connectivity strengths are Gaussian probability distributions centered at zero (which is often referred to as conservative shrinkage priors). The user usually does not need to specify the priors as they are already implemented in the DCM algorithms.

Next, an iterative procedure is used to find the model evidence by maximizing a cost function, a so-called negative free energy (Friston and Stephan, 2007). Negative free energy is a particular cost function that gives a trade-off between model accuracy and complexity (which accounts for correlations between parameters, and for moving away from the prior distributions). During the iterative procedure, the prior probability distributions gradually shift their mean and standard deviation, and converge towards the final posterior distributions. Negative free energy is a more sophisticated approximation of the model evidence when compared to methods such as Akaike’s Information Criterion (AIC, Akaike, 1998) or Bayesian Information Criterion (BIC, Schwarz, 1978); AIC and BIC simply count the number of free parameters (thereby assuming that all parameters are independent), while negative free energy also takes the covariance between the parameters into account (Penny, 2012).

In DCM, causality is modeled as a set of upregulating or downregulating connections between the nodes. During the inference procedure, conservative shrinkage priors can shift towards both positive and negative values, which can be interpreted as effective excitation or effective inhibition (except for self connections, which are always only negative (this self inhibition is mathematically motivated: the system characterizing the fast dynamics of the neuronal network must be stable, and this requires the diagonal terms of the adjacency matrix \( A \) to be negative), Fig. 5.2, (ii), connections marked in blue). During the inference procedure, the neural and hemodynamic parameters of all models postulated for model comparison are optimized. The posterior probability distributions determine significance of all the parameters. The models can contain both uni- and bidirectional connections (Vaudano et al., 2013; Buijink et al., 2015). The estimated model evidence can then be compared. As such, the original DCM is a hypothesis-testing tool working only through model comparison. However, now, a linear version of DCM dedicated to exploratory research in large networks is also available (Frässle et al., 2016b). Testing the immediacy and resilience to confounds in DCM is possible through creating separate models and comparing their evidence. E.g., one can compare the evidence for \( X \rightarrow Y \) with evidence for \( X \rightarrow Z \rightarrow Y \) to test whether or not the connection \( X \rightarrow Y \) is direct or rather mediated by another region \( Z \). Note
that this strategy requires an explicit specification of the alternative models and it cannot take hidden causes into consideration (in this work, I refer to the original DCM implementation, Friston, Harrison, and Penny, 2003, but there are also implementations of DCM involving estimation of time-varying hidden states, such as Daunizeau, Friston, and Kiebel, 2009). However, including extra regions to increase resilience to confounds is not necessarily a good idea. Considering the potentially large number of fitted parameters per region (the minimum number of nodes per region is two hemodynamic parameters and one input/output to connect to the rest of network), this may result in a combinatorial explosion. Also, models with different nodes are not comparable in DCM for fMRI (Friston, Harrison, and Penny, 2003). Extending the models by adding additional nodes not only increases the computation time considerably. The original DCM (Friston, Harrison, and Penny, 2003) is therefore restricted to small networks of a few nodes (as mentioned previously, today, large DCMs dedicated to exploratory research in large networks are also available, Seghier and Friston, 2013; Frässle et al., 2016b).

The proper application of DCM needs a substantial amount of expertise (Stephan et al., 2010; Daunizeau, David, and Stephan, 2011). While ROIs can be defined in a data-driven fashion (through a preliminary standard General Linear Model analysis, Friston et al., 1995), the model space definition requires prior knowledge of the research problem (Kahan and Foltynie, 2013). In principle, the model space should reflect prior knowledge about possible causal connections between the nodes in the network. If a paradigm developed for the fMRI study is novel, there might be no reference study that can be used to build the model space. In that case, using family-wise DCM modeling can be helpful (Penny et al., 2010). Family-wise models group large families of models defined on the same set of nodes to test a particular hypothesis. E.g., one can explore a three node network with nodes $X$, $Y$, $Z$ and compare the joint evidence behind all the possible models that contain connection $X \rightarrow Y$ with the joint evidence behind all the possible models that contain connection $Y \rightarrow X$ (Fig. 5.2, (i)). Another solution that allows for constraining a large model space is Bayesian model averaging (BMA, Hoeting et al., 1999; Stephan et al., 2010) which explores the entire model space and returns average value for each model parameter, weighted by the posterior probability for each model. Finally, one can perform a Bayesian model reduction (Friston et al., 2016), in which the considered models are reduced versions of a full (or, "parent") model. This is possible when the priors can be reduced, e.g., when a prior distribution of a parameter in a parent model is set to a mean and variance of zero.

The number of possible models explodes with the size of the network. To extend the scope of application of the DCM analysis to larger networks, recently the new, large-scale DCM framework for resting state fMRI has been proposed (Razi et al., 2017). This framework uses the new, spectral DCM (Friston et al., 2011) designed for resting state fMRI and which is able to handle dozens of nodes in the network. Spectral DCM is then combined with functional connectivity priors to estimate the effective connectivity in the large-scale resting state networks.

There are a few points that need particular attention when interpreting the results of the DCM analysis. Firstly, in case the data quality is poor, evidence for one model over another will not be conclusive. In the worst
case, it could give a preference to the simplest model (i.e., the model with the fewest free parameters). In that case, simpler models will be preferred over more complex ones regardless of the low quality of fit. It is important, therefore, to include a “null model” in a DCM analysis, with all interesting parameters fixed at zero. This can then act as a baseline, against which models of interest can be compared (Penny, 2012).

Second, the winning model might contain parameters with a high probability of being equal to zero. To illustrate this, let us consider causal inference in a single subject (also referred to as first level analysis). Let us assume that we chose a correct set of priors (model space). The Variational Bayes procedure then returns a posterior probability distribution for every estimated connectivity strength. This distribution gives a measure of probability for the associated causal link to be larger than zero. Some parameters may turn out to have high probability of being equal to zero in the light of this posterior distribution. This may be due to the fact that the winning model is correct, but some of the underlying causal links are weak and therefore hard to confirm by the VB procedure. Also, DCM requires data of high quality; when the SNR is insufficient, it is possible that the winning model would explain a small portion of the variance in the data. In that case, getting insignificant parameters in the winning model is likely. Therefore, it is advisable to check the amount of variance explained by the winning model at the end of the DCM analysis.

The most popular implementation of the DCM estimation procedure is based on Variational Bayes (VB, Bishop, 2006) which is a deterministic algorithm. Recently, also Markov-Chain Monte Carlo (MCMC, Bishop, 2006; Sengupta, Friston, and Penny, 2015) was implemented for DCM. When applied to a unimodal free energy landscape, these two algorithms will both identify the global maximum. MCMC will be slower than VB as it is stochastic and therefore computationally costly. However, free energy landscape for multiple-node networks is most often multimodal and complex. In such case, VB - as a local optimization algorithm - might settle on a local maximum. MCMC on the other hand, is guaranteed to converge to the true posterior densities - and thus the global maximum (given an infinite number of samples).

DCM was tailored for fMRI and, unlike other methods, it explicitly models the hemodynamic response in the brain. The technique tends to return highly reproducible results, and is therefore statistically reliable (Schuyler et al., 2010; Rowe et al., 2010; Bernal-Casas et al., 2013; Tak et al., 2018). Recent longitudinal study on spectral DCM in resting state revealed systematic and reliable patterns of hemispheric asymmetry (Almgren et al., 2018). DCM also yielded high test-retest reliability in an fMRI motor task study (Frässle et al., 2015), in a face perception study (Frässle et al., 2016a), in facial emotion perception study (Schuyler et al., 2010) and in a finger-tapping task in a group of subjects suffering from Parkinson’s disease (Rowe et al., 2010). It has also been demonstrated most reliable when directly compared to GC and SEM (Penny et al., 2004). Furthermore, the DCM procedure can provide complimentary information to GC (Friston, Moran, and Seth, 2013): GC models dependency among observed BOLD responses whereas DCM models coupling among the hidden states generating observations. GC seems to be equally effective as DCM in certain circumstances, such as when the hemodynamic response function (HRF) is deconvolved from
the data (David et al., 2008; Ryali et al., 2011; Ryali et al., 2016; Wang et al., 2016). Importantly, the face validity of DCM was examined on experimental datasets coming from interventional study with the use of rat model of epilepsy (David et al., 2008; Papadopoulou et al., 2015).

On the other hand, proper use of DCM requires knowledge on the biology and on the inference procedure. DCM also has limitations in terms of the size of the possible models. Modeling a large network may run into problems with identifiability - there will be many possible combinations of parameter settings which could give rise to the same or similar model evidence. In other words, strong covariance between parameters will preclude confident estimates of the strength of each connection. One possible remedy for this, in the context of large scale networks, is to impose appropriate prior constrains on the connections - e.g., using priors based on functional connectivity (Razi et al., 2017). Large networks may also give rise to comparisons of large number of different models with varying combinations of connections. To reduce the possibility of overfitting at the level of model comparison - i.e., finding a model which is appropriate for one subject or group of subjects’ data, but not for others - it can be useful to group the models into a small number of families (Penny et al., 2010) based on pre-defined hypotheses. More information on the limitations of DCM can be found in work by Daunizeau, David, and Stephan, 2011.

However, DCM was further developed into multiple procedures including more sophisticated generative models than the original model discussed here. The field of DCM research in fMRI is still growing (Friston et al., 2017). The DCM generative model is continuously being updated in terms of the structure of the generative model (Havlicek et al., 2015), the estimation procedure (Sengupta, Friston, and Penny, 2015), and the scope of the possible applications (Friston et al., 2017).

5.3 Hierarchical network-wise models

The second group of methods involves hierarchical network-wise models: Linear Non-Gaussian Acyclic Models (LiNGAM, Shimizu et al., 2006) and Bayesian Nets (BNs, Frey and Jojic, 2005). Similarly as network-wise methods reviewed in the previous Section, these methods are also multivariate but with one additional constraint: the network can only include feed forward projections (and therefore, no closed cycles). Consequently, the resulting models have a hierarchical structure with feed forward distribution of information through the network.

5.3.1 LiNGAM

The Linear Non-Gaussian Acyclic Model (LiNGAM, Shimizu et al., 2006) is an example of a data-driven approach working under the assumption of acyclicity (Thulasiraman and Swamy, 1992). The model itself is simple: every time course within an ROI \( X_i(t) \) is considered to be a linear combination

\[ X_i(t) = \sum_j A_{ij} X_j(t) + \epsilon_i(t) \]

1A critical note on limitations of DCM in terms of network size can also be found in (Lohmann et al., 2012). See also a response to this article, (Friston, Daunizeau, and Stephan, 2013; Breakspear, 2013).
of all other signals with no time lag:

\[ \vec{X}(t) = B\vec{X}(t) + \vec{\sigma}(t) \]  

(5.5)

where \( B \) - the matrix containing the connectivity weights, \( \vec{\sigma} \) - the neuronal noise. The model is in principle the same as in SEM (Section 5.2.2), but the difference lies in the inference procedure: whereas in SEM, inference is based on minimizing the variance of the residual noise under the assumption of independence and Gaussianity, LiNGAM finds connections based on the dependence between residual noise components \( \vec{\sigma}(t) \) and regressors \( \vec{X}(t) \).

The rationale of this method is as follows (Fig. 5.3). Let us assume that the network is noisy, and every time series within the network is associated with a background noise uncorrelated with the signal in that node. An example of such a mixture of signal with noise is given in Fig. 5.3A. Then, let us assume that \( \vec{X}(t) \) - which is a mixture of signal \( X(t) \) and noise \( \sigma_X(t) \) - causes \( Y(t) \). Then, as it cannot distinguish between the signal and the noise, \( Y(t) \) becomes a function of both these components. \( Y(t) \) is also associated with noise \( \sigma_Y(t) \), however, as there is no causal link \( Y \rightarrow X \), \( X(t) \) is not dependent on the noise component \( \sigma_Y(t) \). Therefore, if \( Y \) depends on the \( \sigma_X(t) \) component, but \( X \) does not depend on the \( \sigma_Y(t) \) component, one can infer projection \( X \rightarrow Y \).

An example of such a simple, directed causal relationship between two variables is demonstrated in Fig. 5.3B: the relationship between age and length in fish. If fish length is expressed in a function of fish age (upper panel), the residual noise in the dependent variable (length) is uncorrelated with the independent variable (age). Therefore, the noise variance is constant over a large range of fish age. On the contrary, once the variables are flipped and fish age becomes a function of fish length (lower panel), the noise variance becomes dependent on the independent variable (length) as it is small for small values of fish length and large for the large values of fish length. Therefore, the first causal model (the fish age influencing the fish length) is correct.

In applications to causal research in fMRI, the LiNGAM inference procedure is often accompanied by an Independent Component Analysis (ICA, Hyvärinen and Oja, 2000) as follows. The connectivity matrix \( B \) in Eq. 5.5 describes how signals in the network mix together. By convention, not \( B \) itself but a transformation of \( B \) into

\[ A = (1 - B)^{-1} \]  

(5.6)

is used as a mixing matrix in the LiNGAM inference procedure. By using this mixing matrix \( A \), one can look at Eq. 5.5 in a different way:

\[ \vec{X} = A \vec{\sigma} \]  

(5.7)

Now, the BOLD time course in the network \( \vec{X}(t) \) can be represented as a mixture of independent sources of noise \( \vec{\sigma}(t) \). This is the well known cocktail party problem, and it was originally described in acoustics (Bronkhorst, 2000): in a crowded room, a human ear registers a linear combination of the noises coming from multiple sources. To decode the components of this cacophony, the brain needs to perform a blind source separation (Comon and
5.3. Causal inference in fMRI, a Review: Hierarchical network-wise models

Figure 5.3: Linear Non-Gaussian Acyclic Model (LiNGAM). A: The noisy time series $\hat{X}(t)$ consists of signal $X(t)$ and noise $\sigma_X(t)$. $Y(t)$ thus becomes a function of both the signal and the noise in $\hat{X}(t)$. B: Causal inference through the analysis of the noise residuals (figure reprinted from http://videolectures.net/bbci2014_grosse_wentrup_causal_inference/). The causal link from age to length in a population of fish can be inferred from the properties of the residual noise in the system. The relationship between age and length in fish. If fish length is expressed in a function of fish age (upper panel), the residual noise in the dependent variable (length) is uncorrelated with the independent variable (age); the noise variance is constant over a large range of fish age (red bars). On the contrary, once the variables are flipped and fish age becomes a function of fish length (lower panel), the noise variance becomes dependent on the independent variable (length) as it is small for small values of fish length and large for the large values of fish length (red bars).

Jutten, 2010): to decompose the incoming sound into a linear mixture of independent sources of sounds. In the LiNGAM procedure, Independent Component Analysis (ICA, Hyvärinen and Oja, 2000) is used to approach this issue. ICA assumes that the noise components $\vec{\sigma}$ are independent and have a non-Gaussian distribution, and finds these components as well as the mixing matrix $A$ through dimensionality reduction with Principal Component Analysis (Jolliffe, 2002; Shlens, 2014). From this mixing matrix, one can in turn estimate the desired adjacency matrix $B$ with the use of Eq. 5.6.

Since the entries $B_{ij}$ of the connectivity matrix $B$ can take any value, LiNGAM can retrieve both excitatory and inhibitory connectivity $1$ of any strength $2$. The author of LiNGAM (Shimizu, 2014) recommends performing significance testing through either bootstrapping (Hyvärinen et al., 2010; Komatsu, Shimizu, and Shimodaira, 2010; Thamvitayakul et al., 2012) or...
permutation testing (Hyvärinen and Smith, 2013). However, LiNGAM requires the assumption of acyclicity, therefore only unidirectional connections can be picked up. Moreover, the connectivity matrix revealed with the use of LiNGAM is meant to pick up on direct connections. The original formulation of LiNGAM assumes no latent confounds (Shimizu et al., 2006), but the model can be extended to a framework that can capture the causal links even in the presence of (unknown) hidden confounds (Hoyer et al., 2008; Chen and Chan, 2013). LiNGAM-ICA’s causal inference consists of ICA and a simple machine learning algorithm, and, as such, it is a fully data-driven strategy that does not involve model comparison. Confidence intervals for the connections $B$ can be found through permutation testing. ICA itself can be computationally costly and its computational stability cannot be guaranteed (the procedure that searches for independent sources of noise can get stuck in a local minimum). Therefore, the computational cost in LiNGAM can vary depending on the dataset. This also sets a limit on the potential size of the causal network. When the number of connections approaches the number of time points (degrees of freedom), the fitting procedure will become increasingly unstable as it will be overfitting the data.

When tested on synthetic fMRI benchmark datasets (Smith et al., 2011), LiNGAM-ICA performs relatively good, but is more sensitive to confounding variables than several other methods discussed in this Section, such as Patel’s tau or GC. However, as LiNGAM performs particularly well for datasets containing a large number of samples, the authors suggested that a group analysis could resolve the sensitivity problem in LiNGAM. The concept was then picked up and developed by at least two groups. Firstly, Ramsey, Hanson, and Glymour, 2011 proposed LiNG Orientation, Fixed Structure technique (LOFS). The method is inspired by LiNGAM, and uses the fact that, within one graph equivalence class, the correct causal model should return maximally non-Gaussian conditional probability distributions. LOFS was tested on the synthetic benchmark datasets, where it achieved performance very close to 100%. Secondly, Xu et al., 2014 published a pooling-LiNGAM technique, which is a classic LiNGAM-ICA applied to the surrogate datasets. Validation on synthetic datasets revealed that both False Positive (FP) and False Negative (FN) rates decrease exponentially along with the length of the (surrogate) time series, however, combining time series of as long as 5,000 samples is necessary for this method to give both FP and FN as a reasonable level of 5%.

Despite promising results obtained in the synthetic datasets, LiNGAM is still rarely applied to causal research in fMRI to date.

### 5.3.2 Bayesian nets

The use of the LiNGAM inference procedure assumes a linear mixing of signals underlying the causal interaction. Model-free methods do not make this assumption: the bare fact that one is likely to observe $Y$ given the presence of $X$ can indicate that the causal link $X \rightarrow Y$ exists (Fig. 5.4). Let us assume the simplest example: causal inference for two binary signals $X(t)$, $Y(t)$. In a binary signal, only two values are possible: 1 and 0; 1 can be interpreted as an “event” while 0 - as “no event.” Then, if in signal $Y(t)$, events occur in 80% of the cases when events in signal $X(t)$ occur (Fig. 5.4A), but the...
opposite is not true, the causal link $X \rightarrow Y$ is likely. Computing the odds of events given the events in the other signal, is sufficient to establish causality. In a model-based approach on the other hand, a model is fitted to the data, to establish the precise form of the influence of the independent variable $X$ on the dependent variable $Y$.

Note that both model-based and model-free approaches contain a measure of uncertainty, but this uncertainty is computed differently. In model-based approaches, p-values associated with the fitted model are a measure of confidence that the modelled causal link is a true positive (Fig. 5.4A, left panel). In contrast, in model-free approaches this confidence is calculated directly by quantifying causal relationships in terms of conditional probabilities (Fig. 5.4A, right panel). In practice, since the BOLD response - unlike the aforementioned example of binary signals - takes continuous values, estimating conditional probabilities is based on the joint distribution of the variables $X$ and $Y$ (Fig. 5.4B). Conditional probability $P(Y|X)$ becomes a distribution of $Y$ when $X$ takes a given value. Bayesian Networks (BNs, Frey and Jojic, 2005) are based on such a model-free approach (Fig. 5.4C).

The causal inference in BNs is based on the concept of conditional independence (a.k.a. Causal Markov Condition, Hausman and Woodward, 1999). Suppose that there are two events that could independently cause the grass to get wet: either a sprinkler, or rain. When one only observes the grass being wet, the direct cause for this event is unknown. However, once rain is observed, it becomes less likely that the sprinkler was used. Therefore, one can say that the variables $X_1$ (sprinkler) and $X_2$ (rain) are conditionally dependent given variable $X_3$ (wet grass), because $X_1$, $X_2$ become dependent on each other after information about $X_3$ is provided. In BNs, the assumption of conditional dependency in the network is used to compute the joint probability of a given model - i.e., the model evidence (once variables $X_i$ are conditionally dependent on $X_j$, the joint distribution $P(X_i, X_j)$ factorizes into a product of probabilities $P(X_j)P(X_i|X_j)$).

Implementing a probabilistic BN requires defining a model: choosing a graph of "parents" who send information to their "children." E.g., in Fig. 5.4C, (i), the node $X_1$ is a parent of nodes $X_4$ and $X_5$, and the node $X_4$ is a child of nodes $X_1$, $X_2$ and $X_3$. The joint probability of the model can then be computed as the product of all marginal probabilities of the parents and conditional probabilities of the children given the parents. Marginal probability $P(X_j)$ is the total probability that the variable of interest $X_j$ occurs while disregarding the values of all the other variables in the system. E.g., in Fig. 5.4C, (i), $P(X_1)$ means a marginal probability of $X_1$ happening in this experiment. Conditional probability $P(X_i|X_j)$ is the probability of a given variable ($X_i$) occurring given that another variable has occurred ($X_j$). E.g., in Fig. 5.4C, (i), $P(X_5|X_1, X_3)$ means a conditional probability of $X_5$ given its parents $X_1$ and $X_3$. 
Chapter 5: Causal inference in fMRI: a Review

FIGURE 5.4: Bayesian nets. A: Model-based versus model-free approach. \( \beta \) - a regressor coefficient fitted in the modeling procedure. \( \sigma(t) \) - additive noise. Both model-based and model-free approach contain a measure of confidence. In the model-based approach, a model is fitted to the data and p-values associated with this fit provide a measure of confidence that the causal link exists (i.e., is a true positive, left panel). In the model-free approach, this confidence is quantified directly by expressing causal relationships in terms of conditional probabilities (right panel). B: Conditional probability for continuous variables. Since BOLD fMRI is a continuous variable, the joint probability distribution for variables \( X \) and \( Y \) is a two-dimensional distribution. Therefore, conditional probability of \( P(Y|X = x) \) becomes a distribution. C: (i) An exemplary Bayesian net. \( X_1, X_2, X_3 \) - parents, \( X_4, X_5 \) - children. (ii) Competitive Bayesian nets: one can define competitive models (causal structures) in the network and compare their joint probability derived from the data. (iii) Cyclic belief propagation: if there was a cycle in the network, the expression for the joint probability would convert into an infinite series of conditional probabilities.

Then, once the whole graph is factorized into the chain of marginal and conditional probabilities, the joint probability of the model can be computed as the product of all marginal and conditional probabilities. E.g., in Fig. 5.4C, (i), the joint probability of the model \( M \) yields

\[
P(M) = P(X_1)P(X_2)P(X_3)P(X_4|X_1,X_2,X_3)P(X_5|X_1,X_2,X_3)
\]

Finally, there are at least three possible approaches to causal inference with BNs:

1. Model comparison: choosing the scope of possible models (by defining their structure a priori), and comparing their joint probability. Mind that in this case, the algorithm will simply return the winning graphical model, without estimation of the coefficients representing connection weights,
2. Assuming one model structure a priori, and only inferring the weights. This is common practice, related to, e.g., Naive Bayes (Bishop, 2006) in which the structure is assumed, and the connectivity weights are estimated from conditional probabilities. In this case, the algorithm will assume that the proposed graphical model is correct, and infer the connection weights only,

3. Inferring the structure of the model from the data in an iterative way, by using a variety of approximate inference techniques that attempt to maximize posterior probability of the model by minimizing a cost function called free energy (Frey and Jojic, 2005, similar to DCM): expectation maximization (EM, Bishop, 2006; Dempster, Laird, and Rubin, 1977), variational procedures (Jordan et al., 1998), Gibbs sampling (Neal, 1993), or the sum-product algorithm (Kschischang, Frey, and Loeliger, 2001).

BNs can detect both excitatory and inhibitory connections $X \rightarrow Y$, depending on whether the conditional probability $p(Y|X)$ is higher or lower than the marginal probability $p(X)$. Like LiNGAM, in general, BNs cannot pick up on bidirectional connections. The assumption of acyclicity comes from the cyclic belief propagation (Fig. 5.4C, (iii)): the joint probability of a cyclic graph would be expressed by an infinite chain of conditional probabilities which usually does not converge into a closed form. In general, this restricts the scope of possible models to Directed Acyclic Graphs (DAGs, Thulasiraman and Swamy, 1992). However, there are also implementations of BNs that cope with cyclic propagation of information throughout the network, e.g., Cyclic Causal Discovery algorithm (CCD, Richardson and Spirtes, 2001). This algorithm is not often used in practice though as it works in the large sample limit, requires assumption on the graph structure and retrieves a complex output. The value of conditional probability $P(Y|X)$ can be a measure of a connection strength. We can consider conditional probabilities significantly higher than chance as an indication for significant connections. In principle, BNs are not resilient to latent confounds. However, some classes of algorithms were designed especially to tackle this problem, such as Stimulus-based Causal Inference (SCI, Grosse-Wentrup et al., 2016), Fast Causal Inference (FCI, Spirtes, Glymour, and Scheines, 1993; Zhang, 2008), and Greedy Fast Causal Inference (GFCI, Ogarrio, Spirtes, and Ramsey, 2016). BNs can either work through model comparison or as an exploratory technique. In the first case, BNs involve model specification which - like in DCM - a priori knowledge about the experimental paradigm is required. In the latter case, the likelihood is intractable and can only be approximated (Diggle, 1984). In principle, networks of any size can be modeled with BNs, either through a model comparison or through exploratory techniques. However, the exploratory techniques typically minimize the cost function during the iterative search for the best model. Since together with the growing network size, the landscape of the cost function becomes multidimensional and complex, the algorithm is more likely to fall into a local minimum.

What can also become an issue while using BNs in practice, is the fact that multiple BN algorithms return an equivalence class of a graph: the set of all graphs that are indistinguishable from the true causal structure on the basis of their sole probabilistic independences (Spirtes, 2010). These
structures cannot be further distinguished without further assumptions or experimental interventions. For finite data, taking even one wrong assumption upon the directionality of causal link in the graph can be propagated through the network, and cause a number of incorrect orientations (Spirtes, 2010). One approach designed to overcome this issue is the Constraint-Based Causal Inference (BCCD, Claassen and Heskes, 2012). In this approach, Bayesian Inference is employed to estimate the reliability of a set of constraints. This estimation can further be used to decide whether this prior information should be used to determine the causal structure in the graph.

BNs cope well with noisy datasets, which makes them an attractive option for causal research in fMRI (Mumford and Ramsey, 2014). Smith et al., 2011 tested multiple implementations of Bayesian nets, including FCI, CCD, as well as other algorithms: Greedy Equivalence Search (GES, Meek, 1995; Chickering, 2002), "Peter and Clark" algorithm (PC, Meek, 1995), and its conservative version (CPC, Ramsey, Zhang, and Spirtes, 2006). All these implementations performed similarly, which was quite well with respect to estimating the existence of connections, but not to the directionality of the connections.

BNs are not widely used in fMRI research up to date, the main reason being the assumption of acyclicity. One exception is Fast Greedy Equivalence Search (FGES, Ramsey, Sanchez-Romero, and Glymour, 2014; Ramsey, 2015; Ramsey et al., 2017), a variant of GES optimized to large graphs. The algorithm assumes that the network is acyclic with no hidden confounders, and returns an equivalence class for the graph. In a recent work by Dubois et al., 2017, FGES was applied with the use of new, computational-experimental approach to causal inference from fMRI datasets. In the initial step, causal inference is performed from large observational resting-state fMRI datasets with the use of FGES to get the aforementioned class of candidate causal structures. Further steps involve causal inference in a single patient informed by the results of the initial analysis, and interventional study with the use of an electrical stimulation to determine which of the equivalent structures revealed by FGES can be associated with a particular subject.

### 5.4 Pairwise inference

The last group of methods reflects the most recent trends in the field of causal inference in fMRI. This family of methods is represented by Patel’s tau (Patel, Bowman, and Rilling, 2006), and involves a two-stage inference procedure. In the first step, functional connectivity is used to find connections, without assessing their directionality. Unlike network-wise methods that eliminate insignificant connections post-hoc, pairwise methods eliminate insignificant connections prior to causal inference. In the second step, each previously found connection is analyzed separately, and the two nodes involved are classified as an upstream or downstream region. These methods do not involve assumptions on the global patterns of connectivity at the network level (recurrent versus feed forward). However, they involve the assumption that the connections are non-transitive: if \( X \) projects to \( Y \), and \( Y \) projects to \( Z \), it does not imply that \( X \) projects to \( Z \). The causal inference is based on the pairs of nodes only, and this has consequences for the interpretation of the network as a whole. As there is uncertainty associated
with estimation of every single causal link, the probability that all connections are correctly estimated decreases rapidly with the number of nodes in the network.

5.4. Transfer Entropy

Transfer Entropy (TE, Schreiber, 2000) is another data-driven technique, equivalent to GC under Gaussian assumptions (Barnett, Barrett, and Seth, 2009), and asymptotically equivalent to GC for general Markovian (nonlinear, non-Gaussian) systems (Barnett and Bossomaier, 2012). Namely, TE is a non-parametric form of GC (or, GC is a parametric form of TE). It was originally defined for pairwise analysis, and later extended to multivariate analysis (Lizier, Prokopenko, and Zomaya, 2008; Montalto, Faes, and Marinazzo, 2014). TE is based on the concept of Shannon entropy (Shannon, 1948). Shannon entropy $H(x)$ quantifies the information contained in a signal of unknown spectral properties as the amount of uncertainty, or unpredictability. E.g., a binary signal that only gets values of 0 with a probability $p$, and values of 1 with a probability $1-p$, is most unpredictable when $p = 0.5$. This is because there is always exactly a 50% chance of correctly predicting the next sample. Therefore, being informed about the next sample in a binary signal of $p = 0.5$ reduces the amount of uncertainty to a higher extent than being informed about the next sample in a binary signal of, say, $p = 0.75$. This can be interpreted as a larger amount of information contained in the first signal as compared to the latter. The formula which quantifies the information content according to this rule reads as follows:

$$H(X) = - \sum_i P(x_i) \log_2 P(x_i)$$  \hspace{1cm} (5.9)\

where $x_i$ are the possible values in the signal (for the binarized signal, there are only two possible values: 0 and 1).

TE builds up on the concept of Shannon entropy by extension to conditional Shannon entropy: it describes the amount of uncertainty reduced in future values of $Y$ by knowing the past values of $X$ along with the past values of $Y$:

$$TE_{X \rightarrow Y} = H(Y|Y_{t-\tau}) - H(Y|X_{t-\tau}, Y_{t-\tau})$$  \hspace{1cm} (5.10)\

where $\tau$ - the time lag.

In theory, TE requires no assumptions about the properties of the data, not even signal stationarity although in most real-world applications, stationarity is required to almost the same extent as in GC. Certain solutions for TE in non-stationary processes are available though (Wollstadt et al., 2014). TE does need an a priori definition of the causal process, and it may work for both linear and nonlinear interactions between the nodes.

TE can distinguish the signum of connections, as the change in the Shannon entropy can be both positive and negative. Furthermore, the absolute value of the drop in the Shannon entropy can provide a measure of the connection strength. TE can also distinguish bidirectional connections, as in this case, both $TE_{X \rightarrow Y}$ and $TE_{Y \rightarrow X}$ will be nonzero. In TE, significance testing by means of permutation testing is advised (Vicente et al., 2011).
Immediacy and resilience to confounds in TE depends on the implementation to a large extent: using a simple Pearson’s correlation to compute functional connectivity increases the amount of spurious (indirect) connections, whereas partial correlation is meant to pick up on direct connections only\textsuperscript{5,6}. The inference in TE is performed through classical hypothesis testing\textsuperscript{7} and is highly cost-efficient\textsuperscript{8}. As in GC, the maximum number of regions in the network divided by the number of shifts can never exceed the number of time points (degrees of freedom)\textsuperscript{9}.

TE is a straightforward and computationally cheap method (Vicente et al., 2011). However, it struggled when applied to synthetic fMRI benchmark datasets (Smith et al., 2011). One reason for this could be the time lag embedded in the inference procedure, which is an obstacle to TE in fMRI research for the same reasons as for GC: it requires at least one full TR. TE is nevertheless gaining interest in the field of fMRI (Sharaev, Ushakov, and Velichkovsky, 2016; Lizier et al., 2011; Ostwald and Bagshaw, 2011; Chai et al., 2009; Montalto, Faes, and Marinazzo, 2014).

5.4.2 Pairwise Likelihood Ratios

As mentioned before, the multi-step procedure to causal inference in fMRI was first proposed by Patel (as Patel’s tau, PT, Patel, Bowman, and Rilling, 2006). The first step involves identifying the (undirected) connections by means of functional connectivity, and is achieved on the basis of correlations between the time series in different regions. This step results in a binary graph of connections, and the edges identified as empty are disregarded from further considerations, because if there is no correlation, there is no causation.

The second step determines the directionality in each one of the previously detected connections. The causal inference boils down to a two-node Bayesian network as the whole concept is based on a simple observation: if there is a causal link \(X \rightarrow Y\), \(Y\) should get a transient boost of activity every time \(X\) increases activity. And vice versa: if there is a causal link \(Y \rightarrow X\), \(X\) should react to the activation in \(Y\) by increasing activity. Therefore, one can threshold the signals \(X(t), Y(t)\), and compute the difference between conditional probabilities \(P(Y|X)\) and \(P(X|Y)\). Three scenarios are possible:

1. \(P(Y|X)\) equals \(P(X|Y)\): it is a bidirectional connection \(X \leftrightarrow Y\) (since empty connections were filtered out in the previous step),
2. The difference between \(P(Y|X)\) and \(P(X|Y)\) is positive: the connection \(X \rightarrow Y\) is likely,
3. The difference between \(P(Y|X)\) and \(P(X|Y)\) is negative: the connection \(Y \rightarrow X\) is likely.

Building on the concept of PT, the Pairwise Likelihood Ratios methodology (PW-LR, Hyvärinen and Smith, 2013) was proposed. The authors improved on the second step of the inference by analytically deriving a classifier to distinguish between two causal models \(X \rightarrow Y\) and \(Y \rightarrow X\), which corresponds to the LiNGAM model for two variables. The authors compared the likelihood of these two competitive models derived under
Chapter 5

5.4. Causal inference in fMRI, a Review: Pairwise inference

LiNGAM’s assumptions (Hyvärinen et al., 2010), and provided with a cumulant based approximation to their ratio. In particular, the authors focused on the approximation of the likelihood ratios with third cumulant for variables $X$ and $Y$, which is the asymmetry between first (the mean) and second (the variance) moment of the distributions of variables $X$ and $Y$ (this version of the method is referred to by the authors as "PW-LR skew"):

$$C_3 = \frac{1}{N} \sum_{i=1}^{N} (X(i)Y(i)^2 - X(i)^2Y(i))$$

(5.11)

Then, if the value of this cumulant is positive, it indicates for the connection $X \rightarrow Y$, and backwards otherwise. Additionally, the authors proposed a modified version of the third cumulant, including a nonlinear transformation of the signal to improve resilience against outliers in the signal (and referred to this modified metric as "PW-LR r skew"). Additionally, the authors also introduced a version based on fourth cumulant (referred to as "PW-LR kurtosis").

PW-LR methods cannot distinguish between excitation and inhibition, but provide a quantitative measure for the strength of the connection. The authors recommended to test significance of PW-LR results through permutation testing (Hyvärinen and Smith, 2013). Following the interpretation from Patel, it is possible to distinguish between uni- and bidirectionality (since scores close to zero might indicate the bidirectionality). The authors proposed using partial correlation instead of Pearson’s correlation in the first step of the causal inference, which aims to focus on the direct connections in the network only. As for the resilience to confounds, PW-LR methods were tested on benchmark datasets for which common inputs to the nodes of the network were introduced (Smith et al., 2011, simulation no 12). PW-LR gave much better performance than the best competitors (LiNGAM-ICA and PT) and reached as much as 84% of correctly classified connections across all the benchmark data. In the original formulation, PW-LR involves a point estimate for the strength of effective connectivity, and lacks estimation of confidence intervals. In such cases, in fMRI studies, estimating confidence intervals is performed in a data-driven fashion. This is typically achieved by means of permutation testing (Hyvärinen and Smith, 2013; Smith et al., 2011, but can also be approached with the use of mixture modeling, Bielczyk et al., 2018). PW-LR, as a closed form solution, is computationally cheap. As the pair-by-pair inferences do not require network fitting procedures, this can easily be applied to larger networks.

On the benchmark datasets, all versions of PW-LR were performing very well, as contrasted with the best competitors: PT and LiNGAM (and, "PW-LR r skew" was giving the best results). In all but one out of 28 simulations PW-LR was performing highly above chance, and in a few cases they even reached 100% accuracy. However, PW-LR has never been validated on the real fMRI datasets.

A number of methods have been discussed, but the search for new ways of extracting causal information from fMRI data is still on, of which I want to highlight four representatives. First of all, one can introduce more prior knowledge into the equation. This is done in laminar analysis where the layered structure of the cortex is assumed to contain information about the signal.
5.4.3 Laminar analysis

Advancements in fMRI acquisition have made it possible to scan at sub-millimetre resolution, which opens up the possibility of a layer specific examination of the BOLD signal. As the different layers of the cortex receive and process feed forward and feedback information largely in different layers (Felleman and Essen, 1991; Bastos et al., 2015, e.g. these different processes could be visible in the laminar BOLD response. In rat studies, the BOLD response was indeed shown to have laminar specificity and have its onset in the input layer of rat motor and somatosensory cortex (Yu et al., 2014). And also in humans, several studies suggest laminar specificity of feedback processes (Kok et al., 2016; Muckli et al., 2015).

These results suggest that human laminar BOLD signal may contain directional and causal information. Hitherto, only single region laminar fMRI has been employed to date, but it may well be worthwhile to investigate how output layers of one region influence the input layer of the other.

5.4.4 Fractional cumulants

Certain new methods take a more statistical approach to neuroimaging data. E.g., characterizing the shape of BOLD distributions by means of fractional moments of the BOLD distribution combined into cumulants (see: Chapter 7) can improve on the classification of the two nodes within one connection into an upstream and a downstream node. As mentioned in Chapter 3, fractional moments of a distribution are a mathematical concept with limited practical interpretation, but could still contain valuable (causal) information.

In this method, a classification procedure using fractional cumulants derived from BOLD distribution is developed. The classifier is informed by the DCM generative model. The initial results show that the causal classification scores similarly or better than competitive methods when applied to low-noise benchmark synthetic datasets (Smith et al., 2011), and its performance is, in general, similar to "PW-LW r-skew." However, the difference shows up after imposing higher level neuronal noise on the network: the fractional cumulant-based classifier is the most robust approach in presence of such natural confounds. However, validation on real fMRI datasets for this method is still pending.

5.4.5 Rendering whole-brain effective connectivity with the use of covariance matrices

Recent approach to causal inference in fMRI involves inferring directionality of information transfer by using a set of covariance matrices with both zero and nonzero time lags (Gilson et al., 2016). The authors build a dynamic model of the brain network and optimize the effective connectivity (adjacency matrix) such that the model covariances reproduce the empirical fMRI/BOLD covariance matrices. In this way, the fitted model best matches the BOLD dynamics with respect to the second-order statistics. The authors validate the model in synthetic datasets, and apply to experimental fMRI datasets, using diffusion-weighted MRI imaging to constrain the network connectivity. The concept of lagged covariance matrices was also used to
evaluate the difference in cortical activation between two behavioral conditions (e.g., while watching movies, Gilson et al., 2017).

As this method incorporates lags, it has similar limitations as other lagged methods (such as GC or TE): it becomes lag-dependent. The authors theoretically demonstrate that for accuracy of the directed connectivity estimation, the time lag must be matched with the time constant of the underlying dynamical system representing the network. How to achieve the accuracy to fulfill this requirement in practice, remains an open research question.

Another recent contribution in this domain by Schiefer et al., 2018 focuses on inferring causal connections from resting state fMRI datasets (and other continuous time series coming from non-interventional studies) based on the assumption that the symmetric, non-lagged covariance matrix derived from the observed activity contains footprints of the direction and the sign of sparse directed connections. This underlying sparse structure is found via L1-minimization with a gradient descent, which allows for obtaining asymmetric output connectivity matrix from the initial symmetric covariance structure. In the process, the method utilizes the fact that in case of a collider present in the network (X and Y projecting to the same node Z), projecting nodes X and Y have a positive covariance which indicates for a particular motif in the covariance structure. The validation on ground truth synthetic datasets derived from a simple Ornstein-Uhlenbeck resulted in impressive results. On the other hand, application to the experimental fMRI datasets brought more vague results. Therefore, the method requires more exploration in the fMRI datasets.

5.4.6 Neural Network Models

Another recent development relevant to the problem of causal inference is the approach of implementing neural-network models to perform a complex task that is emblematic of human cognition (most commonly visual object recognition). It is then possible to study the functional architecture and representational space of such models and attempt to draw insight from optimal model parameters as to how such tasks are implemented in the human brain. In recent years neural-network models designed to recognise objects have reached human levels of performance (Krizhevsky, Sutskever, and Hinton, 2012; Kriegeskorte, 2015) and the potential of using these as models of how biological brains represent object space became a tangible goal. Early studies on feed forward neural networks that has been replicated across multiple studies and have shown that the closer the representational space a model uses resembles inferior temporal cortex fMRI activity the better the model performs (Yamins et al., 2013; Yamins et al., 2014; Khaligh-Razavi and Kriegeskorte, 2014). Of particular interest is the finding that object representations in neural-network models correlate with human brain representations in a hierarchical fashion, a result shown across both spatial and temporal dimensions (Cichy et al., 2016). While care must be taken not to over-interpret the generalisability of such models, these promising findings indicate that neural-network models may be able to provide insight into the fundamental constraints of certain computational processes which in turn can be applied to determining functional (and causal) relationships in human cognition.
5.5 Summary

The characteristics of all the discussed methods are summarized in the following table:

<table>
<thead>
<tr>
<th>Feature</th>
<th>GC</th>
<th>SEM</th>
<th>DCM</th>
<th>LN</th>
<th>BN</th>
<th>TE</th>
<th>PW-LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>group of methods</td>
<td>net</td>
<td>net</td>
<td>net</td>
<td>dag</td>
<td>dag</td>
<td>net</td>
<td>pw</td>
</tr>
<tr>
<td>sign of connections</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>directionality</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>connection strength</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>immediacy</td>
<td>+/-</td>
<td>+/-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+/-</td>
<td>+</td>
</tr>
<tr>
<td>resilience to confounds</td>
<td>+/-</td>
<td>+/-</td>
<td>-</td>
<td>+/-</td>
<td>+/-</td>
<td>+/-</td>
<td>+</td>
</tr>
<tr>
<td>causality through...</td>
<td>c</td>
<td>mc/c</td>
<td>mc</td>
<td>ml+c</td>
<td>mc/ml</td>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td>computational cost</td>
<td>l</td>
<td>l/h</td>
<td>h</td>
<td>l/h</td>
<td>c</td>
<td>l</td>
<td>l</td>
</tr>
<tr>
<td>model-free?</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>prespecify the graph?</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>regression in time</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 5.1: Summary for all the methods discussed in this Chapter. GC: Granger causality, SEM: Structural Equation Modeling, DCM: Dynamic Causal Modeling, LN: LINGaM, BN: Bayesian nets, TE: Transfer Entropy, PW-LR: Pairwise Likelihood Ratios, net: network-wise, dag: Directed Acyclic Graphs only, pw: pairwise, +/-: depends on implementation, mc: model comparison, c: classical hypothesis testing, ml: machine learning, l: low, h: high, n/a: non-applicable. PW-LR is based on the same concept as Patel’s tau (PT), and the inference is the same, therefore I did not add a separate column for PT.

5.6 Discussion

In this Chapter, I focused on discussing methods with respect to the causal structure imposed on the brain. According to this criterion, the methods fall into three categories. Network-wise methods, such as GC or SEM, do not restrict the connectivity patterns whereas Directed Acyclic Graphs (DAGs), such as BNs, assume a hierarchical structure and unidirectional connections. In the latter category, a primary node receives input from outside the network and distributes information downstream throughout the network. This may be a good approximation for many processes, (see, e.g., recent work on the visual cortex, Michalareas et al., 2016). However, the feed forward structure assumes a strictly hierarchical organization which limits its capacity to model communication between different brain networks. Under what circumstances DAGs can be an accurate representation for causal structures in the brain, remains an open question.

Next to network-wise methods and DAGs, I also discussed a third group of methods, referred to as "pairwise." In this approach, the causal inference is performed by splitting the inference into many pairwise inferences. Prior to this, the dimensionality is reduced based on functional connectivity, based on the idea that (partial) correlation is a good indicator for the existence of causal links (Smith et al., 2011) and therefore allows for simplifying the problem, both computationally and conceptually. Since the inference in this class of methods is split into a set of pairwise inferences, it is important to be aware of the fact that the confidence levels are also obtained...
connection by connection. Therefore, for a network represented by a set of connections with p-values $p_i$, the joint probability of the model is roughly $\Pi_i (1 - p_i)$ (in practice, confidence values for the existence of single connections are not independent, therefore this is only a rough approximation of the joint probability). This also means that there is a trade-off between the joint probability of the graph and its density: the joint probability of the whole network pattern can be increased by decreasing the threshold for connectivity at more conservative p-values. Furthermore, one can look at the pairwise inference methods as a sort of model comparison. This is because in the second step of the inference, for every connection only three options are possible to choose from. The difference with DCM procedure lies in the fact that pairwise inference methods are based on the simple statistical properties emerging from causation in linear systems, and do not involve minimizing the cost function - such as negative free energy - as is done in DCM.

In the fMRI community, the DCM family (Friston, Harrison, and Penny, 2003) is currently the most popular approach to causal inference. This is partially due to the fact that DCM was tailor-made for fMRI, and includes a generative model based on the biological underpinnings of the BOLD dynamics (Buxton, Wong, and Frank, 1998). Some of the GC studies also involve estimation of the hemodynamic response function and deconvolving the data before applying the estimation procedure (David et al., 2008; Ryali et al., 2011; Ryali et al., 2016; Hutcheson et al., 2015; Wheelock et al., 2014; Sathian, Deshpande, and Stilla, 2013; Goodyear et al., 2016). This notion of the hemodynamics is both a strength and a weakness: the generative model fits the data well, but only as long as the current state of knowledge about the neurobiology is accurate. Many studies suggest that human hemodynamics is very dynamic and driven by state-dependent processes (Miezin et al., 2000; Handwerker et al., 2012). The influence of this complex behavior on the performance of DCM is hard to estimate.

The DCM procedure performs causal inference through model comparison, and as such, it is restricted to causal research in small networks containing a few nodes - since the computational costs increase like a factorial with the number of nodes. With the rise of research on resting state networks that contain up to 200 nodes, this may prove to be a limiting characteristic (Smith et al., 2009). This issue can be addressed with new methods for pairwise inference such as PT and PW-LR, which do not impose any upper bound on the size of the network.

It is important to remember that there are always two aspects of methods for causal inference. First, the method should have assumptions grounded in a biologically plausible framework, well-suited for the given dataset. E.g., a method for causal inference in fMRI should respect: (1) the confounding, region- and subject-specific BOLD dynamics (Handwerker, Ollinger, and D’Esposito, 2004); and (2) co-occurrence of cause and effect (since the time resolution of the data is low compared to the underlying neuronal dynamics, the causes and their effects most likely happen within the same frame in the fMRI data). The new methods for pairwise inference address this issue by (1) breaking the time order, and performing causal inference on the basis of statistical properties of the distribution of the BOLD samples, and not from the timing of events; (2) using correlation to detect connections. A good counterexample here is GC; GC has been proven useful in
multiple disciplines, and its estimation procedure is impeccable: nonparametric, computationally straightforward, and it gives a unique, unbiased solution. However, there is an ongoing discussion on whether or not GC is suited for causal interpretations of fMRI data. On one hand, theoretical work by Seth, Chorley, and Barnett, 2013 and Roebroeck, Formisano, and Goebel, 2005 suggest that despite the slow hemodynamics, GC can still be informative about the directionality of causal links in the brain. On the other hand, the work by Webb et al., 2013 demonstrates that the spatial distribution of GC corresponds to the Circle of Willis, the major blood vessels in the brain.

Secondly, the estimation procedure needs to be computationally stable. Even if the generative model faithfully describes the data, it still depends on the estimation algorithm whether the method will return correct results. However, the face validity of the algorithms can only be tested in particular paradigms in which the ground truth is known. If in the given paradigm, the ground truth is unknown - which is most often the case in fMRI experiments - only reliability can be tested. One way of assessing reliability of the method is testing for the test-retest convergence. So far, DCM is the only method that has been extensively tested in terms of test-retest reliability in separate studies (Frässle et al., 2015; Frässle et al., 2016a; Schuyler et al., 2010; Rowe et al., 2010; Tak et al., 2018), and performed good overall. In general, it is desirable to have more studies testing the reliability of the methods in experimental fMRI datasets as such validation of multiple methods such as GC or SEM, is still missing.

One last remark about the nature of the different methods: some methods, such as the DCM, were developed for task- (or, event-related) fMRI specifically. Yet, new implementations of spectral DCM for the resting state were proposed as well (Friston et al., 2011). As for the other methods, application to resting-state studies is relatively straightforward, while task fMRI can pose certain constraints on the methods. E.g., lag-based methods such as GC work best when the task is executed in a form of epochs (Deshpande et al., 2008b) rather than as a few second stimulus-response blocks, because it is extremely difficult to fit an AR model to datasets of 1-2 frames in length. For this reason, structural methods (which do not regard the time sequence) such as BNs or PW-LR, will be much more efficient in estimating causality in such cases.

Coming back to the main question posed in this review, can we hope to uncover causal relations in the brain using fMRI? Although there are new concepts in the field, that propose to consider causal interactions in the brain in probabilistic terms (Mannino and Bressler, 2015; Griffiths, 2015), the “traditional,” deterministic models of causality are prevalent in neuroimaging. Within these deterministic models, in the light of the existing literature, the new research directions based on breaking the time order as the axiom of causal inference (such as PW-LR, PT, and LiNGAM), prove more successful than the more “traditional” approaches which take regression in time into account (such as GC or TE, Smith et al., 2011; Hyvärinen and Smith, 2013). Also, Patel’s two-step design to achieve a causal map of connections is very promising, especially once the Pearson correlation is replaced with partial correlation as is done in PW-LR. One note to add is that “success” of any method for causal inference in fMRI depends on the generative model used for creating the synthetic dataset. In the seminal work
Chapter 5

5.6. Causal inference in fMRI, a Review: Discussion

by Smith et al., 2011 I am referring to, multiple methods were evaluated and critically discussed on the basis of simulations of the DCM generative model. However, there are alternatives, e.g., generative model by Seth, Chorley, and Barnett, 2013, which might potentially yield other hierarchy of methods in terms of success rate in discovering the causal links.

In this Chapter, I discuss the topic of inferring causal processes from fMRI datasets on the subject level. One approach that could further contribute to the development of methods for causal inference in fMRI though, is a group inference approach. In such an approach, a prior that different subjects represent similar causal structures, is added to the inference procedure. As lumping the datasets coming from different subjects increases the amount of data to derive the causal structure from, this assumption, in general, facilitates the inference. Multiple algorithms for group inference for effective connectivity in fMRI have already been proposed, including Independent Multiple Sample Greedy Equivalence Search (IMaGES, Ramsey et al., 2010), LOFS algorithm previously mentioned in Chapter 5.3.1 (Ramsey, Hanson, and Glymour, 2011), and Group Iterative Multiple Model Estimation (GIMME, Gates and Molenaar, 2012).

Furthermore, with the current rapid growth of translational research and an increase in use of invasive and acute stimulation techniques such as optogenetics (Deisseroth, 2011; Ryali et al., 2016) or TMS (Kim, Pesiridou, and OReardon, 2009), a rigid validation of methodology for causal inference becomes feasible through interventional studies. Recently, multiple methods for inferring causality from fMRI data were validated using joint fMRI and MEG experiments (Mill et al., 2017), with promising results for GC and BNs. This gives hope for establishing causal relations in neural networks, using fMRI.
Chapter 6

The impact of hemodynamic variability and signal mixing on the identifiability of effective connectivity structures in BOLD fMRI

Multiple computational studies have demonstrated that essentially all the current analytical approaches perform poorly when applied to synthetic fMRI datasets. In this work, I perform a simulation study to gain new insights into the influence of factors such as the slow hemodynamic responses, mixed signals in the networks, and short time series on the effective connectivity studies in fMRI. Firstly, I perform a Linear Discriminant Analysis study and I find that not the hemodynamics itself but mixed signals in the neuronal networks are detrimental to the signatures of distinct connectivity patterns. This result suggests that, deconvolving the BOLD responses is not necessary for methods for effective connectivity based on correlations. Furthermore, as the signal mixing in the networks can be induced by misparcellation, functional rather than anatomical parcellation are recommended as the data preprocessing step. Secondly, I study the impact of neuronal and hemodynamic variability on the inference with the use of lagged methods. I find that both the background noise in the neuronal networks and the local hemodynamic variability provide an upper bound to the success rate of the lagged methods. Therefore, in studies that involve lagged methods for effective connectivity, it is recommended to use the voxel-based region definition combined with hierarchical clustering in effective connectivity studies, and to deconvolve the BOLD time series into the neuronal time series. Furthermore, I demonstrate that downsampling the data to TRs lower than 0.7[s] does not influence the performance of the lagged methods.

Keywords: effective connectivity research, functional Magnetic Resonance Imaging, generative model, hemodynamic response, scale-free noise

This Chapter was published as: Bielczyk, N. Z., Llera, A., Buitelaar, J. K., Glennon, J. C., Beckmann, C. F. (2017). The impact of hemodynamics and signal mixing on the identifiability of causal structures in BOLD fMRI. Brain and Behavior, DOI: 10.1002/brb3.777
6.1 Introduction

Studies on the communication in large scale networks in fMRI were initiated as the functional connectivity (FC) research. FC quantifies the strength of communication between brain regions by means of correlations and therefore without specification of directionality or causality (Heuvel and Pol, 2010).

Extending network research in fMRI from functional to effective connectivity provided a substantial advance to the understanding of brain dynamics in health and disease (Fornito, Zalewsky, and Breakspear, 2015; Sporns, 2014; Friston, 2011; Bielczyk et al., 2015). Effective connectivity in fMRI is a complex research problem that involves not only specification of the presence or absence of connections, but also the directionality of the information flow. It is subject of an ongoing debate whether state-space models (e.g., the DCM, Friston, Harrison, and Penny, 2003 or GC, Granger, 1969; Roebroeck, Formisano, and Goebel, 2011; Seth, Barrett, and Barnett, 2015; Solo, 2016) or structural causal models (e.g., SEM, McIntosh and Gonzalez-Lima, 1994 or BNs, Frey and Jojic, 2005) are better suited for this research problem (Valdes-Sosa et al., 2011). As mentioned in Chapter 5, DCM has become the most popular approach for testing speculative directed network configurations to date. However, as DCM performs the inference through model comparison, its utility for the exploratory network discovery remains elusive.

Furthermore, there are theoretical studies suggesting that GC (Granger, 1969; Seth, Chorley, and Barnett, 2013) can serve for effective connectivity research in fMRI (Seth, Chorley, and Barnett, 2013; Roebroeck, Seth, and Valdes-Sosa, 2011), and that GC can provide a complimentary information to the DCM procedure (Friston, Moran, and Seth, 2013; Seth, Barrett, and Barnett, 2015; Bajaj et al., 2016). GC seems to be equally effective as DCM under certain circumstances, as when the hemodynamic response function is deconvolved from the datasets (David et al., 2008; Ryali et al., 2011; Ryali et al., 2016; Wang et al., 2016). GC was recently empirically validated in a joint fMRI and MEG experiment with very promising results (Mill et al., 2017).

Several properties of fMRI data pose obstacles to the modeling of direct links in the neural networks. Firstly, SNRs are low in fMRI data in general. In grey matter, the intensity change with respect to neurophysiological processes varies from 1-2% at field strengths of $1.5 - 2.0[T]$ (Ogawa et al., 1993; Boxerman et al., 1995) to 5-6% at field strengths of $3.0[T]$. Secondly, the temporal resolution (TR) of MR imaging is generally low, although it is increasingly improving, and the state-of-the art datasets reach TRs as low as $0.70[s]$ (Essen et al., 2013). Thirdly, the fMRI BOLD signal is associated with hemodynamic response and it is not a direct electrophysiological read-out from the neuronal dynamics. The associated response function (Buxton, Wong, and Frank, 1998) acts as a low-pass filter which results in autocorrelations the the BOLD fMRI signal (Logothetis and Wandell, 2003). Furthermore, the hemodynamic responses are known to be region- and subject-
6.1. Impact of hemodynamics and signal mixing: Introduction

specific (Devonshire et al., 2012) and may also differ between different cognitive states (Miezin et al., 2000; Handwerker et al., 2012). This might provide an additional confound to the estimates of the effective connectivity in fMRI.

Understanding factors that govern the performance of methods for effective connectivity has been a subject to multiple computational studies. Previous work used the DCM generative model (as the basis for the DCM inference procedure, Friston, Harrison, and Penny, 2003) to create benchmark synthetic datasets (Smith et al., 2011). Multiple methods for assessing effective connectivity were tested on this synthetic data, including GC (Roebroek, Formisano, and Goebel, 2005), Partial Directed Coherence, PDC (Baccalá and Sameshima, 2001), LiNGAM (Shimizu et al., 2006) or Transfer Entropy, TE (Lizier et al., 2011; Vicente et al., 2011). In general, however, the methods tested in the study did not perform much better than chance even though the testing networks were sparse and relatively small compared to the large-scale resting state networks in the brain (Smith et al., 2009).

In the aforementioned work by Smith et al., 2011, the impact of the jitter in the hemodynamic lags on the effective connectivity estimates in fMRI has also been examined. In this study, hemodynamic lags were defined as varying time to onset in the hemodynamic responses. The authors introduced a Gaussian distribution of the temporal lags and manipulated its width. The study demonstrated that the investigated methods for estimating effective connectivity showed noticeable sensitivity to the jitter of the hemodynamic lags. However, this effect was not as pronounced as sensitivity to a few other factors, e.g., a short session duration and mixed signals in the networks. Other theoretical work has examined the impact of variability of the hemodynamic response on GC (Deshpande, Sathian, and Hu, 2010). The authors demonstrated that in the absence of hemodynamic variability, even in presence of small temporal delays in the physiological range (under 50\[ms\]) in the neuronal communication, causal links between the nodes in the network can be correctly inferred with the use of GC from fMRI datasets. However, when hemodynamic lags are variable, the minimum neuronal delay detectable with the use of GC increases to hundreds of milliseconds. At the same time, for unknown neuronal and hemodynamic delays, the accuracy of detecting the directionality with GC was above chance and, in case of fast sampling, as high as 90%.

In this work, I employ the DCM generative model as implemented in Smith et al., 2011 and perform a simulation study to shed more light on the caveats of effective connectivity studies in fMRI. Based on these simulations, I propose how certain issues can be overcome by the proper data preprocessing and region definition.

Firstly, I investigated the influence of the presence of the slow hemodynamic response, the short signal length, and mixing signals in the network on the effective connectivity research in fMRI. To assess this influence, I

---

3 Actually, a recent intracerebral electroencephalography study by Wang et al., 2017 does not support the results obtained from the benchmark modeling study: Patel’s $\tau$, PT (Patel, Bowman, and Rilling, 2006), a leading method in the Smith’s study, does not perform equally well in experimental conditions.

4 There are also other implementations of the DCM generative model, e.g., Seth, Chorley, and Barnett, 2013. However, I decided to go with the most standard implementation.
investigated under what conditions the distinct network connectivity patterns in the networks can lead to BOLD fMRI time series which can be distinguished in a supervised manner with the use of Linear Discriminant Analysis (LDA, Fisher, 1936) using the correlations between the time series in the nodes of the network as features. Viewing the effective connectivity from such a "discriminative" point of view does not provide practical advice on how to render the directed connections from the data, but can increase our understanding of the core limitations of fMRI data and the limitations of the methods for effective connectivity research.

Then, I compare the results of LDA classification performed on the underlying fast neuronal dynamics with LDA performed on the observed, experimental level. Based on these results, I discuss influence of the slow hemodynamic response on the effective connectivity estimation. Furthermore, I subsample the BOLD response and discuss the influence of TR and the sample length on the effective connectivity research. Lastly, as previous computational studies suggest (Smith et al., 2011), mixed signals in the networks are detrimental to the effective connectivity estimation. Studies on the spectral content of the Local Field Potentials (LFPs) have demonstrated that mesoscale activity in neuronal networks has a scale-free power spectrum (He, 2014; Bédard, Kröger, and Destexhe, 2006; Dehghani et al., 2010) - which is a consequence of mixing signals of various frequencies in the LFP (He et al., 2010), which altogether gives a scale-free dynamics. Therefore, I emulate the effect of mixed signals in the neuronal networks by introducing the structured, scale-free (a.k.a. pink noise) background noise in the communication between the nodes in the network, as opposed to the Gaussian distributed noise typically implemented in the DCM generative model (a.k.a. white noise). Then, I investigate the impact of mixed signals on the effective connectivity estimates by comparing the LDA performance on the BOLD responses obtained from noisy neuronal dynamics, with addition of white versus pink noise. My finding is that the impact of the (fixed) hemodynamic response on the signatures of distinct connectivity patterns in the neuronal and BOLD time series, is small compared to the impact of mixed signals in the networks. This results suggests that effectively, there is no need to deconvolve the BOLD time series in studies using non-lagged methods for effective connectivity in fMRI, i.e., methods based on correlations. I also note that the effect of mixed signals can be partially user-induced. Therefore, it can be controlled in the preprocessing pipeline, e.g., by using functional rather than anatomical parcellation of the brain. Also, the TR has an influence on the classification results but mostly because the number of samples must be sufficient to estimate the features fed into the classifier. In my classification based on pairwise correlations as features, the minimal number of samples was in the range of 200, which is at the lower end of the recordings typically registered in fMRI experiments.

Secondly, I perform a computational study on a very simple, two-node network to investigate the impact of the variability in neuronal activity and in the hemodynamic lags (understood as the time to peak in the hemodynamic response functions) on the effective connectivity estimates with the use of lagged methods (such as GC of TE). In this study, I used lagged crosscorrelation applied to the normalized BOLD responses (El-Gohary and McNames, 2007) to represent the information contained in the sequence of the samples, and therefore to represent all the lagged methods for effective
connectivity. I then investigated how inference with the use of the lagged
crosscorrelation naturally varies for different instantiations of the dynamics
in a simple, two-node network with one connection, and how this depends
on the relationship between hemodynamic lags in the upstream and down-
stream region.

These results suggest that even in relatively simple scenarios, i.e., low
noise in the neuronal communication and strong connectivity between the
nodes, there is a high natural variability in the lagged crosscorrelation scores
derived from the underlying neuronal dynamics which represents studies
with a perfect blind deconvolution (Wu et al., 2013). This result strongly
suggests that it is a good strategy to perform the inference with the use
of lagged methods in a voxel-wise fashion and then integrate the results
over regions of interest (ROIs) with one of the available methods for this
purpose (see: Section 6.4) to overcome this natural uncertainty of the lag-
ged methods. Furthermore, lagged methods are also often used without
the BOLD deconvolution (Zhao et al., 2016; Regner et al., 2016; Chen et al.,
2017). In such case, the utility of lagged methods depends on the variability
in the hemodynamic lags between the upstream and downstream regions.
According to my results, applying the same strategy (i.e., voxel-wise in-
fERENCE AND INTEGRATION OF RESULTS OVER ROIS) can only work if the mean
hemodynamic lags in the upstream region are no more than $200\text{ms}$ higher
than in the downstream region. Otherwise, the lagged methods might swap
the direction of the connection. Therefore, blind deconvolution is strongly
advisable for connectivity research with the use of lagged methods in fMRI.
Additionally, according to these results, decreasing the TR from $0.70\text{s}$ as in
HCP data (Essen et al., 2013) to $0.10\text{s}$ would not further improve the per-
formance of the lagged methods because it would provide quantitative
rather than qualitative difference to the asymmetry in the lagged crosscor-
relation function.

In Section 6.2.1, I introduce the DCM generative model in detail. In Sec-
tions 6.2.3 and 6.3.1, I describe how I set up networks to compute lagged
crosscorrelations and perform the LDA study, respectively. In Section 7.3, I
present the results and in Section 6.4, I discuss these results and their prac-
tical implications on the effective connectivity research in fMRI.

6.2 Materials and methods

6.2.1 The generative model

Over the past decade, multiple generative models have been proposed in
the context of the DCM (Friston, Harrison, and Penny, 2003; Li et al., 2011;
Kiebel et al., 2007; Stephan et al., 2007; Marreiros, Kiebel, and Friston, 2008;
Stephan et al., 2008; Friston et al., 2011; Havlicek et al., 2015; Smith et al.,
2011; Seth, Chorley, and Barnett, 2013). In this Chapter, I chose the original,
single node per region DCM (Friston, Harrison, and Penny, 2003; Smith
et al., 2011). This model operationalizes the generation of BOLD response
from the neuronal networks across two levels: non-observable neuronal
level and the observable hemodynamic level.
The latent neuronal dynamics is described by the simple set of Ordinary Differential Equations:

\[
\frac{d\vec{z}(t)}{dt} = A\vec{z}(t - \tau) + C\vec{u}(t) + \vec{\sigma}(t)
\]  

(6.1)

where \(\vec{z}\) - the current neuronal activity across all the nodes in the network, \(\vec{u}(t)\) - binary inputs (trains of on- and off-states in this case), \(A\) - the adjacency matrix of effective connectivity, \(C\) - connections from (experimental) inputs to the nodes in the network, \(\tau\) - the temporal lag in the neuronal communication, and \(\vec{\sigma}(t)\) - the level of stochasticity on the neuronal level\(^5\). Connectivity (a.k.a. adjacency) matrix \(A\) contains self-inhibition in every node as originally proposed in Friston, Harrison, and Penny, 2003. Additionally, I use small, biologically plausible time lags of 50 ms in the communication between the nodes as implemented in Smith et al., 2011. Therefore the simulated network becomes a system of Delayed Differential Equations in fact (Bocharov and Rihan, 2000).

In this context, the stochastic term \(\vec{\sigma}(t)\) represents neuronal innovations that are not involved in the communication between nodes of the investigated network (Daunizeau, Stephan, and Friston, 2012). This term can either represent the intrinsic dynamics in the given node (other than inhibition), or input from areas outside the investigated network. Strictly speaking, these innovations are not a “noise” (which would mean stochasticity added to the neuronal time series on the top of the simulated dynamics), but rather a background neuronal dynamics that cannot be explained by the given model. However, for the sake of simplicity, I will refer to \(\vec{\sigma}(t)\) as noise in the text below.

The observational level is given by the classic representation of the hemodynamic response, referred to as the Balloon-Windkessel model (Buxton, Wong, and Frank, 1998; Friston, Harrison, and Penny, 2003). In this model, the hemodynamic response is described node-wise: in every node \(i\), it is characterized by the dynamics of four biophysiological variables as follows:

\[
\begin{align*}
\frac{ds_i}{dt} &= \frac{dz_i}{dt} - \kappa_i s_i(t) - \gamma_i(f_i(t) - 1), \\
\frac{df_i}{dt} &= s_i(t), \\
\lambda_i \frac{dv_i}{dt} &= f_i(t) - v_i(t)^{1/\alpha}, \\
\lambda_i \frac{dq_i}{dt} &= f_i(t)E(f_i(t), \rho_i)/\rho_i - v_i^{1/\alpha-1}(t)q_i(t),
\end{align*}
\]  

(6.2)

where \(s_i(t)\) - vasodilatory signal, \(f_i(t)\) - inflow, \(v_i(t)\) - blood volume, \(q_i(t)\) - deoxyhemoglobin content, \(E(f, \rho) = 1 - (1 - \rho)^{1/f}\). The model involves five node-specific constants: \(\kappa\) - rate of signal decay, \(\gamma\) - rate of flow-dependent elimination, \(\lambda\) - hemodynamic transit time, \(\alpha\) - Grubb’s exponent, \(\rho\) - resting oxygen extraction fraction.

Then, the outcome BOLD response yields:

\[
y(t) = V_0(7\rho_i(1 - q_i(t)) + 2(1 - q_i(t)/v_i(t)) + (2\rho_i - 0.2)(1 - v_i(t)))
\]  

(6.3)

where \(V_0 = 0.02\) - the resting blood volume fraction.

---

\(^5\)In this network setup, the modulatory connectivity does not play a role for the research question. Therefore, I set all the modulatory connections \(B\) from the original DCM model, Friston, Harrison, and Penny, 2003, to zero.
I sampled the hemodynamic parameters in the nodes of the network independently and from the distributions derived from physiological experiments (Friston, Harrison, and Penny, 2003). Inputs to the network were simulated as in Smith et al., 2011: as independent trains of on- and off-states with time resolution of $TR = 5[ms]$. The probability of state switches was governed by a Poissonian process of a mean on-state duration of $2.5[s]$, and a mean off-state duration of $7.5[s]$.

To simulate the natural variability in the human hemodynamics, I sampled the parameters from the distributions described previously in (Friston, Harrison, and Penny, 2003). Since this work concerns the effects of the neuronal noise, I did not add the thermal noise to the hemodynamic response as it refers to the quality of the scanning protocol in fMRI. At the end of the modeling pipeline, I subsampled the BOLD to emulate the limited frequency of the fMRI datasets. The TR is one of the parameters in my study, with which I also control the length of the signal. In the typical fMRI experiments, the range of TR is $0.7 – 3.0[s]$, but I make a step beyond this range to better understand the influence of the TR on presence of network signatures in the neuronal dynamics, and reduce it down to $0.10[s]$. The pipeline for generating synthetic fMRI datasets with the use of the DCM generative model was first introduced in Chapter 5, Fig. 5.2.

6.2.2 Impact of the noise and the length of the signal on identifiability of causal structures in fMRI

To investigate under what conditions the problem of effective connectivity research becomes ill-posed, I fix the nodes in the test network (and the associated local hemodynamic parameters) and perturb the connectivity within this network to investigate under what circumstances this perturbation yields detectable effects in the outcome BOLD response. Namely, if the two networks of distinct connectivity patterns yield indistinguishable BOLD response, the effective connectivity problem is ill-posed.

In this work, I restrict myself to the network design comparable to networks investigated in typical fMRI DCM studies. These studies typically involve comparing between small, literature-informed models of $3 – 4$ nodes. I also chose for simple Directed Acyclic Graphs (DAGs, Thulasiraman and Swamy, 1992) presented in Fig. 6.1. This architecture facilitates the feed forward distribution of information throughout the hierarchical connectivity patterns. In Fig. 6.1, I present the three connectivity patterns proposed in this study. The original network ($N_1$) contains the projections $1 \rightarrow 2, 1 \rightarrow 3, 2 \rightarrow 3, 3 \rightarrow 4$. I perturb this original connectivity pattern in two ways:

1. "Flip": exchanging the connection $2 \rightarrow 3$ into $3 \rightarrow 2$ (Fig. 6.1, $N_2$),
2. "Split": substituting the connection $3 \rightarrow 4$ into two weaker connections $1 \rightarrow 4$ and $2 \rightarrow 4$ (Fig. 6.1, $N_3$).

Here, I fix the connectivity strength to $0.15$ which refers to the range of connection strengths typically found in the DCM studies (Volza et al., 2015; Li et al., 2014).

*In addition, all the nodes have inhibitory self-connections (as in every DCM model).
Chapter 6. Impact of hemodynamics and signal mixing

**Figure 6.1:** The test network and its two perturbations. The original network ($N_1$) contains projections $1 \rightarrow 2, 1 \rightarrow 3, 2 \rightarrow 3, 3 \rightarrow 4$. The connectivity flip involves exchanging the connection $2 \rightarrow 3$ into $3 \rightarrow 2$ ($N_2$, red). The connectivity split involves substituting the connection $3 \rightarrow 4$ into two weaker connections $1 \rightarrow 4$ and $2 \rightarrow 4$ ($N_3$, red). This is an extension of the two-node setup presented in Fig. 6.4 to four nodes (inputs and background noise are skipped for simplicity).

Fig. 6.2 presents the dynamics in the 4-node network with perturbations. As anticipated, for both the simulated neuronal time series and for the simulated BOLD responses, the signatures of distinct connectivity patterns vanish along with the decreasing SNR for the pink but not for the white noise.

**Figure 6.2:** Exemplary dynamics for a DAG of 4 nodes, for the white and pink noise, and for two noise levels. A: neuronal time series. B: BOLD response. Red: original network. Green: network with a flipped connection. Blue: network with a split connection. In nodes 2 and 3, there are certain differences between the dynamics in the original network (red) and network with a flip (green), which reflects the flip of connection from $2 \rightarrow 3$ to $3 \rightarrow 2$. In node 4, there is a difference between the dynamics in the original network (red) and network with a split (blue), which reflects the split of connection from $3 \rightarrow 4$ into $1 \rightarrow 4$ and $2 \rightarrow 4$. However, the same magnitude of pink noise yields loss of the systematic differences between the dynamics in the considered networks.
Fig. 6.3 presents the residuals between the time series generated from the original network and the network with a flip (green time series) and split (blue time series). For the neuronal time series, I present 2 iterations (Fig. 6.3A), while for the BOLD time series, I extend to 20 iterations (Fig. 6.3B). I present the results for two SNR levels ($SNR = 100.0$ and $SNR = 0.5$), which reflects neuronal noise of standard deviation equal to 0.01 and 2.0 of the magnitude of the signal, respectively, for the white and the pink noise.

![Figure 6.3: Residuals between signals generated from the original network and from the perturbed networks, for the white and the pink noise, and for two noise levels. A: Neuronal time series, 2 realizations. B: BOLD response, 20 realizations. Green: residuum for a flipped connection. Blue: residuum for a split connection. Since neuronal time series has a very fast dynamics, the structure in the residuum (in white noise regime) is not visible even when time interval is shrunk to 50[s]. For the BOLD, in the white noise regime the structure in the residuals is visible both for the flip (nodes 2 and 3) and for the split (node 4), even under low SNR. This no longer holds in the pink noise regime.](image-url)
When SNR is high, for both white and pink noise one can observe systematic differences between the distributions of neuronal time series and BOLD coming from the original connectivity pattern and its perturbed versions. Once SNR decreases, the results differ with respect to the type of the noise. For the same magnitude of the noise, the dynamics for the pink noise is different than for the white noise. In the neuronal time series, the prevailing component of the dynamics are the slow fluctuations carried by the scale-free, pink noise. The systematic differences between the dynamics in the BOLD generated from the original and perturbed networks are also lost.

In this study, I concentrate on the neuronal noise, or "innovations" $\tilde{\sigma}(t)$: other neuronal activity within a node that is either related to the intrinsic dynamics within the brain region represented by that node, or to the inflow of activity from other regions lying outside this particular network.

In the previous study by Smith et al., 2011, this background noise was set to very low levels\(^7\), whereas in my study, the magnitude of the noise is the variable of interest. I vary this parameter between 0.01 and 2.0 times the value of the signal (which corresponds to $SNR = 100$ and $SNR = 0.50$, respectively). Also, as I am interested in quantifying the strength of the effect of mixed signals in the network on the effective connectivity research, I emulate mixed signals with the use of scale-free, pink noise in the neuronal communication $\sigma(t)$, and compare against Gaussian white noise of the same magnitude. Since the DCM generative model is stochastic in my study, I perform 500 instantiations of the network dynamics for each parameter set. To find out under what circumstances one can properly distinguish the output from the original and perturbed networks, I perform a classification study. For this purpose, I compress a set of four long time series representing the dynamics in the nodes of the networks into six pairwise Pearson correlations (thus, only six features in total)\(^8\). I chose correlations as features because multiple methods for effective connectivity research are built on correlations between the time series, in either direct or in indirect way - as is done in GC, SEM (McIntosh and Gonzalez-Lima, 1994), BNs (Frey and Jojic, 2005), PT (Patel, Bowman, and Rilling, 2006), LiNGAM (Shimizu et al., 2006), etc. Pearson correlation is also a simple, parameter-free feature that can be derived from the data without any additional assumptions.

Then, I investigated the impact of the magnitude and the spectral properties of the noise on the neuronal level and the length of the time series, on the classification accuracy.

### 6.2.3 Influence of hemodynamics on causal discovery

Secondly, I study the particular case of lagged methods for effective connectivity. The lagged methods such as GC, TE and other, new approaches (Hyvärinen, Shimizu, and Hoyer, 2008) assume that there is information preserved in the sequence of the BOLD samples. Therefore, I tested the limitations of the lagged methods for effective connectivity research with respect to the variability in the underlying dynamics and the properties of

---

\(^7\)Namely, to the Gaussian white noise of a standard deviation equal to 0.05 of the signal strength.

\(^8\)Similar feature set but based on partial correlations obtained through Ordinary Least Squares regression yielded very similar results, therefore I skipped it from this Chapter.
the local hemodynamic responses. With the use of the DCM generative model (Section 6.2.1), I simulated the dynamics of a simple, two-node system (Fig. 6.4A). Then, I adapted the lagged crosscorrelation (as proposed by El-Gohary and McNames, 2007), and I proposed a simple quantity $\Delta$ based on asymmetry in the lagged crosscorrelation (Fig. 6.4B,C,D) to quantify the amount of information on effective connectivity contained in the sequence of the samples. Since in this simple model, node 2 receives information from node 1 and the neuronal dynamics is delayed by $50\text{[ms]}$, in the absence of hemodynamic variability the BOLD time series in node 2 should be delayed with respect to the BOLD time series in node 1. Therefore, the BOLD time series in node 2 shifted one sample forward in time $z_2(t + 1)$ ($r_1$, middle row) should correlate with the BOLD time series $z_1(t)$ higher than the BOLD time series in node 2 shifted one sample backwards in time $z_2(t - 1)$ ($r_{-1}$, bottom row). Based on this expected difference, I proposed the variable $\Delta = r_1 - r_{-1}$, and I expected this quantity to be positive for the connection $1 \rightarrow 2$.

Now, since it is expected that $r_1 > r_{-1}$, in the absence of hemodynamic variability between the upstream and the downstream node the value of $\Delta$ for the connection $1 \rightarrow 2$ should be positive. I introduced the variability in the hemodynamic lags and investigated whether the positive sign of $\Delta$ still holds. I also investigated how the TR influences the performance of effective connectivity research with the use of $\Delta$, and whether or not further improving the TR to levels lower than $0.70\text{[s]}$ as implemented in the state-of-the-art HCP data (Essen et al., 2013), would improve the performance of the lagged based methods. Therefore, I compared the results for $TR = 0.70\text{[s]}$ with $TR = 0.10\text{[s]}$.

![Figure 6.4: Defining $\Delta$ for the two-node noisy system. A: The upstream node ($z_1(t)$) is sending information to the downstream node ($z_2(t)$) through a single connection of weight $A_{12}$. Both regions receive a binary signal $u_i(t)$ and neuronal noise $\sigma_i(t)$. B: Computing crosscorrelation between the two time series. Correlation between two time series can be computed without a lag (upper panel), with time series generated in the downstream region shifted one sample ahead in time ($r_1$, middle panel), or backwards in time ($r_{-1}$, lower panel).](image-url)
Chapter 6. Impact of hemodynamics and signal mixing

Figure 6.5: A: The relationship between $\Delta$ and GC obtained from the standard MVGC software (Seth, Chorley, and Barnett, 2013; Barnett and Seth, 2014), for a hundred instantiations of two random time series shifted by $\tau = 50\text{[ms]}$. As GC is computed in different units than $\Delta$, the range of values were aligned to each other on the plot. $\Delta$ and GC are similarly accurate in picking up on the precise lag of the interaction (vertical line). B: The value of $\Delta$ and GC at $t = 50\text{[ms]}$, in a function of SNR. The mean value of $\Delta$ decreases slower than the mean value of GC with respect to the decreasing SNR.

$\Delta$ derived from the asymmetry in crosscorrelation functions is meant to be a model-free and computationally efficient method for rendering effective connectivity that can represent the lagged methods in this simulation study. To support this choice, in Fig. 6.5A, I presented a comparison between $\Delta$ and GC for 100 instantiations of a two random time series shifted by $\tau = 50\text{[ms]}$. GC was computed with the use of the MVGC software (Seth, Chorley, and Barnett, 2013; Barnett and Seth, 2014). As GC is computed in different units than $\Delta$, the range of values were aligned to each other on the plot. As demonstrated in the figure, $\Delta$ characterizes similar type of information as GC as the two methods are similarly accurate in picking up on the precise lag of the interaction (vertical line). In Fig. 6.5B, I presented the mean value of $\Delta$ and GC at $t = 50\text{[ms]}$ in a function of SNR, for 100 instantiations at each SNR level. The mean value of $\Delta$ decreases slower than the mean value of GC along with SNR.

Since in this part of the study, I am only focused on the variability in the hemodynamic lags as a confound to the effective connectivity research, I set the level of neuronal noise to low magnitudes in both nodes (STD=0.01, white Gaussian noise), I set the connectivity strength to a very high value of $w = 0.9$ and I perform a single but long simulation ($T = 3,600\text{[s]}$) of the neuronal dynamics of this two-node system. Then, I convolve the neuronal time series in the upstream node ($z_1(t)$) with a fixed BOLD response (hemodynamic parameters at the mean of the distributions given in Friston, Harrison, and Penny, 2003) which gives a hemodynamic lag of $3.14\text{[s]}$\(^9\). I used 60,000 different BOLD responses to convolve the downstream neuronal time series $z_2(t)$ with. Then, I derive the $\Delta$ value for all pairs of BOLD time series while assuming $TR = 0.70\text{[s]}$ (which refers to the state-of-the-art Human Connectome Project datasets), and a high time resolution of $TR = 0.10\text{[s]}$ for comparison.

\(^9\)Proximity to the $\pi$ constant is accidental.
6.3 Results

6.3.1 Impact of the noise, the hemodynamics and the length of the signal on the presence of the network signatures in the BOLD

The results of the prediction study with the use of LDA are presented in Fig. 6.6A,B. I present the classification accuracy in the function of the SNR. Red and grey lines indicate the pink and the white noise, respectively. Firstly, the split of connection in the network is easier to detect than the flip. This result is intuitive because perturbing the network with a flip involves direct manipulation of one pairwise correlation (pair 2-3), whereas perturbing
Chapter 6. Impact of hemodynamics and signal mixing

the network with a split involves direct manipulation of three pairwise correlations (pairs 1 – 4, 2 – 3, and 3 – 4).

Secondly, the dropout in the accuracy of the LDA classifier in a function of SNR is similar for the neuronal time series and for the full length BOLD time series. This effectively means that, while using the methods of effective connectivity based on correlations between signals, deconvolving BOLD into the neuronal time series is not necessary to perform the inference. The hemodynamic response can even be beneficial to the effective connectivity estimates as it removes the noise-related higher frequencies while leaving experiment-related lower frequencies intact. However, this effect does not hold for the scale-free pink noise: the hemodynamic convolution can no longer filter out the noise that has high power in the slow frequency range.

Importantly, subsampling BOLD time series from very high time resolution of $TR = 5[ms]$ (that gives a total of 120,000 samples for the 10[min] of simulation) into the low $TR = 3.0[s]$ resolution (that only gives 200 samples) does not effectively blur the network signatures contained in the pairwise correlations since LDA performs classification with almost identical accuracy for the BOLD time series subsampled with $TR = 3.0[s]$ as for the full length BOLD time series obtained from convolving the fast neuronal time series (Fig. 6.6, middle columns). This result can be explained by plotting the success rate of the LDA classification for a fixed set of parameters: 120,000 samples of the BOLD time series, pink noise of $SNR = 0.5$, split network, and in the function of the TR (Fig. 6.6C). In this case, $TR = 3.0[s]$ is the critical value of TR which gives performance comparable to the full BOLD time series. This TR relates to the total of 200 samples. For the TRs longer than 3.0[s], however, the performance of LDA gradually drops. This can be explained by the precision in (all six) pairwise correlations over 500 instantiations of the network in a function of TR (Fig. 6.6D) - which is in fact the function of the length of the time series. All the six precision functions are constant until the TR reaches a critical value of $TR = 3.0[s]$, and above that value, the precision in pairwise correlations drops. This effect implies that, given the small network of four nodes with connectivity weights in the range of 0.15, the BOLD time series of about 200 samples should be long enough to estimate pairwise correlations between the nodes.

6.3.2 Influence of hemodynamics on effective connectivity research

In Fig. 6.7, I demonstrated the results of testing $\Delta$ on the two-node system with varying hemodynamic lags in the downstream region. Hemodynamic lags are operationalized as the time to peak of the hemodynamic response. Since in each simulation, the same connection 1 $\rightarrow$ 2 lagged by 50[ms] is present, I expect the time series in the downstream region to follow the time series in the upstream region. Therefore, $\Delta$ should be strictly positive at the lag of 50[ms]. In Fig. 6.7A, the histogram of $\Delta$ values over 500 instantiations of the neuronal dynamics of a two-node network with one connection (Fig. 6.4) is presented. Indeed, all the values are positive. In Fig. 6.7B, values of $\Delta$ for all TRs between 50[ms] and 3[s] and over 500 instantiations of the neuronal dynamics are presented. As expected, both the mean value and the precision of $\Delta$ drop towards zero along with increasing TR (black line). At the point of $TR = 0.70[s]$ (green line), 90% of $\Delta$ values are still positive
6.3. Impact of hemodynamics and signal mixing: Results

This means that in the absence of any hemodynamic lags and at the $TR = 0.70[s]$, $\Delta$ would have empirical accuracy rate of 90%.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure6_7.png}
\caption{Impact of the relative hemodynamic lags computed as the difference between time to peak of the hemodynamic response in the upstream and downstream region on $\Delta$. A: The histogram of $\Delta$ values over 500 instantiations of the neuronal dynamics, for the optimal $TR = 50[ms]$ matching the true delay in the neuronal communication only gives positive values as expected. B: $\Delta$ over 500 instantiations of the neuronal dynamics and in a function of the TR. At $TR = 0.70[s]$, the percentage of positive $\Delta$ values drops down to 90%, which is still high over the chance level. C: $\Delta$ in the function of the difference between hemodynamic lags in the upstream and the downstream region ("relative lag"). D: The relationship between the percentage of positive $\Delta$ values and the relative lags.}
\end{figure}

In Fig. 6.7C, the results for the analysis on the hemodynamic level are presented: $\Delta$ is expressed in the function of the difference between hemodynamic lags in the upstream and the downstream region (referred to as a "relative lag"). The plot presents scatter plot over 60,000 different convolutions with hemodynamic parameters independently sampled from the distributions given in Friston, Harrison, and Penny, 2003. Since the hemodynamic lag in the downstream node can be either higher or lower than the reference hemodynamic lag in the upstream node, the relative lags have positive and negative values (equal lags are marked with orange line). I divided the whole set of relative lags into 30 intervals, and computed the mean for relative lags lying within each interval (Fig. 6.7C, light blue curve). For each of the intervals, I also computed the cut-off for the bottom 10% of all the values (magenta curve).

Firstly, as expected, for positive relative lags, $\Delta$ becomes highly indicative for the connection as the vast majority of the $\Delta$ values are positive. This is because in this case, next to the neuronal dynamics in the downstream
region following the dynamics in the upstream region, the slower hemodynamics further increases the lag in the downstream BOLD time series. The upper 90% of the $\Delta$ values (magenta curve) at the relative lag of zero (orange line) is positive (yellow line), which corresponds with the previous results for the underlying neuronal dynamics (Fig. 6.7B). The negative relative lags on the other make this statistic worse: the lower the value of the relative lag, the higher the percentage of $\Delta$ values that become negative, i.e., the higher the chance of inferring the wrong directionality of the link between the two nodes. Fig. 6.7D demonstrates this relationship. For the critical value of the relative lag of around $-0.22[s]$, the percentage of positive $\Delta$ values drops below 50%.

The last issue is the impact of the TR on these results. In Fig. 6.8A, I demonstrate a reproduction of Fig. 6.7C but with a low $TR = 0.10[s]$. Despite much faster temporal sampling, the results on the lag differences are almost identical, the main difference being the range of absolute values of $\Delta$ (indicated on the y axis). Therefore, the difference is quantitative rather than qualitative. This effectively means that at $TR = 0.70[s]$, the hemodynamics is captured sufficiently well so that decreasing the TR further does not add any extra information that could be used with lagged methods for effective connectivity represented by $\Delta$. This effect is explained in Fig. 6.8B where an exemplary crosscorrelation function between the BOLD response between the upstream and the downstream region is presented (blue curve). Both time series are convolved with the identical, canonical hemodynamic response. The crosscorrelation peaks at $50[ms]$ and then declines as expected. I mirrored the crosscorrelation function around zero (red curve), and $\Delta$ can be computed as the difference between the original value of crosscorrelation and its mirror counterpart (yellow area). TRs of $0.10[s]$ and $0.70[s]$ are marked with vertical black lines. In this low range, the TR has an impact on the absolute value of $\Delta$, but not on its signum.
6.3.3 Summary of the results

As a summary, my findings suggest that:

1. In effective connectivity research with the use of methods based on correlations, deconvolution of BOLD time series into neuronal time series is not necessary,

2. Scale-free, pink noise induced by misparcellation is detrimental to the signatures of distinct connectivity patterns in the BOLD time series, therefore, functional parcellation into ROIs is recommended,

3. The minimal number of samples in the BOLD time series necessary for the causal inference depends on the strengths of the underlying connections, but in typical cases ($w = 0.15$) 200 – 300 samples should be sufficient,

4. Since the neuronal variability provides a limitation to the success rate for effective connectivity estimates involving temporal lags, voxel-wise inference combined with hierarchical clustering is recommended,

5. Since the local hemodynamic variability can inverse the direction of a connection obtained from lagged methods, deconvolution of BOLD fMRI time series into the neuronal time series is recommended in this case,

6. Decreasing the TR from 0.7[s] (as is implemented in HCP datasets (Essen et al., 2013), would not improve the performance of the lagged methods.

6.4 Discussion

The scope of valid assumptions for effective connectivity research in fMRI is a subject to a broad discussion (Roebroeck, Seth, and Valdes-Sosa, 2011). In this study, I conducted a computational study on synthetic fMRI datasets generated with the use of the original DCM generative model (Friston, Harrison, and Penny, 2003; Smith et al., 2011) to demonstrate possible caveats associated with the local hemodynamic variability and region definition, and to point to possible further directions for the effective connectivity research in fMRI.

My LDA prediction study sheds the light on the factors influencing effective connectivity estimates in fMRI. One remark is that LDA does not indicate how to establish the causal relations between the nodes in the network: it just uses the distinct features in the data and, therefore, determines when it is theoretically possible. In this study, I chose pairwise correlations as features because Pearson correlation is non-parametric and in multiple methods for effective connectivity research in fMRI, it serves as the basis for the effective connectivity research (Patel, Bowman, and Rilling, 2006; Hyvärinen and Smith, 2013)\(^\text{10}\). Having that in mind, I compared the impact of the presence of slow hemodynamic response, short time series, and mixed neuronal signals of the background on the LDA results.

\(^{10}\)Also, once correlation becomes indistinguishable between networks, the crosscorrelation will also become indistinguishable which relates to the second part of this study.
Currently, the slow hemodynamics and the low time resolution in the scanning process are viewed as two factors that hinder the performance of methods for causal discovery in fMRI the most. The results of this LDA study suggest that the fixed hemodynamics is not detrimental to the network signatures based on pairwise correlations: the LDA classifier distinguishes networks as easily on the basis on the neuronal time series and based on the BOLD fMRI time series. Moreover, the hemodynamic response can work as a denoiser for the white noise (Fig. 6.6A,B, grey lines). This result is intuitive once you take the spectrum of the white noise into account: a big portion of the power is carried within the high frequency range which is easily denoised by the hemodynamic response. This result demonstrates that, in some cases (namely, when the signal has low frequency, a good portion of the power of the noise is in the high frequency range, and the hemodynamic response is fixed), the BOLD response is not necessarily a detrimental factor to the effective connectivity estimates, but could even benefit the effective connectivity research in fMRI. Therefore, deconvolving the BOLD time series is not necessary while using methods for effective connectivity based on correlations.

Furthermore, I obtained an intuitive result linking the success rate of the LDA classifier to the precision in features estimation. Due to these results, at certain lengths of the BOLD time series, the precision in estimating correlations saturates, therefore further increasing the length of the time series does not improve on the classification accuracy any further. This is a different characteristic from the relationship between the length of the time series and the accuracy of signal detection, where increasing the length of the time series tends to improve the performance as it was found in a theoretical study by Murphy, Bodurka, and Bandettini, 2007.

Furthermore, unlike slow hemodynamics and short time series, mixing signals (represented in the simulations by the transition from white to scale-free, pink noise) hinders the performance of the effective connectivity estimates in fMRI. For this reason, studying scale-free noise and its role for the inference with the use of the DCM - as opposed to the Wiener process typically implemented in the stochastic DCM (Daunizeau, Stephan, and Friston, 2012) - is worth considering. Furthermore, the effect of mixed signals can be controlled to some extent.

There are two sources of signal mixing that can partly be addressed during data preprocessing and before applying any method for effective connectivity research to the fMRI data. The first source is associated with the background neuronal activity in the networks. To control the influence of this background activity, it is strongly recommended to partial out the signal coming from outside the ROIs as an initial step to the effective connectivity study. The second source of the scale-free noise is a possible misparcellation into arbitrary brain areas during the preprocessing pipeline. There is a range of available strategies that give rise to the BOLD time series fed into effective connectivity research. Multiple methods use biology as the starting point: anatomy (Wake Forest University pickatlas, WFU, Maldjian et al., 2003 or Automatic Anatomical Labeling, AAL, Tzourio-Mazoyer et al., 2002), histology (Glasser and Essen, 2011; Eickhoff et al., 2005; Amunts et al., 2013), or white-matter fibre structure (Mars et al., 2011; Mars et al., 2012; Sallet et al., 2013; Neubert et al., 2014; Neubert et al., 2015). Such a biology-informed parcellation can bring confounds to the signal: while
averaging the time series over the ROI, one might mix signals involved in various causal interactions. From the perspective of network analysis, mixing signals of various frequencies is equivalent to inducing a pink noise in the underlying neuronal dynamics. Therefore, one should pursue efforts aiming at best functional segregation into ROIs instead\(^{11}\). The optimal brain parcellation is a widely discussed topic in the field of fMRI (Stanley et al., 2013). As mentioned in previous chapters, there are multiple classes of available functional parcellations, to mention just a few: Bayesian methods (Janssen et al., 2015; Janssen, Jylänki, and Gerven, 2016), mixture modeling (Golland, Golland, and Malach, 2007; Tucholka et al., 2008; Lashkari et al., 2010; Lashkari et al., 2012), Ward algorithm (Ward, 1963), k-means algorithm (Flandin et al., 2002; Yeo et al., 2011; Kahnt et al., 2012), hierarchical clustering (Eickhoff et al., 2011; Michel et al., 2012; Orban et al., 2014; Heuvel, Mandl, and Pol, 2008; Bellec et al., 2006; Bellec et al., 2010; Blumensath et al., 2013) and spectral clustering (Thirion et al., 2006; Chen et al., 2012; Craddock et al., 2012), the new, semi-automated classification technique by Glasser et al., 2016, and hierarchical ICA by Oort et al., 2017.

Another possibility is to consider only the first eigenvariate within the anatomical ROIs as proposed by Sato et al., 2010, and as is implemented in the original version of the DCM inference procedure (Friston, Harrison, and Penny, 2003) instead of averaging activity over full anatomical regions. Lastly, one can also build ROIs in a data-driven fashion based on the patterns of activation only (task localizers, Fedorenko et al., 2010; Heinzle, Wenzel, and Haynes, 2012).

Finally, these results suggest that once the effect of mixed signals is under control (so that only the white noise is present in the neuronal dynamics), the signatures of distinct connectivity patterns are present in the BOLD time series even in a high noise regime. This may encourage further endeavors in the search for new markers of causal connections between brain regions from fMRI datasets.

In the second part of the study, I use lagged crosscorrelation to determine under what conditions the effective connectivity-related information is preserved in the sequence of the samples and could be retrieved with the lagged methods. The main difference between my implementation of the hemodynamic variability and the previous studies is that I concentrated on the *time to peak* of the hemodynamic response as the representation of the hemodynamic lag, as opposed to the *time to onset* of the hemodynamic response as in Smith et al., 2011. If the only difference between two hemodynamic responses was the time to onset (which means just a shift a copy of hemodynamic response forwards and backwards in time), then a lag of 50 [ms] in the underlying neuronal communication would imply that the time to onset in the hemodynamic response in the upstream region 50 [ms] later than the time to onset in the hemodynamic response in the downstream region would automatically flip the sign of \( \Delta \) and therefore also the outcome of the causal inference. However, this definition

\(^{11}\)Most of the recent open-source fMRI datasets such as Human Connectome Project, ADHD-200, and ABIDE, support the functional parcellation of the data (Glasser et al., 2016; Rosenberg et al., 2016; Craddock et al., 2012).
of the hemodynamic lags contains an assumption that hemodynamic response is equal to zero for the initial period of time. Such response profiles, however, are not biophysically plausible as in reality, the hemodynamic response has a positive derivative already at the start. In my study, I employed the classic Balloon-Windkessel model (Friston, Harrison, and Penny, 2003; Buxton, Wong, and Frank, 1998) to generate a natural, distribution of the hemodynamic responses derived from neurophysiological experiments, and I operationalized hemodynamic lags as the time to peak instead. I found that within certain range of the relative hemodynamic lags, the asymmetry in the lagged crosscorrelation carries information about effective connectivity that can be derived from two BOLD time series (since the mean $\Delta$ over the distribution of $\Delta$s derived from biologically relevant priors remains positive).

In this setup, I used a fixed neuronal delay of $50[ms]$, whereas in another computational study on the influence of hemodynamic response on GC by Witt and Meyerand, 2009, this variable was a parameter of interest. In this study, neuronal lags higher than $50[ms]$ would increase the asymmetry of the $\Delta$ in a function of relative lags (Fig. 6.7) on behalf of the positive relative lags. However, I motivate this small neuronal lag by experimental evidence suggesting that the axonal transmission, as the slowest phase of the neuronal transmission\textsuperscript{12}, involves time delays in the range of dozens of milliseconds (Sabatini and Regehr, 1999).

In many applications, lagged methods for effective connectivity in fMRI are applied to the deconvolved BOLD time series, with the example of GC (David et al., 2008; Ryali et al., 2011; Ryali et al., 2016; Hutcheson et al., 2015; Wheelock et al., 2014; Sathian, Deshpande, and Stilla, 2013; Goodyear et al., 2016). However, as demonstrated in Fig. 6.7B, the natural variability in the neuronal dynamics results with an upper bound on the accuracy of the lagged methods: even assuming a perfect deconvolution technique, that would allow for perfect retrieval of the neuronal time series from the BOLD time series, for the $TR = 0.70[\text{s}]$ the accuracy rate of $\Delta$ would not be higher than 90%. This result suggests that it might be worth considering to perform the GC analysis voxel-wise, and to average the resulting GC over the whole ROI instead of computing the GC ROI-wise as is often done in the GC studies (Deshpande et al., 2008a; Chen et al., 2017; Goodyear et al., 2016; Regner et al., 2016; Yang et al., 2017). After voxel-wise application of GC, the result can be averaged over all voxel-wise GC scores between the two regions. Alternatively, the GC score can also be averaged with the LASSO regularization (Tang et al., 2012), through the multi-voxel pattern-based causality mapping as proposed by Kim et al., 2013, or through hierarchical clustering as proposed by Deshpande et al., 2009. This strategic twist was already implemented in some studies (Zhao et al., 2016; Katwal et al., 2013), however, it is not the most common approach in the field\textsuperscript{13}. Applying GC voxel-wise has two major advantages. Firstly, computing the final value of GC as a mean value over GC values derived from a large number of voxels should reduce the natural inaccuracy of the lagged methods when

\textsuperscript{12}The synapses optimize many steps between the firing of a presynaptic cell and the response of the postsynaptic target which shrinks the time necessary for the transmission to a few milliseconds (Sabatini and Regehr, 1999).

\textsuperscript{13}Voxel-wise modeling is, however, increasingly popular for finding activation patterns in cognition from fMRI data (Huth et al., 2016).
applied to the stochastic neuronal dynamics (Fig. 6.7B). Secondly, it neutralizes the bias coming from possible inaccuracies in the blind deconvolution algorithms.

In this part of the study, I also discussed the influence of the local distribution of the hemodynamic lags present in the investigated networks on the performance of the lagged methods for effective connectivity (Fig. 6.7C,D). This part of the analysis refers to the studies in which the lagged methods are applied to the unconvolved BOLD time series, which is an often practice in GC research (Zhao et al., 2016; Regner et al., 2016; Chen et al., 2017). In such case, the utility of lagged methods in fMRI research depends on the variability in the hemodynamic lags. According to my results, under the assumption that the lag in the neuronal communication is in the range of $50\,\text{ms}$, applying a voxel-wise inference followed by integration of results over the ROIs can retrieve a correct directionality of the connection only if the mean hemodynamic lag between all the voxels within the upstream region are no more than $200\,\text{ms}$ higher than the mean hemodynamic lag between all the voxels within the downstream region.

In the seminal work by Handwerker, Ollinger, and D’Esposito, 2004, hemodynamic lags across primary motor and visual cortices as well as the frontal and supplementary eye fields were empirically found in a group of 20 subjects. One of the main findings in this study was that there is higher variability in the hemodynamic lags between than within subjects. However, the local variability of the hemodynamic lags within subjects still seems to be higher than the “safe range” found in my study (namely, $200\,\text{ms}$). Furthermore, a study by Menon, 2012 has demonstrated a large variation in the local hemodynamics depending on the vessel sizes in the voxels. Even vasculature changes across cortical layers have been shown to cause around $400\,\text{ms}$ differences in hemodynamic time-to-peak (Silva and Koretsky, 2002). Therefore, further neurophysiological studies on the variability in hemodynamic responses across the human brain are necessary to advance our ability to characterize causal interactions from BOLD fMRI data, especially with the use of the lagged methods such as GC or TE. According to my results however, the blind deconvolution of the BOLD time series before applying lagged methods for effective connectivity is strongly advisable.

Lastly, my results from this part of the study demonstrate that downsampling the BOLD time series at as little as $TR = 0.10\,\text{s}$ would not significantly improve the ability to retrieve the directionality of a connection with the use of the lagged methods as the results differ from the results obtained for $TR = 0.70\,\text{s}$ only quantitatively but not qualitatively.

My conclusions are concordant with the recent theoretical considerations by Solo, 2016 but differ from conclusions derived by Seth, Chorley, and Barnett, 2013, Barnett and Seth, 2017 and Lin et al., 2014 who conclude that GC estimation improves with the downsampling. The design of this study is different than the previous studies, and the novelty is that I reduced the effective connectivity research with lagged methods to the simplest possible case. First, I quantified the amount of asymmetry in the lagged cross-correlation function between two time series. This is a model-free and probably the simplest operationalization of the lagged methods for effective connectivity research, very precise at picking on the time lag of the interaction, as demonstrated in Fig. 6.4C. Secondly, I focused on investigating
the simplest, two node neural mass model with a single connection, and with neuronal dynamics modeled with the use of the Ordinary Differential Equations as is done in the standard DCM generative model\textsuperscript{14}. Therefore, I have trust that these results give the most reductionist and representative view on how, in general, a lagged method for rendering effective connectivity can be affected by the neuronal and hemodynamic variability, as well as the TR.

\textsuperscript{14}Seth, Chorley, and Barnett, 2013, by comparison, use much more complex design that involves a few nodes with more complex, spiking local neuronal dynamics.
Chapter 7

Increasing robustness of pairwise methods for effective connectivity in Magnetic Resonance Imaging by using fractional moment series of BOLD signal distributions

Estimating causal interactions based on fMRI datasets remains a challenging task. Multiple studies have demonstrated that all the current approaches perform poorly even when applied to synthetic fMRI datasets. Recent advances in this field include methods for pairwise inference, which involve creating a sparse connectome in the first step, and then using a classifier to determine the directionality of connection between every pair of nodes in the second step. In this work, I introduce an advance to the second step of this procedure, by building a classifier based on fractional moments of the BOLD distribution combined into cumulants. The classifier is trained on datasets generated under the Dynamic Causal Modeling (DCM) generative model. The directionality is inferred based upon statistical dependencies between the two node time series, e.g., assigning a causal link from the time series of lower variance to the time series of higher variance. This approach either outperforms or performs as well as other methods for effective connectivity when applied to the benchmark datasets. Crucially, it is also more resilient to confounding effects such as differential noise level across different areas of the connectome.

Keywords: causal inference, effective connectivity, functional Magnetic Resonance Imaging, pairwise causal inference

This Chapter is published in Network Neuroscience as:

7.1 Introduction

In the context of fMRI research, effective connectivity refers to the process of estimating causal interactions between distinct regions within the brain. As introduced in Chapters 5 and 6, several characteristics of fMRI data impose
severe restrictions on the possibility of estimating such effective connectivity (Valdes-Sosa et al., 2011; Friston, 2011; Bielczyk et al., 2019). To briefly summarize the most important points: first, the temporal resolution of the image acquisition is low in fMRI (sampling rate typically < 1 Hz). Furthermore, blood oxygen-level dependent (BOLD) activity is delayed with respect to neuronal firing, with a delay of 3−6 s in the adult human brain (Arichi et al., 2012). The delayed hemodynamic response can also induce spurious cross-correlations between two BOLD time series (Ramsey et al., 2010; Bielczyk et al., 2017b). Both subject-to-subject and region-to-region variability in the shape of hemodynamic response (Devonshire et al., 2012) provide general limitations to the effective connectivity estimates in fMRI: when the hemodynamic response in one region is faster than in another, the temporal precedence of the peak of the hemodynamic response can easily be mistaken for causation. Secondly, fMRI data is characterized by a relatively low SNR. Within grey matter and at field strengths of 3 T, the task-induced signal modulations are within 2−3% of the resting state activity depending on the task (Krüger, Kastrup, and Glover, 2001). Furthermore, the stochastic noise in the brain has been shown to have a scale-free spectral characteristic (He, 2014; Bédard, Kröger, and Destexhe, 2006; Dehghani et al., 2010) which additionally hinders identifiability of causal structures derived from fMRI data (Bielczyk et al., 2017b). Moreover, typical fMRI protocols involve a relatively short time series (a few hundred samples), in which estimation of conditional probabilities between variables, and estimating higher order statistic in the time series becomes difficult. Multiple methods were proposed to estimate effective connectivity from fMRI datasets (Friston, 2011). In the computational study by Smith et al., 2011, a range of methods for effective connectivity were tested on synthetic datasets derived from the DCM generative model (Friston, Harrison, and Penny, 2003). In this study, most methods for estimating causal interactions gave results at a chance level. One method highlighted as relatively successful at identifying causal links is based on Patel’s tau measure (PT, Patel, Bowman, and Rilling, 2006; Smith et al., 2011). PT entails a two-step approach where the first step involves identifying the (undirected) connections by means of functional connectivity, and is achieved based on correlations between the time series in different regions (which is also referred to as Patel’s kappa, Patel, Bowman, and Rilling, 2006; Smith et al., 2011).

Note that Patel’s tau and other pairwise inference procedures assume that causation implies correlation. This assumption is necessary to perform the first step of the inference procedure, i.e., to select reliable connections for further classification into upstream and downstream nodes. However, although this assumption is often true, this is not always the case as under certain circumstances (e.g., in control systems), causation might not be associated with correlation (Kennaway, 2015). In the recent study by Di and Biswal, 2019, the authors investigated task-modulated whole brain functional connectivity with the use of a six block-design and one event-related cognitive task. By using psychophysiological interactions between pairs of ROIs from the whole brain, the authors identified statistically significant task modulations in functional connectivity, and reported that "task modulated connectivity was found not only between regions that were activated
Chapter 7

7.1. Increasing robustness of methods for effective connectivity in fMRI: Introduction

during the task but also regions that were neither activated nor deactivated." This suggests that considering only pairs of regions in which activity is correlated might lead to neglecting some of the underlying effective connections.

There are multiple strategies for implementing and thresholding functional connectivity estimates (Varoquaux and Craddock, 2013). Since Pearson correlation typically returns dense connectomes that contain spurious links (spurious links $X - Z$ appearing as a consequence of links $X - Y$ and $Y - Z$, Aldrich, 1995), most often, partial correlation is employed as a method of choice to build functional connectomes based on fMRI datasets (Marrelec et al., 2006). In partial correlation, for each pair of nodes $X$ and $Y$, the linear input from all the remaining nodes in the network is partialled out from the time series before calculating Pearson correlation on the residuals. For the sake of computational efficiency, partial correlation is often computed as an inverse covariance.

Although partial correlation can be too conservative with respect to the underlying functional links or even induce spurious negative correlations in some cases (an effect known as Berkson’s paradox, Nie et al., 2015; Berkson, 1946), to date, it remains the state-of-the-art approach. Proper thresholding of partial correlation matrices is also an open research problem. A popular strategy for thresholding involves permutation testing (Smith et al., 2011; Hyvärinen and Smith, 2013), in which subject labels are shuffled between samples in multiple random iterations, and a separate null distribution is created for each connection in the network based on these permutations (leading to a single thresholded connectome matrix for the entire subject population). Recently, also a thresholding strategy employing mixture modeling has been proposed (Bielczyk et al., 2018). These techniques allow for creating individualized sparse connectomes for every subject. Other alternative thresholding schemes involve proportional thresholding (Heuvel et al., 2017) and shrinkage estimates, such as Ledoit-Wolf shrinkage (Ledoit and Wolf, 2003) and graphical Lasso (Friedman et al., 2008).

Thresholding a functional connectome obtained with the use of partial correlation results in a binary graph of connections where all other possible edges identified as absent are disregarded from further considerations. The second step determines the directionality in each one of the previously detected connections. In this step, effective connectivity boils down to a two-node Bayesian network. The concept is based on a simple observation: if there is a causal link $X \rightarrow Y$, $Y$ should get a transient boost of activity every time $X$ increases activity. And vice versa: if there is a link $Y \rightarrow X$, $X$ should react to the activation in $Y$. Therefore, one can threshold the signals $X(t), Y(t)$ to obtain a binary series of events $X_1(t), Y_1(t)$, and compute the difference between conditional probabilities $P(Y_1|X_1)$ and $P(X_1|Y_1)$. Three scenarios are possible:

1. $P(Y_1|X_1)$ equals $P(X_1|Y_1)$: it is a bidirectional connection $X \leftrightarrow Y$ (since empty connections were sorted out in the previous step),

2. The difference between $P(Y_1|X_1)$ and $P(X_1|Y_1)$ is positive: the connection $X \rightarrow Y$ is likely,

3. The difference between $P(Y_1|X1)$ and $P(X1|Y1)$ is negative: the connection $Y \rightarrow X$ is likely.
More recently, the Pairwise Likelihood Ratios (PW-LR, Hyvärinen and Smith, 2013) approach was proposed. PW-LR builds on the concept of PT. The authors improved on the second step of the inference by analytically deriving a classifier to distinguish between two models $X \rightarrow Y$ and $Y \rightarrow X$, which corresponds to the LiNGAM model (Shimizu et al., 2006) for two variables. The authors compare the likelihood of these two competitive models derived under LiNGAM’s assumptions (Hyvärinen et al., 2010), and provide a cumulant based approximation to the likelihood ratio. Within the PW-LR family, there are three methods based on a third cumulant built for two signals $X(t)$ and $Y(t)$:

1. "PW-LR skew"; a classic third cumulant,
2. "PW-LR tanh"; an alternative approximation to a third cumulant using a tanh-based nonlinear correlation,
3. "PW-LR r skew"; an alternative approximation to a third cumulant that contains an additional term discounting the outliers,
4. "PW-LR kurt"; a classic fourth cumulant.

The PW-LR approach clearly outperforms all the previously tested methods on the synthetic benchmark datasets (Hyvärinen and Smith, 2013). However, each one of the PW-LR methods is based on a single higher-order cumulant based on the BOLD fMRI distributions of the two signals $X(t)$ and $Y(t)$ (either third or fourth order), it is not robust as lower moments can also account for possible differences in the local SNRs. Furthermore, the SNR magnitude can differ with respect to the various features of the underlying neuronal time series, and if this is the case, these methods can erroneously flip the directionality of the connection.

There are also other statistical methods for pairwise causal inference based on a combination of marginal and conditional probabilities such as information-geometric approach by Janzing et al., 2012 or Unsupervised Inverse Regression by Sgouritsa et al., 2015. However, I will not include these methods in the method comparison as they have not been applied to fMRI datasets to date.

Therefore, I further expand on the concepts of PT and PW-LR by proposing a classifier based on complex cumulants derived from multiple, possibly fractional, moments of the distribution of the BOLD fMRI recordings. I compare the performance of this approach on synthetic benchmark datasets (Smith et al., 2011) relative to other methods for directed functional connectivity often used in fMRI research. Furthermore, I compare performance of the methods using simple two-node simulations generated from the DCM model with varying signal magnitude and noise variance in the projecting (upstream) and the receiving (downstream) node. In this study, I demonstrate that the proposed classifier is more robust to these natural confounders encountered in fMRI experiment than other methods.

In Section 7.2.1, I re-introduce the concept of fractional moments of the distribution, and in Section 7.2.2, I explain the procedure of combining fractional moments into fractional cumulants. In Section 7.2.3, I give details on the classifier built based on a set of fractional cumulants, and in Section 7.2.6, I describe the supervised learning procedure. In Section 7.2.4, I
list popular methods for effective connectivity research in fMRI used for the method comparison. Finally, in Section 7.2.6, I describe the generation of additional synthetic data with confounds often encountered in fMRI datasets, which I further use to benchmark the methods.

In Section 7.3.1, I describe the results of the validation of the new method using synthetic benchmark datasets (Smith et al., 2011). Furthermore, in Section 7.3.2, I describe the results of an additional validation performed with the use of the DCM generative model, but in presence of confounds, i.e., the background noise and the variability in SNRs between upstream and downstream region. In Section 7.4, I critically discuss the results.

7.2 Materials and methods

7.2.1 Fractional moment series of BOLD fMRI signal distributions

In this work, I propose estimating causal links from BOLD fMRI recordings by analyzing the dependence between an expanded set of (fractional) moments of the distribution. I keep the same scheme for the causal discovery as proposed by Patel, Bowman, and Rilling, 2006 and implemented by Hyvärinen and Smith, 2013. However, I improve on the second step of the causal inference - a two-node classification problem - by utilizing all moments of the BOLD fMRI distributions combined into cumulants.

Fig. 7.1 presents a graphical representation of a dynamical system with just two nodes. In this problem, one region ("upstream") is sending information to another region ("downstream") through a connection of weight $A_{12}$. Both regions receive region-specific signal $u_i(t)$. This signal can both relate to experimental input as well as input from other regions. Both nodes also are influenced by a background neuronal noise $\sigma_i(t)$ that has impact on the local SNR. The BOLD fMRI dynamics in such a simple two-node networks can be simulated with the use of the DCM generative model (Friston, Harrison, and Penny, 2003). Given the inputs to the network, the connection strengths and sets of parameters characterizing local hemodynamic responses within the nodes, the DCM generative model makes prediction on the outcome BOLD fMRI dynamics in both nodes. This BOLD fMRI time series can then be normalized, and characterized in terms of its central moments (as previously introduced in Chapter 3):

$$M_k = \frac{1}{N} \sum_{i=1}^{N} \tilde{x}_i^k$$

(7.1)

where $k \in Q$, $\tilde{x}$ - normalized time series, $N$ - the length of BOLD time series.

In this approach, the novelty is that a discrete set of moment orders $k$ typically used to characterize a distribution (in terms of mean, variance, skew, etc.), is converted into a (pseudo-) continuous dimension, by sampling moment orders $k$ from sets of rational numbers within certain range. These fractional moments of the distribution are not isomorphic with the moment generating function as we do not convert the distribution of BOLD values into a probability density function at any stage.

As mentioned in Chapter 3, once the original BOLD time series is normalized to the mean of 0 (and variance of 1), it contains negative values and,
therefore, the fractional moments will become complex. Since the Eq. 7.1 is continuous with respect to $k$, these fractional moments will form a curve in the complex plane (Fig. 7.1B). In Fig. 7.1B, I present a phase diagram for all moments in the range $k \in [0, 5.0]$, for a long simulated BOLD fMRI time series representing a simple two-node network in the noiseless case. The moments are computed separately on the demeaned data, for the upstream region (blue) and the downstream region (red). The curve starts at $(1, 0)$ for $k = 0$. Then, it traverses the upper half-plane and arrives at $(0, 0)$ for $k = 1$. Subsequently, it goes back through the lower half-plane and comes back to $(1, 0)$ for $k = 2$ since the variance is equal to 1. Every time $k$ becomes an integer, the curve crosses the real axis. Note that the imaginary axis characterizes the left half of the BOLD distribution, since fractional moments give nonzero imaginary part for the negative values of distribution of the BOLD values.

**Figure 7.1:** A two node network with one directed connection. A: The upstream node, $x(t)$, is sending information to the downstream node, $y(t)$, through a single connection of weight $A_{12}$. Both regions received a binary signal $u_i(t)$ and neuronal noise $\sigma_i(t)$. The proportion between the amplitude of $s_i(t)$ and the variance of the noise $\sigma_i(t)$ defines the SNR in the network. B: All the fractional moments for $k \in [0; 5]$, for the BOLD fMRI time series from a simulated 2-node network, in the noiseless case. Blue: upstream region. Red: downstream region. The curve starts from $(1, 0)$ for $k = 0$, traverses the upper half-plane and arrives at $(0, 0)$ for $k = 1$. Then, it travels back through the lower half-plane towards $(1, 0)$ for $k = 2$ as the BOLD variance was fixed to 1 through the normalization. Every time the moment order $k$ becomes an integer, the curve crosses the real axis.

### 7.2.2 Complex cumulants of the distribution

For any two time series $x(t)$, and $y(t)$, not only the sole fractional moments but also the asymmetry between the moments can indicate the directionality
of a connection. This asymmetry can be quantified by "fractional cumulants":

\[ C_{kl} = \frac{1}{N} \sum_{j=1}^{N} \left( x^{k}_i y^{l}_j - x^{k}_j y^{l}_i \right) \]  

(7.2)

where \( k, l \in \mathbb{Q} \).

In this particular problem, \( x(t), y(t) \) denote the BOLD fMRI time series in the two-node system. To make a prediction ("upstream" versus "downstream"), one needs to learn the dependencies between moment time series using synthetic datasets derived from the DCM generative model. I have run 1,000 2-node DCM simulations with \( Fs = 200Hz \) for a duration of 10[min]. To marginalize out the influence of the hemodynamic parameters from my results, I sampled the parameters independently for the two nodes, and from the empirical distributions (Friston, Harrison, and Penny, 2003). To marginalize out the effect of different input strengths and frequencies, I also sampled the input magnitudes and frequencies (probabilities of switch from on- to off- state and vice versa) from a Gamma distribution with mean and variance of 1.

---

**Figure 7.2:** A: Mean values for all cumulants, over 1000 simulations of a two node network with one connection (Fig. 7.1), for \( k, l \in [0.0, 5.0] \). Since cumulants are antisymmetric with respect to indexing \( k, l \), the heatmaps for the real and the imaginary component are also antisymmetric. B: A zoom into a smaller range of \([0.0, 3.0]\). The imaginary component is equal to zero for cumulants of integer orders: \( k, l \in \mathbb{N} \), but changes the sign in the intervals between in a systematic way. C: Sign of the cumulants in the majority of 1,000 instances of the generative model. Red: positive. Blue: negative. Green: zero. I further denote the “sign maps” for real and imaginary components as \( S_r \) and \( S_i \).
The input signals driving the upstream and the downstream region were also sampled independently from each other, as trains on- and off- states governed by Poissonian processes. The background neuronal noise was set to 0. To obtain precise estimation of fractional cumulants, I did not subsample the synthetic BOLD fMRI every 2–3 s as is typically done.

I performed this simulation twofold. Firstly, I fed in an empty connection to create a null distribution of cumulants. Secondly, I added a connection $A_{12}$ with a weight of 0.9 to the 2-node system. In Fig. 7.2, I demonstrated the mean values for all cumulants of indexes $k, l \in [0.0, 5.0]$ obtained from the simulations of this connection in Cartesian coordinates.

Fig. 7.2A presents the mean values for all cumulants in the range $k; l \in [0.0, 5.0]$, over 1,000 simulations of a two node network with one connection (Fig. 7.1). Fig. 7.2B shows the same maps zoomed into the range $k; l \in [0.0, 3.0]$. Finally, for every cumulant, Fig. 7.2C presents the sign of this cumulant for the majority of 1,000 instances of the simulation (I further refer to these binary maps as $S_r$ and $S_i$). These binary maps do not represent confidence intervals (or, discriminability) for particular cumulants as majority could mean 51% as well as 100% simulations. To choose cumulants that can best discriminate between a "connection" and "no connection" case, I created distributions of cumulant values across the 1,000 simulations in the null case and compared against the distributions derived from simulations with a non-zero connection. I smoothed these distributions with kernel smoothing function and, for each cumulant, I computed the percentile of samples falling beyond 95th percentile of the null distribution (in case the mean for the given cumulant is negative as in Fig. 7.2C, I took samples falling lower than the bottom 5th percentile of the null distribution, and higher than the 95th percentile otherwise).

**Figure 7.3:** Discriminative power for all cumulants in the range $k; l \in [0.0, 5.0]$, in the ideal case of a very long BOLD time series and no background neuronal noise. Maximal discriminability in both real and imaginary cumulants is achieved in the range of $k + l > 5.0$. Additionally, in real cumulants, there is a region of high discriminability in the range $k + l \leq 2.0$, while in imaginary cumulants there is such a region in the range $2.0 < k + l < 5.0$ (both boundary lines marked with a white line).
7.2. Increasing robustness of methods for effective connectivity in fMRI: Methods

The results of the discriminability analysis in Cartesian coordinates are presented in Fig. 7.3. One can observe that whenever one of the indexes $k, l$ equal to zero, i.e., the cumulant reduces to a simple moment, it has lower discriminative value than the full cumulants. Therefore, I will disregard moments from further analysis and fully concentrate on full cumulants, i.e., the asymmetry between moments ($k, l > 0$).

**FIGURE 7.4:** Cumulants in polar coordinates. A: Mean values for all cumulants over 1000 simulations. The radius and the phase are computed with respect to point (0,0), otherwise the (anti-)symmetry would be lost. Since cumulants are antisymmetric with respect to indexing $k, l$, the heatmaps for the phase are antisymmetric. The radius is always positive. B: Zoom into a smaller range of $[0, 3.0]$. C: Sign of the cumulants. Red: positive. Blue: negative. Green: zero. D: Discriminative power for all cumulants in the range of $k, l \in [0, 5.0]$, in polar coordinates (in the ideal case of a very long BOLD time series and absence of the background neuronal noise). In polar coordinates, cumulants are less informative. Phase is not informative since the discriminative power is relatively high only in the high-moment regime while high moments are hard to estimate for a short time series. Radius is not informative as well, as it is always positive.
Interestingly, the range of high discriminability is different for real and imaginary components. Maximal discriminability in both real and imaginary cumulants is achieved within the range of $k + l > 5.0$. Additionally, in real cumulants, there is a region of high discriminability in the range of $k + l < 2.0$, and in imaginary cumulants, there is such a region in the range of $2.0 < k + l < 5.0$ (Fig. 7.3, both boundary lines marked with a white line). This different characteristic along the imaginary axis illustrates that moving from integer to fractional moments of the distribution provides with an additional predictive power. Note also that, as I am deriving the discriminability maps from the DCM generative model, these maps are fingerprints of the particular problem of effective connectivity in fMRI; when derived from another generative model simulating another dataset, these maps would be different.

I also performed analogous analysis in polar coordinates, however this did not yield a substantial improvement to the classification performance. The results of the analysis are presented in Fig. 7.4.

To investigate how the performance of classification based on single cumulants changes when a single connection is embed in a bigger network, I evaluated their success rate in estimating connectivity for benchmark synthetic datasets (Smith et al., 2011). Fig. 7.5 presents the grand mean success rate achieved with the use of every cumulant separately, across all 28 benchmark synthetic datasets.

![Figure 7.5: Success rate for all the individual cumulants, averaged over 28 simulations from the synthetic benchmark datasets (Smith et al., 2011). White-edged square - a single cumulant used by Hyvärinen and Smith, 2013. The performance of this cumulant is shown in the colorbar, white band. Black-edged squares - cumulants that give the highest performance on this dataset. Their performance is presented in the colorbar, black band. For the cumulants of high indexes $k + l > 4.0$ (white line), the success rate is not as high as the discriminability presented in Fig. 7.3 would suggest. The high success rate is not preserved for high-indexed cumulants that achieved high discriminability on 2-node simulations (low noise case). The maximal grand mean performance equals 0.847 for the real components, and 0.814 for imaginary components.](image)

Fig. 7.5 presents the grand mean success rate achieved with the use of every cumulant separately across all 28 benchmark simulations.
7.2. Increasing robustness of methods for effective connectivity in fMRI: Methods

The maps of simulation-dependent success rate relate to the maps of discriminative power (Fig. 7.3), but they are not identical and differ between iterations of the simulation. One difference is that for the cumulants of high indexes $k + l > 4.0$, the success rate is not as high as the discriminability presented in Fig. 7.3 would suggest. This is because Fig. 7.3 represents the limit of a system of two isolated nodes with infinite SNR, and a very long BOLD fMRI time series, whereas benchmark simulations refer to more realistic case.
In this realistic scenario, for each pair of nodes the BOLD fMRI time series is short, there are confounding signals from other nodes in the network, and there is a certain degree of noise in the neuronal communication. Altogether, these factors cause that the high moments are hard to estimate in practice.
7.2.3 Combining fractional cumulants into a classifier

Further, I propose to combine information contained in multiple cumulants by building the classifier based on a “voting” scheme between the cumulants. This classifier determines whether the map of cumulants obtained for a pair of time series $X(t), Y(t)$ is closer to the benchmark maps presented in Fig. 7.2A (which is an evidence for a connection $X \rightarrow Y$), or their inverse (which is an evidence for a flipped connection $Y \rightarrow X$). Each of the cumulants $C_{k,l}$ votes due to signs $S_{r,k,l}, S_{i,k,l}$ (Fig. 7.2C). If the sign of the cumulant is the same as in Fig. 7.2C, it adds to the evidence for the connection $X \rightarrow Y$, and against this connection otherwise.

Since in realistic conditions (short datasets, large TRs), high index cumulants, $k + l > 3.0$, yield the aforementioned estimation problem, I discount their contribution from the voting by using a nonlinearity of the following form (further referred to as “weighting” throughout the chapter):

$$f(x) = \log(\cosh(\max(x, 0)))$$  \hspace{1cm} (7.3)

Similar function was proposed to discount the outliers present in the BOLD time series in the work by Hyvärinen and Smith, 2013. The final classifier yields:

$$\begin{align*}
\begin{cases}
X \rightarrow Y & \text{if } \sum_{k,l} [S_{r,k,l}f(C_{r,k,l}) + S_{i,k,l}f(C_{i,k,l})] \geq 0 \\
Y \rightarrow X & \text{otherwise}
\end{cases}
\end{align*}$$

in the weighted case, and

$$\begin{align*}
\begin{cases}
X \rightarrow Y & \text{if } \sum_{k,l} [S_{r,k,l}C_{r,k,l} + S_{i,k,l}C_{i,k,l}] \geq 0 \\
Y \rightarrow X & \text{otherwise}
\end{cases}
\end{align*}$$

in the unweighted case.

7.2.4 Supervised learning using synthetic benchmark datasets

In this work, I derive the classifier using sign maps $S_{r,k,l}, S_{i,k,l}$ (Fig. 7.2C) developed using multiple realizations of the 2-node simulation under the DCM generative model - which is a form of supervised learning. In a different application and under a different generative model, these sign maps would look differently.

Cumulants differ in discriminability (Fig. 7.3). The success rate of cumulants differs depending on the range $\text{Ind}_{\max}$. Therefore, I optimize the performance of the classifier with respect to the dimensions $k, l$ by finding a combination that gives the best grand mean performance across the 28 simulations from the synthetic benchmark datasets as this collection of simulations widely represents the variety of experimental conditions encountered in real-life fMRI setups.

Firstly, I fix $\text{Ind}_{\max} = \max(k, l) = 3.1$, and consider cumulants in the triangle $k, l \geq 0.1, k + l < \text{Ind}_{\max}$. Then, I choose only cumulants of discriminability exceeding a particular value to be fed into the classifier. E.g., a cut-off value of 0.1 means the vote from all cumulants for which the discriminative value is not less than 0.1 is included in the classifier. I can then evaluate
the grand mean success rate (as the mean success rate over all 28 benchmark synthetic datasets) in the function of the thresholding discriminability value. Fig. 7.7A, demonstrates that including all the cumulants with a positive discriminative value (all cumulants except for \( k = l \), for which discriminability is always zero) gives the best classification performance.

![Figure 7.7](image)

**Figure 7.7:** Dependence of the grand mean performance when applied to the synthetic datasets on the choice of cumulants. A: Grand mean performance for unweighted cumulants in the range of \( k+l \leq 3.1 \), in the function of the cut-off discriminative value. The higher the cut-off, the fewer cumulants one takes into account while voting for the directionality of the causal link. The results clearly demonstrate that, to maximize the success rate in estimating effective connectivity, all the cumulants should be taken into account (except for the diagonal of \( k = l \)). B: The grand mean performance based on cumulants of indexes \( k, l \) between 0.1 and \( k+l \leq \text{Ind}_{\text{max}} \), in the function of \( \text{Ind}_{\text{max}} \). The optimal performance in the unweighted case equals 0.835 for \([\text{Ind}_{\text{R}_{\text{max}}}, \text{Ind}_{\text{I}_{\text{max}}}] = (2.4, 1.7)\), and 0.886 for \([\text{Ind}_{\text{R}_{\text{max}}}, \text{Ind}_{\text{I}_{\text{max}}}] = (2.1, 3.7)\) in the weighted case.

Secondly, I optimize the window \( \text{Ind}_{\text{max}} \) for indexes \( k, l \) and compare the classifier with and without weighting with the discount function introduced in Eq. 7.3. Since discriminability is generally higher for low indexes \( k, l \) (Fig. 7.3), I evaluate the grand mean performance based on cumulants of indexes between 0 and the maximum \( \text{Ind}_{\text{max}} \), in the function of that maximum. I consider the maximal indexes along real and imaginary dimension separately. The results are presented in Fig. 7.7B. The optimal performance in the unweighted case equals 0.835 for \([\text{Ind}_{\text{R}_{\text{max}}}, \text{Ind}_{\text{I}_{\text{max}}}] = (2.4, 1.7)\), and 0.886 for \([\text{Ind}_{\text{R}_{\text{max}}}, \text{Ind}_{\text{I}_{\text{max}}}] = (2.1, 3.7)\) in the weighted case, which exceeds both the grand mean performance of the "PW-LR r skew" method by Hyvärinen and Smith, 2013, (0.845), and the maximal grand mean performance of any single cumulant in this study (Fig. 7.5, the maximum of
7.2. Increasing robustness of methods for effective connectivity in fMRI: Methods

0.847).

7.2.5 Selection of other approaches for effective connectivity research in fMRI

To benchmark this classifier, I compare the performance against other, established methods for effective connectivity in fMRI. As mentioned in the Introduction, the field of effective connectivity in fMRI is very wide (Smith et al., 2011), therefore, I restricted this study to the most popular approaches (other than the DCM itself, Friston, Harrison, and Penny, 2003, as the synthetic datasets are generated directly from the DCM generative model):

1. State-space implementation of Granger Causality (GC, Granger, 1969; Seth, Barrett, and Barnett, 2015): as introduced in Chapter 5, GC is a multivariate method inferring effective connectivity between a pair of time series under the assumption that both of them can be expressed as autoregressive processes. I used simple version of GC featuring Ordinary Least Square regression with the time lag of 1 frame, implemented in Multivariate Granger Causality Toolbox (Barnett and Seth, 2014), obtained from http://www.sussex.ac.uk/sackler/mvgc. For GC based on vector autoregression (VAR) process as in this study, the state-space implementation is more robust than the spectral GC (Geweke, 1982; Geweke, 1984) because the frequency-domain version has a bias-variance trade-off (a function of the VAR model order that can induce spurious conditional GG causality estimates such as erroneous peaks in the frequency domain, as indicated in the recent work by Stokes and Purdon, 2017). Furthermore, the state space formulation of GC is the most robust, mitigating effects of bias and variance due to the fact that the reduced model is VAR rather than VARMA (Barnett and Seth, 2015). To compare performance with the methods for pairwise inference, I used GC in a bivariate rather than multivariate fashion: by applying GC to each of the previously found connections separately,

2. Partial Directed Coherence (PDC, Baccalá and Sameshima, 2001). PDC is known as a method conceptually close to Directed Transfer Function (DTF, Kaminski and Blinowska, 1991; Baccalá and Sameshima, 2001). Both these methods are derivatives from Geweke spectral measures of GC (Geweke, 1982; Geweke, 1984) and all these methods have similar limitations (Chicharro, 2011). However, PDC is used substantially more often in fMRI studies than the other two methods, especially when compared to the spectral GC. For this reason, I chose PDC as a method representing this class of approaches. I used PDC implementation from the Extended Multivariate Autoregressive Modeling Toolbox (Faes et al., 2013): http://www.lucafaes.net/emvar.html. Similarly as in the case of GC, I applied PDC in a bivariate fashion,

3. Patel’s tau (PT, Patel, Bowman, and Rilling, 2006), implemented similarly as in Smith et al., 2011: by recalculating each time series into the range $[0, 1]$, setting samples under the 10th percentile to 0, over the 90th percentile to 1, and linearly mapping the remaining samples to the range $[0, 1]$. Then, the directionality of the connection is inferred
from the difference between $P(X|Y)$ and $P(Y|X)$. In addition to the previous implementation, however, I also integrate the results over all the possible thresholds to eliminate the thresholding problem while calculating the conditional probabilities $P(X|Y)$ and $P(Y|X)$.


As in Hyvärinen and Smith, 2013, I performed the first step of the inference, i.e., determined the functional connectome, by calculating inverse covariance and thresholding the outcome functional connectome with the use of permutation testing. All the methods, including multivariate methods such as GC and PDC, were then applied in a pairwise fashion (i.e., separately for each two-node node network representing a single connection found in the previous step of the inference).

Furthermore, I did not include the DCM procedure in this comparison, for the same reasons as Smith et al., 2011: DCM is not an exploratory method and it should not be used in this context, namely for exploratory causal research on the set of benchmark synthetic datasets (where the smallest network consists of 5 nodes) as it is not computationally feasible. Furthermore, the characteristics of this synthetic benchmark datasets is that, input signals (Fig. 7.1A) represent random events and can therefore emulate all types of fMRI experiments: classic task-fMRI studies, event-related responses, or resting state BOLD fMRI time series. In DCM, however, the inputs must be strictly specified in the block design, or otherwise, the DCM inference cannot be initiated. Therefore, the assumptions behind DCM do not fit the research problem formulated in this particular way.

### 7.2.6 Testing robustness of the methods against confounds

In addition to evaluating this approach against the benchmark datasets from Smith et al., 2011, I further evaluate the performance under additional yet typical modes of variation in the data. Specifically, I am interested in characterizing the discriminative performance relative to (i) more complex forms of stochastic noise in the data and (ii) unequal levels of SNR per node.

The benchmark synthetic datasets involve temporally uncorrelated white background noise of a low magnitude on the neuronal level (Smith et al., 2011). This type of noise is not physiologically plausible, as it is known from physiological studies that in the neuronal networks, the background noise has a scale-free power spectrum (He, 2014; Bédard, Kröger, and Destexhe, 2006; Dehghani et al., 2010; Bielczyk et al., 2017b). Therefore, I simulated a two node system and injected scale-free (pink) noise to the system. Then, I varied the variance of the noise in the range of $[0.2, 5.0]$ while keeping the amplitude of the inputs $s_i(t)$ fixed to 1.0. I performed 500 realizations of $10\text{[min]}$ simulations at high temporal resolution of $Fs = 200\text{[Hz]}$, for each configuration of the noise variances.
Furthermore, in the original version of the DCM procedure (Friston, Harrison, and Penny, 2003), as well as in the most of the computational studies (Smith et al., 2011), equal stimulus strengths to both nodes $s_i(t)$ are assumed. This assumption might not hold true in the experimental fMRI datasets. Therefore, I performed another, noiseless simulation, in which I varied signal strengths between the upstream and the downstream region. I performed 500 realizations of 10[min] simulation at high temporal resolution of $F_s = 200[Hz]$, for each configuration of input strengths in the range of $[0.2, 5.0]$. 

### 7.3 Results

#### 7.3.1 Supervised learning using synthetic benchmark datasets

The best version of the classifier was obtained for voting between cumulants in the range $[\text{IndR}_{\text{max}}, \text{IndI}_{\text{max}}] = (2.1; 3.7)$, with the discount for high moment indexes (Eq. 7.3). The comparison of this classifier against four other methods, GC, PDC, PT and "PW-LR r skew," on the benchmark simulation no 2 is presented in Fig. 7.8. The violin plots denote the distribution of the Z-scores for connections as compared to the null distribution. Blue dots denote the percentage of correct assignments for the true connections, as in Smith et al., 2011. In most of the other 27 benchmark datasets, this classifier outperforms all the other methods (Fig. 7.9). As in the original study by Smith et al., 2011, the lagged methods, GC and PDC, perform worse than the structural methods. "PW-LR r skew" and fractional cumulants both outperform PT, most probably because this implementation of PT is based on the thresholded signal and therefore contains a free parameter.

![Comparison on synthetic benchmark dataset no 2](image)

**Figure 7.8**: Comparison between the classifier based on the fractional cumulants and several other methods, on the benchmark simulation no 2. The violin plots - the distribution of the Z-scores for connections as compared to the null distribution. Blue dots - the percentage of correct assignments for the true connections (Smith et al., 2011). The difference in performance between the classifier based on fractional cumulants and "PW-LR r skew" (Hyvärinen and Smith, 2013) is small.
In general, in the benchmark synthetic datasets fractional cumulants outperform all other techniques in almost all cases, although the difference between performance of the fractional cumulants and PW-LR methods is small. Similarly as in the original study by Smith et al., 2011, simulation no 7 is relatively the easiest (as it represents a small network composed of 5 nodes, with an extended session duration of 250 minutes) while simulation no 13 is relatively the hardest (as in this simulation, the tested network contains a number of extra backward connections).
7.3. Increasing robustness of methods for effective connectivity in fMRI: Results

In most of the simulations, fractional cumulants give a slight improvement compared to the main competitor, "PW-LR r skew." In simulations 7, 8, 9, 14, 17, 20, and 23 the performance is roughly the same for both methods.

Both GC and PDC do not perform well on these datasets, even in the simplest setting (simulation no 7): the mean success rate is close to the chance level in all the simulations. Additionally, results from PDC are spread; the Z-scores obtained from the PDC estimate have a very high variability. This is not the case for GC, but in this case, the mean success rate is still not
satisfactory to call these results a success. To sum up, the methods rendering causal links from the information contained in the temporal sequence of the samples, do not perform well on these datasets - regardless if they are based on the spectral features of the signals or on fitting autoregressive models to the datasets.

7.3.2 Robustness of the methods with respect to confounds

Fig. 7.10: Robustness of the methods against the background scale-free neuronal noise. The variance of the background noise $\sigma(t)$ differs between upstream and downstream region in the range of $[0,2,5.0]$. As the signal magnitude is constant and equal to 1 in these simulations, the SNR was calculated as $\frac{\text{var}(\sigma)}{\text{var}(\eta)}$. Only the classifier based on fractional cumulants gives a performance above the chance level across the whole parameter space.

Fig. 7.10 presents the comparison between the classifier based on fractional cumulants and various other methods on a 2-node simulation with varying levels of physiologically plausible, scale-free neuronal noise. The results suggest that all the previously tested methods exhibit low levels of robustness towards these additional sources of variability in the data. GC as well as PDC give the lowest performance, and yield results on the chance level across the whole parameter space. PT seems to be fully resilient to the background noise in the downstream node, but not in the upstream node. The performance of "PW-LR r skew" drops down to the chance level with respect to the noise level in both the upstream and downstream node, whereas GC and PDC perform almost equally poorly under any combination of noise variances (probably because the variance of the noise from the fitted autoregressive model is used to establish the directionality of the causal influence in GC).

Fig. 7.11 presents the comparison between the classifier based on fractional cumulants and other methods given noiseless simulation and varying signal magnitudes. The classifier based on fractional cumulants is the
only method where the performance does not descend towards chance level within the chosen parameter space. GC and PDC give performance around the chance level across the whole parameter space, whereas "PW-LR r skew" and PT exhibit certain resilience towards this variability in the inputs. However, the performance breaks down towards the chance level at for the disproportion between the inputs higher than 3.0.

**FIGURE 7.11**: Robustness of different methods to varying SNRs. The variance of the signal differs between upstream and downstream region, both in the range of \([0.2, 5.0]\). GC and PDC yield performance around the chance level across the whole parameter space, whereas "PW-LR r skew" and PT exhibit certain resilience towards the variability in SNRs. However, the classifier based on fractional cumulants is the only method whose performance does not fall towards the chance level within the parameter space explored in this study.

### 7.4 Discussion

#### 7.4.1 Summary

This work provides an advance to the effective connectivity research in fMRI by utilizing the additional information contained in the BOLD fMRI time series with the use of fractional moments of the BOLD fMRI distribution combined into cumulants. Using this additional information (embedded within a classifier) significantly increases the robustness towards plausible sources of variability in fMRI, namely presence of physiologically realistic (scale-free) background noise as well as disproportion in the inputs strengths, either due to the differences in the amount of neuronal activity locally induced and/or due to effective differences induced by, e.g., regional variations in the coil sensitivity profiles. This is where the added value of information coming from fractional cumulants becomes apparent: among the methods tested in this work, only the classifier based on the fractional cumulants gives performance better than chance across the whole parameter space.
Effective connectivity is a research problem directly related to the notion of causality. As previously mentioned in Chapter 5, causality is, in general, difficult both to define (Pearl, 2000) and to measure. In the most basic formulation, if $X$ causes $Y$, it means that without $X$, $Y$ would not occur. In practice, the picture is far less clear in complex dynamic systems such as brain networks: for any event, high number of potential causes can be defined, and these causes most often interfere with each other. This research problem was recently discussed by Albantakis et al., 2017 who decomposed causality into independent dimensions: realization, composition, information, integration, and exclusion. Also, interpretation of a causal component in the given process depends on the context. E.g., respiratory movement is typically considered a confound in fMRI experiments, unless one is interested in the influence of respiration speed on the activity of neuronal populations. Furthermore, in brain networks, temporal ordering of the cause and effect is hard to maintain as information is circulating in recurrent rather than feed forward networks (Schurger and Uithol, 2015).

Furthermore, if one is able to conduct an interventional study, establishing causality becomes straightforward - but this is rarely the case in human research. In human fMRI, all the studies are observational rather than interventional. In such case, causation can never be observed directly, just correlation (Hume, 1772) - and, when correlation is highly stable, we are inclined to infer a causal link. Additional information is then needed to assess the direction of the assumed causal link, as correlation indicates for association and not for causation (Altman and Krzywiński, 2015).

In the simulations, I used multiple iterations of the dynamics for every network pattern. This is because I had to run multiple instances of a noisy system to evaluate the mean success rate of the methods under noisy circumstances. I recorded long samples of the dynamics as I aimed to find an upper bound on the methods’ performance in a function of the SNR disproportion and background noise levels (I did not investigate the effects of the sample length on the results). This is because in this study, I focused on the confounding factors which - unlike the duration of the study - cannot be influenced by the researcher.

As mentioned in Chapter 3, although fractional moments of a distribution, as a mathematical concept, were studied before (Dremin, 1994), this concept was not applied to biomedical sciences to date. One reason for this lack of applications might be that the fractional moments become complex numbers for the normalized time series, and that subsequently, the features characterized by these moments cannot be conceptualized as easily as the features characterized by the integer moments (e.g., skewness can be interpreted as a measure of “asymmetry” of the distribution, and kurtosis can be interpreted as its “flatness”). However, although the fractional moments of a distribution are a mathematical concept with limited practical interpretation, they still contain valuable (causal) information. In this work, I demonstrate that these features of the distribution provide important additional information about the distribution of the BOLD fMRI values. I first derive the classifier via supervised learning performed on the set of benchmark synthetic datasets, and then validate the classifier on 2-node simulations with biologically realistic confounds. Confounding factors such as a physiological background noise of a magnitude varying between the nodes are important to overcome for any method for causal inference in fMRI. This is
because every network in the brain is embedded in larger networks, therefore the background activity coming from other, interconnected networks can be interpreted as "noise" (Deco, Jirsa, and McIntosh, 2011). I demonstrate that this approach can increase the robustness of the methods for pairwise inference in fMRI to the main sources of variability present in the fMRI experiments.

Unlike the previous methods for pairwise inference in fMRI (Hyvärinen and Smith, 2013), the classifier defined in this study is informed by the DCM generative model, therefore it incorporates the priors derived from the neurophysiological studies (Buxton, Wong, and Frank, 1998). Deriving benchmark sign maps for the classifier from the multiple instances of the DCM generative model allows for marginalizing out all the parameters unimportant for the effective connectivity research: the classification procedure focuses on classifying a pair of regions into an upstream and a downstream node instead of fitting all the hyperparameters as is done in the classic DCM inference procedure. Therefore, this approach is (1) a reduction of the problem of effective connectivity in a large network to a two-node classification problem, (2) an extension of the feature space from integer to fractional moments.

The sign maps derived in the process of supervised learning are dependent on the generative model. In this study, I chose the canonical, original formulation of the DCM generative model (Friston, Harrison, and Penny, 2003). There are also newer formulations of the DCM, e.g., the canonical microcircuit approach (Pinotsis et al., 2017) in which layer-specific neuronal populations in the cortex are modelled with the use of neural mass models. In this case, I assume that the inference is performed on a mesoscale level in which ROIs represent brain regions (cortex regions or subcortical nuclei) rather than cortex components. Furthermore, although a version of DCM containing higher-order, nonlinear effects (Stephan et al., 2008) is also developed, I believe that (bi)-linear model is a good simplification to describe the underlying neuronal dynamics as it refers to the linear part of the sigmoidal transfer functions between neuronal populations in the brain (Silver, 2010; Bielczyk et al., 2015). Modeling communication between nodes in the network with the use of linear transfer functions is a common practice in modeling effective connectivity in fMRI, see, e.g., SEM (McIntosh and Gonzales-Lima, 1994) or LiNGAM-ICA (Shimizu et al., 2006; Smith et al., 2011). The bilinear version of DCM, often referred to as the general linear model approach, is still a very popular tool for finding effective connectivity patterns from fMRI in the clinical practice (see, e.g., recent work by Zhang et al., 2018; Nackaerts et al., 2018; Arioli et al., 2018; Pool et al., 2018).

I reproduced the DCM generative model after Smith et al., 2011. This implementation is acknowledged in the field as the benchmark tool for testing new methods for functional and effective connectivity in fMRI (see, e.g., Smith et al., 2013; Hinne et al., 2015; Bielczyk et al., 2018). In this implementation, bilinear effects (namely, modulation of connections by experimental inputs) are not modelled. In pairwise inference, this omission is justified as modulation of connections only affects the strength of the connection weight $A_{12}$ which will influence cumulant values quantitatively but not qualitatively.

Evaluating methods with the use of synthetic datasets as the ground truth is typically the first step in the validation process for any new data
analytic framework. Validating new methods with the use of synthetic datasets is a golden standard across the whole field of neuroimaging, from single cell imaging to whole-brain imaging with the use of fMRI or EEG/MEG. This tradition has a long history, starting from the Nobel-winning Hodgkin and Huxley model for initiation and propagation of action potentials (Hodgkin and Huxley, 1952). Today, methods for effective connectivity between neuronal assemblies measured with multielectrode arrays are still validated on synthetic datasets generated from this classical model, including recent approaches: nonlinear data assimilation (Hamilton, Chen, and Gotlib, 2013) and differential covariance (Lin et al., 2017). In cognitive neuroimaging, testing methods on synthetic datasets from generative models is also a standard. In EEG/MEG research, there are multiple classes of generative models generating different type of dynamics, depending on the purpose of the modeling study, e.g., the nonlinear lumped-parameter model for generating alpha rhythms and its neural-mass extension by David and Friston, 2003, Wong-Wang model for winner-take-all dynamics (Wong and Wang, 2006), Hindmarsh-Rose model for epileptor dynamics (Hindmarsh and Rose, 1984), and DCM for EEG/MEG (Kiebel et al., 2009; Steen et al., 2018; Moran, Pinotsis, and Friston, 2013). Furthermore, the Human Neocortical Neurosolver simulator developed at the Brown University (HNN, https://hnn.brown.edu) is a complex tool simulating Local Field Potentials measured with EEG/MEG by bottom-up modeling of clusters of neurons. All these tools can serve to validate new methods for functional and effective connectivity in EEG/MEG (Valdes-Sosa et al., 2011; Wang and Krystal, 2014).

In fMRI, the selection of generative models is narrower than in EEG/MEG: the DCM Friston, Moran, and Seth, 2013; Smith et al., 2011 achieved a status of a state-of-the-art, standard generative model. With the use of the synthetic data generated from this model, new methods for effective connectivity in fMRI are validated, e.g., the methods based on the third and fourth cumulant by Hyvärinen and Smith, 2013 and artificial immune algorithm combined with the Bayes nets (AIAEC, Ji et al., 2016).

To date, DCM is the most biologically relevant generative model proposed in the field of fMRI research. The implementation of benchmark synthetic datasets based on DCM by Smith et al., 2011 has gained a lot of attention and following in the field. It also received some critics. E.g., according to the Smith’s results, Patel’s tau (Patel, Bowman, and Rilling, 2006) is one of the methods that give the best performance in retrieving directed connectivity patterns from synthetic benchmark datasets. In a recent work, Wang et al., 2017 performed modeling study on datasets derived from experiments by David et al., 2008 in which fMRI activity was measured in genetically modified rats suffering from epilepsy. Activity from the same set of regions was recorded in an associated intracerebral EEG study to uncover the ground truth information flow. The authors chose the primary somatosensory cortex barrel field (S1BF), the thalamus and the striatum (the caudate-putamen; CPu) as ROIs and demonstrated that Patel’s tau is not better than chance in recovering directional connectivity patterns from this data on both raw and deconvolved fMRI datasets. On the contrary, DCM and GC proved to correctly estimate the directionality of the information flow on the group level on the deconvolved data.

There are more caveats with respect to the benchmark DCM simulations
7.4. Increasing robustness of methods for effective connectivity in fMRI: Discussion

by Smith et al., 2011. Firstly, synthetic datasets derived by the authors of the study involve a low-noise condition, in which the background noise in the networks in as low as 5% of the signal magnitude. Given that the background activity in the brain networks includes not only noise but also echo of cognitive processes unrelated to the experiment, 5% of background activity seems to be on the extremely low end of the spectrum of possibilities. Secondly, networks investigated by Smith et al., 2011 are sparse (a number of connections in a network of size $N$ is of order of $N$) and almost acyclic, which also seems to be a very optimistic scenario. Thirdly, Smith et al., 2011 used a TR of $3\text{s}$ and time series of 200 data points. This TR is too long and the time series length too short for generalizing the empirical datasets used today. Today, shorter TRs (1\text{s} or less, e.g., 0.72\text{s} as in the Human Connectome Project datasets, Essen et al., 2013) and longer time series have become the norm (e.g., 4, 800 samples in resting state datasets from Human Connectome Project, Essen et al., 2013). Lastly, Smith et al., 2011 used fixed delays of $50\text{ms}$ in the first layer of the DCM model that represents the underlying neuronal communication. This delay represents synaptic transmission delays and axonal transmission delays between nodes of the network. The constant value of delay is a crude estimation, especially given that, pairs of brain regions positioned at different distances from each other, should have different axonal transmission delays. Also, given polysynaptic connections, effective delays between neuronal populations might be much higher than the aforementioned $50\text{ms}$, e.g., P300 potential appears after $300\text{ms}$ (Polich, 2007) and some other cortical potentials have even slower latencies. This lack of attention towards modeling neuronal delays might favour non-lagged methods (i.e., PT or LiNGAM) in this analysis over the lagged methods. Altogether, there are reasons to believe that benchmark datasets derived by Smith et al., 2011 are, to some extent, not representative of the real fMRI datasets. For these reasons, results of the validation on the benchmark datasets should be interpreted with care. There are also other, competitive generative models in the field, e.g., the model proposed by Seth, Chorley, and Barnett, 2013. In this model, the authors used a simple VAR generative model to simulate neuronal dynamics in the testing network of 5 regions, based on work by Baccalá and Sameshima, 2001. Subsequently, the VAR model output is convolved with five different HRF kernels generated with the use of the difference-of-gamma approach as implemented in SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). Another possibility, would be to use Local Field Potentials (LFPs) instead of neuronal dynamics simulated as a system of differential equations with delay, and convolve LFPs with the hemodynamic response function (Deshpande, Sathian, and Hu, 2010). However, to date, the DCM implementation by Smith et al., 2011 remains the benchmark in the field.

Furthermore, in this work, I performed the inference on the full BOLD responses, without deconvolving the BOLD fMRI time series into the neuronal time series. It has been shown in synthetic and empirical data that incorporating a physiologically based model of spatiotemporal hemodynamic response function into the preprocessing pipeline leads to an improvement in the estimated neuronal activation (Aquino et al., 2014). It was also shown that it is generally difficult to accurately recover true task-evoked changes in the BOLD fMRI time series irrespectively of the method chosen for modeling the hemodynamic response function (Lindquist et al.,
Hence, there is a long-lasting debate in connectomics on whether or not a (blind) hemodynamic deconvolution is necessary to perform the (effective) connectivity research in fMRI (Wu et al., 2013). E.g., SEMs (McIntosh and Gonzalez-Lima, 1994) are often applied without deconvolution to fMRI datasets (Schlösser et al., 2003). My previous theoretical research in synthetic datasets generated from the DCM model suggests that deconvolution is not necessary in effective connectivity research in fMRI if the method used in the study is not lag-dependent (Chapter 6, Bielczyk et al., 2017b). This is because, under the assumption that the underlying signal on the neuronal level is in the low frequency range, the hemodynamic response does not affect the signatures of different connectivity patterns present in the outcome BOLD response (as it is a low-pass filter). In this work, I did not perform the deconvolution step before assessing effective connectivity with any of the tested methods, including GC. In fMRI literature, GC is applied both with (David et al., 2008; Ryali et al., 2011; Ryali et al., 2016; Hutcheson et al., 2015; Wheelock et al., 2014; Sathian, Deshpande, and Stilla, 2013; Goodyear et al., 2016) and without (Zhao et al., 2016; Regner et al., 2016; Chen et al., 2017) the deconvolution step. Recent research suggests that all connectivity methods (including functional connectivity) will improve their estimation accuracy post-deconvolution (Rangaprakash et al., 2018). However, in this work, I chose for the implementation without deconvolution to stay consistent with Smith et al., 2011.

In this place, I would like to mention that in the recent years, a lot of progress has been made in the area of modeling local hemodynamics from fMRI datasets. E.g., Havlicek et al., 2011 proposed a new approach to modeling hemodynamic response functions based on cubature Kalman filtering. Furthermore, Bush et al., 2015 proposed and validated a meta-algorithm for performing semi-blind deconvolution of the BOLD fMRI using bootstrapping. This method allows for estimating the timing of the underlying neural events stimulating the BOLD responses, together with confidence levels. Furthermore, Sreenivasan, Havlicek, and Deshpande, 2015 proposed a nonparametric blind BOLD deconvolution method based on homomorphic filtering.

Lastly, in these simulations I have set the connection strength to a fixed value of $A_{12} = 0.9$. On this stage, the output of the classifier is a binary response, i.e., an indication for a connection, either $X \rightarrow Y$ or $Y \rightarrow X$. This indication is based on a linear combination of the binary sign maps $S_r, S_i$ (Fig. 7.2C) with the values of the cumulants $Cr, Ci$ computed for the given dataset $X(t), Y(t)$ in either weighted or unweighted form. If the connection strength varies, the strength of the coupling between fractional moments in $X(t)$ and $Y(t)$ varies accordingly. Therefore, also the absolute values of the associated fractional cumulants will adjust. If cumulant values scale, then the RHS in the classification formula will scale accordingly, but the sign of this sum should stay the same. Therefore, in the noiseless case, this classifier should return the same output regardless of the connection strength $A_{12}$.

### 7.4.2 Limitations of the method

As mentioned in Chapter 5, it is also important to remember that there are always two independent aspects to a method for causal inference.
7.4. Increasing robustness of methods for effective connectivity in fMRI: Discussion

First, the method should have assumptions grounded in a biologically plausible framework relevant for the given research problem. E.g., a method for causal inference in fMRI should respect (1) the confounding, region- and subject-specific BOLD dynamics (Handwerker, Ollinger, and D’Esposito, 2004) and (2) co-occurrence of cause and effect (since the time resolution of the data is low compared with the underlying neuronal dynamics; the causes and their effects most likely happen within the same frame in the fMRI data). The new methods for pairwise inference, e.g., classification based on fractional cumulants, address this issue by (1) breaking the time order, and performing causal inference based on statistical properties of the distribution of the BOLD fMRI samples, and not from the timing of events; and (2) using correlation to detect connections. A good counterexample here is GC which has been proven useful in multiple disciplines of engineering, economy, and natural sciences. However, there is an ongoing discussion on whether or not GC is suited for the causal interpretations of fMRI data. Theoretical work by Seth, Chorley, and Barnett, 2013 and Roebroek, Formisano, and Goebel, 2005 suggests that despite the slow hemodynamics, GC can still be informative about the directionality of causal links in the brain. However, GC was not performing well in the study by Smith et al., 2011.

Second, the estimation procedure needs to be computationally stable. Even if the generative model faithfully describes the data, it still depends on the estimation algorithm whether the method will return correct results. Unfortunately, the face validity of the algorithms can only be tested in particular paradigms in which the ground truth is known. If in the given paradigm, the ground truth is unknown - which is most often the case in fMRI experiments - only the statistical reliability can be tested.

The new method for detecting the directionality of causal connections in fMRI that I introduced in this chapter, requires certain assumptions. E.g., I assume that on the neuronal level, the effects of directed connectivity are linear. This is also an assumption underlying the original DCM generative model. However, it is known that this is not always the case in the neuronal dynamics. E.g., shunting inhibition (Alger and Nicoll, 1979) is a phenomenon in which excitatory potentials are reduced by division rather than by subtraction. However, effects such as shunting inhibition typically happen in a microscale and should not affect large-scale neuronal dynamics as measured in the BOLD fMRI recordings. Therefore, I do not consider effects such as shunting inhibition as plausible confounds to my approach.

One crucial limitation to the cumulant approach as well as to the previous methods such as pairwise likelihood ratios is that, these techniques only retrieve the net connectivity. Namely, what these methods effectively pick up on is the difference between connectivity strengths: in a scenario where the connectivity strengths $X \rightarrow Y$ and $Y \rightarrow X$ are equal, the outcome cumulant maps for the system will have lower amplitudes than in the case of a unidirectional connection $X \rightarrow Y$ with the same connection strength. The significance of the cumulant values (whether or not the values are significantly different from zero), can be established with the use of permutation testing. However, since in the first step of the inference, I only select strong functional links for the further classification, ambiguous outputs from the classifier can be interpreted as bidirectional connections. Since the quality of the classifier depends on the ability to determine the
directionality of unidirectional connection, I used only unidirectional connections in the validation stage. One point to make here is that, bidirectional connections are hard to disambiguate for many methods for effective connectivity as they represent cyclic nets. This is an issue, e.g., in applications of Bayesian nets (Pearl, 2000) in which, joint probability for a certain graph is not tractable when the propagation of information in the network is cyclic. Also, the methods based on the third- and fourth- cumulant by Hyvärinen and Smith, 2013 fall into this category as the asymmetry between third cumulants in case of bidirectional connections will be equal to zero.

One interesting direction for the method development would be also the classification between excitatory and inhibitory connectivity. This aspect is missing in my study, as I focus on the connections of a positive sign only (deriving a classifier able to distinguish between excitatory and inhibitory connections would require deriving an additional set of benchmark $S_e$ and $S_i$ maps built based on repetitive simulations of an inhibitory connection, and creating a new set of benchmark synthetic datasets - as the original datasets by Smith et al., 2011 involve excitatory links only). The reason why inhibition is not implemented yet, is because the cumulant patterns for inhibitory connection are different from patterns given in Fig. 7.2 (I did not include the pictures in this chapter though). In its current form, the classifier makes a binary decision on whether the connection is in the direction of $X \rightarrow Y$ or $Y \rightarrow X$. The decision is based on whether the cumulant pattern obtained from the given dataset $X(t), Y(t)$ is more similar to the sign maps derived from the DCM generative model for connection $X \rightarrow Y$ (Fig. 7.2C), or to the inverse of these sign maps. To add inhibition to the picture, one would need to extend the number of possible classes by adding sign maps derived from DCM for inhibitory connection $X \rightarrow Y$ and introduce some metrics of distance to excitatory/inhibitory sign maps. This is the next step to take. One note on a side is that, whether or not inhibitory effective connectivity is expected in large scale networks investigated with fMRI, is a matter for a debate. On one hand, it is known that long-distance projections in the brain are mostly excitatory as inhibition is typically local, represented by groups of tightly connected interneurons within single brain regions (Markram et al., 2004, this is also modeled in the DCM generative model with the use of the self-inhibition term, Friston, Harrison, and Penny, 2003). On the other hand, several anatomical and physiological studies indicated that there are also long range inhibitory connections, e.g., between hippocampus and entorhinal cortex in rats (Melzer et al., 2012).

Furthermore, I considered local variability in brain neurophysiology by varying hemodynamic responses between iterations of the simulation and studying the effects of the scale-free background noise on the resulting effective connectivity measures. The DCM generative model summarizes the current state of knowledge about the mechanisms underlying generation of the BOLD fMRI responses (Friston et al., 2017; Havlicek et al., 2015; Havlicek et al., 2017). Therefore, currently there are no efficient ways of incorporating human brain neurophysiology into consideration in any more depth than this model allows for.

However, at the same time, I do not consider the influence of artifacts from the data acquisition, e.g., the effects of movement in the scanner. I believe that the influence of such confounders should be limited by choosing the proper data preprocessing pipeline. E.g., motion artefacts can be
7.4. Increasing robustness of methods for effective connectivity in fMRI: Discussion

reduced with the use of the new, data-driven protocols for motion artefacts removal such as ICA-AROMA (Pruim et al., 2015) or censoring-based artefact removal strategy based on volume censoring (Power et al., 2014). Therefore, developing efficient strategies for controlling such confounders is beyond the scope of this Chapter.

7.4.3 Future research

In the context of fMRI research, increasing the granularity of moments to fractional numbers to better characterize the distribution is an especially useful application because the experimental fMRI datasets are short (a few thousands samples at most). Therefore, the estimation error for the high order integer moments of the distribution becomes high. However, the subsequent cumulants contain information redundant to a certain extent, as they are correlated for any given time series. I chose the granularity that gives relatively smooth patterns of discriminability (Fig. 7.3), which is $\Delta k = 0.1$. Choosing the optimal moment resolution is a subject to future research. However, I believe that increasing index resolution to substantially less than 0.1 would not be beneficial as nearby fractional moments positively correlate (which induces redundancy), yet it would substantially increase the computational cost for the method.

In this work, I validated this approach using synthetic benchmark datasets derived from the DCM generative model. Using generative models is valuable in neuroimaging (e.g., in applied computational psychiatry, Frässle et al., 2018, and in network neuroscience in general, Betzel and Bassett, 2017), in terms of validating new methods as these models enable inference on model parameters which allows for some degree of mechanistic interpretability on the putative processes underlying the studied phenomena. Generative models, especially the DCM, are acknowledged in multiple contexts in the field of cognitive neuroimaging, from method validation to application in clinical datasets.

Here, I faithfully reproduced the research pipeline after Smith et al., 2011. However, since the original version of the DCM (Friston, Harrison, and Penny, 2003) based on the original Balloon-Windkessel model (Buxton, Wong, and Frank, 1998) was published, substantial advancements to the hemodynamic model have been proposed. Firstly, Obata et al., 2004 reported an error in the expression for the outcome BOLD response, and Stephan et al., 2007 proposed a more accurate expression for one of the terms in this formula. Secondly, Heinzle et al., 2016 proposed an updated formula for the hemodynamic response function at the field strengths higher than 1.5[T]. In the future work, I will take these developments into account.

Any method for causal inference should also give predictions testable the clinics (e.g., lead to more reliable estimation of directed causal influences between brain areas induced by cognitive tasks, better stratification of clinical populations in the resting state, etc.), which I will also further investigate in the future studies. Therefore, the approach needs to further be tested and compared against other methods with the use of experimental fMRI datasets, such as, e.g., the Human Connectome Project datasets (Essen et al., 2013; Barch, 2013). Since little is known about causal connections in large scale brain networks, especially in the resting state, such a validation might be quantitative (i.e., by means of statistical reliability)
rather than qualitative. Alternatively, one could perform such a validation in particular pathways in which one can make assumptions about the directionality of information flow, such as the dorsal and the ventral stream in the visual cortex during the visual processing. One possibility is also testing using datasets coming from an interventional study in which neural activity was evoked and, therefore, the ground truth is known: a fused fMRI-optogenetic experiment (Ryali et al., 2016) in which intervention with the use of optogenetics guarantees that there is a causal link between the investigated neuronal populations (which can further be detected with the use of the validated method).

Lastly, one note on bidirectional connections: as introduced above, Patel’s tau approach (Patel, Bowman, and Rilling, 2006) involves a two step procedure. In the first step of this analysis, functional connectivity by means of partial correlation is used to find the position of connections in the connectome (regardless of their directionality). The idea is, once a connection is found by means of functional connectivity, it can either be a uni- or a bidirectional connection. In the second step of the analysis, the directionality of the connection is further determined through the classification of the two nodes into the "upstream" and the "downstream" node. Then, under the assumption that the tested method can correctly estimate the directionality of a connection, if the method does not give a univocal answer, one can assume that the connection is bidirectional (as in the first step of the analysis, the connection was already detected). However, in case there are reciprocal connections between the two nodes, but one of them is stronger than the other, Patel’s tau will only detect the "net" effect, namely, it will indicate the stronger between the two connections. In that sense, reciprocal connection will be reduced to a univocal connection. The same effect holds for other pairwise inference methods, namely the PW-LR approach (Hyvärinen and Smith, 2013). In fact, reciprocal connections are ubiquitous in neuronal networks (Koetter and Stephan, 2003). Therefore, proper representation of these bidirectional connections remains an important challenge in network neuroscience.

7.4.4 Conclusions

The field of effective connectivity research in fMRI is still growing. E.g., in the recent few years, multiple algorithms for the graph network search have been developed and applied to fMRI datasets, including Independent Multiple-sample Greedy Equivalence Search (IMaGES, Ramsey et al., 2010), Group Iterative Multiple Model Estimation (GIMME, Gates and Molenaar, 2012), or Fast Greedy Equivalence Search (FGES, Ramsey et al., 2017). Whether or not bivariate or multivariate inference serves better purpose for effective connectivity research in fMRI, is an open research question. In this work, I focused on the pairwise inference, and I achieved a significant improvement compared to the previous approaches for pairwise estimation of functional causal interactions. Most importantly, the robustness against known sources of variability (i.e., differences between up- and downstream noise magnitude and presence of non-Gaussian scale-free background noise in the networks) significantly increases due to the simultaneous incorporation of multiple aspects of the associated BOLD fMRI distributions. I believe
that, as this approach based on fractional moments of the distribution increases resilience of the methods for pairwise connectivity to potential confounds in the experimental data, it can become a generic method to increase the power in causal discovery studies, both in cognitive neuroimaging and beyond.
Chapter 8

Mapping large scale networks to diagnostic categories as an application of signal detection and causal inference methods for fMRI

Creating large-scale (causal) connectomes of the brain is only the beginning of the journey. The next question is: what are the emergent properties of the revealed networks that can give a better insight into behavior and in particular, into cognitive disabilities? In this work, I propose a new approach, Circuit to Construct Mapping (CCM) that aims to characterize causal relations between the underlying network dynamics (as the cause) and the constructs referring to the clinical symptoms of cognitive disorders (e.g., MDD, as the effect). CCM involves extracting diagnostic categories from behavioral data, linking circuits that are causal to these categories with the use of clinical neuroimaging data, and modeling the dynamics of the emerging circuits with attractor dynamics to provide new, neuroimaging-based biomarkers for cognitive disorders. The CCM approach optimizes the clinical diagnosis and subject stratification. It also addresses the recent demand for linking circuits to behavior and provides new insights into clinical treatment by investigating the dynamics of neuronal circuits underneath cognitive dimensions of psychiatric disorders. CCM can serve as a new regime toward personalized medicine by assisting the diagnosis and treatment.

I explain the concept on the example of the most prevalent cognitive disorder, Major Depressive Disorder (MDD). MDD is a heterogeneous disorder that involves multiple behavioral symptoms on one hand and multiple neuronal circuits on the other hand. In this work, I integrate the literature on cognitive and physiological biomarkers of MDD with the insights derived from the mathematical models of brain networks, especially models that can be used in the fMRI research.

Keywords: modeling, circuit, diagnosis, research domain criteria project, dynamical systems, major depressive disorder

This Chapter was published as:

8.1 Introduction

8.1.1 Major Depressive Disorder

Major Depressive Disorder (MDD), also known as unipolar depression, has a lifetime prevalence that exceeds 16% in the United States (Nestler et al., 2002) and is expected to increase its share in the global disease burden from 4.3% in 2004 to 6.2% by 2030 (WHO, 2012). Treating MDD is costly. In 2010, the total cost of MDD in the European Union was estimated to be 798 billion Eur, of which 60% was the direct cost due to treatment and the remaining 40% was the indirect cost due to the loss of productivity (Olesen et al., 2012). Currently, there is a rich variety of competing biomarker sets, each suggesting different MDD etiology. However, it is unclear how these relate to the current diagnostic criteria. This heterogeneity of biomarkers, behavioral symptoms, and circuit changes in MDD requires the use of multimodal and multidisciplinary approaches together with mathematical modeling to integrate these findings into diagnostic and intervention tools useful in clinical practice.

So far, the search for candidate genes underlying MDD has not yielded a single responsible gene. Instead, genetic models of MDD propose that a large number of genes is involved (Hong and Tsai, 2003), with a small contribution of each of them to the MDD phenotype. Furthermore, these models suggest that epigenetic regulation may underlie gene-environment effects in MDD (Vialou et al., 2013). Epidemiological studies have revealed that genetic factors may account for 40-50% of the risk of developing the disorder (Kessler et al., 2003). Since the definition of an endophenotype involves heritability (Gottesman and Gould, 2003) and can only be used in a family sensitive design (Dubin et al., 2012), it leads to a conclusion that only particular diagnostic categories in MDD can be interpreted as endophenotypes. Therefore, instead of talking about endophenotypes in MDD, I refer to NIH Research Domain Criteria (RDoC) project approach (First, 2013) and to its central concept of a construct as a basic dimension of brain functioning (without the requirement of heritability). While defining constructs, RDoC initiative refers to various units of analysis, from genes to neural circuits and behavior.

In Section 8.2, I review the current state of knowledge on MDD etiology across multiple construct domains, from behavioral through physiological down to neuronal level. Furthermore, I propose a new paradigm to aid the diagnosis of MDD and its clinical management that includes dynamical models of the underlying circuitry and mapping the activity of these circuits onto cognitive constructs diagnostic for MDD. This Circuit to Construct Mapping (CCM) approach can facilitate a personalized approach to MDD and thereby improve the quality of life for MDD subjects.

8.1.2 Causality

Mapping the activity of the underlying circuits onto cognitive constructs diagnostic for MDD involves an assumption that we can point to causal relations between these two domains. In this review, I focus on the altered dynamics of neuronal circuits as the cause of the disrupted behavior. But how can one determine causality? There are two definitions of causality,
both of them often used in research studies in psychiatry. First definition by Lewis, 1973 describes causality in the language of counterfactuals: we may define a cause to be an object followed by another, where, if the first object had not been, the second object never had existed. Based on this definition, in 1986, Holland formulated the no causation without manipulation rule (Holland, 1986) which became the prevailing principle in causal research for another two decades. Today, Woodward’s view at causality through structural equations also becomes popular (Woodward, 2003): assuming that we have an endogenous variable $Y$ produced from variables $X_1, X_2, ..., X_n$, Woodward’s approach involves expressing certain basic counterfactuals in the following form: if it were the case that $X_1 = x_1, X_2 = x_2, ..., X_n = x_n$, then it would be the case that $Y = f(x_1, ..., x_n)$.

However, this is not the only view on causality. Judea Pearl builds on the counterfactual approach and writes in his recent essays (Bollen and Pearl, 2012): “the essential ingredient of causation is responsiveness, namely, the capacity of some variables to respond to variations in other variables, regardless of how those variations came about.” This is an objection to the idea that the establishment of causation necessarily requires manipulation; rather, it is sufficient to observe the system and its natural course. However, the inference of causality based on observational data is a complex research problem, and Pearl developed a comprehensive theory of how to establish causation by means of probabilistic models.

This latter view of causality is beneficial to the causal research in psychiatry as we are not always equipped with the tools to manipulate all the candidate causes in brain networks of interest. E.g., if we are interested in the causal effect of the insular cortex on the emotional states in subjects with MDD and we aim to apply the counterfactual approach to test this hypothesis, we should shut down the activity of the isolated insula and register the observed change in the regulation of emotional states in the cohort. However, since the insula does not occupy the surface of the cortex, it is very hard to non-invasively perturb its activity alone; hence, so far, remote control of deep brain activity is not available in humans. Therefore, in clinical trials, the second definition of causality is typically applied: one compares a population of subjects with high and low activation in the insular cortex, and tries to find systematic differences between these two groups in terms of emotional states. If the effect size is large enough, the causal effect is determined. In the further sections, I will discuss causality in Pearl’s sense, meaning “observation” and “statistical power” rather than “intervention” and “counterfactuals.”

8.2 Etiology of MDD

8.2.1 Constructs in MDD

In case of MDD and other cognitive disorders, causality is a complex research problem because the disorder can be described across various domains, from neurophysiology, through neuronal networks, to behavior. Although causal explanation in MDD can search for relationships between any pair of constructs, from the psychiatric point of view links in which some behavioral construct is the effect, are especially valuable.
Chapter 8. Mapping large scale networks to diagnostic categories

Figure 8.1: Diagnosis, treatment, and brain dynamics in MDD. One can distinguish five classes of drugs (A) based on monoamine receptors that they target (B). fMRI studies reveal that these drugs affect different but overlapping circuits. In C, two exemplary circuits are presented based on imaging studies on two constructs present in MDD: low mood (or, negative affect) and anhedonia (lack of positive affect). The connectivity in the circuits is presented with arrows, solid lines for glutamatergic, and dashed lines for GABAergic projections. Findings in regions activated in negative affect in depressed subjects are summarized in the "negative affect" circuit, left half-circles. The subjects were triggered to fall into low mood by presenting them with scenes of negative emotional valence. Findings on regions up- and down-regulated in anhedonic MDD subjects during presentation of scenes with positive emotional valence are summarized in the "anhedonia" circuit, right half-circles.

Deep red color - overactivation in a given region during the task in respect to healthy controls, light red - hypoactivation, stripes - counteracting evidence in the literature, background color - no data available. Additionally, treatment effects of fluoxetine in the nodes of this network are indicated with black arrows (up: up-regulation in respect to non-medicated subjects; down: down-regulation). Influence of the drug was assessed based on imaging studies that were using experimental tasks focusing on emotion processing (as aforementioned tasks involving presentation with scenes of emotional valence). On the left hand side of each region, influence of fluoxetine treatment in negative affect regime in MDD is indicated, on the right hand side: the same for positive affect regime. On one hand, this figure demonstrates that drugs act on constructs rather than particular brain regions. On the other hand, it shows that circuits underlying constructs are strongly overlapping but not identical. For any given construct, the underlying circuitry, modulated by interplay between the neural substrates within, flows toward a stable activity pattern, the attractor state (D). Stability means that the network will relax to the same stable pattern even after small degree of external stimulation. This phenomenon is pictured with the ball metaphor. Treatment with drugs is most likely to change the subject’s state by reshaping the attractor landscape. This drift results in the change in the particular construct whose circuit is targeted with treatment (E) but may also affect other constructs via circuits overlapping with the targeted one. Then, the behavioral consequences of the treatment are the basis to prescribe a more appropriate drug for the given individual (A).

VMPFC - ventromedial prefrontal cortex, including BA25 (subgenual cortex), BA9 - dorsolateral prefrontal cortex, BA32 - dorsal anterior cingulate cortex, BA33 - part of anterior cingulate cortex, INS - insula, NAC - nucleus accumbens, CN - caudate nucleus, PT - putamen, THA - thalamus, AMY - amygdala.
Fig. 8.1 presents the variety of constructs across multiple levels of description in a typical process of treatment in MDD, with arrows denoting causal relations between them. Firstly, one can distinguish five classes of drugs (Fig. 8.1A) based on monoamine receptors that they target (Fig. 8.1B). A subject diagnosed with MDD is typically prescribed with one or, rarely, with a combination of these drug types. fMRI studies reveal that these drugs affect different but overlapping circuits (Fig. 8.1C). For any given construct, the underlying neuronal circuitry, modulated by interplay between the neural substrates within, reaches a stable activity pattern (a so-called attractor state) - which is pictured with the ball metaphor (Fig. 8.1D). The network specific activation pattern modulates the particular cognitive construct (Fig. 8.1E). The behavior of the subject is subject to repetitive diagnoses which, possibly, can lead to prescription of new, more accurate drugs - which closes the circle. In my understanding, the mechanism underlying MDD is a superposition of multiple circuits, each of them having a causal effect on one of the cognitive constructs present in MDD. Therefore, in these considerations on modeling MDD, I am interested in the causal effect between neuronal circuits (as the cause, C) and behavioral constructs (as the effect, D).

I briefly review the aforementioned levels of description in the following sections. Although the proposed CCM approach only includes mapping from neuronal circuitries straight to the cognitive domain, the physiology underlying MDD is also worth mentioning as the most popular (but not necessarily the most effective) treatments derive from the monoamine theory of MDD and target neuromodulatory receptors in the brain rather than particular circuits.

8.2.2 Cognitive constructs

MDD was originally defined in terms of behavior, therefore, cognitive constructs present in MDD seem to be the right starting point to give full characteristics of this disorder. In DSM-5, diagnostic criteria for MDD are as follows: the subject is diagnosed with MDD if at least five out of nine diagnostic traits are present (Fig. 8.1E), at least one of them being anhedonia or low mood.

Current diagnostic practice for MDD is difficult. First, both DSM-5 and ICD-10 diagnostic criteria allow for creating a broad range of behavioral profiles, all diagnosed with the same clinical condition (Kupfer and Regier, 2011; Organization, 1992). Second, the diagnostic criteria are open to different interpretations, change over time, and are therefore less objective and require review by trained clinicians. E.g., independent symptoms of dysthymia (present in DSM-4 as a self-standing disorder) were recently classified as chronic MDD in DSM-5, because since DSM-4 was released, there was not enough evidence that dysthymia is significantly different from MDD (Cristancho, Kocsis, and Thase, 2012). Third, sometimes new MDD types are distinguished based on specific events triggering the disorder, e.g., grief in the DSM-5 (and in the incoming ICD-11, Kupfer and Regier, 2011; Maj, 2012; Mateen et al., 2012) and premenstrual dysphoric disorder in DSM-5 (Epperson et al., 2012). This change of diagnostic criteria over time leads to differences in the interpretation and is a strong argument for developing an objective approach.
8.2.3 Physiological constructs

As mentioned before, there is a variety of competing biomarker sets, each suggesting different MDD etiology. The catecholamine hypothesis by Schildkraut, 1965 originated in the 60s, advocated that norepinephrine (NE) plays a pivotal role in affective disorders, with a lesser role for epinephrine (E), dopamine (DA), and serotonin (5HT) levels. The hypothesis suggested a reduced level of neurotransmission in E, NE, DA, and 5HT pathways as a possible cause of MDD. Today, it is known that not only DA, NE, and 5HT, but also acetylcholine (AC) has a strong impact on mood (Warner-Schmidt et al., 2012). Nevertheless, the mechanism of the shift from a healthy brain state into MDD and the role of each of these neuromodulators in this process are not yet understood.

Monoamines and AC are not the only neuromodulatory chemicals involved in MDD. Neuroendocrine mechanisms such as the corticotropin-releasing factor (CRF) may also play a role (Pariante and Lightman, 2008). In MDD, this peptide is overproduced in the hypothalamus, which, acting along with arginine vasopressin (AVP), triggers hypersecretion of adrenocorticotropic hormone (ACTH) from the pituitary. ACTH overproduction leads, in turn, to overproduction of glucocorticoids (cortisol in humans, corticosterone in rodents) from the adrenal cortex. This circuit is known as the hypothalamic - pituitary - adrenal (HPA) axis, and as a part of the neuroendocrine system - it controls stress reactions, metabolism, and immunity (Sánchez, Ladd, and Plotsky, 2001). HPA theory of depression corresponds to the evidence that, due to epigenetic mechanisms, early life events can cause HPA overactivation in adult life (Heim and Nemeroff, 2002).

Furthermore, recent observations demonstrate that antidepressant drugs targeting monoamines also modulate synaptic GABA transmission. Additionally, post-mortem studies reveal a dramatic reduction in plasmic GABA concentration in MDD subjects. These findings have implicated GABAergic mechanisms in MDD (Luscher, Shen, and Sahir, 2011), and led to the postulate that the balance of excitation and inhibition (E-I) in brain networks in MDD is disturbed (Wieronska et al., 2012).

Another theory of MDD results from the observation that antidepressants induce plasticity in the synaptic strengths, altering patterns of connectivity in the brain (Castren, 2005). Consequently, it was proposed that MDD may reflect a primary impairment in neuronal information processing caused by a disrupted functional or effective (directed) connectivity rather than by any form of chemical imbalance.

8.2.4 Neuronal constructs

Identification of neuronal circuits underlying MDD with the use of fMRI initially has led to the default mode network (DMN) theory of MDD (Marchetti et al., 2012; Sheline et al., 2009). DMN is a circuit defined by slow, coherent oscillatory activity in a wakeful resting state in humans with eyes closed (Raichle et al., 2001). It mostly involves structures engaged in self-referential processes (parts of the medial prefrontal, posterior cingulate and parietal cortices, and medial temporal lobe), as well as the centers for memory in the brain (hippocampus, parahippocampal gyrus) and limbic structures (amygdala, nucleus accumbens, hypothalamus, Greicius et al., 2003).
Imaging studies reveal that resting-state activity in many of the DMN nodes is altered in MDD (Hamilton, Chen, and Gotlib, 2013). It was found that activity in DMN correlates with mood (Wiebking et al., 2011), therefore, this circuit might be responsible for the affective aspect of the disorder. DMN is just one of many resting-state networks (RSNs) identified so far (Damoiseaux et al., 2006), and methods proposed for identification of MDD based on the resting state fMRI respect not only DMN but also other RSNs. E.g., a recently developed computational diagnostic method utilizing Hurst exponent takes into account the DMN, the right and left fronto-parietal, ventromedial prefrontal, and salience networks (Wei et al., 2013).

Recent evidence suggests that not only RSNs, but also the central executive network (CEN) seems to be impaired in MDD (Anderson and Green, 2001). This network involves a few subdivisions of the prefrontal cortex (PFC), anterior thalamus, and dorsal caudate nucleus. As opposed to RSNs, CEN comes to play during processing that requires cognitive control (Seeley et al., 2007), and is responsible for the executive functions, e.g., response inhibition, reward processing, planning, and working memory. Therefore, as opposed to RSNs, CEN might be involved in such constructs as recurrent thoughts of death and diminished attention. These two families of networks are complimentary and tend to switch activity between each other.

Identification of common patterns of up- and down-regulation in the nodes of RSNs and CEN could serve as new, more robust means to identify network-related biomarkers of MDD (Schlösser et al., 2008). In particular, a construct-based approach would allow for creating individual dynamical profiles for subjects and, therefore, personalized therapy.

8.3. Treatment

Coming back to causality, it is believed that treatments in MDD affect neuronal dynamics and this dynamics in turn triggers the behavioral change. The treatment choice depends on multiple factors, including the course of the disease, prior medical treatment, etc. (Mayberg, 2009). Evidence-based treatment guidelines suggest cognitive based therapy (CBT, Ebmeier, Donaghey, and Steele, 2006) and pharmacology (Beck, 2006) as the first line of choice (Hyler, 2002).

On the other hand, electroconvulsive therapy is only recommended if the aforementioned methods are ineffective for the given subject, whereas deep brain stimulation, as the most invasive method, is not yet approved by the United States Food and Drug Administration (FDA) for treatment-resistant depression. Even though new treatment methods such as repetitive transcranial magnetic resonance (rTMS, a localized, superficial stimulation of the cortex with magnets, Kim, Pesiridou, and OReardon, 2009) and neurofeedback therapy (a combination of cognitive therapy with neurological approach: a real-time feedback of local fMRI signals, Linden et al., 2012) are being tested, they are not established yet.

An example of applying drugs as a treatment procedure affecting the construct-related circuits (that changes the dynamical state of the brain and thus influences the diagnosis) is presented in Fig. 8.1A.
8.4 Circuit for MDD

As mentioned in Section 8.2.1, my viewpoint is that the brain network mechanism underlying MDD is a superposition of multiple circuits, each of them having a causal effect on at least one of the cognitive constructs present in MDD. In fact, the number of these cognitive constructs, and therefore also the underlying circuits, may be much higher than the number of diagnostic categories specified in the DSM-5. Exemplary constructs not mentioned in the DSM-5 but present in a vast majority of MDD subjects include negative bias in attention and memory (Mathews and MacLeod, 2005), a negative view of the world and the future (Beck, 2006), learned helplessness (Li et al., 2011), obsessions, and pathological rumination (Gotlib and Joormann, 2010).

However, to perform causal inference linking circuits to cognitive constructs, one needs to determine which circuits to study in the first place. MDD is a heterogenous disorder, and, as such, it arises from anatomical and functional changes across a wide range of brain regions. The circuits that were first proposed to be responsible for MDD consisted of regions known to be involved in mood. One of these mood generators is the corticomesolimbic loop: one of a few parallel, basal ganglia-thalamo-cortical loops that projects from the ventromedial PFC to the medial dorsal thalamus through the nuclei of the basal ganglia (Alexander and Crutcher, 1990). Yet another mood generator is the aforementioned hypothalamic - pituitary - adrenal axis (HPA) whose dysfunction widely affects monoamine pathways and triggers mood fluctuations. Recently, the viewpoint at MDD and other mental disabilities through the prism of large scale brain networks identified based on fMRI studies (RSNs and subcircuits of the CEN), and interactions between them, has gained in popularity (Menon, 2011; Bressler and Menon, 2010; Li et al., 2013; Speechley, Woodward, and Ngan, 2013; Barch, 2013; Palaniyappan et al., 2013; Manoliu et al., 2014).

I take this large-scale perspective. However, as mentioned above, in my view the search for mechanisms underlying MDD should include zooming into circuits underlying single diagnostic constructs. Large-scale networks are complex and, as such, they might be decomposed into simpler functional circuits. This is definitely the case for the CEN. On one hand, various cognitive constructs could be characterized as different states within the same network. On the other hand, CEN is most probably divided into functional subcircuits which activate while solving particular tasks involving cognitive control, e.g., reward receipt, signal inhibition, decision making or language processing. Another example is the DMN which generates mood. DMN might be composed of a few interacting subcircuits accounting for generation of basic emotions (Vytal and Hamann, 2010; Panksepp, 2010) which do not coexist (Ekman, 1999; Damasio et al., 2000). However, it could also be the case that basic emotions represent various attractors of one large circuit, which is why it is so hard to find specific neuronal underpinnings of basic emotions (Barrett and Wager, 2006; Cacioppo et al., 2000).

In terms of models, so far, RSNs are better characterized than CEN (Honey et al., 2007; Zhou et al., 2006), probably because of stable temporal dynamics that can be easily investigated with fMRI. Interestingly, Deco et al.,
8.5. Modeling MDD

8.5.1 Neural mass models and attractor landscapes

So far, psychiatric disorders have not been properly conceptualized in the language of computational neuroscience (Montague et al., 2012; Wang and Krystal, 2014; Deco and Kringelbach, 2014). Early research in this field was centered on reinforcement learning models that describe behavior as taking actions that maximize predicted rewards (Worgotter and Porr, 2005). Since DA is believed to be involved in prediction (Morris et al., 2006; Schultz,
2002), mostly the disorders linked to DA (e.g., schizophrenia) were modeled with the use of the reinforcement learning (Deserno et al., 2013).

However, since both calculating the odds for possible rewards and taking decisions based on that calculation do not directly correspond to the neuronal activity and physiology of the brain, models based on reinforcement learning are a poor choice when it comes to neuroimaging-based biomarkers of mental disorders. In the last decade, comparing structural and functional connectivity in brain networks in health and disease, in terms of graph theoretic measures, e.g., small-worldness (Watts and Strogatz, 1998) or modularity, (Newman, 2006) has become a popular research direction (Bassett and Bullmore, 2009). These measures have led to multiple new results upon the global properties of brain networks in cognitive disorders (Liu et al., 2008; Zhang et al., 2011b; Wang et al., 2009) including MDD (Borsboom et al., 2011; Peng et al., 2014). However, these measures only take functional connectivity between brain regions into account. The assumption of undirected connectivity yields a conclusion that for every pair of brain regions $A$ and $B$, once the treatment procedure targets region $A$, it has the same impact on region $B$, as if one would target region $B$ with the same treatment and measure the change in activity in region $A$ - which is, in general, an unrealistic assumption. Therefore, graph theoretic measures do not extensively incorporate the information that can be rendered from the neuroimaging datasets and that is of primary importance for assisting diagnosis and treatment in cognitive disorders.

Recently, the concept of attractor networks was proposed as a tool that might explain cognitive disabilities while corresponding to the neural dynamics in the brain. An attractor network is a network of nodes, often recurrently connected, whose dynamics settle to a pattern stable in time: the so-called attractor state. Analysis of the distribution of attractor states and their basins of attraction, a so-called attractor landscape, was effected on a microscale so far. At the microscale, single neurons are the nodes in the network, and stable firing patterns of those neurons constitute an attractor state (Balaguer-Ballester et al., 2011). This approach is present in contemporary computational neuroscience, e.g., in the models of activity in olfactory (Niessing and Friedrich, 2010) and auditory (Bathellier, Ushakova, and Rumpel, 2012) cortices in rodents as well as hippocampal grid cells in humans (Samsonovich and McNaughton, 1997). This concept has also been broadly used in psychiatry. E.g., the PFC has been modeled as attractor network to explain the deficit in short term memory in schizophrenia (Rolls et al., 2008) and compulsions in obsessive-compulsive disorder (Rolls, Loh, and Deco, 2008). Up until now, it is unclear how these models translate to subjects because neither the invasive measurements of a single neuron activity necessary to validate the attractor network models are possible, nor do non-invasive methods have the appropriate resolution.

How about the macroscale? It is now believed that the fMRI research can provide the insight necessary to understand cognitive constructs (Pol-drack, 2006; Castellanos et al., 2013). But is the concept of attractors also applicable for this type of datasets? Here, I propose a conceptual advance to apply mathematical modeling directly to subjects’ data. This proposal involves looking at the large-scale neural circuits to perform attractor landscape analysis on the macroscale.
Figure 8.2: Wilson-Cowan model and a "ball" metaphor. The activity of a single brain area within the network $E$ is a result of the synaptic inputs from other areas, the modulatory tone generated by diffuse projections, and the recurrent connectivity within the brain areas. The activity reflects a specific balance between excitation and inhibition within the area. For simplicity, we describe the activity by one variable, $E$, for which the following equation holds:

$$\tau \frac{dE}{dt} = -E + f(\alpha E + \beta I + \gamma M).$$

The first term on the right tells us that in the absence of any drive (provided by the second term), the activity decays to zero with time scale $\tau$. The second term incorporates the contribution of recurrent connectivity via $E$ itself, input from other areas, represented by $I$, and the level of neuromodulation, represented by $M$. Each of these contributions are weighted by factors: $\alpha$, $\beta$, and $\gamma$ respectively. When the second term is positive, it increases the level of activity. The function $f$ is a response function that translates the sum of activities into a driving term, and is typically sigmoidal (Silver, 2010):

$$f(x) = A \frac{x^2}{\sigma + x^2}.$$ 

In this form, $A$ is the maximum that $f$ can reach for large $x$ values, and $\sigma$ is the value for which $f$ is equal to half its maximum value. In addition, it also specifies how steeply $f$ increases with $x$, a quantity that is also referred to as the gain factor. 

A: In a given region, the sigmoidal input-output (I-O) relationship has three regimes. For small input $y << \sigma$, it increases rapidly. For large inputs, $y >> \sigma$, it saturates. For values in between, it connects these regimes linearly. If the $\sigma$ value, and thus excitability of the region, grows (dashed line), the I-O function is steeper than in the control case (solid line). If the region gets stronger recurrent connectivity, input from other regions or neuromodulation, so that the $\alpha$, $\beta$, and $\gamma$ values grow respectively, I-O function shifts to the left (dotted line). 

B: In an example of two interconnected regions, $E_1$ and $E_2$, this dynamical system has three fixed points that are candidates for attractor states. In this example, two of them are stable (red). C: One may picture attractor states with the ball metaphor. Disease can be represented in two ways. It can mean a change in the landscape of basins of attraction: some attractor states change position and even if the subject occupies the original attractor throughout the process, their brain state gradually changes the attractor state that they occupy. However, it can also mean that, as a result of intrinsic noise in the brain or in response to a particular external input, the brain state in the subject is triggered to switch to another, "wrong" basin of attraction.

Mind that brain circuits are networks of interacting nodes, and therefore can be represented and analyzed as dynamical systems in a similar fashion as networks of single neurons. As opposed to microscale, at the macroscale whole brain areas account for the nodes in the network, and attractor states are stable activity patterns across all nodes within the network. E.g., in case of fMRI datasets, the overall activity in a region of interest can be expressed as the summation over the activity of all the voxels within that region. This
data is very convenient for neural mass models when it comes to modeling cognitive architectures (Deco, Jirsa, and McIntosh, 2013) because the principal idea behind neural mass models is setting the density of neurons to the continuum limit in modeling activity of large neural populations.

This assumption of spatially continuous neural networks thus allows for analytical treatment of such global variables as firing rate in space and time. One example is the classic mean-field model by Wilson and Cowan, 1972. In this model, the activity of neuronal populations (or, brain regions) is represented by dynamical variables. Fig. 8.2 presents a simplified version of the model where spatial patterns of spiking activity are replaced by one dynamical variable. In the model, effectively connected neuronal populations, representing brain regions, interact and are additionally tuned by neuromodulators. Such dynamical systems have a number of stable attractors and, therefore, a number of basins of attraction. The possibility is that in MDD subjects, the shape of the attractor landscape for a particular cognitive construct is different than in healthy controls. However, it can also be that they occupy a "wrong" attractor state (Silver, 2010).

8.5.2 Treatments in the context of dynamical systems

All the available treatments affect the dynamics of large-scale networks and therefore also the associated attractor landscapes (Posner et al., 2013; Abbott et al., 2013; Messina et al., 2013). Therefore, with the use of the Wilson-Cowan model, one can investigate the landscape of basins of attraction in response to the treatment procedures. Antidepressant drugs can reshape the attractor landscape in multiple ways: they can lower the hills of the landscape around the current state of the subject or make the current attractor state shallower to facilitate escaping from the local minimum (Fig. 8.2C, upper panel). The drugs can potentially also modify the level of background neuronal noise, which in turn may affect the probability of occupying different attractor states (Jedynak, Pons, and Garcia-Ojalvo, 2015). On the other hand, stimulation methods that regulate the neural dynamics directly, such as rTMS, ECT, and DBS, can influence the state of the subject by providing a brief pulse to the brain network and thus allowing the brain network to leave the "wrong" attractor state immediately (Fig. 8.2C, lower panel). Interestingly, in the treatment-resistant depression, electrical stimulation through ECT and DBS proves to be highly effective (Mayberg et al., 2005; Tokutsu et al., 2013), which means that, under some circumstances, they perform better than drugs, or even the cognitive therapy which targets the cognitive constructs directly. This provides some hint suggesting that looking at the clinical symptoms of MDD through the prism of neuronal circuits, and targeting treatments at those circuits might be more beneficial than any other treatment - including, paradoxically, even the behavioral treatment centered at specific cognitive traits in MDD.
8.6 Circuit to Construct Mapping

8.6.1 What is CCM

Figure 8.3: Circuit to Construct Mapping (CCM). Causal links between the nodes of the underlying network (nodes $E_1$, $E_2$, $E_3$) and the multidimensional construct space ($t_1$, $t_2$) can go two ways, but we are only interested in neural circuitry as the cause, and cognitive constructs as the effect (A, green lines). Behavioral learning and neuroplasticity can give the backward direction of causality (A, red lines); however, we do not cover this issue in this chapter. I refer to cognitive constructs as $t_i$ because the preliminary step of the CCM includes determining the full list of the involved constructs which can be broader than the list of the DSM-5 diagnostic criteria. Prime - endophenotype, dot - a subject, asterisk - attractor state. Firstly, circuitries involved in these constructs should be linked based on extensive observational research on a large cohort of subjects (A). Secondly, the mapping of these attractors onto cognitive constructs can be done with the use of temperature maps (B). In this example, I plot the value of one continuous construct (which can represent, i.e., mood level) in a three-dimensional space spanned by the attractors of the underlying three-node network. Thirdly, one can track the current state of a subject both in the multidimensional construct space during the treatment (C, a scatter plot in case we want to track multiple subjects at a time). The distribution of subjects in this space may reveal subtypes of MDD ($MDD_1$, $MDD_2$). Moreover, when the network manipulation with treatment is sufficient, it can trigger the subject’s brain state towards a new attractor in some of the construct-related dimensions, and as a result, the subject’s state will flow to another point in the construct space. Since we will create these maps based on the neuroimaging and behavioral datasets coming from a limited number of subjects in the cohort, these temperature maps will not span the whole space of possibilities. Lastly, one may investigate how the treatments $TR_i$ affect the attractors of the underlying networks (D). In this case, we have three nodes in the underlying networks which means that the attractors of this system will become points in a three-dimensional attractor space.
Every subject has a different, individual attractor landscape. This landscape reflects such personal traits as the size of the brain regions involved in MDD, the functional connectivity within DMN and CEN, the baseline concentrations of monoamines, and all the other endogenous chemicals that influence the excitation-inhibition balance in the brain. During rest, DMN and other RSNs are active and the subject occupies stable attractors in their attractor landscapes. On the contrary, during solving cognitive tasks, sub-networks of CEN come to play (depending on the nature of the task) and the brain state jumps to one of its (most probably, also stable) attractors. I can predict that a disturbance of the attractor landscapes within the DMN should account for the cognitive constructs involving affective components of MDD, whereas disturbance of the attractor landscapes within cognition related RSNs (such as fronto-parietal network) and within the CEN should be responsible for the cognitive constructs involving executive functions.

But how do these attractors map onto cognition? Let us consider a brain network consisting of interconnected nodes described by their activities, either in the resting state or in some cognitive process (Fig. 8.3). While looking for causal interactions between neuronal circuitry and the behavioral outcomes, one should perform a mapping from a multidimensional space spanned by patterns of neuronal activity (namely, attractors of the neuronal networks) onto a multidimensional space spanned by the cognitive constructs. This is what I called the CCM approach. The direction of causal inference in CCM goes from circuitries toward behavior because the CCM approach is designed for better treatment, which should ultimately target the diagnostic cognitive constructs in MDD. Therefore, it is essential for the constructs to be compact, but the underlying circuits can be complex as is necessary. The CCM approach involves performing this mapping with the use of joint imaging and psychometric methods on large clinical datasets. Once one identifies the circuits underlying single cognitive dimensions of MDD, one can perturb this construct-related circuits in a single subject with treatments, affecting the neuronal dynamics, and tracking both the resulting position in the cognitive construct space and the dynamical properties in the construct-related circuits.

### 8.6.2 Execution of CCM

Execution of CCM is a multistep process. The preliminary step is to determine an extensive list of constructs involved in MDD. Since the classic diagnostic tools are questionnaires and experimental tasks, this analysis would run through a number of variables, grouping them into dimensions (e.g., based on factor analysis), and with a subsequent sanity check if the outcome constructs have a consistent content. The list of constructs determined in this protocol can be longer than the list of the DSM-5 criteria, thus I denote the constructs as anonymous $t_i$ in Fig. 8.3A. Furthermore, some constructs may be heritable and thus fulfill the definition of endophenotypes, which is especially relevant for executive functions (Friedman et al., 2008). Some other constructs, e.g., recurrent thoughts of death, are not likely to be heritable$^1$.

The second step is to find neuronal mechanisms behind each of the obtained constructs. For every single construct, one should start the procedure

---

$^1$However, this analysis will not reveal whether a given construct is heritable or not.
from the first order analysis: investigating patterns of activation and effective connectivity in a cohort of subjects exhibiting that construct (and, of course, a cohort of controls) to identify the underlying neuronal network and to build a corresponding dynamical system (Fig. 8.3A). Using Pearl’s definition of causality and for the effect size large enough, one can determine causal effects based on this observational study.

If on this first level, the analysis does not identify unique circuitry, and there can be multiple interacting circuitries involved in one construct. In that case, one should perform a second order analysis, e.g., a repeated diagnostic evaluation and repeated fMRI imaging assessment longitudinally within the same subject. Then, by using autoregressive models to analyze the time course of the construct and correlating these independent components with neuroimaging datasets, one could reveal independent components in the circuitry underlying this construct.

I predict that positive correlations between revealed cognitive constructs across subjects are inevitable, which should be reflected in overlaps between circuits underlying the constructs. One can also analyze how the attractors of the dynamical systems map onto cognitive constructs using temperature maps (Fig. 8.3B). Since these maps are created based on the data coming from the limited number of subjects in the cohort, these temperature maps will not span across the whole space of possibilities.

The third step is building dynamical models representing the identified circuitries underlying cognitive constructs. The proposed Wilson-Cowan model can be applied to any clinical data that reveals the distribution of activity in the brain over time, in particular, to the BOLD fMRI or MEG/EEG data. Wilson-Cowan model has some similarities to the dynamical causal modeling (DCM), a well established method for extracting effective connectivity for both fMRI and EEG/EMG data, in a sense that it describes the neuronal communication between brain regions in terms of Ordinary Differential Equations (see: Chapter 5). The major difference is that, in both classical (Friston, Harrison, and Penny, 2003) and recent stochastic version of DCM for fMRI (Daunizeau, Stephan, and Friston, 2012), there is an assumption of linear transfer functions, whereas it is known that large neuronal populations exhibit sigmoidal rather than linear response to the external inputs (Silver, 2010), which is incorporated in the Wilson-Cowan equations (Moran, Pinotsis, and Friston, 2013).

In this procedure, a single subject in a cohort is just an object to the explanatory science. However, once the circuitries underlying cognitive constructs involved in MDD are determined, the subject may become subject to a case study, and receive a personalized treatment. Investigation of the trajectory of the particular subject in the construct space in response to changes in the circuit activity caused by treatments (Fig. 8.3C) might not only provide new biomarkers for MDD and a better insight into the mechanisms of treatments, but also answer the question of how to predict resilience to treatment. This research may also elucidate factors that determine whether a treatment is effective to a particular group of subjects. Furthermore, this analysis might help to address the question if the mental disorders of interest, e.g., MDD, are homogenous or split into subtypes based on the subject trajectories in the construct space. Lastly, one might investigate how the treatments $TR_i$ in the given subject affect the attractors of the underlying networks (Fig. 8.3D).


8.6.3 Conditions to perform CCM

Figure 8.4: Conditions to perform CCM. A: Stable connectivity patterns and connectivity strengths are two sides of the same coin. Let us assume that the sigmoidal transfer function can be approximated as a simple linear function. I provide an example of a network sensitive to changes in connectivity strengths (a small change in connection weights yields a huge change in the value of stable activation patterns) and an example of a network insensitive to changes in connectivity strengths (a huge change in the connectivity strengths yields a small change in stable activation patterns). Therefore, the description of networks by means of dynamical systems provides more versatile description than connectivity strengths in the networks or stable connectivity patterns alone. B: Decomposition of psychiatric disorders into a number of diagnostic traits helps causal inference in the diagnostic process. Networks $N_i$ underlie cognitive constructs $C_i$ diagnostic to psychiatric disorders $D_i$. Disorders $D_1$ and $D_2$ share a common cognitive construct $C_2$, but involve also disorder-specific diagnostic constructs $C_1$ and $C_3$. Let us assume that the joint posterior distribution for every disorder $D_i$ factorizes into posterior probability distributions for single diagnostic constructs. Let us further assume that we find the same, specific pathologies in networks $N_1$ and $N_2$. If we can decompose the $D_1$-related network into a sum of networks $N_1$ and $N_2$ underlying single diagnostic constructs $C_1$ and $C_2$ (upper panel), we collect more evidence for the disorder $D_1$ (pathologies in both networks linked to diagnostic constructs) than for the disorder $D_2$ (pathology in one out of two networks linked to diagnostic constructs). On the other hand, if we are not able to decompose $D_1$ and $D_2$-related networks into networks underlying single diagnostic categories (lower panel), the amount for evidence in favor of both disorders is the same because both networks ($N_1 + N_2$) underlying disorder $D_1$ and ($N_2 + N_3$) underlying disorder $D_2$ are pathological.

CCM brings three new qualities to the table. Firstly, treating networks as dynamical systems allows to extract and characterize global properties of the networks involved in cognitive constructs in a comprehensive and versatile way. So far, research in human neuroimaging was focused on finding particular areas involved in cognitive tasks by virtue of stable activation patterns, or investigating context-dependent connectivity strengths...
between particular areas. These are two out of many angles which one can take to characterize large-scale brain networks. In fact, these are the two sides of the same coin: the distribution of activation patterns in a given network is a global property emerging from behavior of the underlying dynamical systems specified through the connection strengths between areas. Whether the activity patterns are more informative than the connectivity strengths, depends on the circumstances. In Fig. 8.4A, I present a toy example. Let us assume that, in the simplest case, the sigmoidal transfer function can be approximated as a linear function. For some combinations of inputs to the network and connection weights, a small change in connection weights (by 10%) yields an enormous change in the value of stable activation patterns (by 1000%, upper panel). For other combinations of weights and inputs, even huge change in the connectivity strengths (by 300%) yields a small change in stable activation patterns (by 10%). As a consequence, whether activity patterns in the networks are sensitive to changes in connectivity strengths depends on the tuning in the network, e.g., on the balance between connectivity weights in the network and external conditions such as experimental inputs. Therefore, since the dynamical systems incorporate both connectivity (as the cause) and stable activity patterns (as the effect), they integrate the two sorts of information about the circuits into one framework.

Secondly, the decomposition of psychiatric disorders into a number of diagnostic traits allows for the fundamental explanatory research in psychiatry, and therefore also for new, neuroimaging-based biomarkers for cognitive disorders. In terms of causal modeling, gathering clusters of traits into big cognitive paradigms such as psychiatric disorders can be mischievous, given that the disorders strongly overlap in terms of diagnostic criteria. A simple example is provided in Fig. 8.4B. In this example, overlapping networks $N_i$ underlie cognitive constructs $C_i$ which are diagnostic to psychiatric disorders $D_i$. Disorders $D_1$ and $D_2$ share a common cognitive construct $C_2$, but involve also disorder-specific diagnostic constructs $C_1$ and $C_3$. In this toy example, let us assume that the prior probabilities of cognitive constructs $C_i$ are equal and that likelihood of the pathologies in networks $N_i$ given constructs $C_i$ are the same. Let us further assume that in this subject, we find the same, specific pathologies in the networks $N_1$ and $N_2$. If we can decompose the $D_1$-related network into a sum of networks $N_1$ and $N_2$ underlying single diagnostic constructs $C_1$ and $C_2$ (Fig. 8.4B, upper panel), we can perform statistical inference linking specific changes in $N_1$ and $N_2$ with constructs $C_1$ and $C_2$, respectively, and collecting evidence behind the hypothesis that the subject is subject to the disorder $D_1$. Since $C_2$ is also a construct diagnostic to the disorder $D_2$, we also collect some evidence behind the hypothesis that the subject suffers from the disorder $D_2$. However, assuming that the joint posterior distribution for every disorder $D_i$ factorizes into posterior probability distributions for single diagnostic constructs, we collect more evidence for the disorder $D_1$ than for the disorder $D_2$.

On the other hand, if we are not able to decompose $D_1$ and $D_2$-related networks into networks underlying single diagnostic categories (Fig. 8.4B, lower panel), the amount for evidence in favor of both disorders is the same because both networks $(N_1 + N_2)$ underlying disorder $D_1$ and $(N_2 + N_3)$ underlying disorder $D_2$ are pathological, and we are not able to extract any
disorder specific subnetworks that would provide any further evidence in favor of one of the disorders. Therefore, decomposing mental disorders into single diagnostic constructs and linking constructs-specific circuits is of primary importance for explanatory models in psychiatry.

Thirdly, as a modeling procedure that projects neuronal dynamics into behavioral dimensions of MDD, CCM could not only serve as explanatory model when applied to a large cohort of subjects, but also enhance the current treatment selection for individual subjects and make a step toward personalized medicine. To perform explanatory research “in Pearl’s sense,” we need to use neuroimaging along with behavioral data from a large cohort of subjects: to reveal the circuitries underlying MDD-related cognitive constructs, we need to find systematic differences in the circuit dynamics that result in systematic differences in behavior. But once this explanatory research is finished and the circuitries underlying cognitive dimensions of MDD are defined, zooming into the circuit dynamics and its development under treatment in a particular subject would allow for more personalized interventions.

8.7 Limitations of the CCM approach

8.7.1 Plasticity and neurodegeneration

So far, sensory systems are best characterized in terms of underlying circuitries. However, events in sensory systems happen on a millisecond to second timescale whereas the evolution of psychiatric disorders is a few orders of magnitude slower and therefore might be much more complex. MDD may result from traumatic experience or emerge without a particular inducing event, but in any case, the process of falling into a depressive episode lasts for weeks, as opposed to perceptual learning which only takes seconds. Also, some treatment procedures are long-lasting, i.e., MDD pharmacotherapy is primarily monoamine based and typically requires intake for 3-4 weeks prior to symptomatic improvement (with the exception of ketamine). This time course is a major impediment to modeling MDD because imbalance in mood may arise not only on top of changes in neurotransmitter concentrations, but also result from other processes such as structural plasticity and neurodegeneration (Pittenger and Duman, 2008). The mechanisms underlying these two processes are not fully understood, and, in case of structural plasticity, is difficult to track in a living human brain. Neural mass models can only serve to compare between different stages of the disorder in an individual, and between different individuals at the same stage, yet do not provide a framework to demonstrate the real-time evolution of MDD.

8.7.2 Heterogeneity

MDD is a heterogenous disorder. The diagnostic criteria are still evolving, and the recently published DSM-5 diagnostic criteria for MDD allow for a variety of diagnostic combinations of cognitive constructs. Is there a plethora of different MDD types, or rather one prevalent state of mind that manifests itself in various ways depending on the subject? This remains
8.7. Limitations of the CCM approach

an open question. Furthermore, in the literature, there is often no clear distinction between subjects who experience their first depressive episode and those who suffer from recurrent depression whereas, as neurodegeneration proceeds and the severity of symptoms elevates, the course of the disease plays the crucial role in the treatment. This also provides a hindrance to the modeling procedures since the information about the stage of the disease is often missing from databases.

**Figure 8.5:** Attention as an example of a construct with multiple neural mechanisms underneath. Maintaining attention can be disrupted by at least two distinct mechanisms. A: Oversensitivity of the ventral attention network. Imaging studies revealed two systems managing attention in humans. On one hand, there is the dorsal attention system consisting of frontal eye fields (FEF) and intraparietal sulcus (IPS) that controls the voluntary deployment of attention (top-down control). On the other hand, we there is the right-lateralized ventral attention network (VAN) responsible for orienting attention toward sensory stimuli. This network involves temporoparietal junction (TPJ), intraparietal sulcus (IPS) in the parietal cortex, and the inferior frontal gyrus (IFG). As a part of orbitofrontal cortex, IFG receives a strong excitatory input from medial dorsal thalamus (MDT). Since MDT is overactive in MDD, this effect can make ventral attention network oversensitive to stimuli and, as a result, holding attention on salient stimuli becomes difficult to the subject. B: Diminished communication through coherence in the prefrontal cortex. Serotonin produced in the raphe nucleus (RN) modulates gamma oscillations in the prefrontal cortex (PFC), most probably by acting on fast-spiking interneurons expressing serotonin 5-HT2 and 5-HT6 receptors. Gamma oscillations play a key role in higher cognitive processes including attention and working memory. Since serotonergic input to the prefrontal cortex is known to be diminished in MDD, the decrease in gamma power may account for the effect of distractibility in MDD. Both of the above mechanisms lead to a decrease in inhibition within the prefrontal cortex that might explain why the attention managed in the PFC can both be disrupted as a result of hyperactivity of the medial dorsal thalamus and hypoactivity of the raphe nucleus.

Furthermore, complexity of MDD might also project to strongly overlapping construct-related circuits. E.g., it was found that the same brain area may host different circuits which, when activated, have opposing effects on
anxiety (Johansen, 2013). Furthermore, fMRI studies reveal anticorrelated networks to be activated during cognitive tasks (Fox et al., 2005). This is a circumstantial evidence that multiple distinct circuits can underlie single cognitive constructs (Fig. 8.1C). Furthermore, the same constructs can arise from different mechanisms. In Fig. 8.5, I discuss impairment in maintaining attention as an exemplary construct that may develop in the PFC of the MDD subjects from distinct processes.

8.7.3 Application of treatments to the CCM

Some of the possible applications of CCM, e.g., DBS and ECT, require invasive methods that cannot be used in humans on a daily basis and thus require rodent models. Rodent models of MDD is a well-explored discipline. However, whether or not the rodent models in mental disorders are fully translational remains unclear - which presents another difficulty for modeling studies. Whereas anhedonia, weight loss and gain, hypersomnia, or psychomotor retardation can be measured in a rodent, some other constructs such as the presence of recurrent thoughts of death, have no equivalent in rodents. On the other hand, modeling that requires invasive techniques, e.g., electrophysiology, cannot be ethically introduced into living human brains except under certain prescribed neurosurgical situations. However, the TMS-, pharmacotherapy- and neurofeedback-related CCM approach constitutes an adjunct to rodent models and, as a non-invasive method, it is applicable to subjects. Among the emerging treatment methods, neurofeedback seems to be a promising therapeutic procedure for CCM. This method is known to change connectivity in the functional networks (Haller et al., 2013; Koush et al., 2013), but its mechanisms of action are not yet known. Yet the concept of guided self-modulation in a subject in the absence of any third-party tools such as electric current or drugs is tempting. However, CCM can also be paired with all the other treatment procedures.

What can be a hindrance in application of the pharmacotherapy related CCM is that it is difficult to target a given construct with a particular drug because MDD drugs act on monoamine receptors, which are ubiquitous in the brain and present in multiple circuits at a time (Fig. 8.1C). Furthermore, some brain regions are hubs affected in many constructs, thus targeting these nodes with any form of treatment will have broad consequences for the global brain state. E.g., the ventral medial PFC is a major hub in the limbic system known to be involved in low mood (Anand et al., 2005), anhedonia (Sheline et al., 2010), feelings of worthlessness (Fitzgerald et al., 2008), and diminished working memory (Fuster, 1997) in MDD. However, the idea is to provide the online readout for the dynamics of all the involved circuits at a time. Following this approach, the clinician may first apply a specific treatment to target the desired cognitive construct, and then observe how the other construct-related circuits evolve along with the targeted one.

8.7.4 Temporal dynamics in the resting state

Circumstantial evidence suggests that in some aspects, MDD might require deeper insights into the activity of neural networks than the global patterns of activity in the brain as inferred from fMRI studies. E.g., the DBS
has different remission rates depending on the temporal characteristics of the applied current. As it was recently demonstrated that in the Parkinson’s disease, temporally irregular DBS is more effective than oscillatory stimulation (Brocker et al., 2013). This effect suggests that in addition to the modulatory effect on E-I balance, electrical stimulation can change the communication between the targeted region and its efferents by affecting communication through coherence (Fries, 2005). This means that, because of the lack of the strong temporal characteristics in the brain activity, the fMRI datasets might give an incomplete information about mechanisms of MDD. However, CCM is still a substantial progress to the therapy and treatment in mental disorders, and it gives the first insight into the circuits involved in the disorder and opens possibilities for further, more in-depth research.

8.7.5 Effective connectivity in EEG/EMG and fMRI research

So far, there are some applications of Ising models for characterization of the global network properties (as a number of so-called patterns stored in the network, Fraiman et al., 2009). However, Ising models are defined only for undirected networks and, to use the full potential of the CCM, this approach needs taking a step further by making connectivity directional. In fMRI research, determining connectivity strengths between the nodes is hard because of the limited amount of the temporal information in the datasets (see: Chapters 5 and 6). So far, the most established procedure for rendering effective connectivity from fMRI datasets is the DCM (Friston, Harrison, and Penny, 2003). However, this procedure is only applicable to very small networks of 3-4 nodes, and requires specification of a number of network parameters. In addition to that, as an inference procedure, DCM encounters some critics in the field (Lohmann et al., 2012). Since proper region definition is extremely important in causal research in fMRI (Bielczyk et al., 2017b), there is an urge for new, more data driven methods for approaching effective connectivity in these datasets.

In the field of EEG/EMG on the contrary, the problem of causality is orthogonal to the field of fMRI: the DCM procedure is quite successful in finding effective connectivity between the nodes of the network, however the optimal method for defining the nodes (or, ROIs) as the sources of the potentials recorded on the scalp is still an open problem. The three popular approaches are dipole modeling, dynamic imaging of coherent sources, and frequency-domain minimum current estimation (Liljestrom et al., 2005). These methods successfully identify the main sources of oscillations in the brain volume, however, there is still room for improvement in terms of the spatial resolution of reconstructed sources.

8.8 Concluding remarks

As proposed in the RDoC initiative, symptoms diagnostic for psychiatric disorders should be interpreted as psychopathological constructs that need
to be investigated, diagnosed, and treated independently. The CCM approach addresses this demand, and provides a new outlook at clinical treatments in mental disorders. Namely, the treatments not only regulate the levels of neuromodulatory substances in the brain but also change the dynamical state of the brain by regulating excitation-inhibition balance across brain circuits, which can be tracked with the use of the neuroimaging tools. This change in dynamics may be achieved in two ways: either by inducing the structural and functional plasticity that changes the functional connectivity in the circuit (through drug administration), or by providing stimulation/inhibition to the single circuit node(s) and therefore changing the global balance in the brain (through an electrical stimulation).

In this work, I underscore the potential of computational modeling in psychiatry as a tool to unravel mechanisms underlying the diagnostic symptoms, cluster diagnostic cohorts, and customize the clinical treatments in psychiatry. In addition to this, I anticipate that in the near future, new, personalized treatment methods based on non-invasive regulation of specific neuronal populations’ activity with gene therapy or magnetic fields may be possible. These approaches are still in the infancy stage. So far, gene therapy up-regulation of p11 protein in the rodent nucleus accumbens proved to cause a reversal of an anhedonic phenotype (Alexander et al., 2010).

Due to my assumptions, diagnostic symptoms of MDD are caused by (mal)functioning of the underlying neuronal circuits. Therefore, clinical cohorts homogenous in the circuit dynamics should also be responsive to similar treatments. Conducting the diagnosis in terms of circuit defects based on the construct domain will then ensure the clinical groups are clustered and represent more homogenous groups. Furthermore, comparison of depressed subjects and healthy controls in the construct space may assist in investigating whether MDD is a single disorder (and, diagnostic category) or whether it should be split into several diagnostic subtypes. This analysis may also reveal cognitive and neuronal signatures of the treatment-resistance. Tracking the subject’s position in the construct space in response to stimulation/inhibition on one hand, and the evolution of relevant attractor landscapes on the other hand, may provide new insights into the nature of treatments and help to create personalized medicine.
Chapter 9

General discussion

9.1 Summary

In this work, I investigated multiple methodological advances to signal detection, causal modeling, and statistical thresholding of connectomes with the common goal of facilitating correct interpretation of fMRI datasets. In Fig. 9.1, I summarize new contributions introduced in this Thesis.

Firstly, I propose using fractional moments of the BOLD distribution to improve on the algorithms for signal detection with the use of normality testing as described in Chapter 3 (Fig. 9.1, level [1]). These algorithms can, in general, contribute to the signal detection in multiple domains in science where experimental datasets are represented by a time series of poor temporal resolution so that spectral analysis cannot uncover the underlying signal. In the context of fMRI, better signal detection can also contribute to better ROI definition in resting state experiments as described in Chapter 3.

Furthermore, in my research, I use standard techniques to operationalize the concept of functional connectivity, primarily partial correlation (Fisher, 1924) adapted to fMRI by Marrelec et al., 2006. Partial correlation is, effectively, Pearson’s correlation (Pearson and Hartley, 1972) computed on the datasets preprocessed with the use of the Ordinary Least Squares regression (OLS, Edwards, 1976; Hayashi, 2000). As such, it incorporates the assumption that the associations between the nodes in the network are linear - while in fact, it was empirically found that the transfer functions in the communication between neuronal populations are highly nonlinear (Silver, 2010). Certain alternatives, such as assumption-free methods for operationalizing functional connectivity exist, e.g., mutual information (Press et al., 2007). It is a subject for further research whether or not the same concept of fractional moments of the distribution can be used to develop some new, assumption-free methods for finding associations between random variables (which I refer to by putting a question mark in Fig. 9.1, level [2]).

Next, significance of the functional connections needs to be determined; significant connections should be separated from spurious connections that appear as a consequence of computing correlations on a finite time series, or a background noise in the network (Fig. 9.1, level [3]). In this work, I had my little contribution to the subfield of thresholding functional connectomes in fMRI. Even though this research problem sounds very specific, in fact, there is a wide selection of available methods to approach this issue (e.g., permutation testing, Smith et al., 2011; Hyvärinen and Smith, 2013, proportional thresholding, Heuvel et al., 2008, or probabilistic thresholding, Váša, Bullmore, and Patel, 2018). In this work, I proposed a new approach to this
problem, namely thresholding connectomes with the use of mixture modeling (Bielczyk et al., 2018). One advantage of this new approach, is that, unlike in permutation testing, one can reliably perform the thresholding on the first level of the analysis, and the results will converge to the thresholding performed on the second level (Fig. 4.13).

Lastly, I applied the same concept of fractional moments of the distribution to perform more accurate classification into upstream and downstream nodes in the last step of the pairwise causal inference (Fig. 9.1, level [4]). This is achieved by combining pairs of fractional moments into cumulants and developing a classifier trained with the use of DCM (Friston, Harrison, and Penny, 2003) as the canonical generative model in fMRI (Bielczyk et al., 2016). I demonstrated that using this technique allows for extracting more in-depth information from the data, and for more robust classification of the nodes into upstream and downstream nodes with respect to the natural confounds present in the fMRI datasets, i.e., scale-free background noise in the neuronal networks (Bielczyk et al., 2017b) and varying SNRs.

**Figure 9.1:** Contribution to the standard pipeline for pairwise inference in fMRI presented in this Thesis. I proposed an advance to three steps out of four in the Patel’s tau (Patel, Bowman, and Rilling, 2006) data analysis pipeline.

### 9.2 Conclusions: what information can we extract from fMRI data?

As mentioned on multiple occasions throughout this Thesis, fMRI datasets are a difficult input for classic signal analytic methods, especially with respect to uncovering the dynamics low-pass filtered by the slow hemodynamic response (see: Chapters 1, 5 and 6). In my PhD work, I was investigating these intrinsic limitations of fMRI datasets more in-depth. Below, I am listing some general conclusions and questions open for further investigation.

#### 9.2.1 Signal detection

Firstly, signal detection in fMRI is hard. It remains unclear how the real mesoscale patterns of neuronal activity hidden under the slow hemodynamic responses would look like. We can only classify or compare different BOLD fMRI time series by means of the shape of the BOLD fMRI distributions - which strips the signals from the temporal component. It is a controversial subject whether the large-scale resting state networks exhibit any
oscillatory activity. According to the early view at this subject, resting state networks are associated with very slow spontaneous fluctuations in the BOLD fMRI time series at the order of magnitude of $0.01\, [Hz]$ (Niazy et al., 2011; Cordes et al., 2000; Cordes et al., 2001; Cole, Smith, and Beckmann, 2010), which separates them from respiratory ($0.1 - 0.5\, [Hz]$) and cardiovascular ($0.6 - 1.2\, [Hz]$) frequency bands. However, more recent studies suggest that, while most of the power spectrum in resting state BOLD fMRI time series occupies low frequencies indeed, some portion of the spectrum can also be consistently found in the high frequency range (Niazy, Smith, and Beckmann, 2008). Additionally, deconvolving BOLD fMRI time series from hemodynamic response broadens the power spectrum from $0.01\, [Hz]$ to $0.15\, [Hz]$ (Smith et al., 2008). But, are the current methods for estimating hemodynamic responses reliable? As the ground truth remains unknown, this question is hard to answer; one can only compare different methods in terms of reliability, e.g., in test-retest studies.

### 9.2.2 Connection detection

Secondly, how to operationalize functional connectivity and find true functional links between the ROIs? Operationalization of functional connectivity as statistical associations between the ROIs is a broad topic (Smith et al., 2013) that falls beyond the scope of this Thesis. However, in this Thesis, it is considered how to find true positives in the functional connectivity estimates. Mixture modeling proposed here (Bielczyk et al., 2018) is only one of multiple possibilities, next to proportional thresholding (Heuvel et al., 2017), permutation testing (Hyvärinen and Smith, 2013), etc. We need to take into consideration that all these methods give certain output that should also be corrected for multiple comparisons as the number of connections grows with the network size as $O(N^2)$. As mentioned in Chapter 2 while discussing methods for signal detection in fMRI, the most basic version of the correction for multiple comparisons is the Bonferroni correction (Dunnett, 1955), namely dividing the desired threshold $p$-values by the number of tested connections. However, Bonferroni correction is rarely applied in fMRI because its basic assumption, i.e., independence between pairs of time series, is not fulfilled in the brain in general. For this reason, family-wise correction for multiple comparisons using false discovery rates, usually estimated with the use of Benjamini-Hochberg method (Benjamini and Hochberg, 1995) is often a method of choice in connectomic research in fMRI. However, the generic version of Benjamini-Hochberg does not take into account the properties of fMRI data, namely that temporal correlations in the signals induce additional dependency between time series. Recent progress to resolve this issue is the work by Afyouni, Smith, and Nichols, 2018, in which temporal autocorrelations as well as instantaneous and lagged cross-correlations are taken into account to assess dependency between time series. The problem of thresholding functional connectomes is multi-layered, and more attention is needed to propose protocols correctly estimating true positives from the data.
9.2.3 Causal inference in fMRI

Lastly, effective connectivity research in fMRI is highly limited because of the poor temporal characteristics of the BOLD fMRI time series. Even though certain methods such as the DCM (Friston, Harrison, and Penny, 2003) are reported as highly reliable (Schuyler et al., 2010; Rowe et al., 2010; Bernal-Casas et al., 2013; Frässle et al., 2015; Frässle et al., 2016a; Tak et al., 2018; Almgren et al., 2018), yet still, it is a statistical effect. This means that the results are reproducible but not necessarily true as the experimental validation of DCM is still missing. One seminal study by Ryali et al., 2016, attempted to validate so-called state-space multivariate dynamical systems (Ryali et al., 2011) with the use of optogenetic stimulation, and the results were promising. However, it would be beneficial if another validation study is performed by another team, independent from the authors of the method. Furthermore, it would be beneficial if more standard approaches to effective connectivity in fMRI, such as the DCM, are validated in a similar fashion.

9.3 How to use concepts developed in this work in a broader context?

Currently, developments in many branches of science are very dynamic. Perhaps in fifty years from now, fMRI will become an obsolete technique for imaging activity of the human brain, while much more advanced and acute techniques for human neuroimaging will be created and popularized by then. Maybe, one day, tracking the dynamics of every single cell in the human brain in real-time will become a standard.

Having this in mind, rather than developing fMRI-specific methods, e.g., modeling fMRI-specific motion artifacts or hemodynamic response functions, I attempted to create generic methods for data analysis that can be applied or adapted to other types of datasets in neuroscience and beyond. Below, I am listing some ideas for further applications of these methods beyond fMRI research.

9.3.1 Signal detection

Measures for Gaussianity (or, normality) of a distribution are applicable across multiple disciplines of science, from economy to engineering. Normality tests are used to determine whether the sample is drawn from a normal distribution. As mentioned in Chapter 3, this is a necessary step in the data analysis across multiple disciplines of science where normality of the sample distribution is expected. When this assumption is not fulfilled, it is a warning that the data is contaminated, or that there are some (possibly, latent) confounding variables in the dataset. I give some examples below:

1. Genome-wise association studies: the distribution of effect sizes for SNPs across the genome is compared to a normal distribution to determine whether there is a signal in the data or rather, if all the detected effects can be explained as random. Currently, this is usually achieved with the use of a so-called quantile-quantile plots, QQ (Ehret, 2011),
9.3. How to use concepts developed in this work in a broader context?

2. Psychometrics: psychological tests and questionnaires are standardized under the assumption that the distribution of behavioral phenotypes in a cohort is normal. In case the sample is not normally distributed, this can be an evidence that the test/questionnaire measures a combination of two or more independent traits and should either be simplified or factorized into more dimensions.

3. Engineering: Six Sigma rule (Tennant, 2001) is a technique used for process improvement in engineering, standard in large companies in the engineering industry, e.g., General Electric or Motorola. In short, in Six Sigma strategy, the quality of the final product is secured by finding and minimizing variability in the process of manufacturing. In this technique, every parameter of the product is assumed to be normally distributed across the batch. At each stage of production, pieces whose parameters are beyond pre-specified range defined by number of standard deviations below and above the mean in the batch (typically, \([-3.0 \text{STD}, +3.0 \text{STD}\]), hence the name "Six Sigma"), are removed from the sample. Six Sigma is a highly efficient technique to reduce costs, and two-thirds of the Fortune 500 organizations have implemented this strategy in their production pipelines over the course of the past 20 years (Feo and Barnard, 2005).

9.3.2 Connection detection

The new technique for thresholding connectomes proposed in this Thesis (namely, using pseudo-FDR computed with the use of mixture modeling, Bielczyk et al., 2018) is agnostic with respect to the operationalization of functional connectivity: this approach can be applied for thresholding any functional connectome regardless whether they were obtained with the use of Pearson’s correlation, partial correlation, or other methods for quantifying associations between time series, e.g., mutual information. For this reason, the method is highly versatile and thus, applicable in multiple disciplines where correlations in the networks are studied. Below, I give some examples of the disciplines other than fMRI research which feature functional connectivity research in some form:

1. Analysis of the internet structure: since internet is a complex, global network, connectivity and connectivity dynamics can be defined and studied in this network (Pastor-Satorras, Vazquez, and Vespignani, 2002),

2. Economics: correlational analysis to establish factors underlying stock exchange performance has a long history, both with the use of Pearson (Buse and Stefan, 2014; Sharma and Banerjee, 2015) and partial correlation (Kenett et al., 2014),

3. Genetics: correlation between the genetic influences on different phenotypes and traits are studied (Loehlin and Vandenberg, 1968; Purcell, 2002; Kohler, Behrman, and Schnittker, 2011).
9.3.3 Causal inference in fMRI

Further applications are not as straightforward for the research related to effective connectivity introduced in this Thesis. The reason is that, the method proposed in this Thesis relies on the DCM generative model (Friston, Harrison, and Penny, 2003) which is not directly applicable to other datasets. However, the proposed technique to develop a classifier for binary classification can be reapplied to any dataset that exhibits the following two key characteristics:

1. Relatively flat power spectrum in the background noise so that little information is conveyed in the frequencies, e.g., due to poor temporal resolution of the data recordings. For this reason, most of the information hidden in the data, is contained in the distributions of samples rather than in the power spectrum,

2. Generative model for this dataset exists (Frässle et al., 2018; Betzel and Bassett, 2017), so that it is possible to develop a classifier based on forward simulations.

Examples of such research problems might be modeling the following:

1. Climate change with the use of spatial-temporal generative modeling (Lozano et al., 2009),

2. Demographic movements and migration (Spooner, 1971; Sheldon, Elmohamed, and Kozen, 2007),

3. DNA formation (Killoran et al., 2017).

In any case, it should be taken into consideration that discovering causal underpinnings underlying physical phenomena is a complex research problem that requires a reliable generative model, the right assumptions, and good quality datasets.
Bibliography


Bernal-Casas, D. et al. (2013). “Multi-site reproducibility of prefrontal - hippocampal connectivity estimates by stochastic DCM”. In: NeuroImage 82, pp. 555–63. DOI: 10.1016/j.neuroimage.2013.05.120.


Delgado, M. R. et al. (2000). “Tracking the hemodynamic responses to reward and punishment in the striatum”. In: Journal of Neurophysiology 84.6, pp. 3072–7. DOI: 10.1152/jn.2000.84.6.3072.


Bibliography


Li, B. et al. (2014). “Impaired Frontal-Basal Ganglia Connectivity in Adolescents with Internet Addiction”. In: Scientific Reports 4, p. 5027. DOI: 10.1038/srep05027.


Bibliography


Bibliography


Power, J. D. et al. (2014). “Methods to detect, characterize, and remove motion artifact in resting state fMRI”. In: *NeuroImage* 84. DOI: 10.1016/j.neuroimage.2013.08.048.


Tokutsu, Y. et al. (2013). “Follow-up study on electroconvulsive therapy in treatment-resistant depressed patients after remission: a chart review”. In: *Clinical Psychopharmacology and Neuroscience* 11.1, pp. 34–8. DOI: 10.9758/cpn.2013.11.1.34.


Wierońska, J. et al. (2012). “Depression viewed as a GABA/glutamate imbalance in the central nervous system”. In: Clinical, Research and Treatment Approaches to Affective Disorders, pp. 235–66.


Worsley, K. J. (1994). “Local maxima and the expected Euler characteristic of excursion sets of $\mathcal{G}$, F and t fields”. In: Advances in Applied Probability 26.01, pp. 13–42. DOI: 10.1017/S0001867800025970.


Zhang, J. et al. (2011a). “Disrupted Brain Connectivity Networks in Drug-Naive, First-Episode Major Depressive Disorder”. In: Biological Psychiatry 70.4, pp. 334–42. DOI: 10.1016/j.biopsych.2011.05.018.


Summary

In this Thesis, I present the state-of-the-art methods for signal detection and causal inference in functional Magnetic Resonance Imaging, and I introduce my original contributions to this field. One common feature of all the methods I proposed, is that they are based on the distributions of BOLD fMRI values rather than on the original BOLD time series. This is a consequence of the fact that BOLD fMRI time series has a poor time resolution, and due to slow hemodynamics, also poor temporal characteristics (as explained in Chapter 5). Therefore, the sequence of the samples in the BOLD fMRI time series carries a little amount of information that could be useful for signal detection or causal inference. I demonstrated this effect in Chapter 6 using synthetic fMRI datasets, where I showed that as little as 200 ms difference in the local peak of the hemodynamic response between brain regions can invert the directionality of the inferred causal connections when it comes to "lagged methods" that base the causal inference on the sequence of fMRI samples.

Conceptually, my PhD work on causal inference was based on the multi-stage protocol developed by Patel, Bowman, and Rilling, 2006. In his work, Patel and colleagues proposed approaching causal inference in brain networks using fMRI datasets in the following fours steps:

1. Defining the nodes in the network. This seems to be an obvious first step when analyzing networks in the brain. In practice, multiple methods for defining nodes in brain networks exist: this can be done using brain anatomical atlases (a.k.a., anatomical parcellation), or by finding groups of voxels that co-activate, thus might be involved in similar cognitive processes (a.k.a., functional parcellation). In my work, I focused on methods for functional parcellation, and on contributing to these methods using a novel method based on fractional moments of the distributions of the BOLD fMRI values,

2. Computing functional connectivity between nodes of the network. Functional connectivity refers to statistical associations between the nodes of the network. The assumption is that, the co-activation between two nodes indicates that they are likely involved in the same cognitive processes and that potentially, there might be a causal link between them. Multiple methods for computing functional connectivity exist, from full correlation, through partial correlation, to methods imported from information theory such as mutual information. In my work, I chose for partial correlation (computed through inverse covariance) as it focuses on estimating the strength of the direct links between the nodes and partialing out the confounding effects from the other nodes in the network,
3. Sparsifying the functional connectome. While computing functional connectivity from fMRI time series, we will almost certainly obtain a nonzero score even if the two time series are unrelated. Therefore, it is important to determine whether any given functional connection in the network is statistically significant. Again, there are multiple methods for determining significance, and I proposed one new way of approaching this research problem.

4. Having found the significant connections, one can assume that there might be some direct, pairwise causal dependencies in the functional connections. In this thesis, I proposed a new method to establish the directionality of this causal connection based on fractional moments of the BOLD fMRI distribution combined into cumulants.

To sum up, in this thesis, I contributed to points (1), (3), and (4) of this procedure.

Firstly, I discussed the topic of signal detection in fMRI - as this is the basis for defining nodes of the network in a functional way (Chapter 2). I also provided my own contribution to this field. Up to now, the two leading approaches to signal detection in fMRI were the General Linear Modeling (GLM), and the Independent Component Analysis (ICA). GLM is a massive univariate approach based on comparing mean intensity two conditions - usually between experimental condition (e.g., cognitive task), and the resting state - voxel by voxel. As such, GLM can only be used in task fMRI (as opposed to resting state fMRI). ICA, on the other hand, is based on analyzing distributions of BOLD fMRI values rather than BOLD time series. In ICA, voxels in which BOLD fMRI distributions have a similar pattern of non-gaussianity, are grouped together. In this thesis, I explored this direction further, and I proposed new, more sensitive way of computing non-gaussianity of the distribution by using fractional moments of the distribution (Chapter 3). The essence of this concept is that, instead of single discrete markers of non-gaussianity such as skewness or kurtosis (which are, in fact, moments of orders 3 and 4), we look at moments of the distribution as a continuous dimension, and integrate evidence for non-gaussianity along this dimension. This view yields more information and increases sensitivity to non-gaussianity compared to classic methods, that are usually built using a combination of two moments of the distribution as features for the classification: Gaussian vs non-Gaussian. In this work, I demonstrated that 'Momentum,' the test for computing non-gaussianity that I proposed, is more sensitive than the classic tests used in multiple areas of research, i.e., Shapiro-Wolf, Jacque-Bera, and Pearson-d’Agostino. As a generic method for computing non-gaussianity, this method might not only be used to improve the ICA protocols in fMRI but also to test assumptions behind almost any test in data science.

Secondly, in Chapter 4 I proposed a new method for thresholding connectomes with the use of mixture modeling. This method is based on the concept of mixture modeling. The assumption is that the distribution of all connection weights within the network is, in fact, a mixture of two distributions: a null distribution centered at 0 that represents the noise in the
network, and the distribution of “real” connections representing communication between the nodes that we are looking for (in fact, we can take into account two distributions of real connections: one representing excitatory functional connections, and one representing inhibitory functional connections). Given this mixture, pseudo-False Discovery Rate (FDR) can be computed for each value of the threshold. FDR controls the number of errors of type 1: the expected proportion of false “discoveries” (incorrectly rejected-null hypotheses). In this work, I demonstrated that using the method based on mixture modeling and FDR gives an advantage compared to permutation testing (as a competitive method of controlling error of type 1), the most popular method for sparsifying functional connectomes. This method can reliably sparsify a functional connectome in a single subject, while permutation testing can only be used for considerably large groups of subjects. This is especially important when it comes to case studies, and translational studies where experimental groups are usually very small and involve only a few subjects.

Lastly, I used the concept of fractional moments once more - this time, by combining fractional moments into cumulants - to create a method for determining the directionality of causal connections in the connectome built in the previous three steps of the inference procedure (Chapter 7). I used synthetic datasets generated under the state-of-the-art generative model for fMRI, Dynamic Causal Modeling, to develop a classifier that can determine the directionality of the causal links. I then introduced background noise and varying levels of input activity to the synthetic datasets and demonstrated that this new classifier infers the directionality of causal connections with higher accuracy than the state-of-the-art methods for establishing causality such as Granger Causality or Partial Directed Coherence.
Samenvatting

In dit proefschrift presenteer ik de huidige stand van zaken wat betreft methoden voor signaaldetectie en het vinden van oorzakelijk-gevolg relaties (“causal inference”) in functionele kernspintomografie (deze term is beter bekend als de afkorting fMRI, voor het Engelse “functional magnetic resonance imaging”). Daarnaast introduceer ik mijn eigen bijdragen aan dit veld. Een gemeenschappelijk kenmerk van alle methoden die ik heb voorgesteld, is het feit dat ik werkte aan de verdeling van BOLD fMRI-waarden in plaats van aan de originele BOLD-tijdreeksen. Dit was noodzakelijk vanwege het feit dat BOLD fMRI-tijdreeksen vanwege trage hemodynamica een slechte tijdsresolutie en andere temporele kenmerken hebben (zoals uitgelegd in hoofdstuk 5). Daarom bevat de volgorde van de observaties in de BOLD fMRI-tijdreeksen slechts een kleine hoeveelheid informatie die nuttig kan zijn voor signaaldetectie of oorzaak-gevolg analyse. Ik demonstreerde dit effect in hoofdstuk 6. Hier liet ik zien dat causale methoden die juist gebaseerd zijn op de temporele kenmerken van fMRI, de omgekeerde oorzakelijke conclusie trekken wanneer ik een 200ms verschil in de lokale piek van de hemodynamische response introduceerde.

Conceptueel was mijn promotieonderzoek naar causale gevolgtrekking gebaseerd op het multi-stage protocol ontwikkeld door Patel, Bowman, and Rilling, 2006. In zijn werk stelde Patel voor de causale gevolgtrekking in hersennetwerken te benaderen met behulp van fMRI-gegevenssets in de volgende vier stappen:

1. Het definin van netwerkknopen. Dit lijkt een voor de hand liggende eerste stap bij het analyseren van netwerken in de hersenen. In de praktijk bestaan er meerdere methoden voor het definin van knooppunten in hersennetwerken. Dit kan worden gedaan met behulp van hersenanatomische atlassen (anatomische indeling) of door groepen voxels te vinden die tegelijkertijd geactiveerd zijn, en dus betrokken zijn bij vergelijkbare cognitieve processen (ook wel functionele indeling genoemd). In mijn werk heb ik me gericht op methoden voor functionele indeling en op een nieuwe methode op basis van fractionele momenten van de verdelingen van de BOLD fMRI-waarden.

2. Functionele verbindingen tussen de knopen berekenen. Functionele connectiviteit geeft statistische associaties aan tussen de knooppunten van het netwerk. De veronderstelling is dat co-activatie tussen twee knooppunten aangeeft dat ze waarschijnlijk betrokken zijn bij dezelfde cognitieve processen, en dat er mogelijk een oorzakelijk verband tussen hen kan bestaan. Er bestaan meerdere methoden voor het berekenen van functionele connectiviteit, van correlatie via parti correlatie tot methoden georteerd uit de informatietheorie, zoals mutual information. In mijn werk heb ik gekozen voor parti correlatie (berekend via de inverse covariantie) omdat deze zich richt op het schatten
van de sterkte van de directe koppelingen tussen de knooppunten en de verstorende effecten van andere knooppunten in het netwerk compenseert.

3. Het uitdunnen van het hersennetwerk. Wanneer we functionele connectiviteit uit fMRI-tijdreeksen berekenen, zullen we vrijwel zeker geen score van exact nul vinden, zelfs wanneer de twee tijdreeksen volledig onafhankelijk van elkaar zijn. Daarom is het belangrijk om te bepalen of een bepaalde functionele verbinding in het netwerk statistisch significant is. Er zijn meerdere methoden om deze significantie te bepalen, en ik stelde een nieuwe manier voor om dit onderzoeksprobleem te benaderen.

4. Nadat de significante verbindingen zijn gevonden, kan worden aangenomen dat er enkel directe, paarsgewijze causale afhankelijkheden tussen de knooppunten in de functionele verbindingen kunnen zijn. In dit proefschrift heb ik een nieuwe methode voorgesteld om de richting van dit causale verband vast te stellen, gebaseerd op fractionele momenten van de BOLD fMRI-verdeling, gecombineerd in cumulanten.

Samenvattend heb ik in dit proefschrift bijgedragen aan de punten (1), (3) en (4) van deze procedure.

Ten eerste heb ik in hoofdstuk 2 het onderwerp signaaldetectie in fMRI besproken, omdat dit de basis is voor het functioneel definin van knooppunten van het netwerk en mijn eigen bijdrage aan dit veld leverde. Tot nu toe waren de twee belangrijkste benaderingen voor signaaldetectie in fMRI General Linear Modelling (GLM) en Independent Component Analysis (ICA). GLM is een massieve univariate benadering gebaseerd op het vergelijken van gemiddelde intensiteit twee condities - meestal tussen experimentele conditie (bijvoorbeeld cognitieve taak) en rusttoestand - voxel voor voxel. Als zodanig kan GLM alleen worden gebruikt in taak fMRI (in tegenstelling tot rusttoestand fMRI). ICA, aan de andere kant, is gebaseerd op het analyseren van distributies van BOLD fMRI-waarden in plaats van BOLD-tijdreeksen. In ICA zijn voxels waarin BOLD fMRI-distributies een vergelijkbaar patroon van niet-Gaussische-verdeeldheid hebben, gegroepeerd. In dit proefschrift heb ik deze richting verder onderzocht en heb ik nieuwe, meer gevoelige manieren onderzocht om de niet-Gaussische-verdeeldheid van de distributie te berekenen, door fractionele momenten van de distributie te gebruiken (hoofdstuk 3). De essentie van dit concept is dat we, in plaats van afzonderlijke discrete kenmerken van niet-Gaussische-verdeeldheid zoals scheefheid of gewelfdheid (die in feite momenten van orde 3 en 4 zijn), te beschouwen, deze eigenschappen van verdeingen te beschrijven langs een continuum van momenten, en bewijs voor niet-Gaussische verdeeldheid langs deze dimensie te verzamelen. Deze weergave levert meer informatie op, en verhoogt de gevoeligheid voor niet-Gaussische verdeeldheid in vergelijking met klassieke methoden, meestal gebaseerd op een combinatie van twee momenten van de distributie als kenmerken voor classificatie: Gaussiaans versus niet-Gaussiaans. In dit werk
heb ik aangetoond dat "Momentum", de test voor het berekenen van niet-Gaussische verdeelheid die ik heb voorgesteld, gevoeliger is dan de klassieke tests die worden gebruikt in meerdere onderzoeksgebieden, zoals Shapiro-Wolf, Jacque-Bera, en Pearson-d’Agostino. Als een generieke methode voor het berekenen van niet-Gaussische-verdeelheid kan deze methode niet alleen worden gebruikt om de ICA-protocollen in fMRI te verbeteren, maar ook om aannames achter bijna elke test in de wetenschap en dataverwerking te testen.

Ten tweede heb ik in hoofdstuk 4 een nieuwe methode voorgesteld voor het uitdunnen van hersennetwerken met behulp van mixture modelling en het toepassen van een drempelwaarde. Deze methode is gebaseerd op het idee dat de verdeling van alle verbindingsgewichten binnen het netwerk in feite een mengsel is van twee verdelingen: een nullverdeling gecentreerd op 0 die de ruis in het netwerk vertegenwoordigt, en de verdeling van "echte" verbindingen die communicatie vertegenwoordigen tussen de knooppunten waarnaar we op zoek zijn (we kunnen in feite rekening houden met twee distributies van echte verbindingen: een die exciterende functionele verbindingen vertegenwoordigt, en een die remmende functionele verbindingen vertegenwoordigt). Gegeven deze mix kan de False Discovery Rate (FDR) worden berekend voor elke waarde van de drempel. FDR bepaalt het aantal fouten van type 1: het verwachte aandeel van valse "ontdekkingen" (ten onrechte verworpen nulhypothesen). In dit werk heb ik aangetoond dat het gebruik van de methode op basis van mixture modelling en FDR een voordeel heeft in vergelijking met permutatietests (als een alternatieve methode voor het beheersen van type 1 fouten), de meest populaire methode voor het uitdunnen van functionele hersennetwerken. Deze methode kan betrouwbaar een functioneel hersennetwerk in een enkele proefpersoonuitdunnen, terwijl permutatietests alleen voor aanzienlijk grote groepen proefpersonen kunnen worden gebruikt. Dit is vooral belangrijk als het gaat om dierstudies waarbij experimentele groepen meestal erg klein zijn en slechts enkele proefdieren bevatten.

Ten slotte heb ik het concept van fractionele momenten nogmaals gebruikt - dit keer door fractionele momenten in cumulanten te combineren - om een methode te creëren voor het bepalen van de richting van causale verbindingen in het hersennetwerk gebouwd in de vorige drie stappen van de inferentieprocedure (hoofdstuk 7). Ik gebruikte synthetische datasets gegenereerd onder het geavanceerde generatieve model voor fMRI, Dynamic Causal Modelling, om een classifier te ontwikkelen die de richting in de causale koppelingen kan bepalen. Ik introduceerde vervolgens achtergrondruis en verschillende niveaus van inputactiviteit in de synthetische gegevenssets en toonde aan dat deze nieuwe classifier de richting van causale verbindingen met hogere nauwkeurigheid afleidt dan de state-of-the-art methoden voor het vaststellen van causaliteit zoals Granger Causality of Partial Directed Coherence.
Curriculum Vitae

Natalia Bielczyk was born in Katowice in 1986. Since 2004, she was attending the College of Inter-Faculty Individual Studies in Mathematics and Natural Sciences at the University of Warsaw, Poland. Within the College, she obtained three Master titles. In February 2010, she graduated from M.A. in Psychology (specialized in Psychometrics). In her Masters thesis, she showed that the distribution of creativity in the general population is highly non-gaussian, and creative traits do not correlate with other personal traits such as political views or sense of humor. In July 2010, she graduated from M.Sc. in Mathematics (specialized in Applied Mathematics). In her Masters thesis, she was investigating systems of differential equations with delay, and applying those systems to study relations between people. In September 2012, she graduated from M.Sc. in Physics (specialized in Medical Physics). Her Masters thesis was dedicated to a new, automated EEG data analysis with use of Directed Transfer Function. Meanwhile, in September 2011, she also received a title of Young Master of Business Administration from the Warsaw School of Economics.

In 2013, Natalia started a PhD program at the Donders Institute for Brain, Cognition and Behaviour and the Radboud University Nijmegen Medical Center. Her PhD project under supervision of prof. dr. Jan K. Buitelaar and prof. dr. Christian F. Beckmann concerned signal detection and causal inference in functional Magnetic Resonance Imaging. During her PhD, she was working on developing new methods for signal detection, functional and effective connectivity research in functional Imaging Resonance Imaging. During her PhD, she also Served as a Career Development and Mentoring Manager within the Organization for Human Brain Mapping Student and Postdoc Special Interest Group (2017-2019). In 2018, she created Stichting Solaris Onderzoek en Ontwikkeling: a private foundation dedicated to mentoring in academia, and helping researchers in finding jobs outside academia. In 2019, she launched Welcome Solutions: a company helping researchers in self-discovery and finding the right career track in industry. She also authored a book entitled "What is out there for me? The landscape of post-PhD career tracks."

After working hours, Natalia spends her time on blogging, gardening, dancing, painting, learning about economy, spa, hiking and traveling.
List of publications

Publications as a part of the thesis:


Other publications:

A Published:


B Under review / in preparation:


Donders Graduate School for Cognitive Neuroscience

For a successful research Institute, it is vital to train the next generation of young scientists. To achieve this goal, the Donders Institute for Brain, Cognition and Behaviour established the Donders Graduate School for Cognitive Neuroscience (DGCN), which was officially recognised as a national graduate school in 2009. The Graduate School covers training at both Master’s and PhD level and provides an excellent educational context fully aligned with the research programme of the Donders Institute.

The school successfully attracts highly talented national and international students in biology, physics, psycholinguistics, psychology, behavioral science, medicine and related disciplines. Selective admission and assessment centers guarantee the enrolment of the best and most motivated students. The DGCN tracks the career of PhD graduates carefully. More than 50% of PhD alumni show a continuation in academia with postdoc positions at top institutes worldwide, e.g. Stanford University, University of Oxford, University of Cambridge, UCL London, MPI Leipzig, Hanyang University in South Korea, NTNU Norway, University of Illinois, North Western University, Northeastern University in Boston, ETH Zürich, University of Vienna etc. Positions outside academia spread among the following sectors: specialists in a medical environment, mainly in genetics, geriatrics, psychiatry and neurology. Specialists in a psychological environment, e.g. as specialist in neuropsychology, psychological diagnostics or therapy. Positions in higher education as coordinators or lecturers. A smaller percentage enters business as research consultants, analysts or head of research and development. Fewer graduates stay in a research environment as lab coordinators, technical support or policy advisors. Upcoming possibilities are positions in the IT sector and management positions in pharmaceutical industry. In general, the PhD graduates almost invariably continue with high-quality positions that play an important role in our knowledge economy.

For more information on the DGCN as well as past and upcoming defenses please visit:
http://www.ru.nl/donders/graduate-school/phd/
Acknowledgments

My PhD was a long journey. I came to the Netherlands in 2011 as a fresh graduate, unaware of how the academia works, what are my strengths and weaknesses, and what are the pivotal questions in neuroscience. Now I feel like within these few years, I experienced more than for the whole 25 years before. Since I first arrived in Nijmegen, I visited over twenty countries, I met hundreds of amazing people on the way, and acquired a few new hobbies that will probably stay with me for a lifetime. There is a long list of people which made all of this possible to happen, and whom I awe many thanks to. Some names I will mention more than once as I know the people behind these names in a few different roles.

Supervisors
Firstly, I would like to thank my supervisors, Jan Buitelaar, Christian Beckmann, and Jeffrey Glennon.

I would like to thank Christian Beckmann, whose deep belief in my capabilities for solving research problems unsolved for the last fifteen years left me with no other choice than to really try. There were ups and downs in my PhD, and I must say that some of Christian’s points I only understand in retrospection. In any case, he taught me a lot about the craft of designing and validating methods for neuroimaging. Thank you, Christian!

I would like to thank Jan Buitelaar who was always supportive to me. Jan has a very rational outlook on science; he is always looking at the situation objectively and from a perspective, trying to find a neutral ground and picture things as they truly are. I really enjoy this ability in people.

Jeffrey Glennon was actually the first person to see the potential in me and to give me a shot for (actually, re-entering) a PhD program. Jeffrey has an ability to keep on smiling in any circumstances, see the good side of just anything, and he is truly interested in his students’ personal goals.

Lastly, I would like to thank Alberto Llera for helping me in publishing some of the papers I was working on in my PhD.

Coworkers
I would like to thank both labs I was working for: the Translational Psychiatry group by Jeffrey Glennon: Shaghayegh Abghari, Houshang Amiri, Mireille Bakker, Sarita Dam, Josi Driessen, Amanda Jager, Martha Havenith, Sabrina van Heukelum, Katarzyna Kapusta, Euginia Listiani, Charlotte Oomen, Natasha Pasricha, Ilse van de Vondervoort, Peter Zijderveld (and others), and the Statistics in Neuroimaging group by Christian Beckmann: Alberto Arenas, Roselyne Chauvin, Winke Francx, Koen Haak,
Nikki Janssen, Andre Marquand, Maarten Mennes, Peter Mulders, Marianne Oldenhinkel, Erik van Oort, Iza Przeździk, Emma Sprooten, Thomas Wolfers (and others).

Funding sources
I would like to thank Jeffrey Glennon for co-sponsoring my PhD as the part of the TACTICS, the MATRICS, and the OPTIMISTIC consortia. I would also like to thank Christian Beckmann for co-sponsoring the other half of my PhD contract from his personal VIDI grant. I would also like to thank FENS-IBRO for the international travel grant to attend the Brainhack Networks in Paris and the Organization for Human Brain Mapping (OHBM) Annual Meeting in Singapore in 2018. Lastly, I would like to thank FENS for the travel grant that allowed me to attend the FENS meeting in Berlin in 2018.

Department representatives
I would like to hereby thank Arthur Willemsen, Guillén Fernández, Marly Korsten, Femke Maaij, Tanja Doeller, Renee Leclercq, and Claudia Luettk for logistic help and for making my PhD defense happen.

Students
I would like to thank my students, Patrick Ebel, Jirrine Hogenaar, Fabian Walocha, Marein Koningen, and Jesse Zwamborn, as well as the students I was remotely advising: Katarzyna Piskala, Martyna Plomecka. They were enthusiastic, hard working, and, above all, fearless. Once you enter the system and become a PhD student, you might lose this childlike curiosity, and start thinking about other people’s expectations too much. Working with students brought me a lot of joy because they were nowhere near being depressed, intimidated, or anxious. They just embarked on the project, tried to learn something and make new research happen, without hesitation or high hopes - and the results were often beyond any expectations.

Teams I was working in
During my PhD, I had an opportunity to work in a few wonderful teams. Firstly, I would like to thank my dear Causality Club: primarily to Tim van Mourik and Sebo Uithol, and also Peter Vavra, Paul Anderson, and all the other people who contributed to our afternoon discussions on causality. I learned so much about the scientific discussion, and about scientific writing from you guys! (although I am still working on my English articles...)

I would also like to put special thanks for Tim van Mourik here. Tim got genuinely interested in the idea of fractional moments of the distribution, and discussed his insights with me on multiple occasions. This was a great help and encouragement!

Secondly, I would like to thank my colleagues from the DCN PhD council where I spent almost four years between spring of 2012 and fall of 2015: Ruben van Vogt, Tim Schroeder, Anna Geuzebroek, Annemiek Barsingerhorn, Katarzyna Kapusta, Amanda Jager, and others. It was very enjoyable to work with you guys!
I would also like to thank the AoN Brainhack Warsaw team. With this team, we organized a hackathon as a satellite of the Aspects of Neuroscience conference in Warsaw, November 2017. So, cordial thanks to Daniel Borek, Krzysztof Bielski, and Martyna Łempicka, as well as the next committee AD 2019, which we had an opportunity to supervise: Aleksander Molak, Karolina Baranowska, Ada Kochlewskia, and Martyna Plomecka. AoN Brainhack Warsaw was the first major event I ever organized but my peers made it very easy: it was pleasure to work in this team from the first day till the last day, we had a really good vibe, high synergy, and everything went really smoothly in the end. Also, this event opened many doors for me and allowed me to make many new contacts useful for the future.

Furthermore, I would like to thank the OHBM Student and Postdoc Special Interest Group (a.k.a. OHBM Trainees). I was a part of this wonderful committee, serving as Career Development and Mentorship Manager 2017-2019. During my PhD, I had multiple ups and downs, and sometimes I felt that it was very hard to find like-minded people in my daily environment. In the OHBM Student and Postdoc Special Interest Group however, I found wonderful people with whom I shared common aims and goals. We were all restless to improve the quality of life for early career researchers as far as we could with the resources we had. I believe that we got way further than expected. I hope that we will stay in touch for long years after we retire from this council, and that I will encounter this sort of people in academia more often. Lastly, I would like to add that I could not imagine a better platform to work on mentoring, but also to learn about leadership and teamwork. Therefore, I would like to cordially thank Ayaka Ando, Amanpreet Badhwar, and Michele Veldsman, three wonderful and natural-born leaders, but also to all the other members of the committee: Alex Barnett, Chiara Cadinelli, Heidi Foo, Mengxia Gao, Shabnam Hakimi, Amelie Haugg, Leanna Hernandez, Lia Maria Hocke, Kaori Ito, Christian La, Meena Makary, Navot Naor, Aki Nikolaidis, Marzia Scelsi, and Shruti Vij. Working with this council was also so pleasant because of the energetic and cooperative OHBM Executive Staff: Stephanie McGuire, Kayla Stidger and JoAnn Taie as well as our OHBM council liaison, Shella Keilholtz.

I would also like to thank my colleagues from the OHBM Fundraising Committee: Kayla Stidger, Michael Mullaly, Jessica Turner, Abhinav Yadav, Alan Evans, Arthur Toga, Lilianne Mujica-Parodi, Alain Dagher, and M S Zobae. I learned a lot about fundraising from you guys!

In this place, I would also like to thank other persons who do not belong to the committee but with whom I share my interests in mentoring and career development: Casper Albers, Isabelle Augenstein, Chloe Azencott, Lejla Batina, Gerda Bortsova, Aidan Budd, Veronika Cheplygina, Remco van der Drift, Michael Ekstrand, Ece ERCAN, Lynda Hardman, Arjan haring, Felienne Hermans, Matthias Hirth, Stephan heunis, Peder isager, Chris Madan, Mariya Manahova, Daniël lakens, Magda Paschali, Annika reinke, Emma Claire robinson, Raghavendra selvan, Ionica smeets, Chris snijders, Bettina speckmann, Melanie stefan, Kees storm, Tristan de waard, and Kirstie whitaker. In particular, I would like to thank Veronika for introducing me to her own PixelDot community, where I could
meet more people active in the area of self-development in academia and start new projects together - and for helping me in personal development towards what I really love doing. She also taught me invaluable lessons about valuing my time and assertiveness.

I would like to thank Kayson Fakhar, Lisa Schmors, and Dominic Gonischorek, three young and smart researchers with whom I had great discussions about causality.

I would like to thank two teams with whom I had pleasure to work with at Brainhacks. Firstly, I would like to thank the team with whom I pursued project \textit{Functional connectivity: can we find the common ground?} at Brainhack Warsaw 2017: Malgosia Wierza, Grzegorz Link, Paulina Dąbrowska, Agnieszka Dębska, Aleksander Molak, Anna Alińska, Marcin Koculak, Çağdaş Topçu, Agnieszka Tymorek, Juan Camilo, Emilia Kolada, Anastasia Neklyudova and Monika Derda. Secondly, I would like to thank the team with whom I pursued the project \textit{Trends in the resting brain} at Brainhack Networks 2018: Onerva Kohonen, Greg Koehler, Fabrizio Damicelli, Sarah Morgan and Pouya Ghaemmaghami. It was great working with you guys!

I would like to thank Alessandro Crimi and Martyna Plomecka for inviting me to speak at Brainhack Zurich 2019. It was a great opportunity to come over and talk about the post-PhD career tracks!

I would also like to thank the \textit{eLife Ambassadors} community, especially Kora Korzec who gave me a shot to serve as an eLife Associate in the \textit{eLife Community Ambassadors} programme 2019-2020, working with 243 fresh eLife ambassadors on a broad range of topics concerning daily academic life. I would also like to thank other members of the board of the programme: Lotte de Winde, Tracey Weissgerber, and Mark Patterson, and to other associates I was working with: Adriana Bankston, Samantha Hindle, Daniela Saderi, Huajin Wang, and Steven Burgess.

I would also like to thank Stephan Heunis for inviting me for speaking at the \textit{Open MR Benelux} events, Leiden 2019 and Nijmegen 2020, and to people whom I could discuss with at these events, Daniele Marinazzo, Veronika Cheplygina, Gilles de Hollander, Tim van Mourik, and Matthan Caan. I would also like to thank the Open MR community launched and managed by Stephan for keeping this awesome initiative going. I was sad that I did not have any more capacity to get on board but I am sure that Open MR Benelux will flourish!

Furthermore, I would like to thank Cristiano Micheli for inviting me to participate in \textit{Hackademia}, which was an initiative on the edge of science and entrepreneurship oriented at organizing student team projects and executing projects proposed by startups. I would also like to thank the team members on this project: Ian Cameron, Mats van Ees, Tim van Mourik, Robert Oostenveld, Eelke Spaak, and Nick Stracke. Even though in the end, this project had to be put on hold because of the extenuating amount of work and low exposure, I learned a lot about what startups expect from employees from this project.
Last but not the least, I would like to cordially thank Veronika Cheplygina, Aidan Budd, Stephan Heunis, and Malvika Sharan, with whom we coorganized a 5-day long retreat aimed for promoting open science, mentoring, and inclusivity in science, entitled Avengers for Better Science. It was great working on this with you guys!

The Donders FAD tradition
Working on abstract and/or complicated research projects, and going through the long and exhaustive publication process, would not be possible without little moments of relax. I am glad that in the Donders community, there is a tradition of Friday Afternoon Drinks at Cultuur Cafe and whereabouts as it motivates to stay tuned and develop social circles. I would like to thank all the wonderful people I met by that occasion, among others: Paul Anderson, Erika Atutxa, Remco Bastiaanet, Ruben van Bergen, Ruud Berkers, Mirjam Bloemendaal, Miriam de Boer, Ricarda Braukmann, Sybrine Bultena, Ian Cameron, Nadine Dijkstra, Iris Van Dongen, Eelco Van Dongen, Laurence O’Dwyer, Achiel Fenneman, Lisa Genzel, Daniel Gomez, Corina Greven, Leslie Guadron, Chao Guo, Tjerk Gutteling, Simon Jan Hazenberg, Inge Huijsmans, Ronald Janssen, Marpessa Jennifer, Bahadir Kasap, Alina Lartseva, Ashley Lewis, Lieneke Janssen, Lieke van Lieshout, Charl Linssen, Verena Ly, Ili Ma, Thom Marshall, Jeff Martin, Cecilia Mazzetti, Pim Mostert, Jeanette Mostert, Tim van Mourik, Lisa Morano, Jill Naaijen, Tobias Navarro Schroeders, Annelies Van Nuland, Sasha Ondobaka, Antonella Pomante, Catalina Ratala, Andrew Reid, Daniel von Rhein, Joost Rommers, Thomas Rhys Marshall, René Scheeringa, Sanne Schoemakers, Tim Schrooder, Dan Sharoh, Daniel Sheppard, Matthias Sjerps, Eelke Spaak, Laura Speed, Desiree Spronk, Atsuko Takashima, Lara Todorova, Anna Tyborowska, Marvin Uhlmann, Flora Vanlangendonck, Samartha Varma, Susanne Vogel, Katharina Vogt, Linda van Voogd, Johanna de Vos, Isabella Wagner, Mónica Wagner, Marije ter Wal, Emma Ward, Christina Young, Lorijn Zaadnoordijk, Wei Zhang, Xiaochen Zheng, and Fenny Zwart. It was always a good start of the weekend to spend Friday evening chatting with you!

Friends
Firstly, I would like to thank some friends whom I found here in the Donders. Thanks to Thomas Wolfers for our occasional gratitude walks in Goffert. I adapted this habit and I used to do this more often ever since. I learned how to enjoy little things from you! I would also like to thank Joanna Widomska, who helped me to stay cool in hard times I had during my final years of PhD. I really enjoyed goofing around together at Ibiza and in Orlando, and all the other little moments spent together. Thank you for teaching me how to bake a real Polish bread! I would also like to thank Maria Łojowska, a very hard working lady who was my pace-maker during my final year. Next, I would like to thank Lara Todorova with whom we dropped a few events, and spent countless Saturdays working hand in hand. And, Alina Lartseva, with whom we kept on dancing, dancing, and dancing, and who taught me a lot of ugly truth about the Dutch mentality and customs. I would also like to thank Joanna Sierpowska for our chats about ups and downs of academic life. Thank you for your advice
Asia! I would like to thank Max Hinne for helping me with translating the deed of formation for my foundation, and for all the valuable career advice. Thanks to Piray Atsak for our regular late night sessions by a glass of wine and discussions about life and whereabouts. Sorry for falling asleep at random moments! Also, cordial thanks to Veronika Cheplygina and Mattias Hanson for their advice and help. Thank you guys for working on my self-confidence, and for your warm words in hard times!

Second, I would like to thank my other friends in neuroscience, who might be geographically distant but I enjoyed meeting them at conferences and events, and I am glad that the friendships survived over time. They were my fellow sufferers for all this time, sharing their process of growth with me. So, many thanks to Krzysztof Bielski, Jacob Billings, Daniel Borek, Dorottya Cserpán, Adela Desowska, Ann Duan, Astrid Keibler, Birgit Kriener, Milica Markovic, Maciej Jedynak, Rodrigo Ochos, Foteini Protopapa, Janir Ramos, Marcos Socram, Ghaith Tarawneh, and Anne Thoene.

Lastly, I would like to thank my friends from Poland. I still stay in touch with a number of people (mostly physicists and computer scientists) whom I met during my Masters studies in Warsaw. Let’s be clear, physicists are the most intelligent people on Earth, and I feel lucky being able to get a lifeline in necessary, and make some use of their intelligence. So, many thanks to Radek Chrapkiewicz, Marcin Grzybowski, Grzegorz Link, Alicja Luszcak, Piotr Migdal, Marcin and Michal Kotowski, Michal and Agnieszka Krych, Marcin Pomorski, Sebastian Wojtysiak, Krzysztof Zielieniewski, and Bartłomiej Szczygiel. Many thanks also to Anna Stachura, my good old friend from high school and a doctor of medicine. Last but not least, thanks to my best friend and one of the most intelligent people I ever met, Iwona Mochol, who is relentlessly pulling me up ever since we first met in high school.

Mentors

Further, I would like to thank a number of mentors who were guiding me through the process. Tansu Celikel became my mentor in the new mentorship system of the Donders Centre for Cognitive Neuroscience halfway my PhD, and we had a few good discussions on the philosophy of research and becoming an independent researcher. Thanks also to Thomas Nichols who was mentoring me online within the International Mentorship Programme by the OHBM Student and Postdoc Special Interest Group. Thomas has a lot of distance towards his own career, might be because he requalified from another profession (i.e., a photographer), and he gave me a good example on how to brand yourself in a new environment.

Also, many thanks to my teaching couch, Thom Oostendorp who was preparing me for the Basiskwalificatie in Onderzoekstages (BKO) certificate and instructed me how to supervise students properly.

Thanks to Adeel Razi, Monash University and Julien Dubois, California Institute of Technology for organizing a session at Organization for Human Brain Mapping annual meeting in Rome together with me. Even though our proposal was eventually not selected for the OHBM annual meeting 2019,
I have a feeling that this is not the last time we are on to something together.

I would like to thank Harold Bekkering for motivating me to wrap up this Thesis. Furthermore, thanks Abdel Nemri, Martha Havenith-Nemri and Piray Atsak, who gave me a lot of advice about surviving and progressing in science. I would also like to thank Thom Rhys-Marshall for occasionally appearing on the horizon, and sharing random wisdoms.

Thanks to Ewa Roszak, my long-lasting friend and a professional couch who was always on the line for me. I would also like to thank Caroline Pakel who mentored me through Skype and shared advice on how I should work on my sense of self-value as a professional, and how to negotiate for myself.

Furthermore, I would like to thank a few people who gave me a huge credit of trust, and allowed me to join their professional networks. Firstly, I would like to thank Daniele Marinazzo who invited me to join the expert team at the Brainhack Networks event in Paris, June 2018. This was the first event I could contribute to, that was organized in a group of senior researchers. For me personally, it was very important career-wise as it gave me an opportunity for making a transition between organizing "junior events" organized by PhD students and dedicated to PhD students, to "senior events" that include senior faculty and are addressed to all the research community. I would also like to thank Joaquin Goñi who has let me stay on the team remotely at times when I couldn’t afford to travel to the second edition of Brainhack Networks. Thank you for putting that trust in me, and I hope I stood up to the challenge!

Secondly, I would like to thank Daniel Lurie who invited me to work on his collaborative project on the resting state dynamics in fMRI studies, together with Dan Kessler and Vince Calhoun. It as a very important step for me because it was the first major external collaboration for me, and an opportunity to work on an important, fundamental topic with many researchers influential in the field. Sometimes such a lucky strike is a turning point career-wise, and I am very happy that I could take part in this particular project. Thank you for giving me that shot Daniel!

Lastly, I would like to thank Urszula Forys and Joanna Rączaszek-Leonardi for supporting my career by inviting me to give seminars about my PhD work at the Faculty of Mathematics and Faculty of Psychology, University of Warsaw, respectively. It was a great opportunity for me to do networking, to refresh old contacts, and launch new collaborations.

Tenants and neighbors
Thanks also to my tenants. When I was starting the PhD, I was very invested in traveling. But then, at some point, I started feeling more and more connected to Nijmegen, and travelled less and less often. I also realized that movement is relative; you do not need to go travel the world if the world can come to your house. So, many thanks to: Quentin Aubagnac (France), Benjamin Cambronne (France), Everald Suensberg (Netherlands), Pierre
I would also like to thank Roch Kramer who was living in my house in the spring of 2018, and who helped me a bunch in doing a makeover in the house, and to his wise and caring girlfriend, the adorable Ingrid van Uffelen. Thank you for the lessons of Dutch guys!

I would like to also acknowledge my neighbours, Sjef Bink, Lies Bink, Jos Meeuwsen, and Lucienne Meeuwsen, for being so patient and helpful. In my fourth year of studies, I was probably the worst neighbour in the whole street, as I was sitting either at work or at spa till midnight pretty much every day, and I was not engaged in any social life in this little community at all. Gardening is very important to the Dutch people, and my neighbours originally expected me to meet the standards and take care of my garden just like everybody else. But then, at some point, they realized that I am such a poor gardener, and I come back from work so wasted every day that I barely stand on my feet. So they grabbed an electric saw and started trimming my hedge by themselves. "Go to sleep, kid. We will do it" - they said.

Other kind people around
I would like to thank people engaged in the foundation I am leading, Stichting Solaris Onderzoek en Ontwikkeling: Stephan Heunis, Lara Todorova, Alicja Luszczak, Aman Badhwar, Daniel Borek, Patrick Britz, Carsten Bundt, Indranath Chatterjee, Chris Hartgerink, Ali Khatibi, Xenia Kobeleva, Charl Linssen, Daniele Marinazzo, Adeel Razi, Mozhan Soltani, Sonal Sengupta, and Piray Atsak. Thank you for your support and hard work!

I would like to thank all these people who shared information with me, who helped me with networking and looking for jobs, and who were overall supportive to me - especially in times when my PhD contract already expired, and I was looking for my way for the future. Especially, I would like to thank people I have met in the blockchain circles. The blockchain industry is very inclusive, and in many ways, it resembles neuroscience: it
is a Babel tower of sorts: a mixture of people representing multiple professions, all from software engineers, through business developers, to evangelists, influencers, community managers, and bloggers. At the blockchain conferences, I met a number of kind people who taught me about the blockchain industry, but also about entrepreneurship, team work and IT industry in general. I would like to thank Jana Petkanic and Luuk Weber from the Blockchain Talks team. It is very nice to work with you guys, and thank you for sharing what you know about how this industry works! I would also like to give my words of acknowledgment to Richard Kohl for encouraging me to write about crypto. I would not even think that I can do it if Richard did not advise me to try! I would like to thank people I’ve met in the community: Eleonore Blanc, Justin Bons, Olivier de Jong, Dirk Kadijk, Bart van Maarseveen, Natalia Nowakowska, Dax Nagtegaal, Ekin Tuna, Wilhelm Roth, David and Adriana Truong, and Hans Vanmechelen. Lastly, I would like to highlight the good influence of Sanne Groeneveld, Alvin Leito, and Hugo Schoenbeck on my activities in the blockchain community and on my well being in general. They are very emphatic and understanding people, and helped me a lot in hard times when I was at the end of my PhD. I would also like to thank the Qnext team led by Javad van Landewijk and Justin Floeter: Christian Martin, Bram de Lange, Luca Spaapen, Yann van Ewijk, Yacine Bourouba and Hoang Nguyen, (in collab with Jelte Veldmeijer, and Hein-Pieter van Braam). It is great working in this team, and Justin’s and Javad’s networking skills are just outstanding! Also, specials thanks to Javad for teaching me how (not to) play black jack!

I would also like to thank the Impakt Tribe team, Marijn Kegel and Jeroen van der Heide, for teaching me about entrepreneurship and about how to build a good team, and develop how to develop your own company! Also, special thanks to Dirk Kadijk who was also great to work with, and whose systematic approach to writing business plans is just outstanding.

I would like to acknowledge the whole crew of my favourite local spa. I started coming there very often at some point, and this place really helped me in my recovery during my sick leave, and in rewiring my impatient mind in general. I came to so many good research ideas in this place! I am also grateful for all these emergency rescue missions every time I was passing out from exhaustion. A very patient team!

I would also like to thank the wonderful people whom I met at Hackerspace Nijmegen: Jimmy Comack, Ed van Dalen, Sjors Gielen, Michiel Klaarwater, Stefan Marsiske, Sjoerd Timmer, Bas Pals, Bas Wolvers, and others. Hackerspace is an amazing initiative in which anyone is welcome to participate in a community where any technical and art-related projects can be executed in a group. I particularly liked attending Hackerspace because people whom I met there, attend the space with pure intentions to put some ideas to life - and without thinking how these ideas will contribute to their CV or their future career opportunities. It was very relaxing, funny, and useful. A great community to participate in! Thank you for your great patience, and for teaching me about electronics, Arduinos, hacker-jargon, and
other wonders. I definitely also learned a lot about the Dutch culture in general, and about the spirit of decentralized communities from you. I would also like to thank Rob Hermens, Wim Hamersma and other members of Stip Nijmegen Oost who taught me something about gardening.

Lastly, I would like to thank the members of Pecunia Causa, a student investment club at Radboud. It was lovely to join the meetings and I learned a lot! So, thanks to Koen Smeets, Rinze Hartman, Martijn van Os, and all the monthly attendees!

Random people met on the way, who made me think
As a poor student, I was mostly traveling on budget and lived cheap, so I have met some number of non-scientific people on the way. Some of these encounters were truly inspirational. I remember Lady Juliet, a middle-age lady running a cheap hostel in San Diego. When I first met her during the SfN conference AD 2013, she was a slave to the owner of the building, and most of her earnings were vanishing every month as rent. Since she was spending her whole day on looking after the hostel and facebooking in silence, I suggested her to run some online business. When I came back to the conference in 2016, she was already a well-off owner of a website specialized in online hostel bookings. I think that was time when I first realized that what I tell people, really matters, and that I can change reality if I want to.

I would also like to thank people who are completely unimpressed by what I am doing for a living. Sometimes it is better to take some perspective in order to stop stressing and start living. When I was in Chile, I was living with a Chilean family from Mauricio Caviedas for a week, and those guys taught me a lesson. I am used to all kinds of signs of interest - like nodding, curious questions and words of appreciation - when I talk about what I am doing in daily life. But these guys were just sitting at the table and chewing the salad in silence, and when I finished talking, they turned the conversation into considerations over the consistency of avocado. I think that was the point when I realized that I am not as important, I won’t save the world and therefore, stressing about my projects does not really make sense.

Miscellaneous
I would also like to thank the person who supported me at a time I was completing my thesis, Mattie Lafleur, who showed a lot of understanding to me. We met at a time when my professional life was really complicated so I greatly appreciated his patience and supportive attitude.

A few months before I submitted my thesis, I started working on my own innovative project - which will hopefully grow in the future! I would like to thank Alina Bielczyk, Rwik Mukhopadhyay, Moazzem Hossen, and Marcin Grzybowski for their engagement in the project, and their efforts in making it work. I would also like to thank Ewa Bielczyk, Alina Bielczyk, Radosw Chrapkiewicz, Cees van Diemen, Sanne Groeneveld, Jeroen van der Heide, Jacob Issa, Olivier de Jong, Dirk Kadijk, Niels Keurentjes, Michal Krych, Mattie Lafleur, Alvin Leito, Grzegorz Link, Alicja Luszczak, Kornel Maczynski, Piotr Migda Iwona Mochol, and Hugo Schoenbeck.
for their constructive comments on the business plan!

Also, many thanks to Peter Mulders for sharing his disc space with me, and saving my opportunity to do research.

Many thanks to Thom Rhys-Marshall, Eelco van Dongen and Verena Ly for bringing me to the hospital and staying with me overnight when I broke my hand in an accident. You guys are the best!

Thanks to Roselyne Chauvin for inviting me for a hike together by the ocean we happened to both visit California around the same time in early 2017. It was a really nice, relaxing day spent with someone who thinks big.

I would also like to thank the TGs at the Donders Centre for Cognitive Neuroscience, who were helping regardless whether they were just having their default walk-in hours or not. Especially, thanks to Marek Tyc, who was helping me through Skype when I was stuck with some issue (on Sunday!) while sitting with laptop in a bar in Argentina.

I would like to thank Rob Groenendaal and Kristina Tsemirkhanava from Mercator Launch for supporting my business and entrepreneurial activities. Thank you so much for your kind words and advice!

Furthermore, many people who influenced me in a positive way over these few years, were in fact individuals whom I never met in person. Maybe I meet some of them one day, who knows! I mean, e.g., the authors of some books I have read during my PhD: David Allen (Getting Things Done: The Art of Stress-Free Productivity), Stephan R. Covey (7 habits of highly effective people), Seth Godin (Linchpin: Are You Indispensable?), Robert Greene (The 33 Strategies of War), David R. Hawkins (Power vs Force), Napolean Hill (The Law of Success in Sixteen Lessons), Spencer Johnson (Who Moved My Cheese?), Kerry Patterson et al. (Crucial conversations tools for talking when stakes are high), Tony Robbins (Awake the Giant Within), Arnold Schwarzenegger (The Total Recall: My Unbelievably True Life Story), Simon Sinek (Start with Why: How Great Leaders Inspire Everyone To Take Action), Brain Tracy (Eat That Frog! 21 ways to stop procrastinating and Get More Done in Less Time), Emilie Wapnick (How to Be Everything - A Guide for Those Who Still Don’t Know What They Want to Be When They Grow Up), Zig Ziglar (See You at the Top). I would also like to personally thank the authors of a few YouTube channels that were particularly rich in self-development content to me: Evan Carmichael, Patrick Bet David, Alex Ikonn, Krzysztof Król, and Mel Robbins.

Lastly, during my PhD, I learned a lot about science from George R. R. Martin’s books The songs of Ice and Fire. I do not mean this in an ironic way (namely, I don’t mean that science is nothing else by the game of thrones). In fact, what I learned from the books (and from the show) is that, in most instances, the origin of the conflicts between people is that there is not enough resources for everyone, and that the different parties have some conflicting interests. So, they take actions which might be perfectly valid from their own perspective, but are detrimental to others.
Family circles and surroundings
Thanks to Ewa Bielczyk-Maczyńska, my little sister who - without any apparent effort - graduated from University of Cambridge grad school and moved to Stanford as if it was all peanuts. She is a great inspiration for me, and I use to think: "if she can do it, I can do it too." And that helps.

Lastly, cordial thanks to my Mum, Alina Bielczyk, who was sharing her 30-year experience in corporate life with me, always advising me down the line. I have no idea how much time we spent on Skype discussing my PhD - but it were hundreds of hours for sure. I must say that surprisingly, a large portion of my Mum’s wisdom holds in academia. She is also very patient and persistent, a hard-worker and an optimist at the same time. I am grateful that she was always standing by my side, even in the darkest times when I was down and deep in trouble, ambivalent about how to proceed with my life and whether to even pursue science at all. She is just always there, no matter what. Some might say that this is what mothers are for. But honestly speaking, I think no one else has a mom quite like mine, sorry! Also, I would like to have as much energy as she does - she is the most hard working person I have ever met.