

FREQUENCY-RESOLVED DYNAMIC FUNCTIONAL CONNECTIVITY  
AND SCALE STABILITY OF CONNECTIVITY-STATES



Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften  
(Dr. rer. nat.) der Fakultät für Physik  
der Universität Regensburg

vorgelegt von

MARKUS GOLDHACKER aus

Regensburg

im Jahr 2016

Markus Goldhacker: *Frequency-resolved dynamic functional connectivity and scale stability of connectivity-states.*

© 18. August 2016

PROMOTIONSGESUCH EINGEREICHT AM: 24.04.2014

DIE ARBEIT WURDE ANGELEITET VON:

Prof. Dr. E. W. Lang (doctoral advisor)

Prof. Dr. M. W. Greenlee (co-supervisor)

PRÜFUNGSAUSSCHUSS:

Vorsitzender: Prof. Dr. A. Schäfer

1. Gutachter: Prof. Dr. E. Lang

2. Gutachter: Prof. Dr. M. W. Greenlee

Weiterer Prüfer: Prof. Dr. F. Evers

TERMIN PROMOTIONSKOLLOQUIUM: 16.01.2017

DIE ARBEIT WURDE EINGEREICHT AM: 18. August 2016

This thesis was typeset using the  $\LaTeX$  typesetting system originally developed by Leslie Lamport, based on  $\TeX$  created by Donald Knuth. The layout template `classicthesis` was provided by André Miede [www.miede.de](http://www.miede.de), whom I owe a postcard now. For the love of  $\LaTeX$ .

*Dedicated to my parents*

*"They follow the forms [...],  
but they don't get any laws."*

(Richard P. Feynman – on social sciences)

# CONTENTS

---

<b>i</b>	<b>INTRODUCTION</b>	<b>1</b>
<b>1</b>	<b>MATHEMATICAL CONCEPTS</b>	<b>3</b>
1.1	Principle component analysis . . . . .	3
1.2	Independent component analysis . . . . .	5
1.2.1	Infomax . . . . .	9
1.3	Group independent component analysis . . . . .	10
1.4	k-means clustering . . . . .	11
1.5	Empirical mode decomposition . . . . .	12
1.5.1	Plain empirical mode decomposition . . . . .	13
1.5.2	Ensemble empirical mode decomposition . . . . .	15
1.5.3	Multivariate empirical mode decomposition and its noise-assisted extension . . . . .	16
1.5.4	Hilbert-Huang transformation . . . . .	22
1.6	Butterworth filter . . . . .	22
<b>2</b>	<b>FUNCTIONAL MAGNETIC RESONANCE IMAGING AND THE CON- CEPT OF RESTING STATE</b>	<b>27</b>
2.1	Magnetic resonance imaging . . . . .	27
2.2	Functional magnetic resonance imaging . . . . .	36
2.2.1	Preprocessing of fMRI data . . . . .	39
2.3	Resting-state . . . . .	40
2.3.1	Basis of resting state fMRI signal . . . . .	40
<b>3</b>	<b>CONNECTOMICS</b>	<b>43</b>
3.1	Structural connectivity . . . . .	44
3.2	Static functional connectivity . . . . .	45
3.3	Dynamic functional connectivity . . . . .	48
3.4	Connectivity-states . . . . .	50
<b>ii</b>	<b>FREQUENCY-RESOLVED DYNAMIC FUNCTIONAL CONNECTIV- ITY</b>	<b>53</b>
<b>4</b>	<b>DATA SET AND EXTRACTION OF TIME COURSES</b>	<b>55</b>
4.1	Data-set . . . . .	55
4.2	Extracting time courses using group ICA . . . . .	56
4.3	Frequency-resolved time scales . . . . .	58
4.3.1	MEMD . . . . .	58
4.3.2	Butterworth filter banks . . . . .	60
<b>5</b>	<b>FREQUENCY-RESOLVED DYNAMIC FUNCTIONAL CONNECTIV- ITY</b>	<b>65</b>
5.1	Standard dynamic functional connectivity on the Human Con- nectome Project data . . . . .	65
5.1.1	Choosing a boxcar function as window shape for slid- ing window procedure . . . . .	67

5.2	Frequency-resolved dynamic functional connectivity by means of MEMD . . . . .	67
5.2.1	Robustness of connectivity-states over frequency scales	68
5.2.2	Variability of scale stability . . . . .	70
5.2.3	Null-model tests . . . . .	75
5.2.4	Frequency dependent window size . . . . .	75
5.3	Filter banks and dynamic functional connectivity . . . . .	78
5.4	Simulated dynamic functional connectivity . . . . .	79
5.5	On the GLASSO procedure . . . . .	87
iii	<b>DISCUSSION, OUTLOOK, AND CONCLUSION</b>	93
6	<b>DISCUSSION</b>	95
6.1	Dilemma of objective connectivity-state selection . . . . .	95
6.1.1	Generalizing the idea of selecting the number of clusters using frequency-resolved information . . . . .	98
6.2	Filter banks, MEMD, and self-similarity . . . . .	99
6.3	Constant or adapted window size . . . . .	99
6.4	Limitations and future directions . . . . .	100
7	<b>DYNAMIC BACKBONE SIMULATIONS</b>	103
7.1	Introduction . . . . .	103
7.2	Model definitions . . . . .	104
7.2.1	Ising-like model without external magnetic field term	104
7.2.2	Ising-like model with external magnetic field term .	104
7.2.3	Spin-glass model . . . . .	105
7.3	Dynamic functional connectivity as a substrate . . . . .	105
7.3.1	Quasi static backbone . . . . .	105
7.3.2	Dynamic backbone . . . . .	106
7.3.3	Frequency-resolved dynamic backbone . . . . .	107
7.3.4	Combination of backbones of different modalities . .	108
7.3.5	In detail considerations . . . . .	108
7.4	Effects of defects in the connectome backbone . . . . .	109
7.5	Variation of the Ising-model . . . . .	110
8	<b>CONCLUSION</b>	113
A	<b>APPENDIX</b>	117
A.1	Various mathematical concepts . . . . .	117
A.1.1	Pearson's correlation coefficient . . . . .	117
A.1.2	Partial correlation . . . . .	117
A.1.3	Mutual information . . . . .	118
A.2	Connectivity-state arrays of $k = 2, 3, 5, \dots, 10$ . . . . .	119
A.3	Exemplary connectivity-state array: demeaned . . . . .	127
A.4	Exemplary connectivity-state arrays: surrogate . . . . .	128
	<b>BIBLIOGRAPHY</b>	131

## LIST OF FIGURES

Figure 1	Math. concepts: ICA . . . . .	8
Figure 2	Math. concepts: EMD . . . . .	15
Figure 3	Math. concepts: 3D Hammersley sequence . . . . .	18
Figure 4	Math. concepts: 2D Hammersley sequence . . . . .	19
Figure 5	Math. concepts: ensemble noise-assisted MEMD . . . . .	21
Figure 6	Math. concepts: Butterworth filter . . . . .	26
Figure 7	MRI: $T_2$ relaxation . . . . .	30
Figure 8	MRI: spin echo . . . . .	31
Figure 9	MRI: frequency encoding . . . . .	33
Figure 10	MRI: spin echo sequence . . . . .	34
Figure 11	MRI: k-space . . . . .	35
Figure 12	MRI: EPI sequence . . . . .	38
Figure 13	Connectomics: structural connectivity . . . . .	45
Figure 14	Connectomics: functional connectivity . . . . .	46
Figure 15	Connectomics: dynamic functional connectivity . . . . .	48
Figure 16	Connectomics: dFC pipeline . . . . .	49
Figure 17	Connectomics: connectivity-states . . . . .	52
Figure 18	Extracting time courses: dynamic range and LF to HF ratio – theory . . . . .	57
Figure 19	Extracting time courses: dynamic range and LF to HF ratio – result . . . . .	57
Figure 20	Extracting time courses: frdFC pipeline . . . . .	59
Figure 21	Extracting time courses: time course characteristics – MEMD . . . . .	61
Figure 22	Extracting time courses: covariance structure IMFs . . . . .	62
Figure 23	Extracting time courses: time course characteristics – filter bank . . . . .	63
Figure 24	frdFC: tapered window . . . . .	67
Figure 25	frdFC: connectivity-state array $k = 4$ – MEMD . . . . .	71
Figure 26	frdFC: theoretical evolution of $\langle I_{\Delta f}^{sim} \rangle(k)$ . . . . .	73
Figure 27	frdFC: scale stability – MEMD . . . . .	74
Figure 28	frdFC: scale stability – null-models . . . . .	76
Figure 29	frdFC: adjusted window size . . . . .	77
Figure 30	frdFC: scale stability – adjusted windows . . . . .	78
Figure 31	frdFC: scale stability – filter banks . . . . .	80
Figure 32	frdFC: scale stability – selected simulations . . . . .	81
Figure 33	frdFC: scale stability – simulation $k_{inh} = 2$ . . . . .	82
Figure 34	frdFC: scale stability – simulation $k_{inh} = 4$ . . . . .	83
Figure 35	frdFC: scale stability – simulation $k_{inh} = 6$ . . . . .	84
Figure 36	frdFC: scale stability – simulation $k_{inh} = 8$ . . . . .	85
Figure 37	frdFC: scale stability – simulation $k_{inh} = 10$ . . . . .	86

Figure 38	frdFC: scale stability – simulation $k_{inh} = 4$ filter bank . . . . .	88
Figure 39	frdFC: GLASSO validation – Fisher transformed . . .	89
Figure 40	frdFC: scale stability – MEMD – non-GLASSO . . .	90
Figure 41	frdFC: scale stability – selected simulations – non-GLASSO . . . . .	91
Figure 42	frdFC: GLASSO validation – non-Fisher transformed	92
Figure 43	frdFC: scale stability – selected simulations – non-GLASSO – non-Fisher . . . . .	92
Figure 44	Discussion: cluster validity 1 . . . . .	96
Figure 45	Discussion: cluster validity 2 . . . . .	97
Figure 46	Discussion: cluster validity 3 . . . . .	97
Figure 47	Discussion: cluster validity 4 . . . . .	98
Figure 48	Dynamic backbone simulations: quasistatic backbone	106
Figure 49	Dynamic backbone simulations: dynamic backbone	107
Figure 50	Dynamic backbone simulations: defects in functional connectomes . . . . .	110
Figure 51	Dynamic backbone simulations: spreading dynamics on the human connectome . . . . .	111
Figure 52	Dynamic backbone simulations: model selection for Ising-model variation . . . . .	112
Figure 53	Appendix: connectivity-state array $k = 2$ – MEMD	119
Figure 54	Appendix: connectivity-state array $k = 3$ – MEMD	120
Figure 55	Appendix: connectivity-state array $k = 5$ – MEMD	121
Figure 56	Appendix: connectivity-state array $k = 6$ – MEMD	122
Figure 57	Appendix: connectivity-state array $k = 7$ – MEMD	123
Figure 58	Appendix: connectivity-state array $k = 8$ – MEMD	124
Figure 59	Appendix: connectivity-state array $k = 9$ – MEMD	125
Figure 60	Appendix: connectivity-state array $k = 10$ – MEMD	126
Figure 61	Appendix: connectivity-state array $k = 4$ – MEMD demeaned . . . . .	127
Figure 62	Appendix: connectivity-state array $k = 4$ – shuffled	128
Figure 63	Appendix: connectivity-state array $k = 4$ – phase randomized . . . . .	129

## LIST OF TABLES

---

Table 1	Extracting time courses: filter order adjusted Butterworth filter bank . . . . .	62
---------	--	----

## ACRONYMS

---

<b>BOLD</b>	blood-oxygenation-level dependant
<b>dFC</b>	dynamic functional connectivity
<b>DMN</b>	default mode network
<b>EEG</b>	electroencephalography
<b>EMD</b>	empirical mode decomposition
<b>EPI</b>	echo planar imaging
<b>FID</b>	free induction decay
<b>frdFC</b>	frequency-resolved dynamic functional connectivity
<b>gICA</b>	group independent component analysis
<b>HHT</b>	Hilbert-Huang transformation
<b>IC</b>	independent component
<b>ICA</b>	independent component analysis
<b>IMF</b>	intrinsic mode function
<b>fMRI</b>	functional magnetic resonance imaging
<b>MEMD</b>	multivariate empirical mode decomposition
<b>MRI</b>	magnetic resonance imaging
<b>PCA</b>	principle component analysis
<b>ROI</b>	region of interest
<b>rs-fMRI</b>	resting state functional magnetic resonance imaging
<b>RSN</b>	resting state network
<b>TE</b>	echo time
<b>TR</b>	repetition time



## Part I

### INTRODUCTION

In this introductory part, I introduce the theoretical background and state-of-the art. The text is written to be comprehensible without consulting secondary literature. In the first chapter (ch. 1), I describe the mathematics, machine learning and data-driven concepts. This includes the description of the matrix factorization techniques principal component analysis (PCA) (sec. 1.1), independent component analysis (ICA) (sec. 1.2), and its group extension (sec. 1.3). The next section describes clustering procedures focusing on the k-means algorithm (sec. 1.4). Those methods represent mostly feature extraction and dimension reduction. The following section deals with empirical mode decomposition (EMD) and its extensions (sec. 1.5) as means of data-driven mode extraction from uni- and multivariate time-series. In the next chapter (ch. 2), functional magnetic resonance imaging (fMRI) as the data generating method (sec. 2.2) and resting state functional magnetic resonance imaging (rs-fMRI) as the data generating paradigm (sec. 2.3), are described. The next chapter (ch. 3) elaborates on the state-of-the art in terms of connectomics reaching from structural (sec. 3.1), to static (sec. 3.2), and dynamic functional connectivity (dFC) (sec. 3.3). This cumulates into the introduction of connectivity-states (sec. 3.4), which requires most of the introduced mathematical concepts.



*"Physics is to mathematics  
what sex is to masturbation."  
(Richard P. Feynman)*

### 1.1 PRINCIPLE COMPONENT ANALYSIS

In this section, I describe the concept of PCA relying on [54, 58]. PCA is a data reduction technique represented as a linear model. For its usage there has to be redundancy in the data under consideration. This redundancy has to be reflected in terms of correlations, since the transformed data is uncorrelated after the application of PCA. The starting point for PCA is a number of random variables representing the data. Let  $\mathbf{x}$  be an  $N$ -dimensional random vector, which represents our data. In general, vectors are represented as column vectors in my thesis. The first step is to subtract the mean of each random variable  $x_i$  with  $i = 1, \dots, N$  in  $\mathbf{x}$ . The linear transformation, or rather projection  $\mathbf{y}$ , of  $\mathbf{x}$  is represented by

$$\mathbf{y} = \mathbf{w}^T \cdot \mathbf{x}, \quad (1)$$

where  $\mathbf{w}$  is the so-called weight vector of  $\mathbf{x}$ . The PCA criterion to reach uncorrelated data  $\mathbf{x}$  is to successively maximize the variance  $\sigma^2(\mathbf{y})$  of projections  $\mathbf{y}$ . The data  $\mathbf{x}$  is projected onto a rotated, orthogonal coordinate system that PCA derives from the data with respect to the original one. In other words, the projection of  $\mathbf{x}$  onto the first axis of the transformed coordinate system has to have maximum variance, the projection onto the second axis the second largest variance, and so on. Thus, we can formulate the PCA criterion for the first projection  $y_1$  of the data by using the cost function

$$\Phi_{PCA} = \sigma^2(y_1). \quad (2)$$

This function can be further rearranged by using the linear representation 1 and the statistical momentum notation of variance to

$$\Phi_{PCA} = \sigma^2(\mathbf{w}_1^T \cdot \mathbf{x}) = \mathbb{E} [(\mathbf{w}_1^T \cdot \mathbf{x})^2]. \quad (3)$$

The weight vector  $\mathbf{w}_1$  can be written outside the momentum yielding the notation employing the covariance matrix  $\mathbf{C}$ , which is also the correlation matrix, since the data has been demeaned

$$\Phi_{PCA} = \mathbf{w}_1^T \cdot \mathbb{E}(\mathbf{x} \cdot \mathbf{x}^T) \cdot \mathbf{w}_1 = \mathbf{w}_1^T \cdot \mathbf{C} \cdot \mathbf{w}_1. \quad (4)$$

This term, which has to be maximized, depends on  $\mathbf{w}_1$ . If we do not constrain  $\mathbf{w}_1$ , the maximization depends solely on it, since it can take arbitrary values.

Therefore, the constraint of  $\|\mathbf{w}_1\| = 1$  is chosen. Since this is a constraint for the maximization problem itself, the method of Lagrangian-multipliers is used defining the function

$$\Lambda_1 = \mathbf{w}_1^T \cdot \mathbf{C} \cdot \mathbf{w}_1 - \lambda_1 (\mathbf{w}_1^2 - 1). \quad (5)$$

Since we are looking for the correct choice of  $\mathbf{w}_1$  to maximize the variance of  $y_1$ , the function  $\Lambda_1$  is derivated with respect to  $\mathbf{w}_1$  and set to zero

$$\frac{d\Lambda_1}{d\mathbf{w}_1} = 2\mathbf{C} \cdot \mathbf{w}_1 - 2\lambda_1 \mathbf{w}_1 = 0. \quad (6)$$

For the minuend, the symmetry of  $\mathbf{C}$  was exploited. Further on, this results in an eigenvalue equation

$$\mathbf{C} \cdot \mathbf{w}_1 = \lambda_1 \mathbf{w}_1. \quad (7)$$

Therefore, it can be concluded that the first weighting vector projecting  $\mathbf{x}$  into the new coordinate system is an eigenvector of the covariance matrix  $\mathbf{C}$ . From the original PCA criterion of equation 4 we can derive exploiting the unit norm of  $\mathbf{w}_1$

$$\Phi_{PCA} = \mathbf{w}_1^T \cdot \mathbf{C} \cdot \mathbf{w}_1 = \lambda_1 \mathbf{w}_1^2 = \lambda_1, \quad (8)$$

and see that it is maximized for the largest eigenvalue  $\lambda_1$ . Therefore,  $\mathbf{w}_1$  is the eigenvector corresponding to the largest eigenvalue  $\lambda_1$ .

For the next weighting vector  $\mathbf{w}_2$ , we look for the projection

$$y_2 = \mathbf{w}_2^T \cdot \mathbf{x}, \quad (9)$$

where the variance  $\sigma^2(y_2)$  takes the second largest value possible. Since one aspect of PCA is the reduction of redundancy in the data set, a new constraint comes into play demanding

$$\text{cov}(y_1, y_2) = 0. \quad (10)$$

Exploiting again the demeaning of the data, the momentum notation, and writing the weight vectors outside of the momentum yields

$$\text{cov}(y_1, y_2) = E(y_1 y_2) = E[(\mathbf{w}_1^T \cdot \mathbf{x})(\mathbf{w}_2^T \cdot \mathbf{x})] = \mathbf{w}_1^T \cdot \mathbf{C} \cdot \mathbf{w}_2 = 0. \quad (11)$$

Since we also want to optimize for the maximum variance under the already known constraint of unit norm for our weight vector, the whole maximization problem for the second weighting vector is

$$\Lambda_2 = \mathbf{w}_2^T \cdot \mathbf{C} \cdot \mathbf{w}_2 - \lambda_2 (\mathbf{w}_2^2 - 1) - \gamma (\mathbf{w}_1^T \cdot \mathbf{C} \cdot \mathbf{w}_2). \quad (12)$$

The derivation of this equation with respect to  $\mathbf{w}_2$  and setting to zero yields

$$\frac{d\Lambda_2}{d\mathbf{w}_2} = 2\mathbf{C} \cdot \mathbf{w}_2 - 2\lambda_2 \mathbf{w}_2 - \gamma \mathbf{w}_1^T \cdot \mathbf{C} = 0. \quad (13)$$

By multiplying this equation from the left with  $\mathbf{w}_1$  and using equation 7 yields

$$\frac{d\lambda_2}{d\mathbf{w}_2} = 2\mathbf{w}_1^T \cdot \mathbf{C} \cdot \mathbf{w}_2 - 2\lambda_2 \mathbf{w}_1^T \cdot \mathbf{w}_2 - \gamma \lambda_1 \mathbf{w}_1^2 = 0. \quad (14)$$

Employing all the known constraints, a simple equation results

$$\gamma = 0. \quad (15)$$

Using this in the derivated maximization problem yields another eigenvalue equation

$$\mathbf{C} \cdot \mathbf{w}_2 = \lambda_2 \mathbf{w}_2. \quad (16)$$

Since we want a successive maximization of the variance of the projected data, following the above introduced logic 8 analogously yields the conclusion for  $\lambda_2$  being the second largest eigenvalue with its corresponding eigenvector  $\mathbf{w}_2$ .

For every other projection  $y_k$ , the derivation is analogously with additional constraints of zero covariance with the already projected data

$$\text{cov}(y_k, y_i) = 0, i = 1, \dots, k-1 \quad (17)$$

and unit norm of the weighting vector under consideration  $\|\mathbf{w}_k\| = 1$ . This means, each weighting vector  $\mathbf{w}_k$  is an eigenvector of  $\mathbf{C}$  with  $\lambda_1 \geq \dots \geq \lambda_k \geq \dots \geq \lambda_K$ , where usually  $K < N$ . This information can be condensed in a matrix notation

$$\mathbf{y} = \mathbf{W} \cdot \mathbf{x}, \quad (18)$$

where  $\mathbf{W}$  is the matrix containing the above derived eigenvectors as row vectors in the specified order. If we extend the product on the right-hand side to

$$\mathbf{z} = \mathbf{D}^{-\frac{1}{2}} \cdot \mathbf{W} \cdot \mathbf{x} \quad (19)$$

with  $\mathbf{D}$  being the diagonal matrix having the corresponding eigenvalues on its diagonal, then the entries in  $\mathbf{z}$  are called whitened. This means that the random variables in  $\mathbf{z}$  have unit variance and are mutually uncorrelated. Whitening is often an important preprocessing step for ICA algorithms.

## 1.2 INDEPENDENT COMPONENT ANALYSIS

In this section, I describe ICA relying on Hyvärinen et al. [54]. ICA is a blind source separation approach for extracting source signals from so-called mixtures of those source signals. In the case of fMRI, mixtures are represented by fMRI images. Consider an fMRI experiment with  $T$  time-points and  $N$  voxels sampled at every time point. To apply ICA on such a dataset we have to rearrange the collected data to a matrix  $\mathbf{X}$  with dimensions  $T \times N$ . Every

row corresponds to a whole fMRI image resulting in increasing time as being the row dimension. Column  $i$  depicts the time-evolution of one single voxel. ICA decomposes this data matrix  $\mathbf{X}$  in a mixing matrix  $\mathbf{A}$  and a source matrix  $\mathbf{S}$  in a linear fashion

$$\mathbf{X}_{T \times N} = \mathbf{A}_{T \times L} \cdot \mathbf{S}_{L \times N} \rightarrow \widehat{\mathbf{S}}_{L \times N} = \mathbf{B}_{L \times T} \cdot \mathbf{X}_{T \times N}, \quad (20)$$

where  $\mathbf{B} = \mathbf{A}^{-1}$  is the so-called demixing matrix and  $\widehat{\mathbf{S}}$  are the estimated sources. We have to distinguish between real and estimated sources, since ICA is only capable of reconstructing the underlying sources approximately and, furthermore, since there is an uncertainty concerning the arrangement and scaling of the estimated sources compared to the underlying ones. The arrangement of the data matrix  $\mathbf{X}$  in equation 20 – time as being the first, space the second dimension – results in a so-called spatial ICA. Equation 20 is the general matrix notation of ICA – not only applicable in the context of fMRI. The algorithm decomposes  $\mathbf{X}$  under the constraint of resulting in maximally, statistically independent sources in  $\mathbf{S}$  – the independent components (ICs). Every row in  $\mathbf{S}$  represents a spatial intensity distribution in the same space as the brain images in  $\mathbf{X}$ . These mutually independent intensity distributions can be linearly combined by multiplying the mixing matrix  $\mathbf{A}$  by the source matrix  $\mathbf{S}$ . Consider the first row of  $\mathbf{A}$  being multiplied by every column vector of  $\mathbf{S}$  representing the first row of the data matrix  $\mathbf{X}$ . We can look at the components of the first row vector of  $\mathbf{A}$  as being the weights of the ICs in  $\mathbf{S}$  to represent the first fMRI image in  $\mathbf{X}$ . Following this logic for every row in  $\mathbf{A}$  we end up with the time evolution of the weights necessary to represent the images in  $\mathbf{X}$ . One column vector in  $\mathbf{A}$  is often referred to as the activity of an IC over time.

In general, ICA is a data-driven matrix decomposition method with very few and weak preconditions on the data, or rather the sources underlying the mixtures represented by the data. In the following, we switch to the notation using random variables instead of matrices to be consistent with the notation introduced in section 1.1. To extract sources  $\mathbf{s}$  from mixtures  $\mathbf{y}$ , the single sources  $s_i$  in the random vector  $\mathbf{s}$  have to be statistically independent and non-Gaussian. Statistical independence is given by

$$p(\mathbf{s}) = p(s_1) \dots p(s_L), \quad (21)$$

where  $p(\mathbf{s})$  is the joint probability density of the sources  $\mathbf{s}$  and  $p(s_l)$  the marginal probability density of source  $s_l$ . This means that statistical independence is realized if and only if the joint probability density can be expressed as the product of the marginal densities. Gaussianity on the other hand is represented by the fourth statistical moment, which is called kurtosis

$$\text{kurt}(s_l) = E(s_l^4) - 3 [E(s_l^2)]^2. \quad (22)$$

For the sources to be non-Gaussian, the kurtosis has to fulfill  $\text{kurt}(s_l) \neq 0$ . Actually, a weaker but for ICA sufficient precondition on Gaussianity is that

at most one source in  $\mathbf{s}$  has a Gaussian distribution. In comparison to PCA, ICA is usually not used for dimension reduction. This task is accomplished by PCA used as a preprocessing step before ICA is conducted. It is worth noting that PCA 'only' results in uncorrelated, transformed data, whereby ICA results in the stronger condition of statistically independent, transformed data. Statistical independence implies uncorrelatedness, but not vice versa. Using PCA as whitening procedure before ICA is common, since it reduces the complexity of the problem. Consider a square mixing matrix  $\mathbf{A}$  with  $n$  columns and  $n$  rows. Then the number of components to be calculated by the ICA algorithm is  $n^2$ . It can be shown that this reduces to  $\frac{n(n-1)}{2}$  in the case of an orthogonal matrix. If the data is whitened beforehand, the mixing matrix can be identified as an orthogonal matrix. This can be derived from the covariance matrix of the whitened data  $\mathbf{y}$ , expanding it to the ICA mixing model from equation 20, and combining the matrices

$$\begin{aligned} E(\mathbf{y} \cdot \mathbf{y}^T) &= E\left[\left(\mathbf{D}^{-\frac{1}{2}} \cdot \mathbf{W} \cdot \mathbf{A} \cdot \mathbf{s}\right) \cdot \left(\mathbf{D}^{-\frac{1}{2}} \cdot \mathbf{W} \cdot \mathbf{A} \cdot \mathbf{s}\right)^T\right] = \\ &= E\left[\tilde{\mathbf{A}} \cdot \mathbf{s} \cdot \mathbf{s}^T \cdot \tilde{\mathbf{A}}^T\right]. \end{aligned} \quad (23)$$

Again, it can be shown that  $\tilde{\mathbf{A}}$  can be written outside the momentum. Exploiting the uncorrelatedness of sources  $\mathbf{s}$  resulting from their statistical independence, this yields

$$E\left[\tilde{\mathbf{A}} \cdot \mathbf{s} \cdot \mathbf{s}^T \cdot \tilde{\mathbf{A}}^T\right] = \tilde{\mathbf{A}} \cdot E(\mathbf{s} \cdot \mathbf{s}^T) \cdot \tilde{\mathbf{A}}^T = \tilde{\mathbf{A}} \cdot \tilde{\mathbf{A}}^T = \mathbf{I}. \quad (24)$$

The last step is known, because the covariance matrix of the whitened data  $E(\mathbf{y} \cdot \mathbf{y}^T)$  has to be the identity matrix  $\mathbf{I}$ . Hence, we can see that the mixing matrix of whitened data  $\tilde{\mathbf{A}}$  is an orthogonal matrix.

In a next step, it is of concern to think about a way to evaluate the statistical independence of estimates of sources  $\hat{\mathbf{s}}$ . The concept of statistical independence can be illustrated by toy data drawn from a joint probability distribution of two uniformly, but statistically independent random variables (see box 1). In this case, the marker for statistical independence are the edges of the data cloud being parallel to the coordinate axes and mutually orthogonal. A linear transformation by an mixing matrix  $\mathbf{A}$  yields a parallelogram. From the edges of this parallelogram, the mixing matrix can be derived and the blind source separation problem is solved in this simple case. In this example, two uniformly distributed sources have been used. In real data cases, the solution is more complicated, since we are usually dealing with higher-dimensional data and more complicated probability distributions. There are many ICA algorithms solving this task of estimating the most statistically independent sources from data in various ways. I want to focus on the so-called Infomax-principle, since the analysis of my data is based on this way of applying ICA.

## BOX 1: EXAMPLE ICA

The concept of statistical independence is illustrated by drawing 200 data points from two random variables  $s_1$  and  $s_2$  each with the probability density

$$p(s_i) = \begin{cases} \text{const.}, & \text{if } 0 \leq s_i \leq 1 \\ 0, & \text{else.} \end{cases} \quad (25)$$

Plotting  $(s_1, s_2)$  in a two-dimensional coordinate system results in the plot shown in figure 1(left). Those two random variables are statistically independent – the projections of  $(s_1, s_2)$  onto  $s_1$  do not contain any information about the projections onto  $s_2$ , and vice versa. By multiplying  $\mathbf{s}$  with mixing matrix

$$\mathbf{A} = \begin{pmatrix} 2 & 6 \\ 2 & 2 \end{pmatrix} \quad (26)$$

the scatter plot shown in figure 1(right) results. The mixtures  $x_1$  and  $x_2$  of the statistically independent random variables now convey informations of each other. Using edge detection, it is now possible to decipher the mixing matrix  $\mathbf{A}$  from the transformed data. The columns of  $\mathbf{A}$  are represented by the vectors defined by the edges of the parallelogram. In this simple case, the source separation problem would be solved with  $\hat{\mathbf{s}} = \mathbf{B} \cdot \mathbf{x}$ .

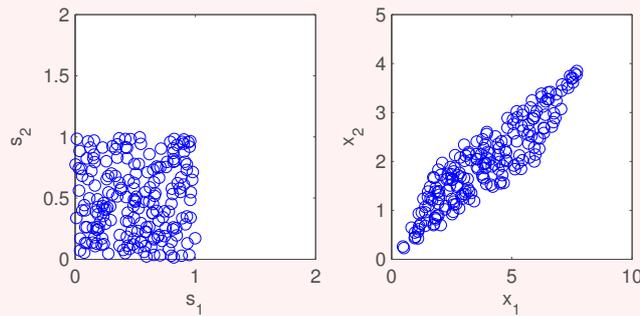


Figure 1: (Left) This panel shows the scatter plot of random vector  $\mathbf{s}$  composed of two uniformly distributed random variables  $s_1$  and  $s_2$ . (Right) Here, the matrix product  $\mathbf{A} \cdot \mathbf{s}$  is depicted.

## 1.2.1 Infomax

To understand the Infomax algorithm, we first have to look at maximum likelihood estimation of the ICA problem. For the mixing model of equation 20 it can be shown that the probability density of the mixtures is

$$p(\mathbf{x}) = |\det \mathbf{B}| \prod_{l=1}^L p_l(s_l), \quad (27)$$

since the  $s_l$  have to be mutually independent. When considering  $T$  independent realizations of  $\mathbf{x}$  as  $\mathbf{x}(t)$ ,  $t = 1, \dots, T$ , then the Likelihood-function can be defined as

$$L(\mathbf{B}) = \prod_{t=1}^T |\det \mathbf{B}| \prod_{l=1}^L p_l(\mathbf{b}_l \cdot \mathbf{x}(t)), \quad (28)$$

with  $\mathbf{B} = (\mathbf{b}_1^T, \dots, \mathbf{b}_L^T)^T$  and  $\mathbf{b}_l$  being the  $l$ -th row vector of  $\mathbf{B}$ . This yields the log-Likelihood function

$$\log L(\mathbf{B}) = T \log |\det \mathbf{B}| + \sum_{t=1}^T \sum_{l=1}^L \log p_l(\mathbf{b}_l \cdot \mathbf{x}(t)). \quad (29)$$

This can also be written as

$$\frac{1}{T} \log L(\mathbf{B}) = \log |\det \mathbf{B}| + E \left[ \sum_{l=1}^L \log p_l(\mathbf{b}_l \cdot \mathbf{x}(t)) \right]. \quad (30)$$

After having estimated the probability densities  $p_l(s_l)$ , a gradient descent method can solve the optimization problem of maximizing  $L(\mathbf{B})$  using a step width of the form

$$\Delta \mathbf{B} \propto \frac{\partial \log L(\mathbf{B})}{\partial \mathbf{B}}. \quad (31)$$

Maximizing  $L(\mathbf{B})$  is then equivalent to finding optimal parameters for the probability densities of the sources to best fit the estimated versions of them, which is the same as optimizing for statistical independence.

The formulation of the Infomax algorithm is based on a neural network

$$\hat{s}_l = f_l(\mathbf{b}_l \cdot \mathbf{x}) + \mathbf{n}, \quad (32)$$

where  $\hat{s}_l$  are estimates of the sources underlying the mixtures  $\mathbf{x}$ , the  $f_l$  are the transfer functions, and  $\mathbf{n}$  is the noise vector representing **white Gaussian noise**. The algorithm maximizes the mutual information  $I(\mathbf{x}; \mathbf{f}(\mathbf{B} \cdot \mathbf{x}))$  between the input units  $\mathbf{x}$  and the output  $\mathbf{f}(\mathbf{B} \cdot \mathbf{x})$ . The mutual information can also be written as (see eq. 110)

$$I(\mathbf{f}(\mathbf{B} \cdot \mathbf{x}); \mathbf{x}) = H(\mathbf{f}(\mathbf{B} \cdot \mathbf{x})) - H(\mathbf{f}(\mathbf{B} \cdot \mathbf{x})|\mathbf{x}). \quad (33)$$

Bell and Sejnowski [12] explain that  $H(X|Y)$  from equation 110 can be seen as the entropy of remaining information not conveyed by the input neurons,

**White Gaussian noise** refers to a temporally uncorrelated stochastic process drawn from an underlying Gaussian distribution that exhibits constant intensities over all frequencies [54].

and hence for equation 32  $H(X|Y) = H(\mathbf{n})$  holds. Bell and Sejnowski [12] mention that the algorithm uses a gradient decent  $\frac{\partial I}{\partial \mathbf{b}_l}$  and since noise  $\mathbf{n}$  is not depending on system inherent parameters the derivation yields

$$\frac{\partial I}{\partial \mathbf{b}_l} = \frac{\partial H(\mathbf{f}(\mathbf{B} \cdot \mathbf{x}))}{\partial \mathbf{b}_l}. \quad (34)$$

Therefore, the maximization of mutual information is equivalent to the maximization of the entropy of the output units in this case. It can be calculated that the entropy of our output units can be written as

$$H(\mathbf{f}(\mathbf{B} \cdot \mathbf{x})) = H(\mathbf{x}) + \log |\det \mathbf{B}| + E \left[ \sum_{l=1}^L \log f'_l(\mathbf{b}_l \cdot \mathbf{x}) \right]. \quad (35)$$

We can see that the last two summands resemble equation 30, if we choose the probability densities for  $f'_l$ . Therefore, the Infomax principle can be reduced to the estimation of the ICA model in the context of maximum likelihood.

### 1.3 GROUP INDEPENDENT COMPONENT ANALYSIS

ICA applied on the single subject level has the problem that comparability of the extracted ICs is not guaranteed. Actually, the scaling and order of the ICs resulting from ICA algorithms is arbitrary. This issue can be tackled by using group independent component analysis (gICA), which allows to make group inferences on ICs. I focus on the gICA approach introduced in Calhoun et al. [19], which I elaborate on in the following section. Since it is an approach introduced for the evaluation of fMRI data, I switch to matrix notation again. First, the authors apply PCA on a single subject level

$$\mathbf{Y}_i = \mathbf{W}_i \cdot \mathbf{X}_i, \quad (36)$$

where – analogously to section 1.1 –  $\mathbf{Y}_i$  is the  $K \times N$ -dimensional transformed, or reduced, version of the  $T \times N$ -dimensional data matrix  $\mathbf{X}_i$  of session  $i$ , and  $\mathbf{W}_i$  is the corresponding matrix of eigenvectors having dimension  $K \times T$ . In a next step, they apply PCA on a group level, which means that the matrices resulting from the single subject reduction step are concatenated in the row dimension yielding

$$\tilde{\mathbf{Y}} = \begin{pmatrix} \mathbf{W}_1 \cdot \mathbf{X}_1 \\ \vdots \\ \mathbf{W}_n \cdot \mathbf{X}_n \end{pmatrix}, \quad (37)$$

where  $\tilde{\mathbf{Y}}$  is the concatenated group matrix of dimension  $nK \times N$  with  $n$  being the number of used sessions. Their approach results in a new, group matrix reduced in the row dimension

$$\tilde{\mathbf{Y}}_G = \mathbf{W}_G \cdot \tilde{\mathbf{Y}} \quad (38)$$

having the dimensionality  $L \times N$  and  $\mathbf{W}_G$  having  $L \times nK$ . In a next step, they set up the linear mixing model of ICA

$$\tilde{\mathbf{Y}}_G = \mathbf{A} \cdot \mathbf{S}, \quad (39)$$

where  $\mathbf{A}$  is the  $L \times L$  mixing matrix and  $\mathbf{S}$  the  $L \times N$  source matrix on the group level. In a next step, the authors partitioned the matrices of the model

$$\begin{pmatrix} \mathbf{W}_1 \cdot \mathbf{X}_1 \\ \vdots \\ \mathbf{W}_n \cdot \mathbf{X}_n \end{pmatrix} = \begin{pmatrix} \mathbf{W}_{G,1}^{-1} \\ \vdots \\ \mathbf{W}_{G,n}^{-1} \end{pmatrix} \cdot \mathbf{A} \cdot \mathbf{S}, \quad (40)$$

where  $\mathbf{W}_{G,i}^{-1}$  is the  $K \times L$ -dimensional  $i$ -th part of the inverse of the group reduction matrix  $\mathbf{W}_G$  corresponding to session  $i$ . By using back reconstruction algorithms described in Erhardt et al. [27], time courses and spatial maps for each session can be estimated separately

$$\mathbf{X}_i = \mathbf{W}_i^{-1} \cdot \mathbf{W}_{G,i}^{-1} \cdot \mathbf{A} \cdot \mathbf{S}_i = \mathbf{T}_i \cdot \mathbf{S}_i, \quad (41)$$

where the matrix  $\mathbf{T}_i$  contains the time courses as column vectors of each corresponding spatial map in  $\mathbf{S}_i$  reflected as row vectors.

#### 1.4 k-MEANS CLUSTERING

Clustering procedures offer means of data reduction and feature extraction. Clustering procedures like the  $k$ -means algorithm are unsupervised learning techniques that assign to each data point of a data set a cluster index. This ideally results in the assignment of each data point to its data inherent cluster. Unsupervised in this case means that the number of extracted clusters has to be predefined, since the data inherent number of clusters is not known beforehand. The representatives of a cluster are called centroids. In the following, I elaborate on the  $k$ -means clustering algorithm and its extension, which is called  $k$ -means++. For the  $k$ -means algorithm, I rely on MacKay [69].

The starting point is a data matrix  $\mathbf{X}$  of dimension  $T \times N$ . The  $k$ -means algorithm groups the  $T$  observations in matrix  $\mathbf{X}$  into  $k$  predefined clusters. Those clusters are represented by the centroids  $\mathbf{c}_z$  with  $z = 1, \dots, k$ . The algorithm has two steps: an assignment step, where all data points are assigned to their nearest centroids, and an update step, in which the new centroids are calculated from the newly assigned data points. In the assignment step, the distances between the centroids  $\mathbf{c}_z$  and each data point  $\mathbf{x}_t$  with

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_t \\ \vdots \\ \mathbf{x}_T \end{pmatrix} \quad (42)$$

are calculated. Afterwards, each data point is assigned to its nearest centroid

$$\hat{z}_t = \arg \min_z (d(\mathbf{c}_z, \mathbf{x}_t)), \quad (43)$$

where  $\hat{z}_t$  represents the assigned cluster index to data point  $\mathbf{x}_t$  and  $d(\mathbf{c}_z, \mathbf{x}_t)$  is the distance between centroid  $\mathbf{c}_z$  and data point  $\mathbf{x}_t$ . Different distance measures can be selected for  $d(\mathbf{c}_z, \mathbf{x}_t)$ . From this assignment of each data point  $\mathbf{x}_t$  to its nearest centroid, the new centroids are calculated. This is done by evaluating the mean of all data points belonging to a certain index  $\hat{z}_t$ . For a mathematical expression of this, we define the so-called cluster membership

$$u(t, z, \hat{z}_t) = \begin{cases} 1, & \text{if } \hat{z}_t = z \\ 0, & \text{else.} \end{cases} \quad (44)$$

By using this quantity, we can express the updating step, or rather calculation of the new centroids, by

$$\mathbf{c}_z \leftarrow \frac{\sum_{t=1}^T u(t, z, \hat{z}_t) \mathbf{x}_t}{\sum_{t=1}^T u(t, z, \hat{z}_t)}. \quad (45)$$

Next, the assignment step is repeated followed by the updating step again. This procedure stops as soon as the algorithm converges, which means that there is no change between the newly calculated centroids and the ones one step beforehand. The initialization of the  $k$  centroids is chosen at random for the original  $k$ -means algorithm. There is an extension to the so-called  $k$ -means++ algorithm, which changes this circumstance.

In this extension to the common  $k$ -means algorithm, Arthur et al. [9] propose to change the initial seeding procedure to elevate accuracy of the algorithm. They suggest to replace the procedure of randomly choosing the initial centroids by selecting them depending on a certain probability distribution. They define this probability to be

$$p_{\mathbf{c}_z}(\mathbf{x}_t) = \frac{\left( \min_z (d(\mathbf{c}_z, \mathbf{x}_t)) \right)^2}{\sum_{t=1}^T \left( \min_z (d(\mathbf{c}_z, \mathbf{x}_t)) \right)^2}. \quad (46)$$

They select the initial centroids then sequentially, choose the first one randomly, and the next one with probability  $p_{\mathbf{c}_z}(\mathbf{x}_t)$  from all data points, and so on. Their extension of the  $k$ -means algorithm improves the original one in accuracy and speed.

### 1.5 EMPIRICAL MODE DECOMPOSITION

In the following paragraphs I will describe EMD and its extensions. Since my thesis is based mostly on the multivariate version of EMD, the introductory part of univariate EMD is kept compact.

## 1.5.1 Plain empirical mode decomposition

EMD is a data-driven approach of decomposing e. g. time-series, which can both be non-stationary and non-linear. The approach was first introduced by Huang et al. [49]. The decomposition results in so-called intrinsic mode functions (IMFs). At first, in the plain EMD case, we are dealing with one-dimensional signals. Let  $x(t)$  be a one-dimensional time-series. The decomposition of  $x(t)$  by EMD results in  $F$  IMFs  $u_f(t)$ ,  $f = 1, \dots, F$  and a non-oscillating trend  $r(t)$  [45, 49, 70]

$$x(t) = \sum_{f=1}^F u_f(t) + r(t). \quad (47)$$

Note that EMD results in a complete decomposition of the signal, i. e., summing up all IMFs and the residual results in the original time course. To result in such a decomposition, the algorithm has to rely on predefined conditions identifying an IMF as such. EMD is an iterative algorithm. The iteration process is called sifting and after each iteration the resulting time course is checked, whether it meets the condition for being an IMF. Sifting stops when the current time course meets the criterion for being a residuum. In the following, I describe the algorithm in detail.

The algorithm starts by selecting all minima and maxima of  $x(t)$ , respectively. The sets of maxima and minima are connected separately via **spline interpolation** to form envelopes of the time course  $x(t)$ . This results in an envelope  $e_{\max}^0(t)$  for the interpolated set of maxima and  $e_{\min}^0(t)$  for the set of minima. The superscript indices stand for the number of elapsed iterations. In a next step, the mean of the envelopes is subtracted from  $x(t)$  to form a proto time course  $y^1(t)$ . The first sifting iteration is completed. The algorithm now checks, if  $y^1(t)$  meets the prerequisites for being an IMF. The first condition for  $y^1(t)$  for being an IMF is that the number of zero-crossings of  $y^1(t)$  and the number of extrema differ at most by one. For the whole algorithm, this means

**CONDITION 1:** For a proto time course  $y^i(t)$  resulting from the sifting process to be recognized as an IMF, its number of zero-crossings and number of extrema has to differ at most by one.

This condition ensures that there are no riding waves in the resulting IMFs meaning that there are no extrema with values above or below the zero line [45, 49]. The second condition for  $y^1(t)$  to be an IMF is that its envelopes  $e_{\max}^1(t)$  and  $e_{\min}^1(t)$  have to be symmetric, i. e., their local mean has to be zero. This has to hold for a certain amount of iteration steps [70, 105]. Again, for the whole algorithm, this means

**CONDITION 2:** For a proto time course  $y^i(t)$  resulting from the sifting process to be recognized as an IMF, its envelopes  $e_{\max}^i(t)$  and  $e_{\min}^i(t)$  have to be symmetric.

**Spline interpolation** (the following from Levy [67]) is a special case of the interpolation problem of finding a function  $Q(x)$  connecting given data points  $x_i$  in a smooth manner. It is characterized by piece-wise polynomial interpolation between  $x_i$  with  $Q(x), x \in [x_i; x_{i+1}]$ .

If  $y^1(t)$  does not meet those two conditions, it is not considered an IMF. The algorithm then looks for all extrema in  $y^1(t)$  and the sifting process proceeds with  $y^1(t)$  as the input. Consider  $m$  sifting steps done. The proto time course  $y^m(t)$  now meets the two conditions for being an IMF. Then the algorithm assigns  $u_1(t) \leftarrow y^m(t)$ , subtracts  $y^m(t)$  from the original time course  $x(t)$  resulting in the next, first proto time course  $y^1(t) \leftarrow x(t) - y^m(t)$ , with the sifting process starting all over again until the next IMF is extracted. The whole EMD algorithm (alg. 1) stops when the residual  $r(t)$  is a non-oscillating function [70] resulting in the decomposition of equation 47. A proto time course  $y^i(t)$  to be identified as an IMF has to meet so-called

---

**Algorithm 1** Plain EMD; from Mandic et al. [70]

---

**Input:**  $x(t) \in \mathbb{R}$

**Output:**  $u_f(t)$ , with  $x(t) = \sum_{f=1}^F u_f(t) + r(t)$

- 1: Find extrema of  $x(t)$
  - 2: Interpolate the sets of maxima and minima separately using spline interpolation resulting in a lower  $e_{\min}(t)$  and an upper envelope  $e_{\max}(t)$
  - 3: Create the mean of the two envelopes  $\bar{e}(t)$
  - 4: Create pseudo time course  $p(t) \leftarrow x(t) - \bar{e}(t)$
  - 5: Check, if  $p(t)$  satisfies the criteria for being an IMF – local mean of the envelopes equals zero and number of extrema and zero crossings differ at most by one – and set  $u_f(t) \leftarrow p(t)$ , else  $x(t) \leftarrow p(t)$  and go to step 1
  - 6: Stop algorithm, if  $u_f(t)$  is a non-oscillating function and assign  $r(t) \leftarrow u_f(t)$ , else assign  $x(t) \leftarrow x(t) - u_f(t)$  and go to step 1
- 

stopping criteria. Mandic et al. [70] mention that the accuracy of EMD has to be sensible meaning that over and under sifted IMFs have to be avoided. One common stopping criterion they mention is that condition 2 has to hold for  $Q$  sifting iterations with  $4 \leq Q \leq 8$  [50]. An exemplary decomposition is shown in box 2.

## BOX 2: EXAMPLE PLAIN EMD

To illustrate the EMD procedure and its output, in this example the decomposition of a signal combined of two sine waves with periods  $2\pi$  and  $10\pi$  with added uniformly distributed noise is shown (fig. 2). For this demonstration, scripts from Al-Subari et al. [6] are used, which are provided online<sup>a</sup>. The added noise has max. 20% of the signal's amplitude and the sampling rate of the signal is  $f_s = \frac{\pi}{10}$ . The implementation replicates the algorithm of Huang et al. [49].

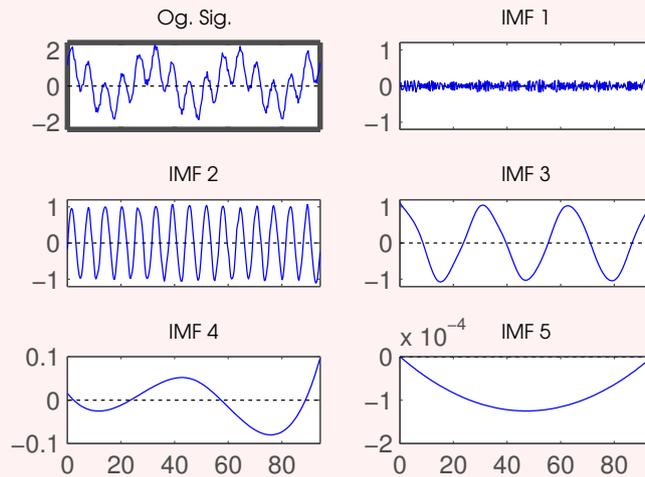


Figure 2: This figure shows a decomposition of an artificial signal with added noise (Og. Sig). The other panels depict the IMFs resulting from the plain EMD algorithm.

This EMD decomposition results in five IMFs and the two components of the signal can be clearly found in IMFs 2 and 3, and noise in IMF 1.

<sup>a</sup> [http://scn.ucsd.edu/eeglab/plugins/EMDLAB\\_Plugin.zip](http://scn.ucsd.edu/eeglab/plugins/EMDLAB_Plugin.zip)

### 1.5.2 Ensemble empirical mode decomposition

Mandic et al. [70] mention at least two popular shortcomings of plain EMD. First, it is not ensured that modes are found in just one IMF, i. e., modes with a physical meaning could spread over several IMFs yielding the loss of physical meaning for the single IMFs. This condition is termed mode mixing. A characteristic of mode mixing is the appearance of significantly different local frequencies in one IMF. Second, they mention a problem with so-called end effect artifacts – to create proper envelopes a sufficient number of extrema is needed. As time courses being decomposed have finite range, the density of extrema tends to decrease near the edges of the sampled time interval. Therefore the fit of the envelopes is more error-prone at the beginning and end of the time courses. Compared to the review of the state-of-the-art of EMD approaches the method introducing paper of Wu and Huang [109]

elaborates on these issues and their connection to ensemble EMD in more detail. Wu and Huang [109] rate the principle of their introduced approach as simple. They extend the algorithm of plain EMD to a so-called noise-assisted version. Following their way of thinking they introduce white noise to create ensembles of IMFs in order to average over the realizations of such an ensemble to overcome shortcomings of the original approach of plain EMD. Their extension of the algorithm starts by creating a white noise background for the signal being decomposed, i. e., adding white noise to the original signal with finite amplitude. Then the signal with added noise is decomposed by EMD (alg. 1) and the IMFs are saved. This step is repeated with newly generated white noise for every iteration resulting in an ensemble of IMFs. Afterwards, the IMFs of same index number are averaged resulting in a physically more meaningful IMF than getting the same IMF from plain EMD, as they say. Their extension is summarized in algorithm 2. Mandic et al. [70] introduce the algorithm for ensemble EMD using white Gaussian noise. They explain the usage of white Gaussian noise by exploiting the fact that this type of noise enables the algorithm to show the behavior of a **dyadic filter bank** [82, 31, 108].

**Dyadic filter bank**  
behavior of IMFs  
means that their  
number of zero  
crossings  $n_z$  has an  
exponential  
dependence  
 $n_z \propto b^{-f}$ , where  
 $b \approx 2$  and  $n_z$  is an  
estimate for the  
mean frequency [31].  
The term dyadic is  
deduced from the  
value of the base.

---

**Algorithm 2** Ensemble EMD; from Rehman and Mandic [82], Wu and Huang [109]

---

**Input:**  $x(t) \in \mathbb{R}; \{w_j(t)\}_{j=1, \dots, J} \sim \mathcal{N}(0, \sigma^2)$ , Gaussian noise ensemble

**Output:**  $u_f(t)$ , with  $x(t) = \sum_{f=1}^F \bar{u}_f(t) + \bar{r}(t)$

- 1: Initialize one realization of white Gaussian noise  $w_j(t)$  and add it to the time course  $x_n^j(t) \leftarrow x(t) + w_j(t)$
  - 2: Apply plain EMD from alg. 1 on  $x_n^j(t)$  and save all IMFs  $u_f^j(t)$  and the residual  $r^j(t)$
  - 3: Repeat steps 1 and 2 until desired size of ensemble is reached
  - 4: Average all IMFs and the residual resulting in representative IMFs  $\bar{u}_f(t) = \frac{1}{J} \sum_{j=1}^J u_f^j(t)$  and residual  $\bar{r}(t) = \frac{1}{J} \sum_{j=1}^J r^j(t)$
- 

Mandic et al. [70] also say that another problem of doing plain EMD on e. g. single channels in a multi-channel setup is that over several channels IMFs of same indices can differ in certain aspects like frequency, so these IMFs fail in aligning. Rehman and Mandic [82] show that this issue is tackled by the approach of multivariate empirical mode decomposition (MEMD) – introduced in the next section.

### 1.5.3 Multivariate empirical mode decomposition and its noise-assisted extension

MEMD [81, 70] is the multivariate extension of plain EMD introduced above. Interpreting plain EMD as an univariate approach of decomposing one-dimensional time courses, then MEMD sees the response of a system from several channels as a multivariate signal with every channel – like electroencephalography-channels (EEG) or time courses of ICs – represent-

ing one dimension in a multidimensional space. The time course can be seen as a trajectory propagating in this multidimensional space and MEMD decomposes this signal in multidimensional IMFs [81, 32]. Rehman and Mandic [81] do this by generalizing the aspect of creating envelopes around the time course to  $n$  dimensions by introducing a set of  $n$ -dimensional direction vectors, projecting the signal onto these direction vectors, getting the extrema of those projections, re-projecting them into the  $n$ -dimensional space, and using them as the base of constructing an  $n$ -dimensional envelope. The approach of Fleureau et al. [32] starts with the generalization at an earlier stage. They generalize the method of looking for extrema to  $n$  dimensions by using the generalized principle of a tangent vector. I do not elaborate on the later, because I use the approach of Rehman and Mandic [81] in my thesis.

To prevent confusion, it has to be mentioned that the multidimensional EMD introduced by Wu et al. [110] and applied e. g. by Al-Baddai et al. [5] in combination with ICA to analyze combined EEG/fMRI data is a completely different approach. It does not treat the signal in a **multivariate** manner. It is used to decompose an univariate, but **multidimensional** signal, e. g. images like fMRI data in univariate, but multidimensional IMFs. This approach can also be extended to  $n$ -dimensional signals, but those signals still being univariate [4]. For the investigation of the temporal evolution of a system – like the activities of several different brain areas representing the system brain as a whole – the just mentioned approach is not suitable. For this kind of analysis it is important to treat the signal as a trajectory in an  $n$ -dimensional space with all ICs contributing to the evolution of the whole system. I am not aiming to decompose fMRI pictures into IMFs inherent to the picture *per se*, rather I am interested in the temporal evolution of the system brain as a whole. Therefore, the multivariate approach is more suitable in my case.

Rehman and Mandic [81] propose to start MEMD by generating a Hammersley sequence to sample the  $n$ -dimensional space of the multivariate signal  $\mathbf{x}(t)$  by a set of direction vectors  $\{\hat{\mathbf{d}}_i\}_{i=1,\dots,I}$  in a low-discrepancy manner (see box 3 for details). The authors aim to cover the  $n$ -dimensional space uniformly by the generation of such a set of Hammersley sequenced direction vectors, which is equivalent to looking for a low-discrepancy distribution of points on a hypersphere representing the direction vectors. They mention that a natural way of sampling an  $n$ -dimensional space would be to do angular sampling. But they highlight in the three-dimensional case shown in figure 3a that the density of points increases when approaching the poles of the spheres. This renders angular sampling an inappropriate choice for uniform sampling of an  $n$ -dimensional space. Therefore, Hammersley sequenced direction vectors (fig. 3b) are preferred over angular sampled ones.

The algorithm starts by generating a Hammersley sequenced set of direction vectors  $\{\hat{\mathbf{d}}_i\}_{i=1,\dots,I}$ . Then, for every  $i$ , the  $n$ -dimensional time course  $\mathbf{x}(t)$  is projected on  $\hat{\mathbf{d}}_i$  generating a one-dimensional signal. One of the difficulties of creating envelopes in  $n$ -dimensional space is the selection of proper extrema or even the definition of an extremum [32]. After having

*A **multivariate** signal in the case at hand can be formulated as  $f(\mathbf{x}, t)$ , with  $\mathbf{x} \in \mathbb{R}^n$ , and a **multidimensional** signal mentioned here can be written as  $\mathbf{x} \in \mathbb{R}^n$ . This means the values of  $f(\mathbf{x}, t)$  depend e. g. on several spatial locations  $\mathbf{x}$  and time  $t$ , and in the multidimensional case the signal is just an  $n$ -dimensional tuple  $\mathbf{x}$ .*

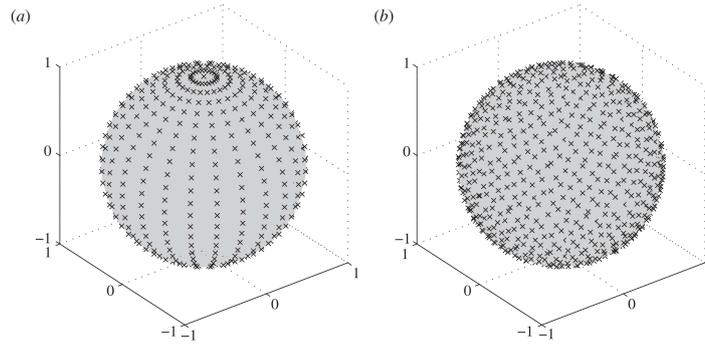


Figure 3: This figure from Rehman and Mandic [81] depicts angular- (a) and Hamersley sequenced (b) sets of points on the three-dimensional unit-sphere.

projected  $\mathbf{x}(t)$  onto the first direction vector, a one-dimensional signal  $\chi_1(t)$  arises. For a trajectory propagating in one dimension there is a proper definition of extrema. Using this common definition Rehman and Mandic [81] look for the maxima and minima in  $\chi_1(t)$ . They project the extrema into the  $n$ -dimensional space again and connect the maxima and minima separately to form  $n$ -dimensional envelopes  $\mathbf{e}_{\max}^1(t)$  and  $\mathbf{e}_{\min}^1(t)$  both realized by spline interpolation. This is repeated for all direction vectors  $\{\hat{\mathbf{d}}_i\}_{i=1,\dots,I}$  covering the whole  $n$ -dimensional space and every realization of the envelopes is saved. Thus, now one can imagine the ensemble of envelopes estimating an envelope around the  $n$ -dimensional signal  $\mathbf{x}(t)$ . The next step analog to the original EMD approach is averaging the envelopes. In the case of MEMD all created  $n$ -dimensional envelopes  $\mathbf{e}_{\max}^i(t)$ ,  $\mathbf{e}_{\min}^i(t)$ ,  $i = 1, \dots, I$  are averaged. Afterwards, the same procedure as in plain EMD (alg. 1) is applied meaning that the original time course and the mean of the envelopes are subtracted resulting in a proto time course, which is checked concerning the criterion for being an IMF and the iterative sifting process takes over with the above mentioned additions for compensating the problems of a multivariate signal (alg. 3). Rehman and Mandic [81] mention that the criterion for a time course to be an IMF is similar to that of the plain EMD algorithm, besides the comparison of number of zero crossings and number of extrema does not hold anymore, as there is no proper definition of zero crossings in  $n$ -dimensional space.

As for the plain EMD approach Rehman et al. [83] propose an ensemble noise-assisted method. It is important here to distinguish between noise-assisted and ensemble approaches. For plain EMD these two approaches are synonymous. In case of MEMD it has to be mentioned, whether only a noise-assisted or an ensemble noise-assisted MEMD is applied. The noise-assisted MEMD approach has been introduced by Rehman and Mandic [82] and extended by Rehman et al. [83] to the ensemble noise-assisted MEMD method. Mandic et al. [70] deduced the multivariate approaches from the original EMD principle. Compared to MEMD the noise-assisted MEMD introduces noise channels in addition to the signal channels spanning the  $n$ -dimensional space of the multivariate signal [70, 82]. The authors propose

## BOX 3: HAMMERSLEY-SEQUENCE AND UNIFORM SAMPLING

Hammersley sequences are generated by employing the prime base representation of an integer number  $k = \sum_{i=0}^r a_i p^i$  and the subsequent definition of function  $\Phi_{p_j} = \sum_{i=0}^r \frac{a_i}{p_j^{i+1}}$  [107], where  $p_j$  are primes. Wong et al. [107] introduce a Hammersley sequence as the points in a  $d$ -dimensional space with  $(\frac{k}{n}, \Phi_{p_1}(k), \dots, \Phi_{p_{d-1}}(k))$ . This means for each dimension another prime base is needed. This procedure results in a low-discrepancy sequence of points in a  $d$ -dimensional space and therefore results in a uniform sampling of it [81]. Figure 4 shows a Hammersley sequenced point set in two dimensions. The uniform sampling character is quite obvious.

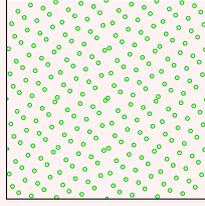


Figure 4: In this figure, a two-dimensional example of a Hammersley sequenced point set is depicted. From [106].

---

**Algorithm 3** MEMD; from [81, 70]
 

---

**Input:**  $\mathbf{x}(t) \in \mathbb{R}^n$ ,  $\{\hat{\mathbf{d}}_{i=1, \dots, I}\} \subset \mathbb{R}^n$  Hammersley sequenced

**Output:**  $\mathbf{u}_f(t)$ , with  $\mathbf{x}(t) = \sum_{f=1}^F \mathbf{u}_f(t) + \mathbf{r}(t)$

- 1: Project  $\mathbf{x}(t)$  on direction vector  $\hat{\mathbf{d}}_i$  resulting in  $x_i(t)$
  - 2: Get maxima and minima of  $x_i(t)$ , project them to  $n$ -dimensional space and connect them, both realized via spline interpolation resulting in envelopes  $\mathbf{e}_{\max}^i, \mathbf{e}_{\min}^i$
  - 3: Repeat step 1 and 2  $\forall i$  and save the envelopes
  - 4: Average the sets of envelopes  $\bar{\mathbf{e}}(t) = \frac{1}{2I} \sum_{i=1}^I \mathbf{e}_{\max}^i(t) + \mathbf{e}_{\min}^i(t)$
  - 5: Check if  $\mathbf{p}(t) \leftarrow \mathbf{x}(t) - \bar{\mathbf{e}}(t)$  meets the criterion of being an IMF and assign  $\mathbf{u}_f \leftarrow \mathbf{p}(t)$ , else set  $\mathbf{x}(t) \leftarrow \mathbf{p}(t)$  and go to step 1
  - 6: Stop algorithm, if the number of all  $x_i(t)$  is smaller than 3<sup>1</sup> and assign  $\mathbf{r}(t) \leftarrow \mathbf{u}_f(t)$ , else assign  $\mathbf{x}(t) \leftarrow \mathbf{x}(t) - \mathbf{u}_f(t)$  and go to step 1
- 

not to add noise to the already existing signal. In fact, they suggest to add  $l$  separate channels entirely consisting of white Gaussian noise. They mention that one of the benefits of doing MEMD in general is that – besides avoiding mode mixing – scales of same index IMFs align (see also [81]). Rehman and Mandic [82] mention that introducing white Gaussian noise channels exploits the dyadic filter bank property of MEMD for this type of noise. They argue that the added white Gaussian noise channels enforces the dyadic filter bank structure and with it a better mode alignment. Furthermore, they

<sup>1</sup> Deduced from their code available at [http://www.commsp.ee.ic.ac.uk/~mandic/research/memd/memd\\_version\\_2.zip](http://www.commsp.ee.ic.ac.uk/~mandic/research/memd/memd_version_2.zip) (file: memd.m, line 180)

emphasize that mode mixing is also reduced by introducing noise to the MEMD approach.

In a recent extension to noise-assisted MEMD the ensemble principle like for plain EMD was introduced to MEMD, or rather noise-assisted MEMD [83]. They suggest to just create ensembles of IMFs like in algorithm 2 by using different realizations of noise for several noise-assisted MEMD runs – keep in mind that those realizations of noise are in separate channels. The authors suggest to use a noise power of 2 – 10% of the power of the original signal. This algorithm does not have to be mentioned separately here, as the principle is the same as for ensemble EMD in the univariate case. Since this approach is the most sophisticated and most recent one, I am using that method introduced by Rehman et al. [83] in my thesis. An exemplary decomposition of the ensemble noise-assisted MEMD algorithm is presented in box 4. For my studies, I chose to use the ensemble noise-assisted MEMD approach. Whenever I am referring to MEMD in the following text, I specifically mean the ensemble noise-assisted version of it.

## BOX 4: EXAMPLE ENSEMBLE NOISE-ASSISTED MEMD

In this example, the ensemble noise-assisted MEMD algorithm from Al-Subari et al. [6] is used to decompose a trivariate signal in the form of a spiral with

$$\begin{aligned} z &= t \\ x &= \sin(t) \\ y &= \cos(t) \end{aligned}$$

where  $t \in [0; 10\pi]$  and the sampling rate  $f_s = \frac{\pi}{50}$ <sup>a</sup>. There is one prominent rotational mode in the signal. It is reflected in a rotary trajectory in the  $yz$ -plane. The algorithm was applied with 30 realizations of white Gaussian noise. The noise amplitude was chosen to have 6% power of the original signal. Two noise channels were added to the trivariate signal. The signal and one characteristic trivariate IMF is shown in figure 5. This IMF depicts the characteristic  $yz$ -mode in the signal.

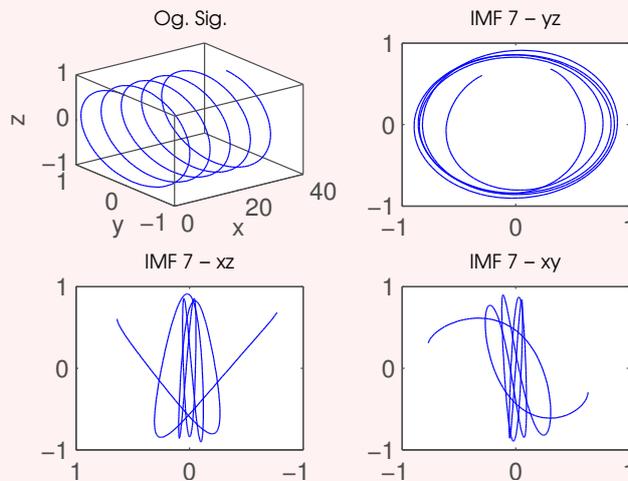


Figure 5: The upper left panel shows the original signal, which is decomposed by ensemble noise-assisted MEMD. The other three panels show the  $yz$ -,  $xz$ -, and  $xy$ -view of the IMF containing the rotational  $yz$ -mode, respectively.

<sup>a</sup> <http://de.mathworks.com/help/matlab/ref/plot3.html>

#### 1.5.4 Hilbert-Huang transformation

Huang et al. [51] describes the Hilbert-Huang transformation (HHT) as a purely empirical transformation applicable to non-linear and non-stationary signals. He mentions that the HHT is suitable for extracting instantaneous frequency from a mono-component signal, and therefore the IMFs resulting from EMD are proper signal components for estimating it. Since I am using only the instantaneous frequency aspect of the HHT in my thesis, I focus on this part of the transformation. Huang et al. [51] mentions that with the HHT any signal  $x(t) \in \mathbb{R}, \forall t$  can be transformed into an analytical form by adding just an imaginary part, which is defined as

$$y(t) = \frac{1}{\pi} \text{P} \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau = x(t) * \frac{1}{t} \quad (48)$$

transforming it to the complex notation

$$z(t) = x(t) + iy(t) = a(t)e^{i\theta(t)}, \quad (49)$$

where  $a(t)$  is the instantaneous amplitude and  $\theta(t)$  the instantaneous phase at time  $t$ . The Hilbert transform of equation 48 is suitable for this purpose, because of its uniqueness, like mentioned in Huang et al. [49]. The Cauchy principal value in equation 48 is used to ensure the integrability of the integrand. In Huang et al. [49] it was mentioned that equation 48 resembles a convolution with an hyperbolic function, which emphasizes the local features of  $x(t)$ . Huang et al. [49] explain the value of getting an analytical transformation of  $x(t)$  (eq. 49) is that such a form yields a unique way of defining the instantaneous frequency

$$\omega(t) = -\frac{d\theta}{dt}, \quad (50)$$

with

$$\theta(t) = \arctan\left(\frac{y(t)}{x(t)}\right). \quad (51)$$

#### 1.6 BUTTERWORTH FILTER

Since Butterworth filter banks were used as *post hoc* investigations of my MEMD results and were not the focus of this thesis, I elaborate only on particular aspects to understand my approaches. The following is taken from Johnson [57] and Oppenheim and Schaffer [77]. To understand the concept of a Butterworth filter, the general notation of discrete time systems has to be introduced. In signal processing, the discrete input  $x(n)$  and output  $y(n)$  of a system are connected by mapping the former to the latter one

$$y(n) = \mathcal{J}\{x(n)\}, \quad (52)$$

where the general notation of a discrete-time signal is

$$x(n) = \sum_{k=-\infty}^{\infty} a_x(k)\delta(n-k), \quad (53)$$

with  $\delta(n-k)$  being the so-called unit sample sequence at time point  $n$  with shift  $k$  and  $a_x(k)$  the amplitude at this particular shift. For the unit sample sequence

$$\delta(n) = \begin{cases} 1, & \text{if } n = 0 \\ 0, & \text{else} \end{cases} \quad (54)$$

holds. This is the most general form for representing any discrete-time signal. If this system has linear behavior, the superposition and scaling principle yields

$$\mathcal{J}\{x(n)\} = \mathcal{J}\left\{\sum_{k=-\infty}^{\infty} a_x(k)\delta(n-k)\right\} = \sum_{k=-\infty}^{\infty} a_x(k)\mathcal{J}\{\delta(n-k)\}. \quad (55)$$

This mapping of the shifted delta distribution is called the impulse response function

$$h(n, k) = \mathcal{J}\{\delta(n-k)\}. \quad (56)$$

This dependency can be further simplified by introducing time invariance to our system. Time invariance is defined as

$$y(n-m) = \mathcal{J}\{x(n-m)\} = \sum_{k=-\infty}^{\infty} a_x(k)h(n-k) = \sum_{k=-\infty}^{\infty} a_x(n-k)h(k) \quad (57)$$

meaning that a change in the input signal is immediately transferred to the output signal. The latter equality can be understood when substituting the summation index by  $m = n - k$ . When assuming that  $h(n, k)$  is the response to  $\delta(n-k)$ , then applying the time invariance principle from equation 57 yields

$$h(n-k) = \mathcal{J}\{\delta(n-k)\}. \quad (58)$$

The response to the delta distribution input is directly translated to the output. Therefore, a linear time-invariant system is completely described by its response to  $\delta(n)$ .

To reach our goal of understanding the concept of a filter, two more concepts have to be introduced. At first, we have to look at the so-called  $z$ -transformation in the context of discrete-time signals. This transformation of a signal  $x(n)$  is formulated like

$$X(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n}. \quad (59)$$

This series converges for certain  $z$  and otherwise diverges. To be further on able to analyze systems mathematically, we have to narrow down the possible systems space. In the following, we consider systems obeying a difference equation of order  $N$

$$\sum_{k=0}^N a_y(k)y(n-k) = \sum_{m=0}^M a_x(m)x(n-m), \quad (60)$$

which is usually seen as a causal system. Applying the  $z$ -transformation from equation 59 yields

$$\sum_{k=0}^N a_y(k) \sum_{n=0}^{\infty} y(n-k)z^{-n} = \sum_{m=0}^M a_x(m) \sum_{n=0}^{\infty} x(n-m)z^{-n}. \quad (61)$$

By using the so-called shift property of the  $z$ -transform, we can write

$$\sum_{k=0}^N a_y(k) \sum_{n=0}^{\infty} y(n)z^{-(n+k)} = \sum_{m=0}^M a_x(m) \sum_{n=0}^{\infty} x(n)z^{-(n+m)}, \quad (62)$$

which yields

$$Y(z) = \frac{\sum_{m=0}^M a_x(m)z^{-m}}{\sum_{k=0}^N a_y(k)z^{-k}} X(z), \quad (63)$$

where the coefficient of  $X(z)$  is defined as the systems function  $H(z)$ . We can also see that the systems function is the  $z$ -transformed version of the impulse response function by using the delta impulse as the input to equation 63. From tables, it is known that the  $z$ -transformation of  $\delta(n)$  is unity. Therefore  $Y(z) = H(z)$  holds and yields

$$H(z) = \sum_{n=-\infty}^{\infty} h(n)z^{-n}. \quad (64)$$

The above deduction was done to be able to understand that the stability of a filter system depends on its pole behavior. A pole of a system can be defined by rearranging the systems function  $H(z)$  to its factored form

$$H(z) = A \frac{\prod_{i=1}^M (1 - z_i z^{-1})}{\prod_{j=1}^N (1 - p_j z^{-1})} \quad (65)$$

Stability in this sense means that the output of a system has a finite magnitude given the input is bounded  $|x(n)| \leq M$ , with  $M < \infty$ . To see that a system is stable when its poles lie in the unit circle, we investigate the

magnitudes of the response of a linear time-invariant system introduced in equation 57

$$\begin{aligned} |y(n)| &= \left| \sum_{k=-\infty}^{\infty} a_x(n-k)h(k) \right| \leq \sum_{k=-\infty}^{\infty} |a_x(n-k)||h(k)| \leq \\ &\leq M \sum_{k=-\infty}^{\infty} |h(k)|. \end{aligned} \quad (66)$$

This means that the system is stable, if condition

$$\sum_{k=-\infty}^{\infty} |h(k)| < \infty \quad (67)$$

is fulfilled. Extending this to the system level, we investigate the magnitude of the systems function

$$|H(z)| = \left| \sum_{n=-\infty}^{\infty} h(n)z^{-n} \right| \leq \sum_{n=-\infty}^{\infty} |h(n)||z|^{-n}. \quad (68)$$

If we consider the area outside the unit circle  $|z| \geq 1$ , then

$$|H(z)| \leq \sum_{n=-\infty}^{\infty} |h(n)| \quad (69)$$

holds. From equation 67 we see that this is finite for a stable system. Considering the factored form of a systems function from equation 65 we see that for  $|z| \geq 1$  no poles are allowed. Because of this circumstance, the filter banks used in this thesis are optimized to have no poles with  $|z| \geq 1$ .

Butterworth filters are systems with a systems function magnitude

$$|H(i\omega)|^2 = \frac{1}{1 + \left(\frac{\omega}{\omega_c}\right)^{2N}}, \quad (70)$$

where  $\omega_c$  is the so-called cut-off frequency. The behavior of  $|H(i\omega)|$  is depicted in figure 6, where its dependency on the filter order can be seen. Frequencies with  $\omega \geq \omega_c$  are called the stopband and otherwise the passband. Using the terminology introduced above the systems function of a Butterworth filter transforms an input in a way that its frequency spectrum is modulated according to equation 70.

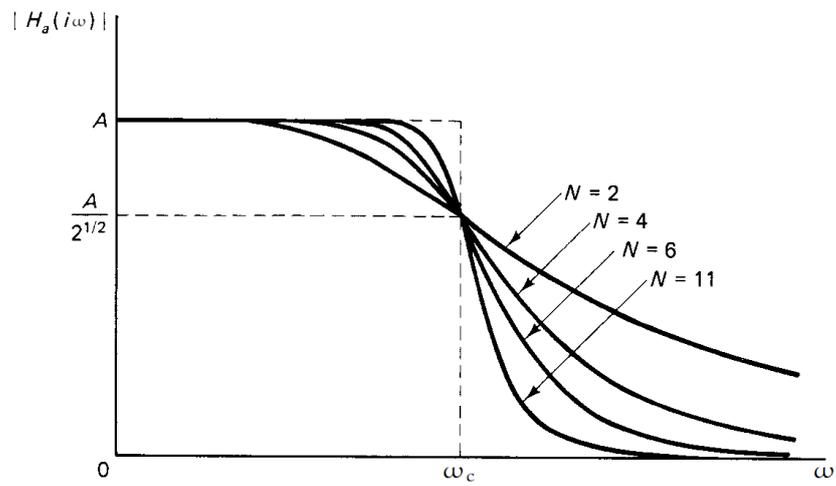


Figure 6: In this adjusted figure taken from Johnson [57], the transfer function of a Butterworth filter is depicted for different filter orders  $N$ .  $\omega_c$  is the cut-off frequency and  $A$  is the magnitude.

## FUNCTIONAL MAGNETIC RESONANCE IMAGING AND THE CONCEPT OF RESTING STATE

---

In the next sections, I introduce the concept of magnetic resonance imaging (MRI) (sec. 2.1) and its extension to fMRI (sec. 2.2). For both sections, I rely on references [48, 53, 44] and <http://www.imaios.com/en/e-Courses/e-MRI>.

### 2.1 MAGNETIC RESONANCE IMAGING

MRI is a non-invasive medical imaging modality. It is based on the fundamental property spin, which conveys angular momentum and a magnetic moment. MRI exploits a particular spin – the nuclear spin. Each nucleotide has spin  $s = \frac{1}{2}$  and with it a magnetic moment  $\mu$ . For MRI, the most important atomic nucleus is the hydrogen nucleus, because of its frequent occurrence in the human body. In the following text, I refer to those nuclei as protons. The functioning of MRI systems is based on the interaction of a radio frequency pulse with the magnetic moments of the protons. For this purpose, the subjects to be investigated have to be exposed to strong magnetic fields. One of the most common field strengths used for MRI imaging today is 3 T. For a reliable investigation, this field strength has to be kept homogeneous over a relatively large cross section (e. g.  $\approx 30$  cm in diameter in a SIEMENS Trio head scanner). To accomplish that, a superconducting coil is cooled down to 4.2 K by liquid helium allowing for resistance-free current flowing in the coil material. This current induces the homogeneous magnetic flux density<sup>2</sup>  $\mathbf{B}_0$  within the examination area. The direction of the magnetic field vector is parallel to the axis defined by the subject lying in the scanner. In this framework, this axis is defined as the z-axis pointing from feet to head of the subject, which means that the homogeneous magnetic field vector is  $\mathbf{B}_0 = B_0 \hat{\mathbf{e}}_z$ . The xy-plane is perpendicular to the z-axis with the y-axis pointing from back to chest of the subject.

When placing a subject in the magnetic field  $\mathbf{B}_0$ , the z-component of the magnetic moments  $\mu$  of the protons align either parallel, or anti-parallel to  $\mathbf{B}_0$ . This dichotomous discretization is due to the quantum mechanical nature of those magnetic moments conveyed by the fundamental property of particles with spin  $s = \frac{1}{2}$ . The alignment in parallel to the magnetic field is the energetically preferred state according to the Boltzmann statistics applicable in this case [48]

$$\frac{N^-}{N^+} = e^{-\frac{\epsilon}{k_B T}}, \quad (71)$$

---

<sup>2</sup> In the following, the magnetic flux density  $\mathbf{B}$  is referred to as magnetic field for readability reasons.

where  $N^-$  is the number aligning antiparallel to  $\mathbf{B}_0$ , and  $N^+$  parallel. Therefore, there is an excess of  $\mu_z$  of the protons aligned in parallel to the magnetic field. This yields a so-called netto magnetization  $\mathbf{M}^{\text{netto}}$  of the tissue under consideration. This netto magnetization is parallel to the homogeneous magnetic field  $\mathbf{B}_0$  yielding

$$\mathbf{M}^{\text{netto}} = (N^+ - N^-)\mu_z\hat{\mathbf{e}}_z. \quad (72)$$

For the following theoretical derivations, the tissue under consideration is subdivided into so-called spin-packets. Each spin-packet is a three-dimensional cuboid containing several protons. If  $N^+$  and  $N^-$  now refer to the number of protons in a spin-packet, then equation 72 describes the netto magnetization  $\mathbf{M}^{\text{netto}}$  of one spin-packet. The next, more coarse-grained subdivision of the tissue is a so-called voxel, which also is a cuboid consisting of several spin-packets. A common spatial elongation of a voxel in anatomical MRI is  $1 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$ , and in fMRI  $3 \text{ mm} \times 3 \text{ mm} \times 3 \text{ mm}$ . When dealing with MRI, voxels define the spatial resolution and are the fundamental building blocks. Spin-packets, on the other hand, are only of theoretical concern for deducing the physical framework.

Consider as initial condition the subject lying in the scanner and the netto magnetization of each spin-packet aligning with the homogeneous magnetic field yielding

$$\mathbf{M}^{\text{netto}} = M^{\text{netto}}\hat{\mathbf{e}}_z. \quad (73)$$

To interact with the tissue, the MRI scanner applies electromagnetic high frequency pulses in the radio frequency range to it. This procedure exploits a phenomenon called nuclear magnetic resonance. A spin system in a magnetic field has a certain precession frequency called the Larmor-frequency

$$\omega_L = \gamma B_0, \quad (74)$$

which is proportional to the homogeneous magnetic field  $B_0$  with the proportionality constant  $\gamma$  representing the so-called gyromagnetic ratio. If the frequency of the electromagnetic pulses applied on the spin system is tuned to  $\omega_L$ , then there is maximum energy transfer – the spin system is excited in resonance. The radio frequency pulse can be designed in such a way as to flip the netto magnetization about a certain angle. This depends on the duration  $t$  of the puls and can be described by

$$\theta = 2\pi\gamma t B_1, \quad (75)$$

where  $B_1$  is the amplitude of the magnetic field component of the puls. Here, I want to highlight that this angle can take arbitrary values, but the magnetic moments of each proton can only have two discrete values  $\boldsymbol{\mu} \propto \pm\hat{\mathbf{e}}_z$ . This discrepancy is resolved by the so-called Bloch equations, which introduce a semiclassical view of such spin systems allowing to employ Euler's equations of motion to the precessing netto magnetization vectors. This in

turn presents a mathematical justification of the continuous angle of the netto magnetization resulting from discrete states of the magnetic moments  $\mu$ . Two important flip angles are  $90^\circ$  and  $180^\circ$ . When a radio frequency pulse is applied flipping  $\mathbf{M}^{\text{netto}}$  about  $90^\circ$  into the  $xy$ -plane,  $\mathbf{M}^{\text{netto}}$  precesses around the  $z$ -axis. This rotation of  $\mathbf{M}^{\text{netto}}$  can induce an oscillating current in a nearby coil according to Faraday's law of induction. This is the first step towards the extraction of a signal from the tissue under consideration – this signal is induced in a receiver coil. After the excitation pulse, the signal starts to decay in a process called relaxation. The energy transferred to the spin system by the radio frequency pulse is released into the surrounding by spin-spin and spin-lattice interactions, causing spins to flip back to their initial configurations. This relaxation process causes the netto magnetization vector to turn back to its initial configuration – from precessing in the  $xy$ -plane to aligning parallel to the  $z$ -axis – in a spiraling motion. With this relaxation process, also the signal induced in the receiver coil decreases. This decremental process of the signal is called the free induction decay (FID). To better understand this process, we have to look at relaxation in more detail.

At first, the relaxation process is subdivided into so-called  $T_1$ - and  $T_2$ -relaxation. The former one describes the relaxation of the  $z$ -component of  $\mathbf{M}^{\text{netto}}$ .  $T_1$  is called the longitudinal relaxation time constant, and is defined as the time it takes for  $M_z^{\text{netto}}$  to reach  $1 - \frac{1}{e}$  of the initial value. Its temporal evolution is described by

$$M_z^{\text{netto}} = M^{\text{netto}} \left(1 - e^{-\frac{t}{T_1}}\right). \quad (76)$$

This  $T_1$  process is mainly driven by spin-lattice interactions. The  $T_2$ -relaxation describes the decay of the  $xy$ -component of  $\mathbf{M}^{\text{netto}}$  and is called the transversal relaxation process. This decay can be best understood when dealing with spin-packets as the building blocks of voxels. After a  $90^\circ$ -pulse, besides precessing in the  $xy$ -plane, the  $\mathbf{M}^{\text{netto}}$  of all spin-packets are also in phase. Mostly because of spin-spin interactions, the netto magnetization vectors begin to dephase. Figure 7 illustrates this process. The transversal netto magnetization vectors of three spin-packets are depicted as black arrows, which sum up to the transversal netto magnetization vector of one voxel depicted as a red arrow. Directly after a  $90^\circ$ -pulse, the transversal netto magnetization vectors of the spin-packets are in phase and result in the maximum netto magnetization in the  $xy$ -plane of this voxel. With time passing, the transversal netto magnetization vectors of the spin-packets lose their phase – they dephase. And with this increasing dephasing the transversal netto magnetization of the voxel also decreases.  $T_2$  is defined as the time, at which the transversal netto magnetization has decayed to  $\frac{1}{e}$  of the maximum value. The temporal evolution of this relaxation is described by

$$M_{xy}^{\text{netto}} = M^{\text{netto}} e^{-\frac{t}{T_2}}, \quad (77)$$

where  $M_{xy}$  is the  $xy$ -component of  $\mathbf{M}^{\text{netto}}$ .

The signal decay after a  $90^\circ$ -pulse is too fast [62] as to serve as the signal for reconstructing the desired image. Therefore, a phenomenon called

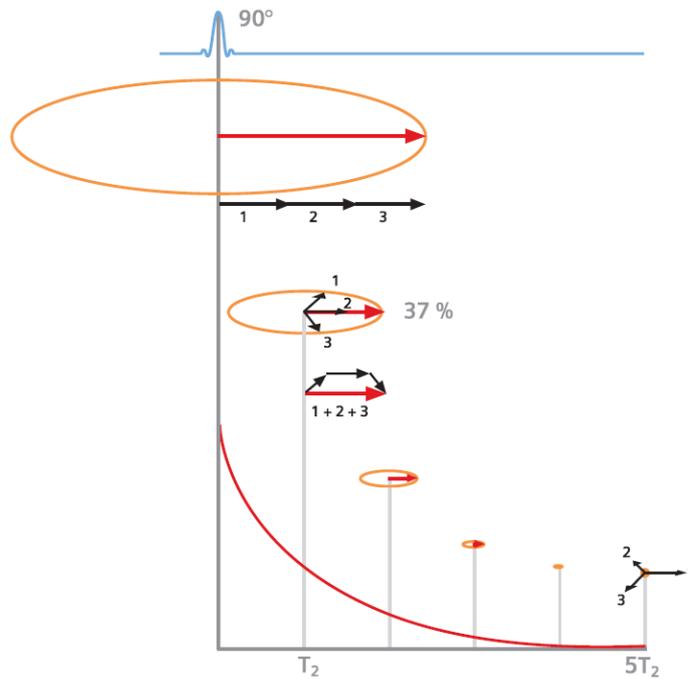


Figure 7: This figure depicts the transversal relaxation process. After the 90°-pulse, the transversal component of the exemplary netto magnetization vectors of three spin-packets (black arrows) are in phase. This yields the maximum transversal component of the netto magnetization of the voxel (red arrow) having those spin-packets as building blocks. With increasing time, the vectors of the spin-packets lose their phase causing the transversal netto magnetization of the voxel to decay.  $T_2$  is the transversal relaxation time constant marking the time when the transversal netto magnetization of the voxel is decayed to  $\approx 37\%$  of the initial value. From Hendrix [44].

rephasing is exploited. The dephasing netto magnetizations of spin-packets can be brought into phase again by applying a  $180^\circ$ -pulse. The  $180^\circ$ -pulse flips the netto magnetization about  $180^\circ$ . When applying this pulse to a spin system, a dephasing process is reversed and the transversal components of netto magnetization vectors of e. g. spin-packets start to decrease their phase difference again – they rephase. This process yields almost the initial magnitude of the signal. This recurrence of the signal is termed spin echo. The time between the  $90^\circ$ -pulse and the peak of the spin echo – or rather the acquisition of the signal – is called echo time (TE). The initial magnitude of the signal cannot be retrieved completely. When applying several  $180^\circ$ -pulses in a row, the amplitude of the echoed signal decays with  $T_2$ -relaxation constant. The decay of the FID happens much faster. The dephasing is accelerated by time-independent inhomogeneities of the magnetic field yielding a  $T_2^*$ -relaxation of the FID. Since the amplitude of the signal induced in the receiver coil depends on the magnitude of the transversal netto magnetization, the shape of the FID in figure 8 can be comprehended. Furthermore, the decay of the magnitude of the spin echo is also depicted in this figure.

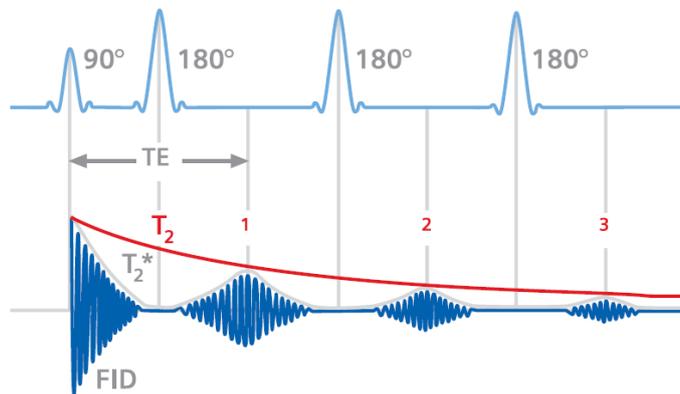


Figure 8: This figure depicts the FID and three spin echoes. The decaying signal induced in the receiver coil of the MRI system after the  $90^\circ$ -pulse depends on the magnitude of the precessing transversal netto magnetization (red arrow), which decays over time because of dephasing processes. The FID decays on a faster time scale ( $T_2^*$ ) than the magnitudes of the spin echoes ( $T_2$ ). The time between the  $90^\circ$ -pulse and the peak of the spin echo is called echo time (TE). From Hendrix [44].

Applying a  $90^\circ$ -pulse with frequency  $\omega_L$  to the whole tissue yields an excitation of all spins with Larmor-frequency  $\omega_L$ . Since one main aspect of MRI is to spatially resolve the distribution of tissues in the sample under consideration, this type of excitation is inappropriate. The first step towards this spatial resolution is the introduction of a so-called gradient coil in  $z$ -direction. As soon as this gradient coil is switched on, it alternates the field strength of the homogeneous field to show a linear increase in  $z$ -direction. This introduces a linear dependency of the overall magnetic field  $B \propto z$ . Taking equation 74 into account results in a linear dependency of  $\omega_L \propto z$ . After

this step, each  $xy$ -plane, or rather slice of the tissue, has a unique  $\omega_L$  allowing for the excitation of single slices. This means that, the gradient coil has to be simultaneously switched on with the radio frequency pulse to excite one slice instead of the whole sample. Since the radio frequency pulse has a finite bandwidth, the thickness of each excited slice is also finite. As can be seen on the voxel dimensions mentioned above, a common slice thickness for anatomical images is 1 mm and for functional scans 3 mm. This procedure with a gradient coil in the  $z$ -direction encodes one dimension of the three-dimensional problem. For the other two dimensions, two additional gradient coils are needed.

Without loss of generality, consider the  $x$ -direction the next dimension to be encoded. To accomplish that, another gradient coil has to be switched on – also introducing a linear dependency  $\omega_L \propto x$  in the  $x$ -direction. For simplicity, consider one row of voxels in the  $x$ -direction first. When the gradient coil in  $x$ -direction is on, each voxel has a unique  $\omega_L$  as precession frequency. To be encoded in the signal, this so-called frequency encoding gradient is switched on simultaneously with the acquisition of the signal. By applying Fourier transformation on the mixture of signals induced by all of those voxels, the intensity and spatial information of this row of voxels is decoded, which yields the spatial and intensity information. This procedure is illustrated in figure 9. Since the problem in each slice is two-dimensional, we need another gradient coil to encode the third dimension. This coil encodes the  $y$ -direction and is called the phase encoding gradient. Instead of leaving the gradient coil on during signal acquisition, the phase encoding gradient coil is switched on before signal acquisition. Depending on the duration this gradient coil is switched on, the phases between the transversal netto magnetizations of voxels in  $y$ -direction increases. This procedure has to be repeated for each row of voxels in the slice of the image – but with different phase angles. In other words, the phase encoding gradient is switched on with different magnitude for each row of voxels. This information can be condensed in a pulse-sequence diagram. The gradient patterns introduced above can be summed up as a so-called spin echo sequence (fig. 10). In this case, the time between the acquisition of two slices is called repetition time (TR). The slice selection gradient causes the netto magnetizations of the excited slice to dephase in  $z$ -direction, therefore the slice selection gradient has to be switched on again after the  $90^\circ$ -pulse with opposite polarity [44].

The signal generated by pulse sequences is not directly stored as coordinates in Euclidean space. The most basal format used by an MRI scanner is the so-called  $k$ -space representation. In the  $k$ -space, every coordinate can be represented by a lattice with varying phase and spatial frequency. Figure 11 illustrates the  $k$ -space for a two-dimensional slice acquired by MRI. The closer to the origin, the lower the spatial frequency of the representative lattice pattern gets. The intensity of each coordinate in  $k$ -space represents the weight each lattice pattern contributes to the image in Euclidean space. The pattern in this figure is created by the slice of an MRI image. One can imagine overlaying all weighted lattice patterns to create the image in Eu-

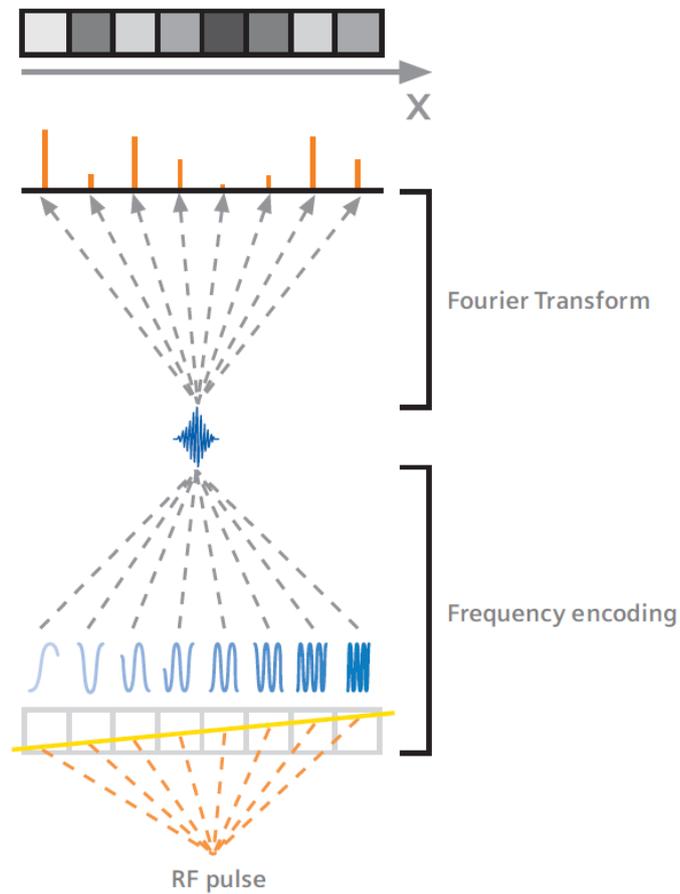


Figure 9: This figure depicts the illustrative example of frequency encoding on a single row of voxels. The radio frequency (RF) pulse excites the spins in those voxels. The gradient coil introduces different Larmor-frequencies for all voxels. The combined signal from all voxels is then decoded via Fourier transformation, which then gives the spatial and intensity information of the voxels. From Hendrix [44].

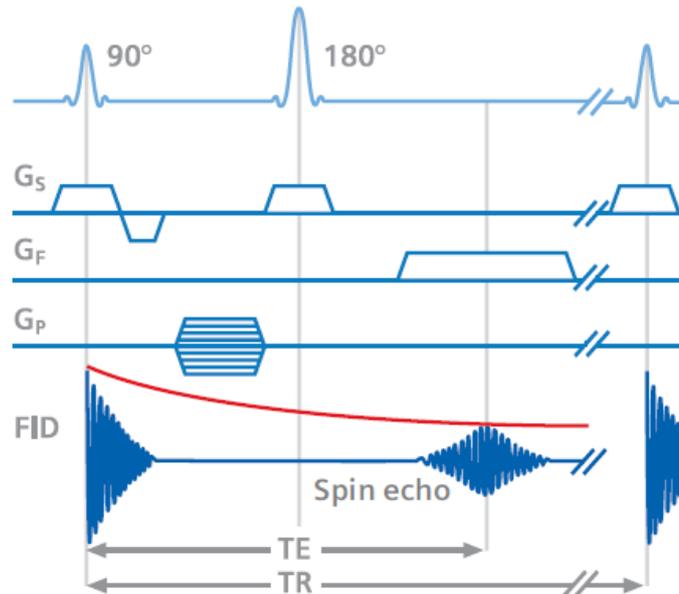


Figure 10: In this figure, the so-called spin echo sequence is illustrated. After a  $90^\circ$ -pulse with simultaneously applied slice-selection gradient ( $G_S$ ) (followed by rephasing) the phase encoding gradient ( $G_P$ ) is switched on for a certain time interval and with a distinct magnitude for each row of voxels. Afterwards, the re-phasing  $180^\circ$ -pulse is applied in parallel to the slice selection gradient again. The frequency encoding gradient ( $G_F$ ) is switched on during signal acquisition, which happens after a time interval called echo time ( $TE$ ). The time between acquisition of two slices is called repetition time ( $TR$ ). The free induction decay (FID) is depicted in the bottom row. From Hendrix [44].

clidean space. The transformation from  $k$ - to Euclidean space is done by a two-dimensional Fourier transformation.

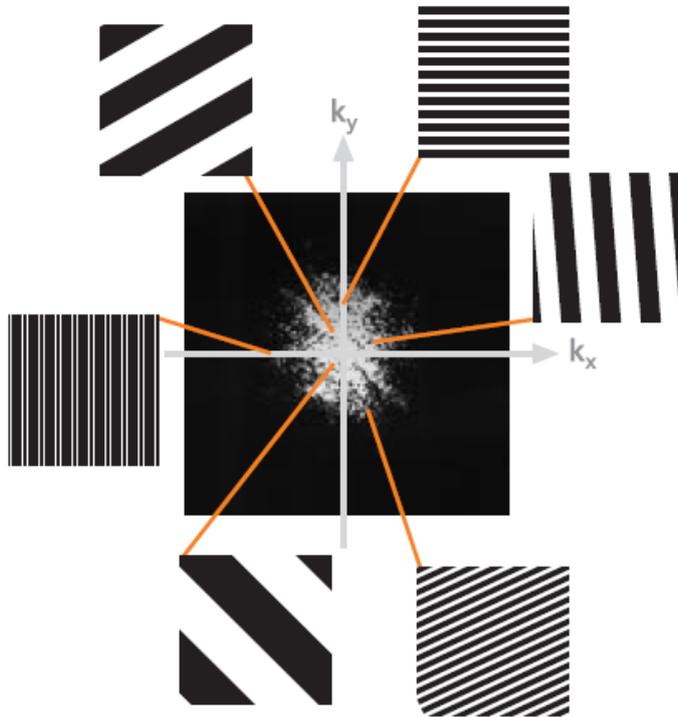


Figure 11: This figure illustrates the concept of  $k$ -space representation. Every coordinate can be represented by a lattice with varying spatial frequency and phase. The closer to the origin the lower the spatial frequency gets. From Hendrix [44].

The contrast between different tissues of an MRI image is based on the fact that spins in various tissues have different relaxation time constants. This can be exploited to distinguish tissues depicted on an MRI image. I elaborate on two contrasts: the  $T_1$ - and  $T_2$ -weighted contrast. The  $T_1$ -weighted contrast bases on different  $T_1$  constants for spins in different tissues. For this contrast, a short TR and short TE is chosen. This combination ensures to maximize the differences in longitudinal relaxation of various tissues. Short TRs emphasize differences in  $T_1$  between various tissues, and short TEs inhibit effects based on different  $T_2$  constants. The latter is true, because it takes time for differences in transversal magnetization based on  $T_2$ -effects to develop. This in turn means that choosing a long TE emphasizes effects based on  $T_2$  differences, which is exploited by a  $T_2$ -weighted contrast. Additionally, in this contrast, a long TR is chosen to suppress effects based on varying  $T_1$  constants. Those two contrast types have distinct features when it comes to distinguishing tissues. The  $T_1$ -weighted contrast images are characterized by high intensities in white matter tracts, lower ones in gray matter, and very low intensities in cerebro spinal fluid. For  $T_2$ -weighted contrast images it is the opposite – high intensities in cerebro spinal fluid, lower ones in gray matter, and even lower intensities in white matter. Another contrast

is called the  $T_2^*$ -weighted contrast, which has an analog mechanism as a  $T_2$ -weighted contrast basing on the  $T_2^*$  time constant and which is important when it comes to fMRI.

## 2.2 FUNCTIONAL MAGNETIC RESONANCE IMAGING

fMRI is based on the fact that hemoglobin has different magnetic properties depending on whether it is oxygenated or deoxygenated. Oxygenated hemoglobin has diamagnetic and deoxygenated paramagnetic properties. This dependency of the magnetic properties of hemoglobin on its oxygenation level can be captured by the  $T_2$ -relaxation time constant – or rather the  $T_2^*$ -relaxation time constant.  $T_2$  increases with oxygenation level, whereby  $T_1$  is unaffected by it. An increase in  $T_2$  yields a longer lasting transversal magnetization, which in turn results in higher MRI signals for oxygenated than deoxygenated blood. This dependency of the MRI signal on the oxygen level of blood is called blood-oxygenation-level dependant (BOLD) effect. This effect is used to infer indirectly on neuronal activity in the human cortex.

Neurons consume oxygen from the capillary bed for their metabolism. When idle, neurons have a baseline metabolism, which results in a constant transformation of oxygenated to deoxygenated hemoglobin in the capillary bed. This means that in the venules there is a constant amount of deoxygenated blood, when neuronal activity is lacking. With increasing activity level the oxygen consumption rises resulting in more deoxygenated hemoglobin in the venules. The body reacts to this decrease in available oxygen with a mechanism called supercompensation, which increases the supply of oxygen exceeding the consumed amount. This yields a displacement of deoxygenated by oxygenated hemoglobin in the venules of the capillary bed. The former process of decreasing oxygen level results in a decreased BOLD signal, which is called the initial dip. The latter process of supplying more oxygenated blood than is actually consumed resulting in an increased BOLD signal in the venules is called overshoot. After the overshoot, the BOLD signal falls below the initial, or rather baseline level. This process is called undershoot and completes the so-called hemodynamic response. The shape of the hemodynamic response can be explained by the so-called balloon-model. This model considers two sources that can vary the amount of oxygenated blood in the capillaries – the cerebral blood flow and the cerebral blood volume. In the process of supercompensation in the capillary bed, both the cerebral blood flow and volume are increased, which results in the rise and overshoot of the hemodynamic response. The cerebral blood volume is varied by changing the cross-sectional area of the capillaries, and has a delay compared to the cerebral blood flow. This results in a longer lasting expansion of the capillaries than increase in blood flow. This increased blood volume combined with rising deoxygenated hemoglobin levels yields the undershoot of the hemodynamic response. This physiological response is the correlate of neuronal activity. In other words, the variations

in intensity level of a voxel over time induced by BOLD fluctuations are referred to as activity time course of this particular voxel. However, this initial dip is too short in duration and too weak in intensity to serve as a proper measure of brain activity. Therefore, fMRI exploits the overshoot as another correlate of brain activity. Hence, fMRI images show the counter-intuitive feature of rising intensity values in an active voxel.

To capture this correlate, the so-called echo planar imaging (EPI) sequence has to be considered. Relying on [48, 44] for the following, the EPI sequence starts with a  $90^\circ$ -pulse combined with a slice selection gradient. Then both the frequency and phase encoding gradient is shortly switched on to start at one corner of k-space. After the  $180^\circ$ -pulse, the typical EPI pattern of frequency and phase encoding gradient starts. The frequency encoding gradient alternates in polarity, and before each switch a phase encoding gradient is switched on. The signal is recorded during the frequency encoding phase of either polarity. For every phase encoding step, one row in k-space is covered. By this combination of alternating frequency and short phase encoding gradients it is possible to record one slice in around 100 ms, which is the time span of the fast decaying FID. Therefore one complete image can be acquired within one TR. Since the FID decays with  $T_2^*$ -relaxation constant, EPI images have a  $T_2^*$ -weighted contrast, which has comparable contrast features to the  $T_2$ -weighted contrast – bright cerebro spinal fluid, dark white matter. Figure 12A shows the EPI sequence diagram with a detailed version of the combination of frequency and phase encoding gradient in figure 12B. Figure 12C shows how the k-space of one slice is covered by the EPI sequence.

A common way of analyzing fMRI images is to employ the so-called general linear model to the temporal evolution of EPI images. For every voxel time course separately, a general linear model

$$\mathbf{x} = \mathbf{D} \cdot \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (78)$$

is estimated, where  $\mathbf{x}$  is a vector of one voxel time-series containing intensity values of its activity time course,  $\mathbf{D}$  is the so-called design matrix containing at least the experimental time courses as regressors,  $\boldsymbol{\beta}$  is the vector of predictors, and  $\boldsymbol{\varepsilon}$  is the remaining error term after estimating the model. The experimental time courses in  $\mathbf{D}$  reflect the experimental design. E. g. in a **block-design** experiment, they are defined as boxcar functions convolved with the so-called hemodynamic response function, which is a mathematical estimate of the hemodynamic response course. This illustrates that for the common way of analyzing fMRI data an experimental design is needed to compare the activity time courses of voxels with. Since I am dealing with rs-fMRI data in this thesis, I do not go into further mathematical detail concerning this so-called task-based analysis by applying the general linear model. rs-fMRI data is *per definitionem* task free. However, task-based fMRI and task-free rs-fMRI data share certain preprocessing steps. Preprocessing is a necessary part in fMRI investigations, since raw EPI images do not convey the necessary quality to result in useful results after analysis.

*A block design experiment is characterized by on- and off-times of a stimulus, where the stimulus is either present or absent, respectively. Usually, both time blocks have the same length.*

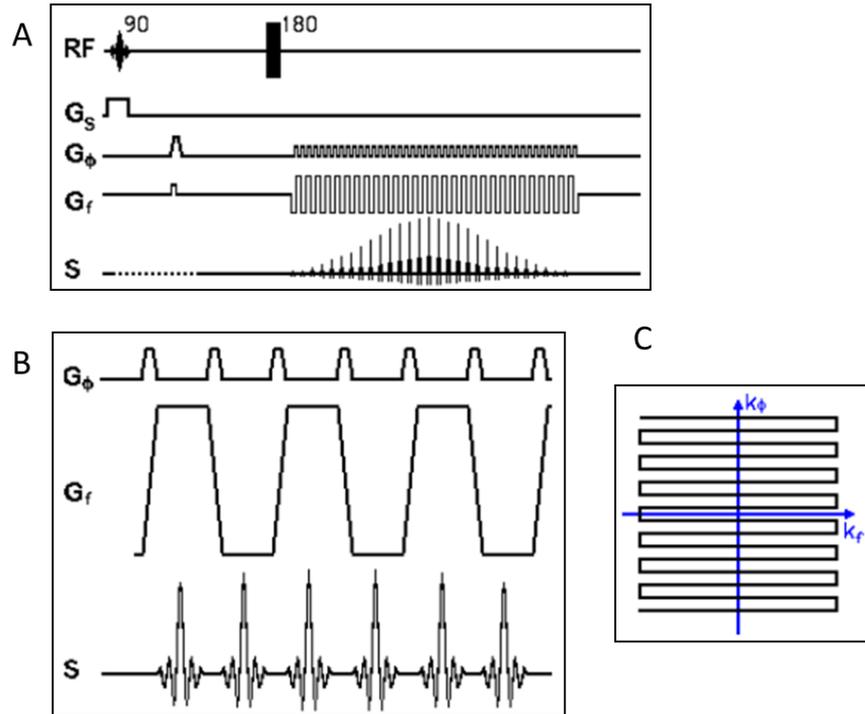


Figure 12: (A) This panel depicts the EPI sequence starting with the  $90^\circ$ -pulse and slice selection gradient  $G_S$ . Afterwards, one frequency and phase encoding step, by switching on the frequency ( $G_f$ ) and phase encoding gradient ( $G_\phi$ ) simultaneously, focuses the acquisition to one corner of the  $k$ -space. Then the  $180^\circ$ -pulse is applied to induce the spin echo. Afterwards, the alternating frequency and phase encoding gradients (detail can be seen in (B)) traverse the whole  $k$ -space in a pattern shown in (C). Assembled from Hornak [48] and adjusted.

### 2.2.1 Preprocessing of fMRI data

In the following, I elaborate on common preprocessing steps for fMRI data [53]. Task-based and rs-fMRI share the following first steps in a preprocessing pipeline.

**SLICE TIMING CORRECTION:** brain images are usually acquired slice by slice, which means that there are temporal delays between the acquisition of each slice. To correct for this error, adjacent time points are used to interpolate between intensity values of two neighboring slices.

**REALIGNMENT:** head movement cannot be avoided in fMRI scanning. Therefore, this preprocessing step is very crucial. It is assumed that the brain is a rigid body correcting for displacement in the six possible degrees of freedom (three translational; three rotational). By using a cost function like the sum of squared intensity differences of two images, all images are registered to match one reference image best. This procedure results in the evolutions of displacement over time for each degree of freedom, which are called realignment parameters.

**COREGISTRATION:** to also match the alignment of anatomical and functional images, this preprocessing step is employed. Since those types of images use different contrasts, a cost function using differences of intensity values cannot be applied. Instead, mutual information (sec. A.1.3) of the to be aligned images is calculated and minimized.

**NORMALIZATION:** this step corrects for the large spatial variability of different brains. It maps each brain into a common space (e. g. the so-called MNI space), where ideally the structures of all brains used in the analysis match. The transformation allows – besides the usual six degrees of freedom – for stretching, warping, and squeezing of the brain images to match a template representing the common space. This step is crucial for doing a gICA analysis, since this method assumes that each voxel in all images correspond to the same structures.

In task-based fMRI, many trials of the same stimulus type are presented and averaged, which eliminates noise – this averaging is absent in rs-fMRI and therefore additional preprocessing steps are needed with this modality [33]. There are many variations and additional preprocessing steps, but the following are the most widely used and accepted steps.

**LOW-PASS FILTERING:** Since most of the information for functional connectivity lies in the low frequency range of  $\omega < 0.1$  Hz [22], an important preprocessing step for rs-fMRI data is low-pass filtering with common cut-off values of  $\omega_c = 0.08$  Hz or  $\omega_c = 0.1$  Hz [33].

**TEMPORAL DETRENDING:** Like in Allen et al. [8], the removal of linear, quadratic, or cubic trends is also common to account for different sources of noise (e. g. scanner drift).

*A region of interest concerning human neuroscience is either a hypothesis driven or exploratorily defined region in the brain, which is treated as a separate entity to other ROIs. In this way, the cortex can be divided into distinct areas. This procedure is called **parcellation**.*

REGRESSING OUT NUISANCE VARIABLES: an important preprocessing step is to regress out nuisance variables to account for non-neuronal noise and artifacts [33]. Commonly, representative time courses of white matter and cerebro spinal fluid in addition to the six realignment parameters are included in a regression model, which is applied to every voxel or **region of interest** (ROI) time course, where the error term of this calculation represents the cleaned up version of the time course of interest [34, 65]. To get representative time courses from e. g. white matter areas, time courses of voxels attributed to white matter are averaged [112] – analogously for other nuisance areas.

### 2.3 RESTING-STATE

As an experimental – or rather pseudo experimental – paradigm, rs-fMRI is very simple. When conducted in the setting of fMRI, subjects either rest with eyes open fixating a fixation cross, or with eyes closed. Usually, subjects are instructed not to fall asleep and to let the mind wander. Contrary to the simplicity of the paradigm, the validation of the neuronal source of the signal underlying rs-fMRI is complex. In the following, I elaborate on the discussion about whether this signal is of physiological or neuronal origin.

#### 2.3.1 Basis of resting state fMRI signal

In their review article, van den Heuvel and Hulshoff Pol [102] elaborate on the origin of rs-fMRI signals. They mention that the real origin is still discussed and that there has been a debate on whether rs-fMRI bases on physiological or neuronal effects. For the physiological hypothesis, they refer to the possibility of aliasing effects introduced by higher frequency signal sources like respiration or heart rate. For the neuronal hypothesis, they highlight three main points: supportive evidence from neuroanatomy, frequency considerations, and electrophysiology. I illustrate on those four perspectives by elaborating on studies they refer to in more detail.

RESPIRATORY/HEART RATE ORIGIN: Birn et al. [13] investigate the effect of respiration on rs-fMRI data. In particular, they construct a measure for variability of breathing depth

$$V_B = \frac{\Delta_{\text{belt}}}{T_{\text{breath}}}, \quad (79)$$

where  $\Delta_{\text{belt}}$  is the difference between maximal and minimal lift of the chest of subjects, and  $T_{\text{breath}}$  is the breathing period. They find that this measure shows a mean frequency of 0.03 Hz, which lies in the typical frequency regime of rs-fMRI signals. Their analysis shows that the regions, which have the highest correlation with  $V_B$ , are default mode network (DMN) areas [17], and correction for breathing influence only reduces this effect slightly. They conclude that DMN

regions could be involved in respiratory function, have dense vascular anatomy, or that those regions have such a high blood volume and metabolism that the DMN just reflects these features. In another study [93], the heart rate variability is investigated in a similar way in rs-fMRI. The authors measure the heart rate of each subject and assign each EPI image a heart rate value. They define the point-wise heart rate as  $r = \frac{1}{\Delta t}$ , where  $\Delta t$  is the time interval from peak to peak in the cardiac measurement. The evolution of  $r$  represents the variability in heart rate. They find that the heart rate variability is found mostly in low frequencies  $\omega < 0.1$  Hz, which is a relevant frequency range when it comes to analyzing rs-fMRI data. Their results suggest that this predictor shows strong correlations in gray matter, and in particular also in DMN regions.

**NEUROANATOMICAL EVIDENCE:** In their study, Salvador et al. [89] perform an **atlas-based** analysis of functional connectivity by extracting representative time courses from predefined ROIs. They use partial correlation (sec. A.1.2) as association measure between the 90 ROIs. One of their investigations employs **hierarchical clustering** using  $d_{ij} = 1 - \tilde{\rho}_{ij}$  as dissimilarity measure, where  $\tilde{\rho}_{ij}$  represents the partial correlation between ROIs  $i$  and  $j$ . They find that at the lowest hierarchy level, the dendrograms show symmetrical connections between **homologous, bilateral** regions, and at the highest level they identify six large clusters that have been attributed to known neuroanatomical systems. Also at an intermediate level of clustering, they identify that homologous regions can be attributed to **gyral** domains. They conclude that the functional connectivity architecture is closely related, or shaped, by the underlying anatomical structure.

**ELECTROPHYSIOLOGICAL EVIDENCE:** Shmuel and Leopold [92] investigate in their monkey study, whether BOLD fluctuations acquired during rest correspond to neurophysiological measures. To accomplish that they simultaneously extract neurophysiological measures via electrode and BOLD signals. To make those signals comparable, they apply Fourier transformation on time windows of 1 s of the neurophysiological signal. They extract three time courses on three different frequency bands, and additionally the spiking rate for the 1 s time windows. With this procedure, they assure that the BOLD signal and the neurophysiological signal have a comparable temporal resolution. Besides similarities in the fluctuation distribution of all measures and in their spectra, their main finding is that the BOLD signal shows significant correlations with the neurophysiological measures. Furthermore, those correlations, they find, peak, if the BOLD signal lags behind the neurophysiological measures several seconds. They conclude that BOLD fluctuations seem to be more neuronal than physiological origin strengthening the method of functional connectivity with rs-fMRI data.

*Conducting **atlas-based** connectivity analysis is done by parcellating the cortex into functionally or anatomically segregated/specialized areas based on a brain atlas in a standard space. Representative time courses of each of those areas serve as the substrate for the analysis.*

***Hierarchical clustering** employs a distance measure between nodes to categorize them in a successive manner into more and more coarse-grained clusters.*

*The corresponding brain region of a bilaterally present one is called **homologous**.*

***Bilateral** brain areas have two symmetrically arranged counterparts in each brain hemisphere.*

*The surface of the brain can be roughly divided into **gyri** and **sulci**, where the former are elevations and the latter are sinks.*

*Seed-based correlation analysis uses the activity time course  $a_{v_i}(t)$  of node  $v_i$  as a reference to correlate the activity time courses of all other nodes  $v_j$  for  $j \neq i$  with, where high correlation values represent high functional connectivity with the seed node [10, 14].*

EVIDENCE FROM FREQUENCY CONSIDERATIONS: Cordes et al. [22] investigate the contribution of different frequencies to correlation maps resulting from **seed-based** rs-fMRI analysis. They calculate the correlation coefficient in the Fourier space resulting in a decomposition [21]

$$\rho = \sum_{\omega=1}^{N-1} \rho_{\omega} \quad (80)$$

where

$$\rho_{\omega} = \frac{N [\Re(\omega_n)\Re(\lambda_n) + \Im(\omega_n)\Im(\lambda_n)]}{\sigma_i \sigma_j} \quad (81)$$

with  $\omega_n$  and  $\lambda_n$  being the complex frequencies resulting from the inverse Fourier transform,  $N$  the number of time points,  $\omega$  running over  $N - 1$  discrete frequencies, and  $\sigma_i$  being the standard deviation of activity time course of voxel  $v_i$ . Cordes et al. [22] use task-based fMRI to localize regions in the auditory, motor, and visual cortices. They build seed-based correlation maps from seeds in those regions and calculate the average correlation coefficient spectra  $\langle \rho_{\omega} \rangle$  over all significant voxels of each map. They find that for seeds in the auditory, motor, and visual cortices the contribution to the correlation coefficient comes almost exclusively from low frequencies  $\omega < 0.1$  Hz and physiological noises sources are almost absent in those spectra. They also report that the correlation coefficient spectra of seed voxels placed in a ventricle, artery, or vein result in different frequency patterns compared to the gray matter analogues.

Taken together, the majority of findings attribute neuronal origin to rs-fMRI signals. This tendency is becoming more and more evident, since there is almost a shift in the fMRI community to rs-fMRI studies at the moment.

*"Everything  
hangs together."  
(Günther Oettinger)*

In this chapter, different physiological and methodological ways of investigating connectivity in the human brain are elucidated. Connectivity concerning the human brain is a generic term embracing all possible ways of how connections can emerge in this neural system. These connections can be reflected by pure physiological ones, like e. g. white matter tracts. This type of connectivity is called structural connectivity, since it deals with anatomical structures. This is also the most canonical form of connectivity – brain area A connects to brain area B via axons and exchanges information. A less canonical form of connectivity is called functional connectivity. It is based on the assumption that brain areas exhibiting a similar temporal evolution of activity share information and are connected in a functional manner. This can be investigated statically and dynamically. Connectivity-states is a concept resulting from dFC by extracting features from the data. The field in neuroscience, which deals with connectivity, is called connectomics.

The description of connectomics in this thesis is by no means exhaustive. I mostly focus on connectivity types that are necessary for the comprehension of my proposed methods. For the sake of completeness, however, I elaborate on structural connectivity. One major topic dealing with so-called effective connectivity is left out. It describes the estimation of directionality in neural systems – which brain area feeds into what is extracted from functional data. Besides the fact that my proposed methods do not rely on effective connectivity, it is worth mentioning that the estimation of directionality in connectomics is problematic in general with present approaches reported by one study [94] investigating an artificial neural system with known ground truth.

Before going into detail, a general scheme can be deduced from the different connectivity approaches. Each mentioned form of connectivity deals with associations between different nodes in the brain. On the anatomical level, these nodes can range from physiological entities like neurons as the basic unit, over banks of neurons, to brain areas, which can be defined e. g. by histological differences [16]. On the functional level, focusing on the neuroimaging method fMRI (sec. 2.2), the basic unit, which can be interpreted as a node, is a voxel. The next, more coarse-grained version of a node is a non-anatomically and non-functionally defined ROI. The most commonly used shape is a spherical ROI containing a certain amount of voxels. But there are also other definitions like using a cubic shape [28]. The most coarse-grained version of a node in this context is a ROI defined by anatomical atlases like

**Diffusion MRI** (the following from Stieltjes et al. [97]) is an imaging modality extracting directionality information of e. g. white matter fibers in the brain. It exploits the fact that diffusion of water molecules in parallel to fibers is facilitated and restricted perpendicular to it. By using phase encoded sequencing, this diffusion information can be extracted and used for the estimation of white matter fibers alignment. **Diffusion spectrum imaging** is a submodality of diffusion MRI. **Tractography** uses algorithms to estimate white matter tracks by constructing paths from the alignment of the diffusion tensor ellipsoids. A **backbone** in the context of dynamical systems represents a substrate, on which the system is evolving in time, represented e. g. by the coupling constant in an Ising model [71].

the AAL [85]. To accomplish that, the brains under consideration have to be normalized to fit into a standard space (sec. 2.2.1). In this space, functionally distinct regions are predefined and can be mapped onto the single brains resulting in ROIs for each subject. One ROI represents one functionally distinct brain area. This can also be accomplished by machine learning approaches like ICA (sec. 1.2) and gICA (sec. 1.3) with the advantage of ROIs defined on a single subject level. Each of those different types of ROIs can be used to represent nodes in connectivity considerations.

Let  $v_i$  and  $v_j$  be nodes on the human cortex and a connection between them shall be denoted as  $\chi_{ij}$ . Throughout all connectivity types, a matrix notation of this circumstance is common. The connection  $\chi_{ij}$  can be interpreted as the  $ij$ -th component in a connectivity matrix. Since connectivity matrices are the substrate for my following analyses, let the notation for this type of matrices be  $\mathbf{X}$  – analogously to the data substrate notation in the introduction part (part i). A connectivity matrix is symmetrical, if there is no directionality information present, otherwise  $\chi_{ij} \neq \chi_{ji}$  is possible. Thus, the most common scheme of representing connectivity in human neuroscience is by using a connectivity matrix both for calculation and visualization. The connectivity matrices  $\mathbf{X}$  are often referred to as connectomes, which are, in other words, the entities resulting from connectivity analyses.

### 3.1 STRUCTURAL CONNECTIVITY

When focusing on MRI in the context of structural connectivity, then nodes are mostly defined by **diffusion MRI**. I want to highlight the concept of structural connectivity on the basis of two publications in more detail. One study uses the structural connectome for simulations [71] and the other [43] offers the basis for it via **diffusion spectrum imaging**.

Hagmann et al. [43] reach a fine- and coarse-grained version of a structural connectome by the following five steps. First, they acquire a  $T_1$  (sec. 2.1) weighted image and DSI of each subject. Second, they do a white and gray matter segmentation of the  $T_1$ -weighted image, which are used for **tractography** and **parcellation** of the cortex, respectively. Third, they estimate the elongation of white matter fibers. Fourth, the cortex is parcellated into  $n = 66$  and  $n = 998$  ROIs. Fifth, the network is constructed by assigning a link  $\chi_{ij}$  to each pair of ROIs, i. e. nodes,  $v_i$  and  $v_j$  with at least one fiber connecting them. The strength, i. e. value, of connections  $\chi_{ij}$  is determined by the number of fibers per unit surface, with  $\chi_{ij}$  being the entries of the structural connectivity matrix  $\mathbf{X}^S$ . The two versions of their structural connectome are shown in figure 13. Like introduced above, the structural connectome is represented by a symmetrical  $n \times n$  matrix.

As an example of how to do physics using structural connectivity, I elaborate on the following investigation. In the study of Marinazzo et al. [71] the above introduced two versions of a structural connectome are used as a **backbone** in the simulation of an Ising-system. This means that the authors

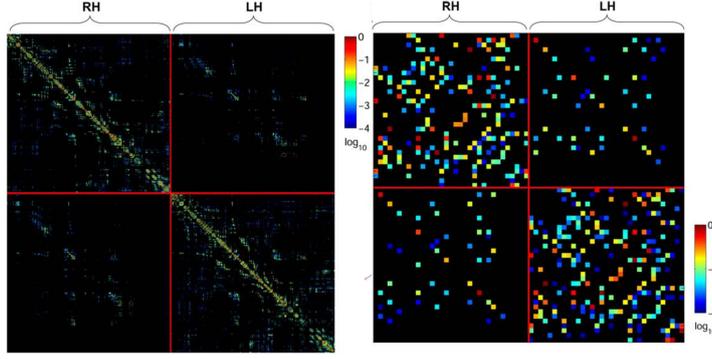


Figure 13: This figure assembled from Hagmann et al. [43] depicts their structural connectomes with 66 (right) and 998 (left) nodes. Each row and column corresponds to a certain ROI. Each matrix entry represents the color coded fiber density connecting certain ROIs. The matrices are divided into right (RH) and left hemispheres (LH).

simulate an Ising-model

$$H_i = -s_i \sum_j J_{ij} s_j \quad (82)$$

$$J_{ij} \propto x_{ij}, \quad (83)$$

where the coupling constant  $J_{ij}$  between spins  $s_i$  and  $s_j$  is proportional to the entries of the structural connectivity matrix. Two of their findings are that an Ising-system with such constraints also shows a phase transition on both scales when sweeping the temperature, and they reveal nodes, which are bottlenecks of Information flow. Simulating Ising-systems on structural connectivity matrices as backbones is a recent field in neuroscience. One study switched to the static functional connectome as a backbone in a spin-glass simulation [52]. This type of analysis is elaborated on in more detail in chapter 7. In the next sections, I introduce functional connectivity in its several variants.

### 3.2 STATIC FUNCTIONAL CONNECTIVITY

Functional connectivity is a way of revealing statistical dependencies between distinct locations in the brain [96]. Those dependencies are referred to as functional connections. To get them, the activity time courses  $a_{v_i}(t_k)$  and  $a_{v_j}(t_k)$  of nodes  $v_i$  and  $v_j$  are used to calculate association measures. The most common approach is to use Pearson's correlation coefficient (sec. A.1.1)

$$\rho_{ij} = \frac{\sum_p a_{v_i}(\tau_p) a_{v_j}(\tau_p)}{\sigma_i \sigma_j}, \quad (84)$$

where  $\sigma_i$  and  $\sigma_j$  refer to the standard deviation of the activity time courses of nodes  $v_i$  and  $v_j$ , respectively. In this notation, the activity time courses are demeaned. Note that we are dealing with discrete time steps  $\tau_p$ , since

the sampling of fMRI cannot be considered a continuous measurement procedure. Consider  $P$  the last image in an fMRI session. If  $p = 1, \dots, P$ , then value  $\rho_{ij}$  is termed static functional connectivity between nodes  $v_i$  and  $v_j$ . Thus, the static functional connectivity matrix  $\mathbf{X}^{\text{SF}}$  has  $x_{ij} = \rho_{ij}$  as entries. A typical example of a static functional connectivity matrix  $\mathbf{X}^{\text{SF}}$  is shown in figure 14. It is derived from the time courses of a rs-fMRI session of one

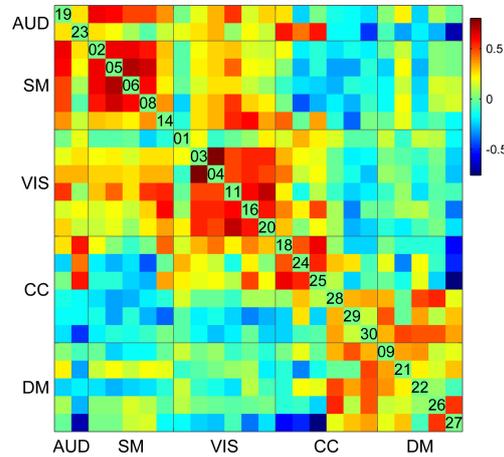


Figure 14: This figure depicts a typical example of a static functional connectivity matrix derived from an exemplary subject from the Human Connectome Project. The time courses resulted from one rs-fMRI session and are correlated with each other. This results in the shown symmetric static connectivity matrix. The time courses resulted from a gICA analysis (sec. 1.3) and the numbers in the trace reflect the component number. The components are grouped into different coarse-grained brain networks. (AUD) Auditory, (SM) somato-motor, (VIS) visual, (CC) cognitive control, (DM) default mode. The colorbar encode correlation coefficient values. Script adopted from P. Keck.

subject in the Human Connectome Project used in my analyses below. The time courses are derived from a gICA analysis and the whole time courses are correlated with each other. This means that each matrix entry represents a color coded correlation coefficient. This matrix is solely shown for exemplary purposes and to give the reader an idea of a static functional connectivity matrix. The technical details of constructing connectivity matrices will be explained in part ii, mostly focusing on dFC.

*A spike train of a neuron is the discrete firing pattern that emerges, when action potentials are recorded. Each spike represents one action potential. Note that neurons have resting activity. Thus, spike trains are not empty.*

**DIFFERENT NETWORK MEASURES** For the sake of completeness and for a better understanding, I elaborate on several other association measures used for doing static functional connectivity analysis. I highlight the advantages and disadvantages of the different measures by referring to a study from Smith et al. [94]. In this study, the authors investigate the most common measures used for static functional and effective connectivity. I concentrate solely on their functional connectivity results. They simulate the BOLD response of networks with known ground truth by estimating a balloon model (sec. 2.2) on **spike trains** occurring at each node. I elaborate on

their findings about partial correlation and mutual information compared to Pearson's correlation coefficient.

Instead of plain correlation introduced above partial correlation (sec. A.1.2) can also be used as an association measure between activity time courses of pairs of different nodes. Smith et al. [94] mention that partial correlation has the advantage of discarding **spurious correlations** by regressing out the activity time courses of nodes, which are not under consideration. This means, the functional connectivity matrix  $\mathbf{X}$  is constructed with the reduced, i. e. partial, correlation values  $x_{ij} = \tilde{\rho}_{ij}$ . In comparison to plain correlation, the authors find that partial correlation performs slightly better when it comes to the detection of the simulated network structure. They note that partial correlation as a static functional connectivity measure results in less false positives than plain correlation, because spurious information is regressed out. Another association measure for static functional connectivity is mutual information (sec. A.1.3). Smith et al. [94] mention that it has high values, if one random variable helps predicting another, and that it employs higher order statistics, which lack in plain correlation. Since mutual information is symmetric in commuting the two random variables under consideration, also the static functional connectivity matrices are symmetric. They find that mutual information performs significantly worse in detecting the underlying network structure than the other two measures mentioned here. To be more specific, they found that in most of their simulations partial correlation as a static functional connectivity measure finds more than 90% of the true connections, slightly less with plain correlation, and less than 50% with mutual information. Following their analyses, the widely used Pearson's correlation coefficient as *the* representative of functional connectivity is a robust and reliable measure, when it comes to network detection.

**Spurious correlations** result from indirect correlations e. g. by a modulating variable. Consider a network, where A feeds B feeds C, then A and C are spuriously correlated [94].

**EMERGENCE OF STATIC FUNCTIONAL CONNECTIVITY** Static functional connectivity is interwoven with rs-fMRI as a method. It has been the first method to be applied on the first rs-fMRI study conducted by Biswal et al. [14]. The authors apply seed-based correlation analysis on a single axial slice covering the left and right primary motor area. They calculate a complete cross-correlation matrix for all voxels of this slice lying in gray matter of a rs-fMRI scan in eleven subjects. At that early point in static functional connectivity in the context of rs-fMRI, the authors have been interested in the functional connectivity of specific areas – the bilateral primary motor areas. Those ROIs are determined by a finger tapping task. As a measure of static functional connectivity between the two primary motor cortices, they counted the number of voxels in each of those two areas significantly correlating with the voxels from the contra-lateral area. They find that even in the rs-fMRI condition a significant number of voxels are correlated, i. e. functionally connected, between hemispheres in those two regions. These findings have been the first hints towards a non-noise source in the brain at rest and the first application of static functional connectivity to rs-fMRI data.

## 3.3 DYNAMIC FUNCTIONAL CONNECTIVITY

In recent years, there has been a shift from the static analysis of functional connectivity to so-called dFC. One of the main studies introducing a dynamical view on functional connectivity has been done by Allen et al. [8]. The authors used rs-fMRI data from 405 healthy subjects with a session duration of 5 min 4 s and a TR = 2 s resulting in 152 EPI images for each subject. The following description of dFC is extracted from this study.

Instead of calculating Pearson correlation  $\rho_{ij}$  using whole activity time courses originating from two distinct brain regions, like introduced above, dFC is based on the segmentation of activity time courses  $a_{v_i}(\tau_p)$ . With  $\tau_p, p = 1, \dots, P$  representing the whole time course of an rs-fMRI session, segmentation is done by using a subset  $\tau_{p+\xi t}, p = 1, \dots, w$ , where  $w < P$ ,  $\xi \in [1; w]$ , and  $t = 1, \dots, T$  such that  $w + \xi T \leq P$ . This approach is also called a sliding window approach, which is applied in the study of Allen et al. [8]. The window  $w$  segments the activity time courses into more or less overlapping parts. In their study,  $\xi = 1$  meaning that the window is shifted by one TR every step. After segmenting the activity time courses, Pearson correlation is calculated for each  $t$  and node pair  $v_i$  and  $v_j$

$$\rho_{ij}(t) = \frac{\sum_{p=1}^w a_{v_i}(\tau_{p+\xi t}) a_{v_j}(\tau_{p+\xi t})}{\sigma_i(t) \sigma_j(t)}, \quad (85)$$

with  $\sigma_i(t)$  being the standard deviation of segment  $t$ . This results in changing connectivity over time  $x_{ij}(t) = \rho_{ij}(t)$ , or rather correlation functions for each  $x_{ij}(t)$ , and eventually in a set of correlation matrices  $\{\mathbf{X}(t)\}_{t=1, \dots, T}$ . In figure 15, the time evolution of such a set and single correlation functions are depicted.

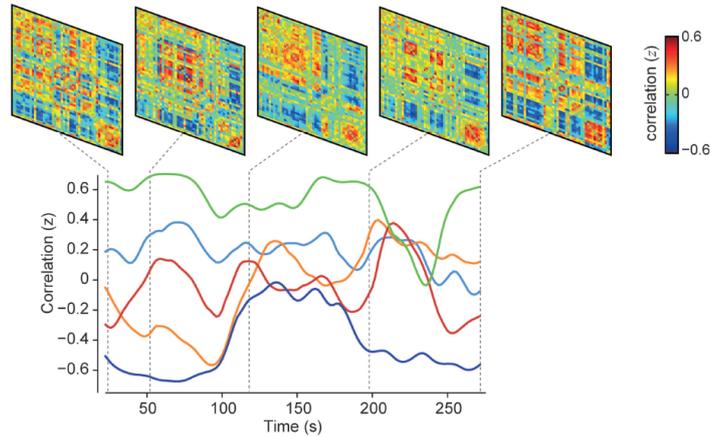


Figure 15: This figure from Allen et al. [8] depicts the temporal evolution of functional connectivity when a sliding window approach is applied. (Top) For each window a correlation matrix is calculated, which results in a set of correlation matrices varying over time. (Bottom) Each of the matrix entries is represented by a correlation function. Here, the evolution of five different entries over time is depicted.

Since I rely on the analysis pipeline introduced by Allen et al. [8], I illustrate their concept of dFC analysis on time courses resulting from a gICA analysis on rs-fMRI data here, and elaborate on it in more detail in section 4.2 when using the data set employed in this thesis. Their concept is to represent the single subject data matrices by a decomposition resulting from gICA  $\mathbf{Y}_i = \mathbf{T}_i \mathbf{S}_i$ , where the  $\mathbf{T}_i$  contain the time courses corresponding to the spatial ICs in  $\mathbf{S}_i$ . This means, by applying gICA on the EPI images they get spatial maps of certain brain areas, representative for the whole group of subjects. Then, by back reconstructing to the single subject level, they get the time courses  $\mathbf{T}_i$  corresponding to the spatial ICs  $\mathbf{S}_i$  for each subject separately. Afterwards, on those time courses a sliding window procedure is applied resulting in a set of correlation matrices  $\{\mathbf{X}_i(t)\}_{t=1, \dots, T-w}$  for each subject. The temporal variability of functional connectivity is captured in those sets of correlation matrices. Figure 16 illustrates the approach of Allen et al. [8].

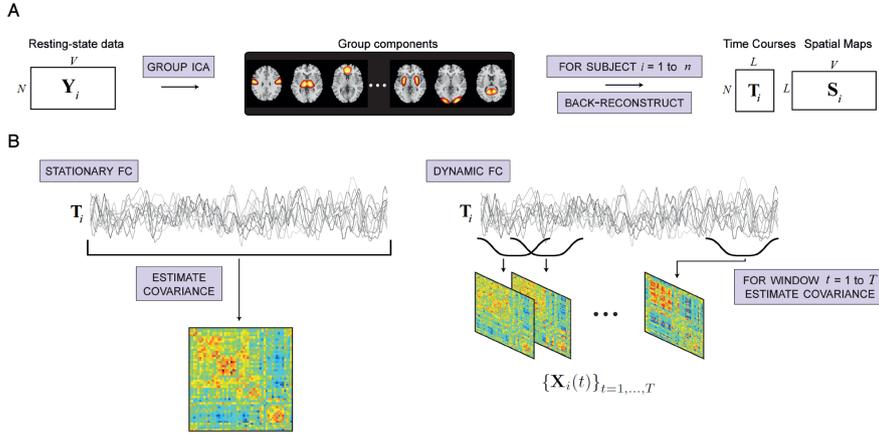


Figure 16: From Allen et al. [8] and adjusted. (A) summarizes the application of gICA on rs-fMRI data from the single subject data matrices  $\mathbf{Y}_i, N \times V$ , over the group ICs representing certain brain areas, to the representation of the single subject data matrices via back reconstructed single subject time courses  $\mathbf{T}_i$  and spatial ICs  $\mathbf{S}_i$ . (B) Here, stationary, or rather static, functional connectivity is contrasted to dFC. In the static case, whole time courses are used for constructing the representative correlation matrix. In the dynamic case on the right, the application of the sliding window approach on the time courses  $\mathbf{T}_i$  is depicted. A correlation matrix is derived from each segment of the time course resulting in a set of correlation matrices for each subject  $\{\mathbf{X}_i(t)\}_{t=1, \dots, T-w}$ .

In summary, independent of the type of nodes  $v_i$ , there are two crucial steps for tracking the variability of functional connectivity over time. First, segmenting the activity time courses of interest  $a_{v_i}(\tau_p)$ , and second, calculating association measures between pairs of nodes  $v_i$  and  $v_j$  for each segment separately. For my analysis below, I decided to stick with nodes represented by brain areas resulting from gICA applied on rs-fMRI data.

## 3.4 CONNECTIVITY-STATES

The concept of so-called connectivity-states, on which I am elaborating in this section, has also been introduced by Allen et al. [8]. Concerning the following text, I rely on their description of a connectivity-state.

Consider  $\mathbf{X}_i(t)$  one connectivity matrix derived from time course segment  $t$  having dimensions  $C \times C$ , with  $C$  being the number of used nodes. As introduced above, the association measure for calculating connectivity shall be Pearson correlation. This results in a symmetrical matrix  $\mathbf{X}_i(t)$  with  $R = \frac{C(C-1)}{2}$  unique entries. Each of those matrix entries can be interpreted as a dimension in an  $R$ -dimensional space. By the temporal variability of functional connectivity introduced by dFC analysis, the correlation coefficients of each entry  $x_{ij}(t)$  in  $\mathbf{X}_i(t)$  varies over time, which is equivalent to a change of the value on the corresponding dimension in the  $R$ -dimensional space. This means that each correlation matrix  $\mathbf{X}_i(t)$  can be represented as one data point in this  $R$ -dimensional space and the set of correlation matrices from all subjects  $\{\mathbf{X}_i(t)\}_{t=1, \dots, T-w; i=1, \dots, n}$  populates this space. This is the starting condition after dFC analysis to extract connectivity-states.

The idea behind this concept is that each subject traverses a certain trajectory in this high-dimensional space, where each  $\mathbf{X}_i(t)$  represents one data point on this trajectory, and that there are stable patterns of connectivity common to all subjects. Consider  $\mathbf{C}$  a stable pattern. In the view of dFC represented in a high-dimensional space, a stable connectivity pattern means that  $\mathbf{X}_i(t) = \mathbf{C}$  for a considerable number of time steps  $t$ . This would be the ideal case, where the time evolving connectivity  $\mathbf{X}_i(t)$  stays constant for a certain time interval, defining one stable pattern. When dealing with real data, such an ideal case cannot be expected. Assuming that there are common intrinsic stable patterns, noise and variability between subjects yields a blurring of those patterns deviating from a single point in  $R$ -dimensional space. This means that a stable pattern is not defined as a single point  $\mathbf{C}$  in  $R$ -dimensional space, but rather as a cluster of several connectivity matrices  $\mathbf{X}_i(t)$  in the vicinity of  $\mathbf{C}$ .

Following this line of thought, assuming that there are such stable patterns present in the data, an appropriate way of extracting them is to apply clustering algorithms, like  $k$ -means clustering, on the set of correlation matrices  $\{\mathbf{X}_i(t)\}_{t=1, \dots, T-w; i=1, \dots, n}$ . The  $k$ -means algorithm (sec. 1.4) extracts a predefined number of clusters  $k$ . Those clusters are represented by centroids  $\mathbf{C}_z, z = 1, \dots, k$ , which are then termed connectivity-states. Here, we stick with matrix notation instead of the introduced vector notation of a centroid  $\mathbf{c}_z$ , since our substrate consists of connectivity matrices. The logic is the same as in the introductory section. Each connectivity matrix and connectivity-state can also be written as a column vector with  $R$  entries.

In their study, Allen et al. [8] suggest that there are seven connectivity-states inherent to their data set. They attribute distinct function to each of those states. In figure 17, those seven connectivity-states are shown. Each matrix represents a centroid  $\mathbf{C}_z$  resulting from  $k$ -means clustering. They

arrange their **resting state networks** (RSNs) resulting from gICA in subcortical (SC), auditory (AUD), somatomotor (SM), visual (VIS), cognitive-control (CC), default mode (DM), and cerebellar (CB) networks. The authors describe state 1 as the connectivity-state resembling the static functional connectivity, which also has the highest occurrence probability. They found that most of the states exhibit anti-correlations of the default mode RSNs with sensorimotor RSNs, and that the latter RSNs often fuse with the other sensory RSNs (AUD + VIS) to a large module. They mention states 6 and 7 showing positive correlations of some default mode regions with sensorimotor regions, and state 4 shows a positive correlation of some frontal with default mode RSNs, which are typically un- or anti-correlated. Moreover, they describe an additional feature of discrimination between the connectivity-states, which they found in the functional connectivity of subcortical and cortical regions. They highlight that state 3 exhibits anti-correlation of subcortical and sensorimotor RSNs in contrast to the remaining states, where this correlation tends to be positive, and that the typical default mode behavior vanishes in state 3.

*In the sense of ICA, a **resting state network** consists of brain regions that are intrinsically connected and therefore show up in one IC [8].*

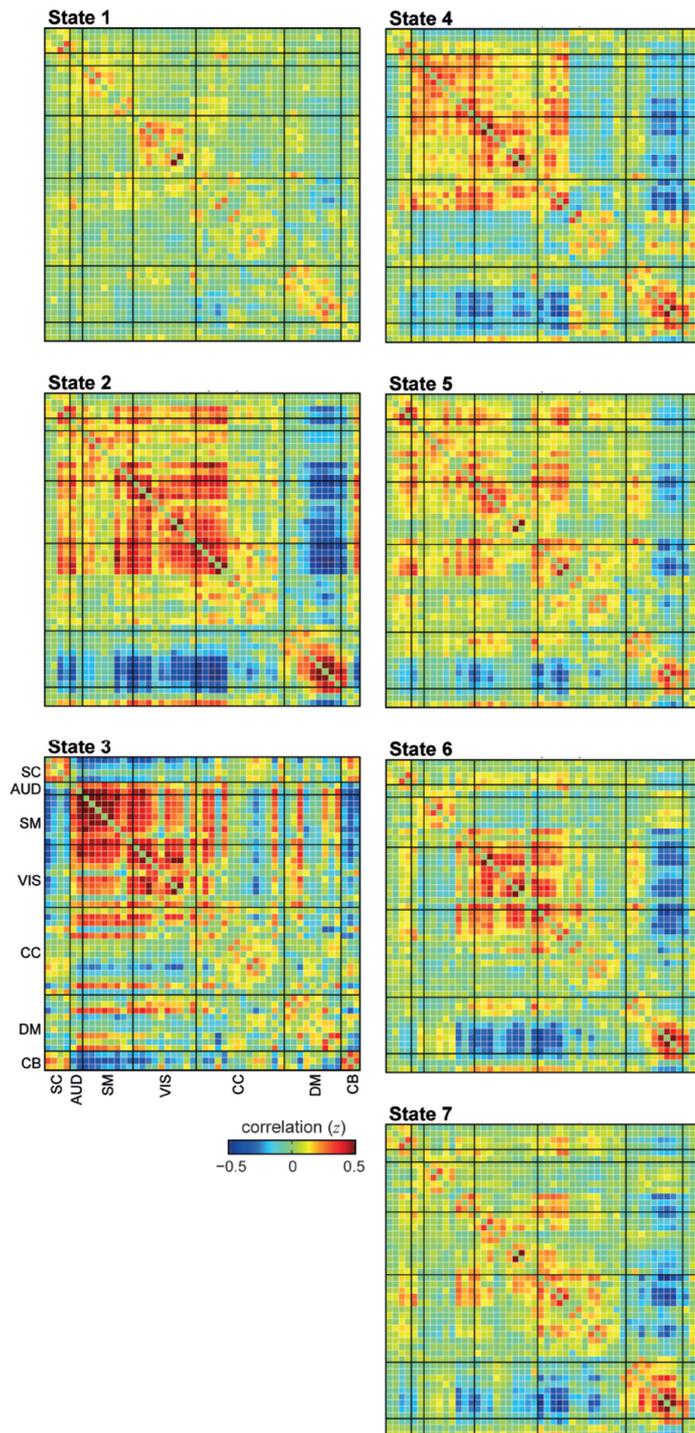


Figure 17: Taken from Allen et al. [8]. In this figure, the seven connectivity-states from the study of Allen et al. [8] are shown. Each matrix represents a centroid resulting from k-means clustering. The RSNs are arranged in sub-cortical (SC), auditory (AUD), somatomotor (SM), visual (VIS), cognitive-control (CC), default mode (DM), and cerebellar (CB) networks.

## Part II

### FREQUENCY-RESOLVED DYNAMIC FUNCTIONAL CONNECTIVITY

The following part reports the main ideas and results of my developed approaches. I start by describing the process of creating time courses by means of gICA and frequency-resolving them using MEMD and Butterworth filter banks (ch. 4). Next, I introduce different frequency-resolving methods to the field of dFC. The resulting frequency-resolved dynamic functional connectivity (frdFC) identifies connectivity-states as a multiscale entity, besides the already known multivariate properties. I apply a sliding window procedure on the MEMD decomposed time courses of rs-fMRI data (sec. 5.2). Using this method, I develop the approach of entering frdFC. Since MEMD has properties of a dyadic filter bank [82], in a *post hoc* analysis I investigate the application of a sliding window procedure on time courses decomposed by Butterworth filter banks on the same datasets (sec. 5.3). Both ways of entering frdFC reveal scale-invariance features of data-inherent connectivity-states. This claim is strengthened by simulations of dFC traversing artificial connectivity-states (sec. 5.4). The usage of GLASSO procedure is justified in section 5.5. The main results and ideas of this part can be found in an arXiv.org publication [40], which serves as a backbone for the following chapters. For my studies I rely heavily on the GIFT toolbox<sup>3</sup> and its functions. I use, adopt, and change scripts conveyed with this toolbox – and also use code snippets of GIFT toolbox functions for assembling code.

---

<sup>3</sup> <http://mialab.mrn.org/software/gift/>



*"What I cannot create,  
I do not understand."  
(Richard P. Feynman)*

In this chapter, the data set and extraction of time courses by means of gICA are explained. The pipeline of this extraction – starting from the EPI images entering gICA to the preprocessing of the time courses  $T_i$  resulting from this decomposition – is adopted from the study of Allen et al. [8].

#### 4.1 DATA-SET

The data used in this thesis originates from the so-called Human Connectome Project<sup>4</sup> [103], which offers freely available rs-fMRI data for download. Because subjects are acquired over the course of five years, there are several versions of the data. I use, from the so-called S500 release, a preselected bundle of 100 non-related subjects. Those subjects have undergone four rs-fMRI sessions á 14.4 min resulting in  $n = 400$  sessions. In each session, 1200 images with a repetition time of  $TR = 0.72$  s and an echo time of  $TE = 33.1$  ms are acquired using a gradient-echo EPI sequence. Flip angle for the EPI images was  $52^\circ$  with an field of view of  $208 \text{ mm} \times 180 \text{ mm}$ , 72 slices are acquired with a slice thickness of 2 mm yielding a voxel resolution of  $2 \text{ mm} \times 2 \text{ mm} \times 2 \text{ mm}$ . For this purpose, customized 3 T MRI scanners at Washington University have been used employing multi-band technique with 8 band factor [74, 29, 90, 111]. Three versions of preprocessed data are available for download: unprocessed, minimally preprocessed, and fully preprocessed including ICA-FIX denoising [39, 55, 56, 30, 95, 88, 42]. I choose the latter preprocessing level for the data, to have the most noise free material available. This includes structural preprocessing, and regressing out movement covariates and artifact time courses extracted by means of ICA from voxel time-series. The acquired EPI images are spatially smoothed by a Gaussian kernel with an full width at half maximum of  $5 \text{ mm} \times 5 \text{ mm} \times 5 \text{ mm}$  using SPM<sup>5</sup> software package for MATLAB. To account for magnetization saturation effects, the first five images of each session are discarded yielding  $T = 1195$  images in total for each session.

<sup>4</sup> <http://humanconnectome.org/>

<sup>5</sup> <http://www.fil.ion.ucl.ac.uk/spm/>

## 4.2 EXTRACTING TIME COURSES USING GROUP ICA

The EPI images enter the gICA processing pipeline described in section 1.3. For the application of gICA on the used rs-fMRI data, I use the GIFT Toolbox<sup>3</sup> and with it relying the following workflow on the one introduced in the study of Allen et al. [8]. The data matrix  $\mathbf{X}_{nT \times N}$  consists of the T EPI images from all n sessions concatenated in the row dimension, and the column dimension represents the  $N = 193965$  voxels of the three-dimensional images reshaped to a row vector. This spatial dimension is further reduced by applying a mask to the functional data only including important voxels. The voxel time courses are normalized in variance by linear detrending and z-scoring as a preprocessing step. In the single session data reduction step, each data matrix  $\mathbf{X}_{i,T \times N}$  with  $i = 1, \dots, n$  is reduced to  $K = 45$  PCs in the row dimension resulting in the reduced single session data matrices  $\mathbf{Y}_{i,K \times N} = \mathbf{W}_i \cdot \mathbf{X}_i$ . Those matrices are concatenated in the row dimension yielding  $\tilde{\mathbf{Y}}_{nK \times N}$  and then the group data reduction step is applied on the resulting matrix extracting  $L = 30$  PCs yielding  $\tilde{\mathbf{Y}}_{G,L \times N} = \mathbf{W}_G \cdot \tilde{\mathbf{Y}}$ . For PCA the expectation maximization algorithm [87] is used. On the group reduced data matrix, the Infomax algorithm [12] is applied extracting L ICs  $\mathbf{S}_{L \times N} = \mathbf{A}^{-1} \cdot \tilde{\mathbf{Y}}_G$  employing the ICASSO procedure [47] for estimating robust ICs by repeating the Infomax algorithm ten times with different initial conditions. The back reconstruction step using the GICA3 approach [27] yields the representation  $\mathbf{X}_i = \mathbf{W}_i^{-1} \cdot \mathbf{W}_{G,i}^{-1} \cdot \mathbf{A} \cdot \mathbf{S}_i$  with  $\mathbf{W}_{G,i}^{-1}$  resulting from matrix partitioning. For each of the 30 ICs in  $\mathbf{S}_i$  there is a corresponding time course in  $\mathbf{T}_i \leftarrow \mathbf{W}_i^{-1} \cdot \mathbf{W}_{G,i}^{-1} \cdot \mathbf{A}$ . Some of the 30 ICs, which I refer to as RSNs, contain mostly signal representing activity of certain brain areas. However, there are ICs representing so-called artifactual sources originating from sources like blood vessels, ventricles, movement, or white matter. A generic problem of ICA decompositions is to classify signal ICs from artifactual. I follow an approach used by Allen et al. [7] and introduced by Robinson et al. [84]. The authors propose to use two characteristics derived from the spectra of the ICs. They suggest, for each IC in  $\mathbf{S}$ , to calculate the so-called high-to-low-frequency ratio (LF to HF ratio) using the average spectra of the corresponding time courses from all sessions. They define the ratio as the fraction of two areas under the average spectrum – the area of the low-frequency band ranging from 0 Hz to 0.1 Hz, and of the high-frequency band ranging from 0.15 Hz to 0.25 Hz. As a second classification criterion, they introduce the so-called dynamic range, which represent the difference of the maxima of the curve in the low- and high-frequency band. The two measures are depicted in figure 18. Allen et al. [7] use those measures as two dimensions in a plot shown in figure 19, where each IC is represented as one point. This figure shows the result of the classification of the ICs resulting from the decomposition of the data used in this thesis. Visual inspection (fig. 20A) of ICs 7, 15, and 17 allow to conclude that these are artifactual ICs representing ventricles (IC 7), blood vessels (IC 15), and white matter (IC 17) as sources. ICs 10, 12, and 13 represent basal ganglia, cerebellar, and visual

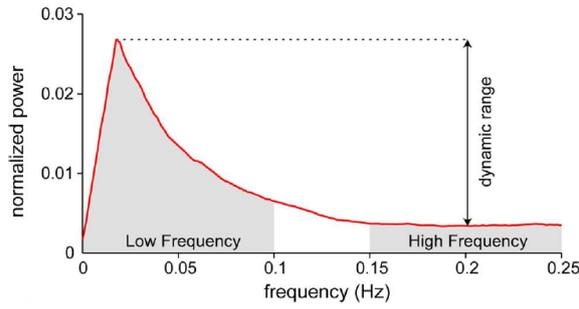


Figure 18: This figure depicts the two measures used by Allen et al. [7], Robinson et al. [84] to classify signal and artifactual ICs on an exemplary spectrum. The authors define the dynamic range as the difference of the maxima of the two frequency bands, and the low-frequency-to-high-frequency ratio as the fraction of the area under the curve of those two bands. From Allen et al. [7].

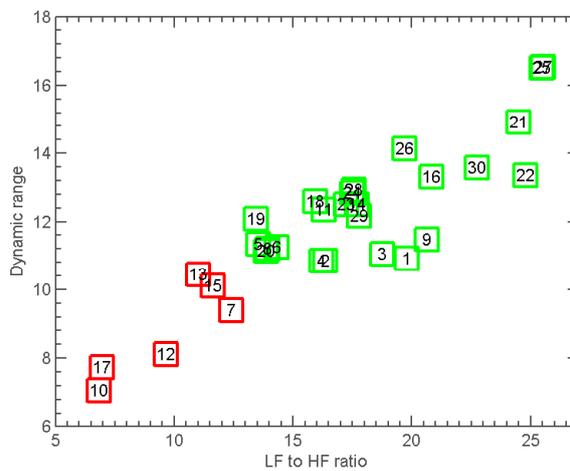


Figure 19: This figure depicts the classification scheme from Allen et al. [7], Robinson et al. [84] applied on the data used in this thesis. The ICs are classified into 24 RSNs and 6 artifactual/noisy components. Numbers represent the different ICs from **S**.

cortex sources, respectively, but their spectral characteristics are comparably bad or worse than those of the 'best' artifact IC. To decide, which of the ICs showing low values in both measures to discard, I evaluate my approach described below for data sets containing 27, 26, and 24 RSNs. In the data set with 27 RSNs, ICs 10, 12, and 17 are discarded as artifactual. In the data set with 26 RSNs, ICs 10, 12, 17, and 13 are discarded as artifactual. And in the data set with 24 RSNs, ICs 10, 12, 17, 13, 15, and 7 are discarded as artifactual. For all three types of data sets, comparable preliminary results show up. Therefore, I decide to use the most conservative data set containing 24 RSNs in order to deal with noise free data. The time courses of those 24 RSNs are the substrate for the following analyses. Additionally, the spatial maps  $\mathbf{S}_i$  are z-scored and are shown in figure 20. Before resolving time scales into separate frequency bands and before doing the standard dFC approach,  $\mathbf{T}_i$  undergo certain preprocessing steps. To get rid of trends in the time courses, I apply detrending up to cubic polynomial degree. Additionally, so-called de-spiking is conducted, where outliers in intensities, presumably induced by motion, are removed by spline interpolation using clean parts of the remaining time course. A similar approach can be found in the study of Power et al. [79]. Last, a low-pass filter is applied to all time courses in each  $\mathbf{T}_i$  with a cut-off frequency of  $\omega_c = 0.15$  Hz.

### 4.3 FREQUENCY-RESOLVED TIME SCALES

In the next sections, I describe different methods of decomposing  $\mathbf{T}_i$  into separate frequency scales. I adopt the term frequency scale from EMD literature (see e. g. [68, 3, 23]) and also extend this term to time-series covering different frequency ranges resulting from other decomposition methods. The decomposed time scales are referred to as  $\mathbf{U}_i^f$ , which are IMFs in the case of MEMD as decomposition method and time-series evolving on different frequency ranges otherwise.

#### 4.3.1 MEMD

The first way of getting frequency-resolved time scales is conducted by applying the MEMD approach introduced in section 1.5.3 on each  $\mathbf{T}_i$ . To accomplish that, I modify the script<sup>6</sup> provided by Rehman and Mandic [81] to conduct the ensemble noise-assisted MEMD instead of plain MEMD. The preprocessed  $\mathbf{T}_i$  enter the MEMD pipeline and IMFs are extracted. For each IMF  $\mathbf{U}_i^f$  an ensemble of 30 realizations is created adding 4 noise channels to the 24 signal channels. Noise is generated for each realization with a noise amplitude of 6% of the signal power averaged over all signal channels. The low-pass filtering restricts the time courses  $\mathbf{T}_i$  to similar frequency ranges for each  $i$ . This yields a comparable number of IMFs over all sessions and realizations. The most common number of IMFs is 10 in preliminary runs.

<sup>6</sup> [http://www.commsp.ee.ic.ac.uk/~mandic/research/memd/memd\\_version\\_2.zip](http://www.commsp.ee.ic.ac.uk/~mandic/research/memd/memd_version_2.zip)

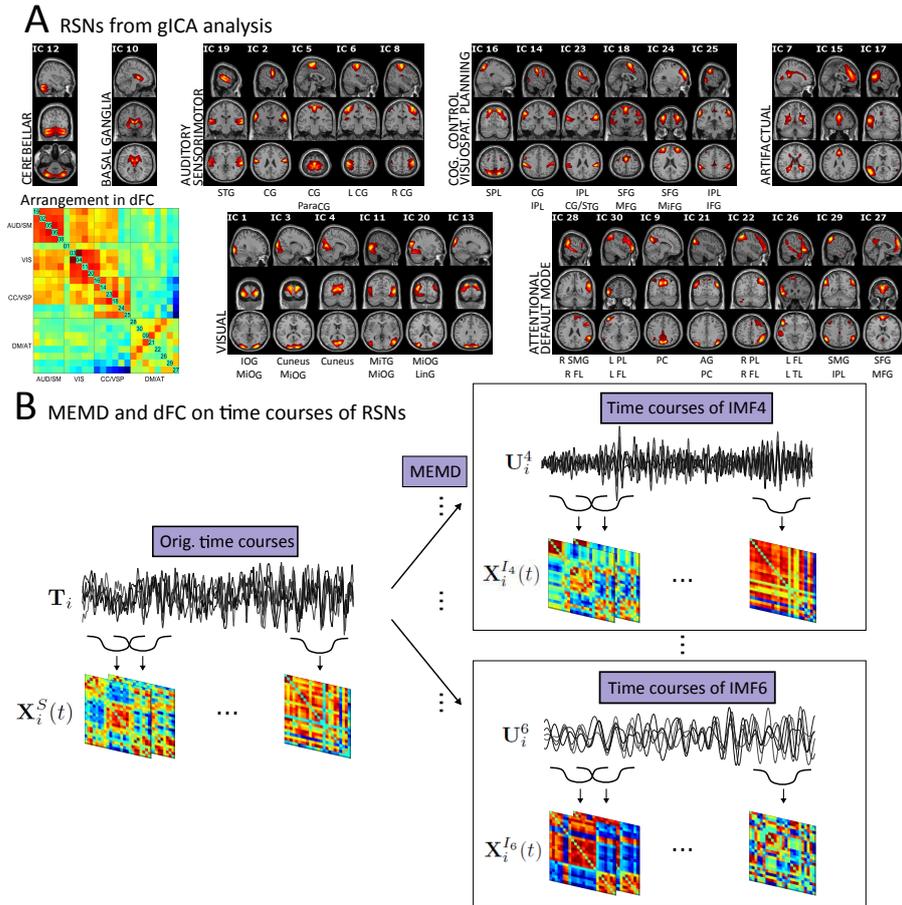


Figure 20: This figure depicts the result of the gICA decomposition of the 400 rs-fMRI sessions and the MEMD decomposition of the time courses corresponding to those RSNs. (A) shows three slices of each RSN intersecting in the peak activation voxel. The threshold for activated regions is chosen to be  $z > 2$  and the color range depends on the largest voxel value of the spatial activation map. Each IC represents the z-scored average over all 400 corresponding single-subject components. The RSNs are grouped into cerebellar, basal ganglia, auditory/sensorimotor (AUD/SM), cognitive control/visuospatial planning (CC/VSP), visual (VIS), and attentional/default mode networks (DM/AT). Additionally, the three artifactual components are also shown. In the bottom left of panel (A), the arrangement of the RSNs is shown in the average dFC matrix over all sessions. The numbers on the diagonal correspond to the IC numbers. Abbreviations: Superior Temporal Gyrus (STG), Central Gyrus (CG), Superior Parietal Lobule (SPL), Inferior Parietal Lobule (IPL), Superior Frontal Gyrus (SFG), Medial Frontal Gyrus (MFG), Middle Frontal Gyrus (MiFG), Inferior Frontal Gyrus (IFG), Inferior Occipital Gyrus (IOG), Middle Occipital Gyrus (MiOG), Middle Temporal Gyrus (MiTG), Lingual Gyrus (LinG), Supramarginal Gyrus (SMG), Frontal Lobule (FL), Parietal Lobule (PL), Precuneus (PC), Angular Gyrus (AG), Temporal Lobule (TL), Superior Frontal Gyrus (SFG). (B) Shows the MEMD decomposition of the original time courses  $\mathbf{T}_i$  on the left with the applied sliding window approach resulting in correlation matrices  $\mathbf{X}_i^S(t)$ , and on the right this is analogously depicted for two exemplary IMFs ( $\mathbf{U}_i^4$  and  $\mathbf{U}_i^6$ ) resulting in frdFC  $\mathbf{X}_i^{I_4}(t)$  and  $\mathbf{X}_i^{I_6}(t)$ . All connectivity matrices in this figure depict non-transformed correlation coefficients.

I use this information to adjust the algorithm to result in at least  $F = 10$  IMFs by discarding realizations with less than 10 extracted IMFs [38]. The discarded realizations are replaced by ones with at least 10 extracted IMFs. This procedure ensures that no residual was taken into account when creating the ensemble averaged IMF  $\mathbf{U}_i^f$  for each index  $f = 1, \dots, 10$ . The noise channels are discarded from the data as well as IMFs with indices above 10 before creating the ensembles. The procedure is depicted in figure 20B, where **projections** of two exemplary IMFs ( $\mathbf{U}_i^4$  and  $\mathbf{U}_i^6$ ) are shown. All 10 projections from one IC of one IMF is shown in figure 21A.

*Since MEMD results in multivariate IMFs, I distinguish between **IMF projections** and **IMFs**. The former represents the projection of the multivariate IMF onto one of the dimensions, which can be seen analogously to an IMF resulting from plain EMD. In my case, such a projection represents the time course of one RSN/IC.*

I investigate the spectral characteristics of the MEMD decomposition. Like mentioned by Mandic et al. [70], I find that mode alignment works well with each IMF index resulting in a narrow frequency band (fig. 21B). This is estimated by calculating the instantaneous frequencies for all projections of  $\mathbf{U}_i^f$  using HHT<sup>6</sup> (sec. 1.5.4) and averaging over all time points and RSNs for each  $f$ . This results in a characteristic frequency for each  $f$  and  $i$  resulting in 400 data points for calculating the depicted box-plots. Additionally, I calculate the power spectra of each RSN time course in each  $\mathbf{U}_i^f$  and average the spectra over all sessions and RSNs. The resulting spectra are normalized to the maximum value to investigate the extent of overlapping. Visual inspection of figures 21B and C allows to conclude that the resulting frequency bands are narrow and minimally overlapping. Only scales 1 and 2 show overlapping spectra, which can be attributed to the fact that IMFs of these indices mostly contain noise. Additionally, I calculate the average cross-correlation and covariance matrices of all IMF projections. This was accomplished by calculating the covariance and correlation coefficient of each possible pairing of IMF projections of different scales. Afterwards, the resulting covariance or correlation matrices for each IMF projection is averaged – over RSNs and sessions. The resulting matrices (fig. 22) illustrate that there is only slight covariance/correlation between IMFs of adjacent scales, which also confirms the low extent of mode mixing.

#### 4.3.2 Butterworth filter banks

As a next approach, I use Butterworth filter banks to decompose the time courses  $\mathbf{T}_i$  into frequency-resolved time scales  $\mathbf{U}_i^f$ . To accomplish that I rely on advice, suggestions, and scripts provided by Prof. Dr. A. Tomé. One important difference between using MEMD and Butterworth filter banks as frequency decomposition methods is the data-driven manner of the former one. Since filter banks are not data-driven, several parameters have to be pre-defined: filter bandwidth, filter order, number of bands  $f$ . Preliminary results show that the narrow frequency scales shown in figure 21B – especially in the low frequency scales – cannot be emulated properly using a filter bank. Therefore, filter banks with constant bandwidth and varying number of extracted bands  $F$  are used. In general, for the lowest frequency band, a low-pass filter is used. All other bands are extracted applying band-pass filters. Additionally, two types of filter banks are employed. One with constant or-

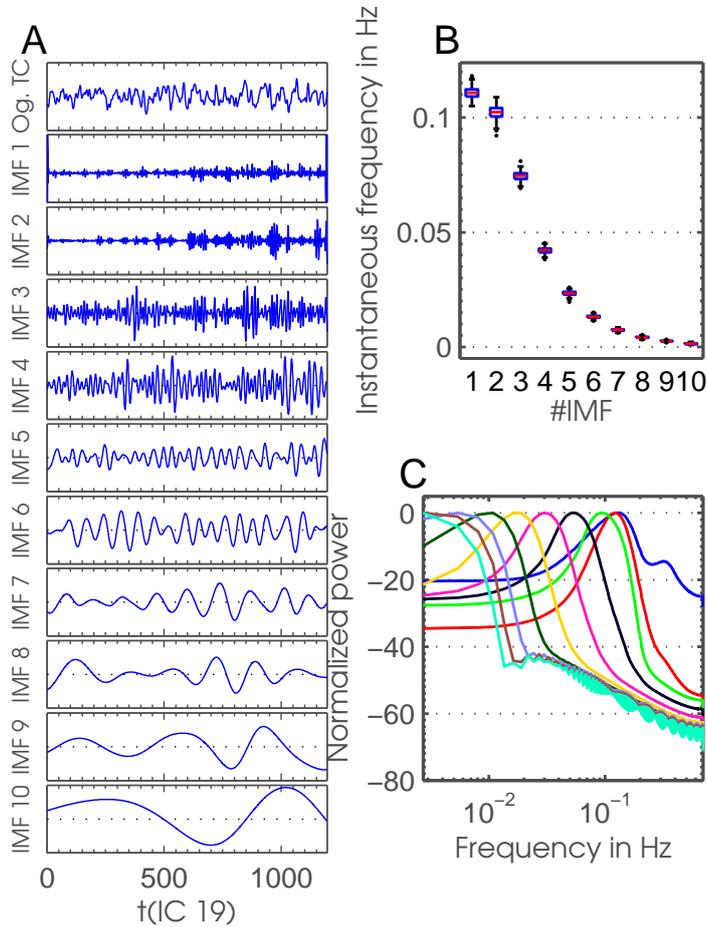


Figure 21: (A) This panel shows the result of the decomposition using the MEMD approach. At the top, the original time course is shown followed by the projections of  $\mathbf{U}_{322}^f$  for  $f = 1, \dots, 10$  representing IC 19. (B) depicts box-plots of the representative instantaneous frequency for each scale #IMF separately. Every box-plot was derived from 400 data points, each representing one session as the average instantaneous frequency over all RSNs and time points. (C) shows the average power spectra of the scales derived from all IMFs. Each spectrum represents the average over all sessions and RSNs. The spectra are normalized to the maximum value for investigating the extent of overlapping.

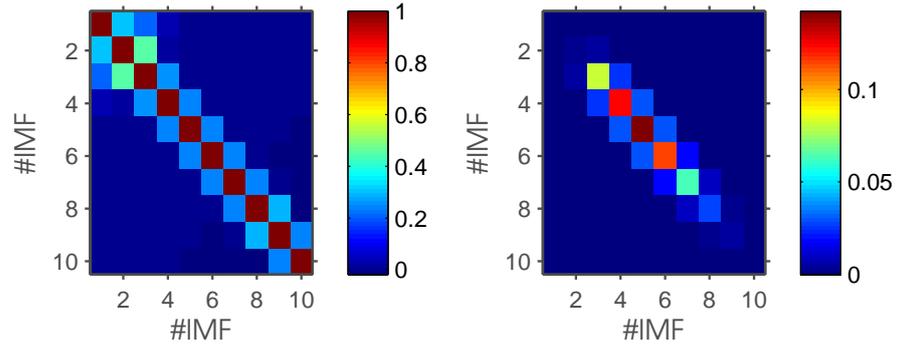


Figure 22: (Left) This panel shows the average of the correlation matrices resulting from correlating each possible pairing of IMF projections between different scales  $\#IMF$ . This means for each IMF projection one correlation matrix results, and averaging is done over all RSNs and sessions. (Right) This panel shows the same for covariance of IMF projections between different scales.

der for all filters and one with adjusted filter order. In the constant case, one order is chosen for all filters to result in a stable pole behavior. In the adjusted case, filter order is chosen to result in merely stable pole behavior. In other words, for each band a filter order is chosen to result in highly narrow frequency profiles of  $\mathbf{T}_i$ .

I choose to use filter banks with a number of extracted frequency scales of  $F = 5, 8, 10, 12, 15$ . The bandwidth  $\Delta\omega$  of each filter bank is selected to separate the frequency range  $\omega = [0; 0.15 \text{ Hz}]$  of the original time courses  $\mathbf{T}_i$  into disjunct frequency scales  $f$  with constant bandwidth  $\Delta\omega = \frac{0.15 \text{ Hz}}{F}$ . In the case of a constant filter order for each extracted band, filter order is chosen to be 10 for  $f = 5, 8$ ; 8 for  $f = 10, 12$ ; 6 for  $f = 15$ . For the adjusted case, the filter orders are shown in table 1. In figure 23 the time-series and

Table 1: In this table, the filter order used for the Butterworth filter banks in the adjusted case are shown. Note that the first frequency scale  $f = 1$  is realized by a low-pass filter and all other  $f$  by band-pass filters.  $F$  is the number of scales that are extracted by applying filter banks on the time courses  $\mathbf{T}_i$ .

		f														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
F	5	12	6	6	7	8										
	8	10	4	5	6	6	7	7	7							
	10	9	4	5	5	5	6	6	6	6	7					
	12	9	4	4	5	5	5	6	6	6	6	6	6			
	15	8	3	4	4	5	5	5	5	5	5	5	6	6	6	6

spectrum characteristics resulting from the application of a filter bank with constant filter order extracting 10 frequency scales is depicted. In panel 23B an estimation of the frequency range over all sessions is analogously calcu-

lated to the estimation done in figure 21B. Since the filter bank procedure does not ensure stationary time courses on different frequency scales, the HHT is not applicable. To circumvent this issue, I calculate the Fast Fourier transform of the time courses of each RSN in  $U_i^f$  and average the resulting spectra over all RSNs. This yields a representative spectrum for each of the 400 sessions and 10 frequency scales. The data points for each box-plot in figure 23B are the frequencies, at which the representative 400 spectra peak. This means that for each frequency scale again 400 data points are taken into account when building the box-plots. The equidistant, constant bandwidth of the extracted frequency scales is evident.

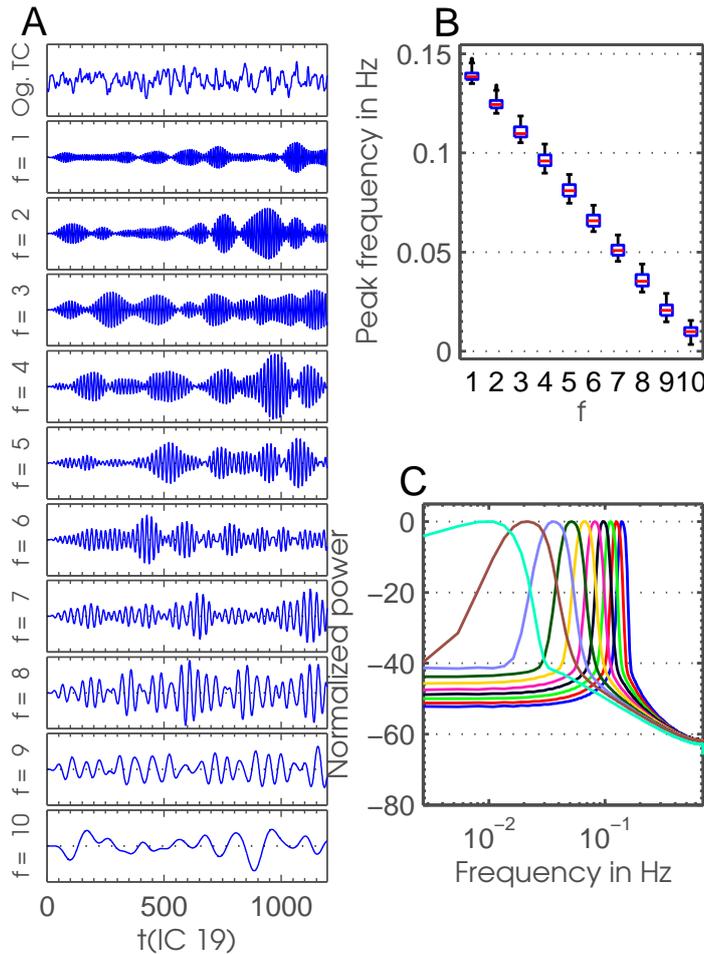


Figure 23: (A) This panel shows the result of the decomposition using a Butterworth filter bank with constant filter order and 10 extracted frequency scales. At the top, the original time course is shown followed by the scales  $U_{322}^f$  for  $f = 1, \dots, 10$  representing IC 19. (B) depicts box-plots of the representative peak frequency for each scale  $f$  separately. Every box-plot is derived from 400 data points, each representing one session as the peak frequency of the average spectra over all RSNs. (C) shows the average power spectra. Each spectrum represents the average over all sessions and RSNs. The spectra are normalized to the maximum value for investigating the extent of overlapping.



## FREQUENCY-RESOLVED DYNAMIC FUNCTIONAL CONNECTIVITY

---

*"I don't know anything,  
but I do know that everything is interesting,  
if you go into it deeply enough."  
(Richard P. Feynman)*

In this chapter, I describe the dFC procedures used in this thesis. First, I replicate the approach of Allen et al. [8] on the data introduced in section 4.1. In the next sections, I apply the same sliding window procedure on frequency-resolved time scales resulting from MEMD and Butterworth filter banks. On each of those sets of correlation matrices resulting from different time-series, k-means clustering is employed to extract connectivity-states.

### 5.1 STANDARD DYNAMIC FUNCTIONAL CONNECTIVITY ON THE HUMAN CONNECTOME PROJECT DATA

In section 3.3, I have introduced the concept of dFC. In this section, I describe in detail the parameters used for applying dFC on the Human Connectome Project data set, and in parallel the detailed method Allen et al. [8] uses, on which I rely on. This means that the workflow introduced here bases on the study of Allen et al. [8] and the given sequence from the MATLAB toolbox GIFT, which has a dFC plugin. Furthermore, the following results also serve as replication of both, the standard dFC procedure, and the subsequent application of k-means clustering to extract connectivity-states.

The sliding window procedure is applied with a window size of  $w = 80TR$ , which resembles a time interval of 57.6 s. I choose to use a boxcar shape for the used windows instead of the tapered, Gaussian windows used in the original study. I investigate the influence of switching from tapered to boxcar windows and find similar results in the evolution of dFC between the two types of windows. The sliding window procedure is first applied on the time courses resulting from gICA  $T_i$  with a step size of  $\alpha = 1TR$ . This results in  $T - w = 1115$  time steps, or rather, correlation matrices in total for each of the  $n = 400$  sessions<sup>7</sup>. The correlation matrices are estimated using the so-called GLASSO procedure for estimating precision matrices [36]. This implementation in the GIFT toolbox is based on the fact that in dFC the time courses used for creating correlation matrices are relatively short, which can lead to problems when estimating correlation coefficients [104].

---

<sup>7</sup> Because of a rounding error, the first window is always discarded when the sliding window procedure used in this thesis is applied. Holds whenever the sliding window method is applied in this thesis.

Additionally, the correlation coefficients are Fisher transformed by applying  $\text{atanh}$ .

After the application of the sliding window procedure on the preprocessed  $\mathbf{T}_i$ , the k-means clustering algorithm is applied on the complete set of correlation matrices  $\{\mathbf{X}_i(t)\}_{t=1, \dots, T-w; i=1, \dots, n}$ . The clustering algorithm is started with initial conditions derived from a single k-means++ [9, 38] run. For the following analyses,  $k = 2, \dots, 10$  centroids, or rather connectivity-states, have to be extracted. The k-means algorithm runs up to maximally 200 iterations using a city-block distance measure. Centroids are calculated using the component-wise median. The extracted connectivity-states for each  $k$  can be seen in the top row of figures 53, 54, 25, 55, 56, 57, 58, 59, 60. The connectivity-states resulting from the standard dFC procedure are arranged concerning their occurrence probability beginning with the most frequent connectivity-state on the left-hand side.

In the following, I compare those connectivity-states with the ones derived by Allen et al. [8] depicted in figure 17. I refer to their connectivity-states with the prefix A. Since frdFC and scale stability analysis below show that four connectivity-states seem to be the most valid number of data inherent connectivity-states, I restrict the comparison to the four connectivity-states shown in figure 25. I can identify state A1 in all k-means runs, which Allen et al. [8] refer to as the state resembling static functional connectivity. This state can be found in column S1b in figures 53, 54, 25, 56, 57, 58, 59, 60. State A2 can also be identified in the data used in this thesis. It has a typical pattern showing mostly DMN components anti-correlating with most of the other components. This state can be found in column S2b in figure 25, and column S3b in figures 54, 55, 58, 60. State A3 resembles the states in column S2b in figures 53, 54; S4b in figure 25; S5b in figures 55, 56; S7b in figure 57; S8b in figures 58, 59; S10b in figure 60. In this state the anti-correlation of the DMN components is absent, which serves as a unique characteristic assigning those states to state A3. State A4 shows slight anti-correlation of some cognitive control network components with mostly auditory, somatomotor, and visual components. This behavior can be found in states corresponding to column S3b in figure 25; S4b in figures 55, 56; S6b in figures 58, 59; S10b in figure 60. I want to highlight the most dominant anti-correlation found for the components 24 and 25 in the data used in this thesis resembling the brain areas intraparietal lobule, middle frontal gyrus, inferior frontal gyrus (see fig. 20A, bottom left). Those brain areas can also be found in the above mentioned slightly anti-correlating cognitive control network components of state A4.

This results suggest that standard dFC analysis applied on the data set used in this thesis yields comparable results with the used parameters and settings. This is encouraging for the following alterations introduced by frdFC analysis.

### 5.1.1 Choosing a boxcar function as window shape for sliding window procedure

The choice of a boxcar shape for the sliding window procedure is justified by its application on time-series evolving on low frequency scales and its implementation in the GIFT toolbox. In the GIFT toolbox, the sliding window procedure is implemented by convolving a boxcar with a Gaussian function resulting in a tapered window. This function is defined on the whole time course as a vector instead of restricting it to the values defining the window, which means that there are zero entries in this vector besides the non-zero entries defining the window. Since the size of the window  $w = 80$  is small compared to the whole time course  $T = 1195$ , the additional zeros in the vector introduce spurious correlation depending on the mean of the time course segment in the window. This spurious correlation is introduced, because the time courses are z-scored before correlation. This yields non-zero values for entries with value zero, which in turn introduces the above mentioned spurious correlations. Additionally, in the GIFT toolbox, the tapered window erroneously overlaps in the beginning and at the end of the time course introducing unnecessary noise to the derived functional connectivity. To avoid both the additional zeros and the additional noise, I decide to use a boxcar function restricted only to the segment of the time course of interest. Figure 24 illustrates both issues on the basis of the implementation of a tapered window for the first time step.

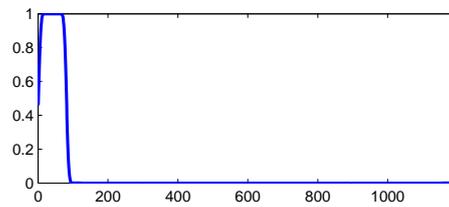


Figure 24: This figure depicts a tapered window implementation of the GIFT toolbox. The two issues of additional zeros and overlapping parts of the window in the beginning and at the end of the vector are obvious. This is the implementation of a window used for the first time step. The erroneous overlapping is also a problem for the windows of the last few time steps.

## 5.2 FREQUENCY-RESOLVED DYNAMIC FUNCTIONAL CONNECTIVITY BY MEANS OF MEMD

The next step in this thesis is to apply a frequency decomposition of time courses  $T_i$  before employing the sliding window procedure. This is done by using MEMD (sec. 1.5.3) first, followed by *post hoc* analysis using Butterworth filter banks. Using a data-driven approach like MEMD as the first means of choice is sensible, since the number of data inherent frequency scales is unknown.

The extraction of time courses used for entering frdFC have been described in section 4.3.1. For each session  $i$  frequency scales defined by the resulting IMFs  $\mathbf{U}_i^f$  result, with  $f = 1, \dots, 10$ . On each  $\mathbf{U}_i^f$  a sliding window procedure with the same parameters introduced in section 5.1 is applied. Correlation values are also Fisher transformed. This results in frequency-resolved sets of correlation matrices  $\left\{ \mathbf{X}_i^{I_f}(t) \right\}_{t=1, \dots, T-w}$  for each session  $i$  and frequency scale  $f$ .

On the whole sets of correlation matrices  $\left\{ \mathbf{X}_i^{I_f}(t) \right\}_{t=1, \dots, T-w; i=1, \dots, n}$  for each frequency scale  $f$ ,  $k$ -means clustering is applied to extract connectivity-states on each frequency scale. Analogously to the application on the original time courses  $\mathbf{T}_i$ ,  $k$ -means is applied extracting  $k = 2, \dots, 10$  connectivity-states. Analogously to the standard dFC case, the  $k$ -means algorithm runs up to maximally 200 iterations using a city-block distance measure, and centroids are calculated using the component-wise median. The clustering algorithm is initialized ten times with varying initial seeds. The seeds are derived by one run of the  $k$ -means++ algorithm before each of the ten main runs. This procedure reduces the risk that results from the clustering approaches are dependent on the initial seeds. For each  $k$ -means run,  $F \times k$  connectivity-states are extracted. Those  $F \times k$  plot arrays are shown in figures 25, 53-60. In those figures, non-transformed correlation values are shown.

#### 5.2.1 Robustness of connectivity-states over frequency scales

Visual inspection of the  $F \times k$  connectivity-states for each  $k$  over all  $F$  frequency scales reveals similarities in their structure over separate scales. The first step towards the quantification of the extent of similarity of connectivity-states over scales is the application of ordering algorithms to the arrays of the  $F \times k$  connectivity-states (see fig. 25 as an example with  $k = 4$ ). I employ a self-developed, but suboptimal, ordering algorithm and the optimal solution of such an assignment problem by using the so-called Hungarian algorithm [61]<sup>8</sup>. Applying two different ordering algorithms strengthens my results concerning reliability. Before the ordering procedure, the correlation values of the used connectivity-states are Fisher back transformed applying  $\tanh$  to compare the original correlation structure.

The suboptimal algorithm (alg. 4) correlates – in a second-order correlation manner – each correlation matrix  $\mathbf{X}_j^{I_{f+1}}$  of scale  $f + 1$  and column  $l$  in the plot arrays with the average of the correlation matrices  $\overline{\mathbf{X}}_j^{I_{1, \dots, f}}$  over scale  $1, \dots, f$ , which then is the representative of column  $j$  of the  $F \times k$  array. By exploiting this information, the algorithm checks, whether it is possible to uniquely assign each connectivity-state of scale  $f + 1$  to one connectivity-

<sup>8</sup> For the application of the Hungarian method, I rely on scripts offered by <http://www.mathworks.com/matlabcentral/fileexchange/6543-functions-for-the-rectangular-assignment-problem>. Using the Hungarian algorithm was suggested by a journal reviewer.

**Algorithm 4** Ordering states over different frequency scales**Input:**  $\mathbf{X}_j^{I_f}$  correlation matrix of scale  $f$  and column  $j$ 

- 1: **for**  $nIter = 1 : numIter$  **do**
- 2:   Shuffle scales
- 3:   Check, if unique assignment of states from scale  $f + 1$  to the average of the preceding scales is possible by
 
$$l_j = \arg \left\{ \max_{\forall l} \left[ \text{corr} \left( \bar{\mathbf{X}}_j^{I_{1 \dots f}}, \mathbf{X}_l^{f+1} \right) \right] \right\}, \forall j$$
- 4:   Assign  $l_j$  to corresponding  $j, \forall j$  with  $l_m \neq l_n, \forall m, n$
- 5:   Get all states with  $l_m = l_n$  and assign  $l_m$  with maximum correlation coefficient, go to step 2) with remaining unassigned states
- 6:   After all states are assigned go to step 2) and increase scale index  $f$  by one
- 7:   After final scale is aligned, save configuration and go to step 1)
- 8: **end for**

state of the preceding scale  $f$  concerning their similarity. For this assignment problem, the average over the preceding  $1, \dots, f$  scales is taken to avoid outlier sensitivity. If it is not possible to uniquely assign all states in the first run, then there are several connectivity-states of scale  $f + 1$  showing highest correlation values on the same connectivity-state of scale  $f$ . In this case, the algorithm assigns the connectivity-state of scale  $f + 1$  correlating highest with the corresponding connectivity-state of scale  $f$ . Afterwards, it correlates the remaining correlation matrices  $\mathbf{X}_j^{I_{f+1}}$  from scale  $f + 1$  with the averages of the remaining columns  $\bar{\mathbf{X}}_j^{I_{1 \dots f}}$ , and tries to find a unique assignment. Otherwise, this procedure is repeated until all connectivity-states from scale  $f + 1$  are assigned. This is done for all  $F$  scales and then the configuration is saved. The averaging procedure of the first  $1, \dots, f$  scales in each column  $j$  introduces statistical dependencies. Furthermore, the assignment procedure also depends on the configuration of scales. To account for these two aspects, the algorithm shuffles frequency scales before assignment starts using only non-repeating configurations. For each shuffling run  $nIter$ , the configuration after assignment is saved yielding  $numIter = 500$  ordered configurations of  $F \times k$  connectivity-states for further analysis. If  $F! < 500$ , then all possible permutations  $numIter = F!$  are used. The plot arrays shown in this thesis depict the results of assignment without shuffling frequency scales using one realization of seeds, but including the connectivity-states from standard dFC into the averaging procedure for aligning according to those states.

Steps 3 – 5 constitute my suboptimal version of an ordering algorithm. These steps can be replaced by the optimal solution provided by the so-called Hungarian method [61, 75]. Munkres [75] introduces the rating matrix  $\mathbf{R}$  of an assignment problem, with the elements of this matrix  $r_{ij}$  resembling the performance rating of assignment of  $M_i$  to  $J_j$ , where  $M_i$  is a worker and  $J_j$  is a job in this case, and  $i, j = 1, \dots, n$ . He formulates the assignment

problem, which is solved by the Hungarian method, as the selection of  $n$  independent  $r_{ij}$  with the constraint of maximizing their sum, where independent in this case means that none of those selected  $r_{ij}$  are in the same row or column. This assignment problem can be exactly mapped to the assignment of connectivity-states from scale  $f + 1$  to scale  $f$ , where instead of assigning a worker  $M_i$  to a job  $J_j$  a connectivity-state  $\mathbf{X}_i^{f+1}$  from scale  $f + 1$  is assigned to the representative  $\bar{\mathbf{X}}_j^{1\dots f}$  of column  $j$ . The rating matrix  $\mathbf{R}$  in my case is represented by the correlations of connectivity-states from scale  $f + 1$  with the average of the connectivity-state matrices over scales  $1, \dots, f$ . In other words, the entry  $r_{ij}$  of the rating matrix is represented by the distance measure  $1 - \rho_{ij}$ , where  $\rho_{ij}$  is the correlation coefficient resulting from the second-order correlation of connectivity-state  $i$  from scale  $f + 1$  with the average of the connectivity-state matrices over scales  $1, \dots, f$  of column  $j$ . This means, for each frequency scale  $f$  in each  $F \times k$  array of connectivity-states, the Hungarian method has to solve the assignment problem for a rating matrix  $\mathbf{R}$  of the size  $k \times k$ . For all figures in my thesis, where an ordering algorithm has been necessary, the Hungarian method is used. Comparing some results of my algorithm and the Hungarian method show that outcomes hardly differ.

In figure 53, the result of extracting  $k = 2$  connectivity-states from different frequency scales resulting from MEMD and subsequent ordering by applying algorithm 4 is depicted. State S1b is robust up to scale  $f = 8$  and S2b up to scale  $f = 7$ , but with more pronounced heterogeneity. For  $k = 3$  in figure 54, the first two states resemble the same states found for  $k = 2$ . Additionally, a third connectivity state (S3b) shows up. In this array, all three states show high robustness over scales. In the following, I refer to this robustness of states over scales as scale stability. Also for  $k = 4$  a new state (S4b) shows up being robust over scales, with scale stability of the other three states still being preserved. When extracting one more connectivity-state from each scale, this pattern of scale stability breaks down. In figure 55, a vast loss in scale stability can be observed. Three connectivity-states still show robust scale stability, which is lost for the connectivity-state found in the first column of figure 25. This loss in scale stability is also preserved when extracting  $k > 5$  connectivity-states (see figs. 56-60). The tendency of increasing correlation coefficients evident in the histograms shown in each of those figures, can be explained by the increasing period of the time courses and the constant window size used for the sliding window approach. In the next section, I introduce a measure to quantify this scale stability and investigate its variability.

### 5.2.2 Variability of scale stability

Looking at connectivity-states over frequency scales more quantitatively reveals that connectivity-states of  $k$ -means runs are highly variable concerning their similarity within states, i. e. columns in the ordered  $F \times k$  arrays, over scales. To look at scale stability of states, correlation of adjacent connec-

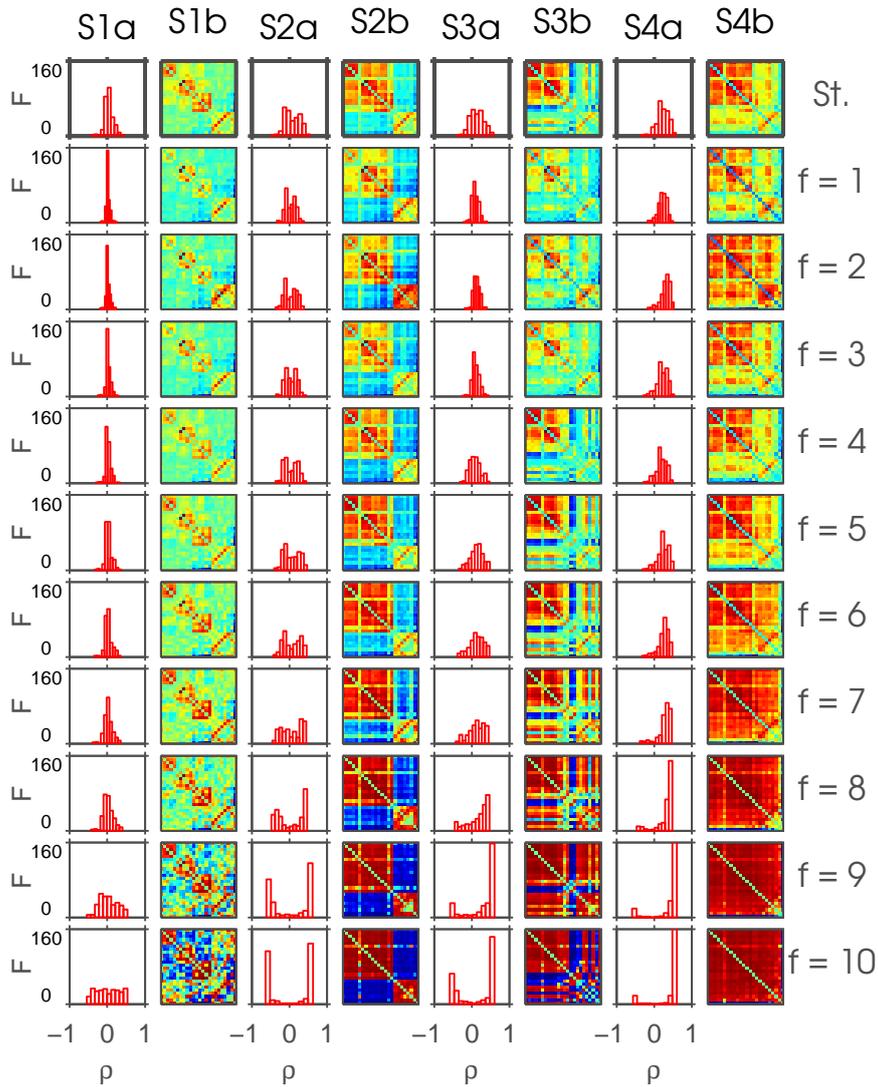


Figure 25: This figure depicts the result of the ordering algorithm 4 applied on connectivity-states resulting from frdFC procedure using  $\mathbf{U}_i^f$  extracted by MEMD. This is a realization of the ordering procedure without shuffling the frequency scales  $f$  and  $k = 4$  extracted connectivity-states on each frequency scale. In the top row, the connectivity-states resulting from the standard (St.) dFC procedure are shown. Below, each row shows the connectivity-states from frequency scales  $f = 1, \dots, 10$  with increasing frequency defined by the IMFs resulting from MEMD. Each column  $S^*b$  represents one connectivity-state, if it could be found robustly over frequency scales. Color represents the range from minimum (blue) to maximum (red) value of correlation to highlight the common structure over scales. The columns  $S^*a$  show histograms of the absolute frequency of the correlation coefficients of the corresponding connectivity-states.

tivity-states is calculated representing their similarity over nearby frequencies, i. e., after ordering connectivity-states for each  $k$ -means run, patterns of frequency scale  $f$  and  $f + 1$  are correlated. I refer to this "distance" of frequency scales as  $\Delta f$ , where in this case  $\Delta f = 1$ . This means, for each of the  $k$  columns there are  $F - 1$  correlation coefficients. To get a measure for scale stability at this particular frequency distance of  $\Delta f = 1$ , the average of all those correlation coefficients is taken. Then  $\Delta f$  is increased step by step to look at scale stability over more distant frequency scales. Correlation is done in a second-order manner looking for highest association between patterns resulting in a scale stability measure for a particular  $k$  and  $\Delta f$

$$I_{\Delta f}^{\text{sim}}(k) = \frac{1}{(F - \Delta f)k} \sum_{i=1}^k \sum_{f=1}^{F-\Delta f} \text{corr}(\mathbf{X}_i^f, \mathbf{X}_i^{f+\Delta f}). \quad (86)$$

To get a global measure of scale stability in the sense that this measures represents scale stability for each  $k$ , the average is taken over  $\Delta f$ . This results in the average scale stability index  $\overline{I_{\Delta f}^{\text{sim}}}(k)$  for each  $k$ -means run with  $k$  extracted connectivity-states. As mentioned above, the ordering algorithms are applied to `numIter` shuffled realizations of frequency scales. For each of those realizations,  $\overline{I_{\Delta f}^{\text{sim}}}(k)$  is calculated. Afterwards, the average of those `numIter` values of  $\overline{I_{\Delta f}^{\text{sim}}}(k)$  are averaged resulting in the final  $\langle \overline{I_{\Delta f}^{\text{sim}}}(k) \rangle$  global scale stability measure for a  $k$ -means run with  $k$  extracted connectivity-states. As mentioned above,  $k$ -means clustering procedures are initialized several times with varying initial seeds. This means, for each of those realizations  $\langle \overline{I_{\Delta f}^{\text{sim}}}(k) \rangle$  is calculated and its evolution with  $k$  can be plotted including error bars representing the consistency over all those realizations.

For this measure of scale stability  $\langle \overline{I_{\Delta f}^{\text{sim}}}(k) \rangle$ , the theoretical behavior can be deduced. The following idealizations are deduced from the findings in figures showing  $F \times k$  arrays. Assume, the clustering procedure always finds connectivity-states inherent to the data  $k_{\text{inh}}$ , regardless of the number  $k$  of extracted connectivity-states. In other words, when  $k$  connectivity-states are extracted by  $k$ -means and  $k_{\text{inh}} < k$  states are inherent to the data, then  $k_{\text{inh}}$  columns in the  $F \times k$  arrays show this states in a scale stable manner. Assume further that the remaining  $k - k_{\text{inh}}$  columns contain either connectivity-states with no scale stability or noise. If  $k \leq k_{\text{inh}}$  then we assume that the first  $k$  data inherent connectivity-states show up in the  $k$  columns of the  $F \times k$  array. If connectivity-states are completely scale stable, then the correlation in equation 86 shall be  $\text{corr}(\mathbf{X}_i^f, \mathbf{X}_i^{f+\Delta f}) = 1, \forall f$ . For cases with  $k \leq k_{\text{inh}}$ , this results in  $\langle \overline{I_{\Delta f}^{\text{sim}}}(k) \rangle = 1$ . As soon as  $k > k_{\text{inh}}$  columns with no scale stability show up in the  $F \times k$  arrays. These columns shall be represented by  $\text{corr}(\mathbf{X}_i^f, \mathbf{X}_i^{f+\Delta f}) = 0, \forall f$ . This means, in the case of  $k > k_{\text{inh}}$  the scale stability measure  $\langle \overline{I_{\Delta f}^{\text{sim}}}(k) \rangle$  drops. The sum over  $f$  combined with its normalizing factor  $F - \Delta f$  either results in 1 or 0, depending on whether  $i$  is

a column representing a connectivity-state or not. With this consideration in mind it is obvious that equation 86 reduces to

$$\overline{\langle I_{\Delta f}^{sim} \rangle}(k) = \begin{cases} 1, & \text{if } k \leq k_{inh} \\ \frac{k_{inh}}{k}, & \text{else.} \end{cases} \quad (87)$$

This means that the ideal evolution of  $\overline{\langle I_{\Delta f}^{sim} \rangle}(k)$  has a distinct behavior, namely, staying constant until the data inherent number of connectivity-states  $k_{inh}$  is reached and then has a drop off according to  $\frac{k_{inh}}{k}$ . In figure 26, this ideal evolution is depicted for the case with  $k_{inh} = 4$  data inherent connectivity-states.

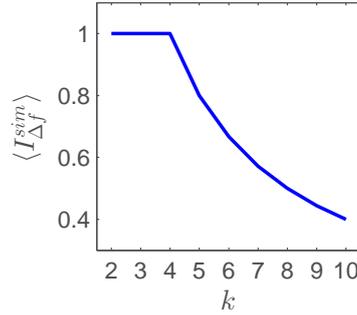


Figure 26: This figure depicts the evolution of  $\overline{\langle I_{\Delta f}^{sim} \rangle}(k)$  in the ideal, or rather theoretical, case with  $k_{inh} = 4$  data inherent connectivity-states.

In figure 27A1, the scale stability measure  $\overline{\langle I_{\Delta f}^{sim} \rangle}(k)$  is plotted for the MEMD procedure. The error bars depict the standard deviation of  $\overline{\langle I_{\Delta f}^{sim} \rangle}(k)$  for the ten realizations of  $k$ -means with varying initial seeds. Some features of the theoretically predicted behavior of  $\overline{\langle I_{\Delta f}^{sim} \rangle}(k)$  can be seen, namely a drop off from  $k = 4$  to  $k = 5$  with a constant value beforehand. This evolution of  $\overline{\langle I_{\Delta f}^{sim} \rangle}(k)$  points to  $k_{inh} = 4$  connectivity-states in the data used in this thesis. In figure 27B1, the measure from equation 86 is plotted for each  $\Delta f$ . In a recent study, Leonardi et al. [66] investigated dFC and connectivity-states by suggesting that dFC matrices should be demeaned before entering the clustering procedure resulting in better clustering. By demeaning they understand to subtract the temporal mean of each correlation function  $X_{ij}(t)$  from its time course, i. e. each entry of the dFC matrices is temporally demeaned. In figure 27A2, the scale stability measure  $\overline{\langle I_{\Delta f}^{sim} \rangle}(k)$  is shown for this type of procedure applied to the data used in this thesis. The demeaning procedure is applied after correlation coefficients have been Fisher transformed. This results in a peak at  $k = 3$ , but with a sharp drop off evident from  $k = 4$  to  $k = 5$ . Looking at the  $F \times k$  array for  $k = 4$  of the demeaned data (fig. 61), it can be seen that the state resembling static functional connectivity is absent, and the other three states are similar to the remaining states from figure 25. Also the plots of  $I_{\Delta f}^{sim}(k)$  are shown in figure 27B2 for the demeaned data. The investigation of this demeaning procedure has been suggested by a journal reviewer.

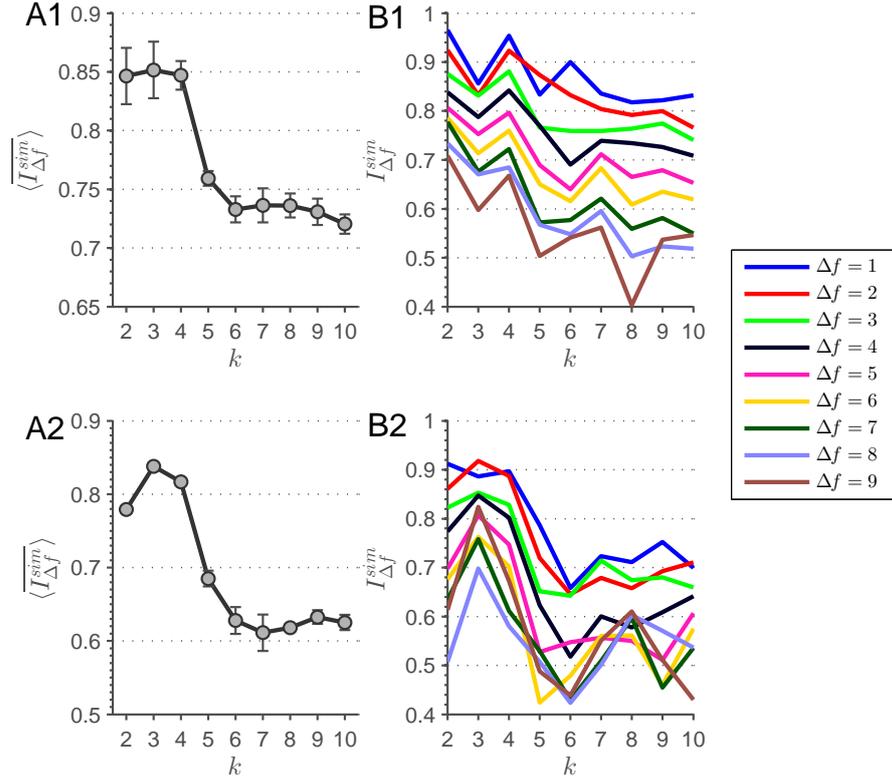


Figure 27: (A1) Here, the evolution of  $\overline{\langle I_{\Delta f}^{sim} \rangle}(k)$  is depicted for the MEMD procedure. A clear drop off is visible from  $k = 4$  to  $k = 5$  extracted connectivity-states with a constant  $\overline{\langle I_{\Delta f}^{sim} \rangle}(k)$  beforehand. In panel (B2),  $I_{\Delta f}^{sim}(k)$  is depicted for each  $\Delta f$  from which  $\overline{\langle I_{\Delta f}^{sim} \rangle}(k)$  of panel (A2) is derived. Analogously for panels (A2) and (B2), but with the alteration of demeaning the connectivity-state matrices before applying the clustering procedure. Error bars depict the standard deviation.

### 5.2.3 Null-model tests

In order to check, whether the results from the above introduced scale stability measure are not due to statistical effects, I apply my approaches to two sets of surrogate time-series. First, I randomly shuffle the time courses  $\mathbf{T}_i$  in the time domain, and second, I apply so-called phase randomization to the time courses  $\mathbf{T}_i$  similar to Allen et al. [8], but relying on the work of Theiler et al. [100], who introduced univariate phase randomization<sup>9</sup>. Allen et al. [8] applied phase randomization preserving the cross-correlation structure of the data [80], which results in the same correlation matrices as in the non phase randomized data *per definitionem*. Using univariate phase randomization yields deviating correlation matrices resulting from time courses with very similar features compared to the original time courses.

This phase randomization procedure is done in a similar way like Allen et al. [8] has done their inconsistent Fourier phase shift, which they have investigated additionally to the above mentioned one. For both procedures of generating surrogate time series,  $\mathbf{T}_i$  with  $i = 1, \dots, 400$  sessions are shuffled either in the time or phase domain. This means that each of the 24 time courses in  $\mathbf{T}_i$  is shuffled separately. Those sets of surrogate time courses undergo MEMD decomposition, sliding window, and clustering procedure similar to the original time-series. Also the scale stability measure is calculated in the same way as for the standard approach. One exemplary plot of the  $F \times k$  array for  $k = 4$  can be found in figures 62 (shuffled) and 63 (phase randomized). It can be seen that the patterns of the connectivity-states from both surrogate approaches have random appearance. Calculating scale stability results in the evolution of  $\langle \overline{I_{\Delta_f}^{sim}} \rangle(k)$  depicted in figures 28A1 and A2. Note that for the randomized time courses in the time domain, the procedure results in  $F = 11$  frequency scales. This can be explained by the introduction of higher frequencies to the time courses by the shuffling procedure. For both surrogate time series approaches,  $\langle \overline{I_{\Delta_f}^{sim}} \rangle(k)$  shows very small values of scale stability  $\langle \overline{I_{\Delta_f}^{sim}} \rangle(k) \leq 0.0294$  (shuffled) and  $\langle \overline{I_{\Delta_f}^{sim}} \rangle(k) \leq 0.0322$  (phase-randomized), and additionally a slight increase in  $\langle \overline{I_{\Delta_f}^{sim}} \rangle(k)$  with growing  $k$  is found instead of a drop off in the original data – also theoretically predicted.

### 5.2.4 Frequency dependent window size

The investigations of this section are based on suggestions made by a journal reviewer. For the sliding window procedure a constant window size of  $w(\overline{\omega}) = 57.6$  s has been used for all above mentioned approaches. In this section, I investigate the effect of varying window size on the results generated with constant  $w(\overline{\omega})$ . As a reference for adjusting the window size, I calculate the average number of periods contained in a window of size  $w(\overline{\omega})$ .

<sup>9</sup> For the implementation of phase randomization, I used the script provided by <http://www.mathworks.com/matlabcentral/fileexchange/1597-chaotic-systems-toolbox/content/phaseran.m>

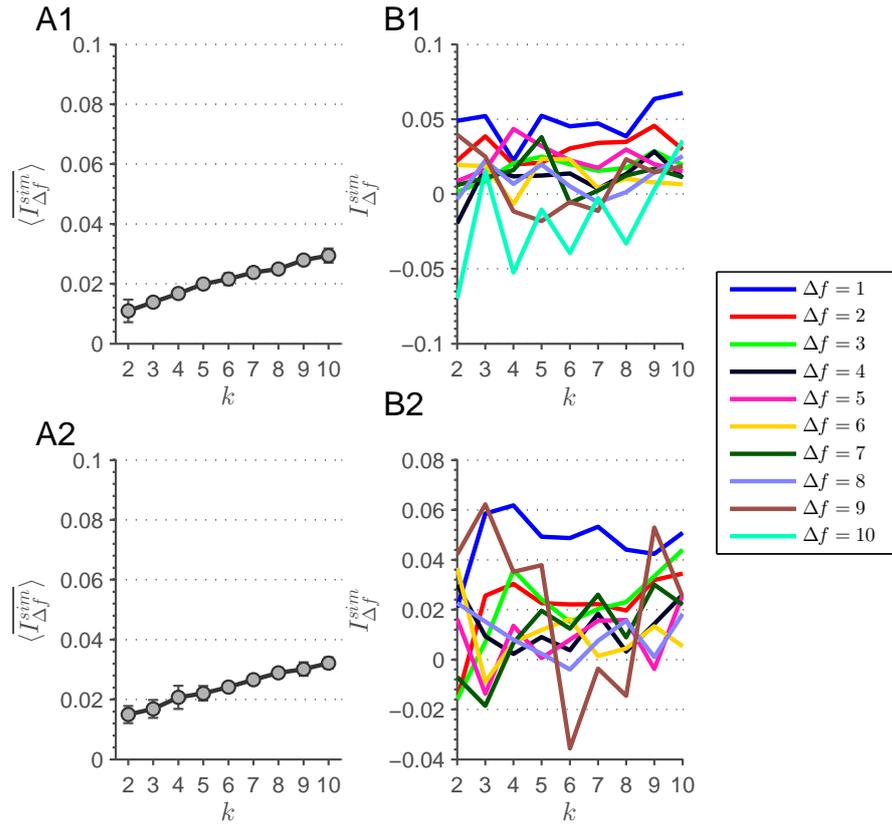


Figure 28: (A1) depicts  $\langle I_{\Delta f}^{sim} \rangle(k)$  from the shuffled and (A2) from the phase-randomized time courses. (B1) and (B2) show  $I_{\Delta f}^{sim}$ , from which  $\langle I_{\Delta f}^{sim} \rangle(k)$  is derived, for the corresponding time courses. Error bars depict the standard deviation.

To estimate this value, I calculate Fourier transformation on each RSN time course in each  $T_i$  averaging the resulting spectra afterwards. From this average spectrum, I estimate the weighted average frequency  $\bar{\omega} \approx 0.0548$  Hz, and with this the average number of periods falling into a window of constant size is  $\bar{n}_T \approx 3.157$ . I choose to investigate a continuum from constant window size to a window size containing  $\bar{n}_T$  periods at all frequency scales. This is done by calculating

$$\Delta w^\alpha(\omega_f) = w(\bar{\omega}) \left( \frac{\bar{\omega}}{\omega_f} - 1 \right) \cdot \alpha, \quad (88)$$

where  $\Delta w^\alpha(\omega_f)$  represents the adjustment in window size,  $\omega_f$  the average instantaneous frequency (see sec. 4.3.1) of the corresponding scale  $f$ , and  $\alpha$  the coefficient to sweep between a window size of constant length ( $\alpha = 0$ ) and constant period within each window over scales ( $\alpha = 1$ ). The size of the window is represented as  $w^\alpha(\omega_f) = w(\bar{\omega}) + \Delta w^\alpha(\omega_f)$ . The event time courses are then convolved by a hemodynamic response function. To probe the effect of changing window size, I choose  $\alpha \in \{0.05, 0.1, 0.2, 0.3, 0.5, 1\}$ . Figure 29 shows the used window sizes across scales, i. e. the evolution of  $w^\alpha(\omega_f)$  over scales. In the cases of  $\alpha = 0.5, f = 10$  and  $\alpha = 1, f = 9, 10$ ,

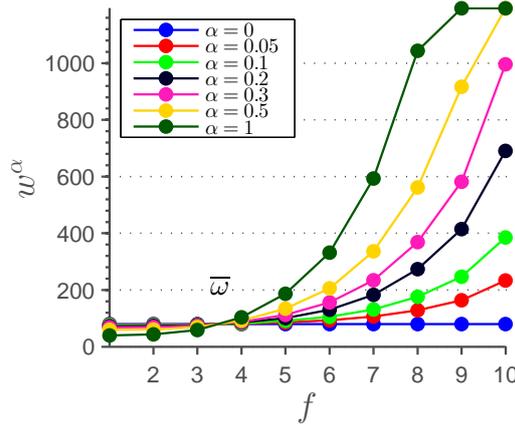


Figure 29: This figure depicts the evolution of the frequency dependent window size  $w^\alpha(\omega_f)$ . As can be seen for  $\alpha = 0.5, f = 10$  and  $\alpha = 1, f = 9, 10$ ,  $w^\alpha$  is set to  $T - 1$ , since the adjusted window size would have been larger than  $T$ .  $f$  represents the indices of frequency scales defined by IMFs resulting from MEMD, and  $\bar{\omega}$  is the weighted average frequency of all RSN time courses.

$w^\alpha$ , window size  $w^\alpha$  would be larger than the maximum number of time points  $T$ , if calculated by equation 88. Therefore, it was set to  $w^\alpha = T - 1$ .

The above definitions of window sizes are used for applying the sliding window approach on the time courses  $U_i^f$  resulting from MEMD (sec. 4.3.1). On the sets of correlation matrices, k-means clustering is applied with the same settings introduced in section 5.2. This was done for each  $\alpha$  and frequency scale  $f$ . It is worth noting that by varying window size the number of correlation matrices can differ drastically between different sets. In the

extreme case of  $\alpha = 1$ , we have 462000 matrices for scale  $f = 1$  and 400 for  $f = 10$ . After the clustering procedure, scale stability measure  $\langle I_{\Delta f}^{sim} \rangle(k)$  is calculated. The results for each  $\alpha$  are shown in figure 30. The curves for

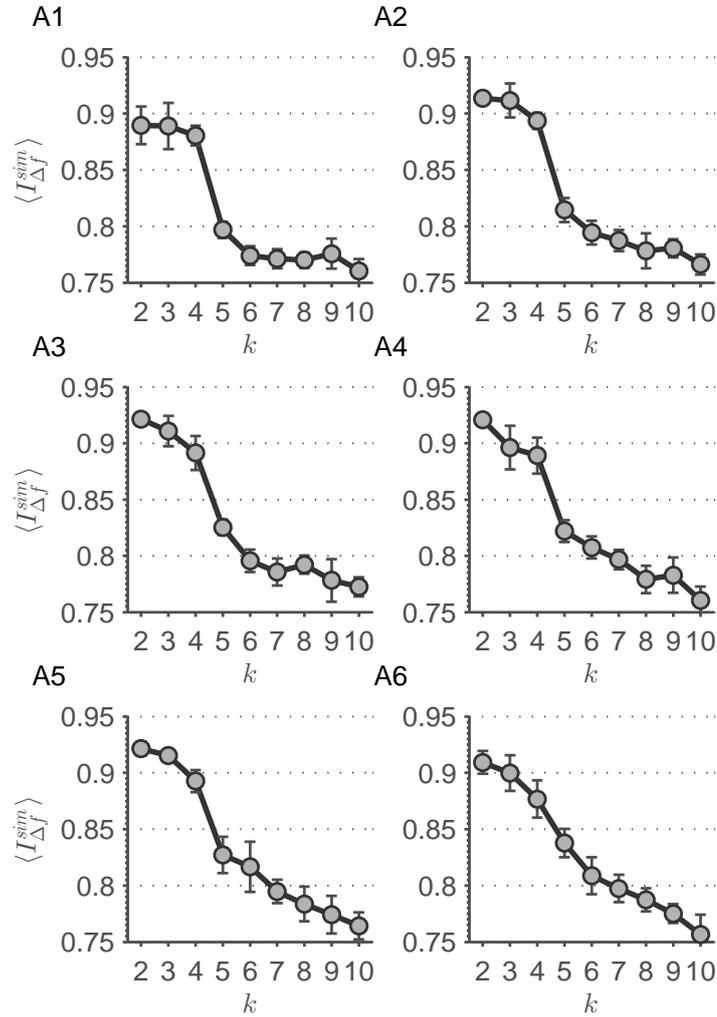


Figure 30: In this figure, the results from the frequency dependent window size procedure are shown. Each panel represents the evolution of the scale stability measure  $\langle I_{\Delta f}^{sim} \rangle(k)$  for each value of  $\alpha \in \{0.05, 0.1, 0.2, 0.3, 0.5, 1\}$ . (A1)  $\alpha = 0.05$ , (A2)  $\alpha = 0.1$ , (A3)  $\alpha = 0.2$ , (A4)  $\alpha = 0.3$ , (A5)  $\alpha = 0.5$ , (A6)  $\alpha = 1$ . Error bars represent the standard deviation of  $\langle I_{\Delta f}^{sim} \rangle(k)$ .

$\alpha = 0.05$  and  $\alpha = 0.1$  show a very similar behavior, which is comparable to the result found by using a constant window size (fig. 27A1). An overall result is that with increasing  $\alpha$  the drop from  $k = 4$  to  $k = 5$  shrinks, until there is almost a linear behavior of  $\langle I_{\Delta f}^{sim} \rangle(k)$  for  $\alpha = 0.5$  and  $\alpha = 1$ .

### 5.3 FILTER BANKS AND DYNAMIC FUNCTIONAL CONNECTIVITY

The investigations of this section are based on suggestions made by journal reviewers. The next step in validating the results found in section 5.2.2

is to vary the number of frequency scales. This is done by extracting time courses on different frequency bands by applying Butterworth filter banks to  $T_i$  resulting in  $F = 5, 8, 10, 12, 15$  scales. Also two ways of designing the filter banks is introduced referred to as constant and adjusted order approach, respectively. The procedure is described in section 4.3.2. On the resulting time courses  $U_i^f$  of the two approaches and five filter banks, the sliding window procedure and k-means clustering are applied with the same parameters used for time-series resulting from MEMD (sec. 4.3.1). Afterwards, the introduced scale stability measure  $\langle I_{\Delta_f}^{sim} \rangle(k)$  (eq. 86) is calculated for each  $F \times k$  array. Figure 31 shows the results of this approach. With increasing number of extracted frequency bands, the drop from  $k = 4$  to  $k = 5$  found by the MEMD approach (fig. 27) becomes more evident. For  $F = 5$  and  $F = 8$  there is almost no drop-off. Furthermore, differences between the constant and adjusted filter order approach are very subtle. Comparing panels A3-5 with A8-10 shows a greater uncertainty in terms of larger standard deviations adjacent to the drop off for the adjusted order approach. This means, when taking the results found in figure 27 as the desired outcome, panel 31A5 shows the clearest result with least uncertainty.

#### 5.4 SIMULATED DYNAMIC FUNCTIONAL CONNECTIVITY

The investigations of this section are based on suggestions made by journal reviewers. The above results suggest that with the combination of frequency resolution and clustering it seems to be possible to detect the number of data inherent connectivity-states. To further validate this suggestion, I create simulated time courses using a script from the study of Allen et al. [8]<sup>10</sup>, which employs the the SimTB [26]. With the SimTB from Erhardt et al. [26] it is possible to simulate rs-fMRI time courses and the mentioned script allows to simulate those time courses traversing artificial connectivity-states. I simulated 24 RSNs, with  $T = 1200$  time points, and a  $TR = 0.7$  s. In this script, first, the module structure of the simulated connectivity-states is built. This means each of the 24 simulated RSNs is assigned to a module, and those modules are defined whether they are correlating or anti-correlating with each other. Each RSN is assigned to one of three modules. This information is then used in the script to define the sequence of events for each RSN, whereas simultaneously occurring events yield high correlation. There are two event types to be defined: the simultaneous events within a module with an occurrence probability  $p_S$ , and unique events independent from those module events effecting only single RSNs. The latter had occurrence probability  $p_U$  and event amplitude  $\alpha_U$  in comparison to the amplitude of module events. The unique events can be seen as noise introduced to the time courses. Those three parameters are the degrees of freedom that have to be adjusted. Since those parameters cannot be derived from theory or empirical results and because of restrictions concerning computability, I choose an *a priori* parameter space  $p_S, p_U, \alpha_U \in \{0.6, 0.8, 1\}$ . All possible interactions (permutations)

<sup>10</sup> [http://mialab.mrn.org/software/simtb/docs/create\\_toysimulation.m](http://mialab.mrn.org/software/simtb/docs/create_toysimulation.m)

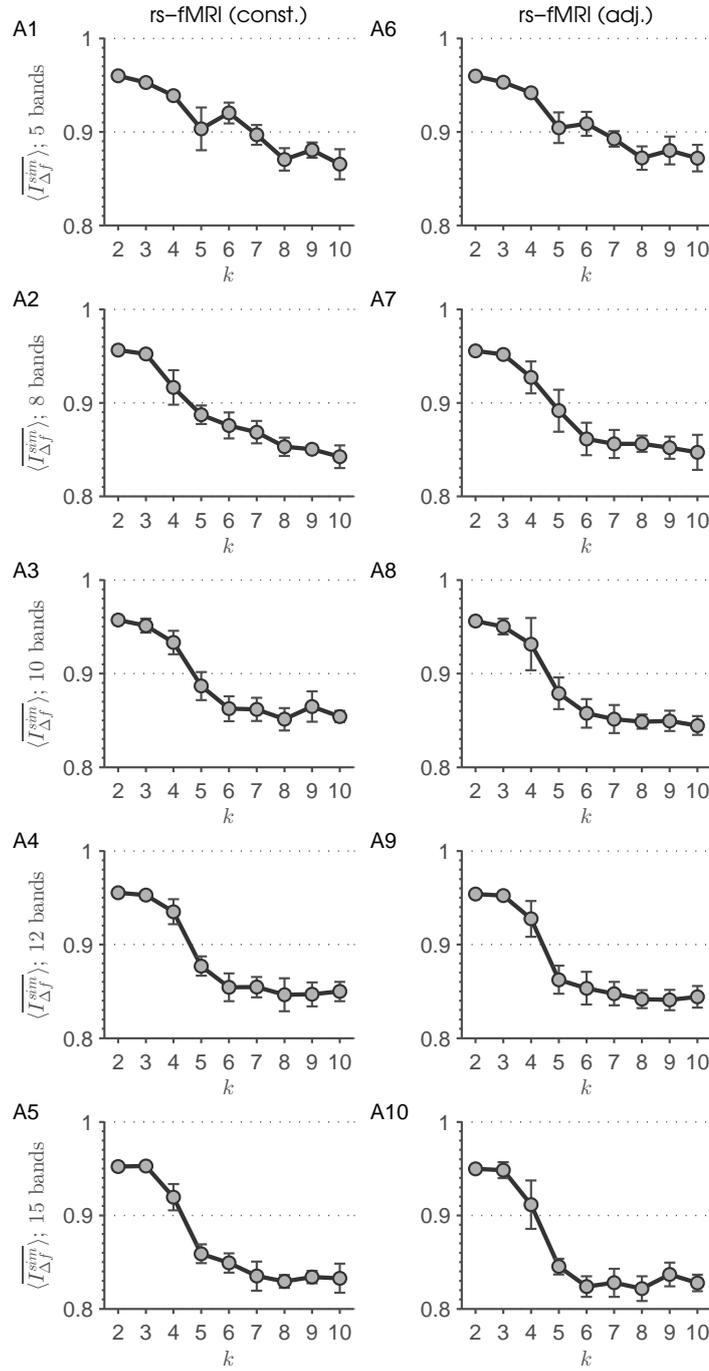


Figure 31: This figure shows evolutions of  $\overline{\langle I_{\Delta f}^{sim} \rangle}(k)$  for filter bank data of rs-fMRI time courses from the Human Connectome Project data set. (A1-5) represent results from the constant (A6-10) from the adjusted order approach. Error bars depict the standard deviation.

and main effects (keeping two parameters at 0.8 varying the remaining one) are investigated. To probe the potential of my approaches as a detection tool for data inherent connectivity-states  $k_{inh}$ , I simulated  $k_{inh} = 2, 4, 6, 8, 10$  states depicted in figure 32. The duration of each state is set to  $\Delta T = 150$  TR,

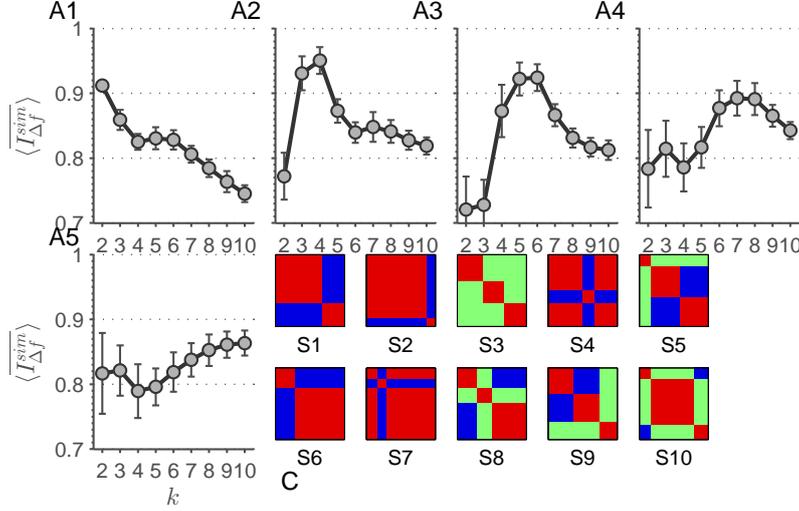


Figure 32: (A1-5) In these panels, selected results of  $\langle \overline{I_{\Delta f}^{sim}} \rangle(k)$  are shown for  $k_{inh} = 2$  (A1),  $k_{inh} = 4$  (A2),  $k_{inh} = 6$  (A3),  $k_{inh} = 8$  (A4), and  $k_{inh} = 10$  (A5) simulated connectivity-states. Error bars depict the standard deviation. (C) depicts the artificial states (S1-10) the simulated rs-fMRI time-series traverse. Each correlation matrix has the dimension  $24 \times 24$ . Blue color represents anti-correlation, red correlation, and green no correlation. The states are created by the script from Allen et al. [8]<sup>10</sup>.

and the sequence of them is random with the constraint of non-repetition of the same state (besides for  $k_{inh} = 2$ , where the states are alternating). For sake of consistency, the number of simulated sessions depends on  $k_{inh}$ , i. e. to have comparable cluster sizes over different numbers of inherent connectivity-states  $k_{inh}$ . For  $k_{inh} = 2, 4, 6, 8, 10$  the number of sessions is chosen to be  $n = 24, 48, 72, 96, 120$ , respectively. For creating the correlation matrices with a sliding window approach a window size of  $w = 80$  TR is used.

$\langle \overline{I_{\Delta f}^{sim}} \rangle(k)$  is plotted for each simulation run in figures 33-37. It can be seen that the evolution of  $\langle \overline{I_{\Delta f}^{sim}} \rangle(k)$  is distinct for different  $k_{inh}$  overall simulations (A1-13) between each figure. In particular, for  $k_{inh} = 2, 4$  the maximum of  $\langle \overline{I_{\Delta f}^{sim}} \rangle(k)$  can always be found at  $k = k_{inh}$ . Defining  $k = k_{inh}$  as the strong criterion for detecting data inherent connectivity-states, then for  $k_{inh} = 2, 4$  detection works for all parameter combinations. In the cases with more data inherent connectivity-states, the strong criterion cannot be met for detection over all parameter sets. For  $k_{inh} = 6$ , only parameter combination in panel 35A7 meets the hard criterion peaking at  $k = 6$ . For all other combinations the peak can be found at  $k = 5$ . For  $k_{inh} = 8$ , no peak in  $\langle \overline{I_{\Delta f}^{sim}} \rangle(k)$  corresponds to the right number of data inherent connectivity-states. For each parameter combination, the peak can be found at  $k = 7$ . In the case of  $k_{inh} = 10$ , the hard criterion is met for parameter

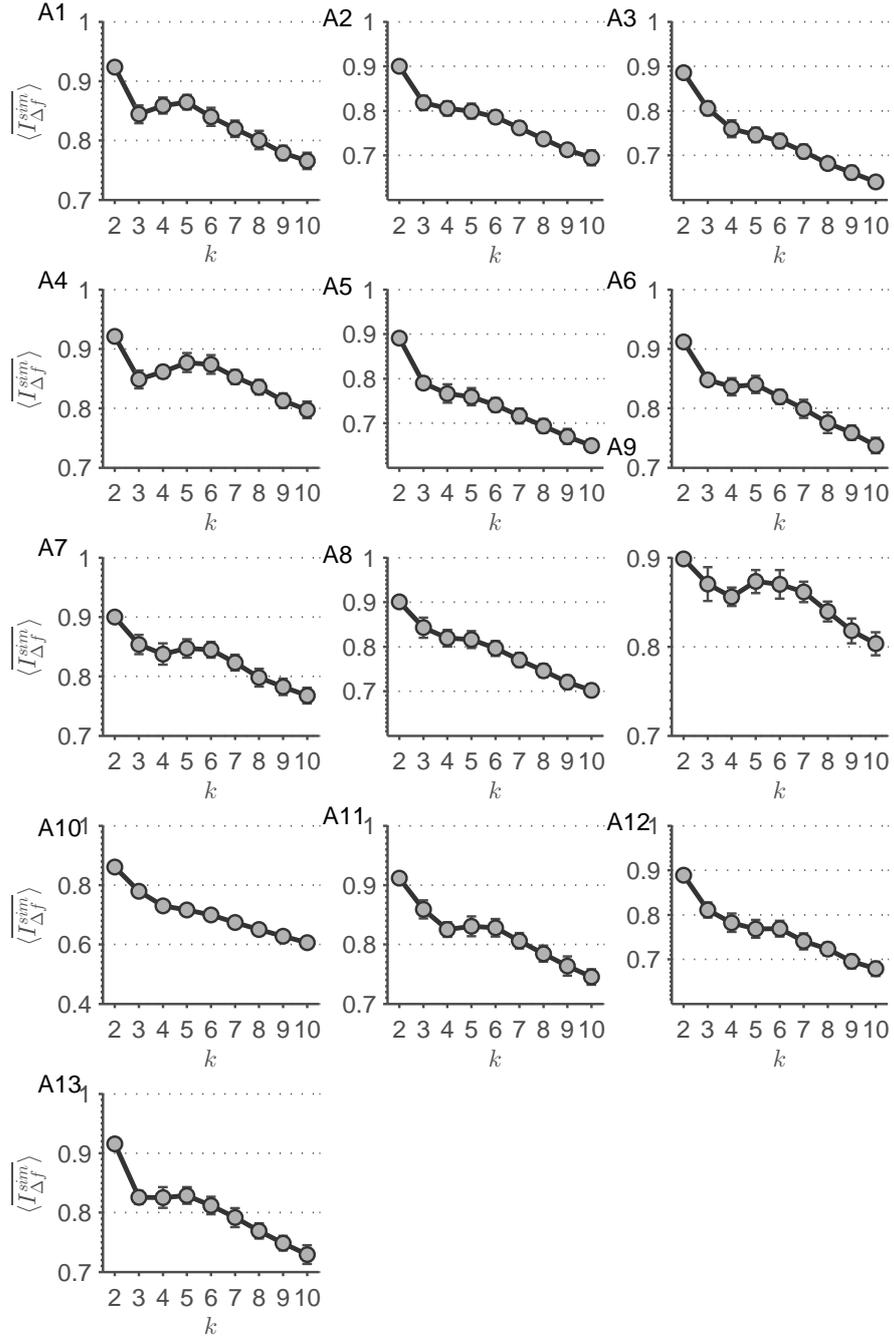


Figure 33: In this figure,  $\langle I_{\Delta f}^{\text{sim}} \rangle(k)$  is plotted for each simulation run for  $k_{\text{inh}} = 2$  data inherent connectivity-states. Error bars represent the standard deviation. Each panel represents a certain parameter combination: (A1)  $p_U = .6; a_U = .8; p_S = .8$ , (A2)  $p_U = 1; a_U = .8; p_S = .8$ , (A3)  $p_U = .8; a_U = 1; p_S = .8$ , (A4)  $p_U = .8; a_U = .6; p_S = .8$ , (A5)  $p_U = 1; a_U = .8; p_S = .6$ , (A6)  $p_U = 1; a_U = .6; p_S = .8$ , (A7)  $p_U = .6; a_U = .8; p_S = 1$ , (A8)  $p_U = .6; a_U = 1; p_S = .8$ , (A9)  $p_U = .8; a_U = .6; p_S = 1$ , (A10)  $p_U = .8; a_U = 1; p_S = .6$ , (A11)  $p_U = .8; a_U = .8; p_S = .1$ , (A12)  $p_U = .8; a_U = .8; p_S = .6$ , (A13)  $p_U = .8; a_U = .8; p_S = .8$ .

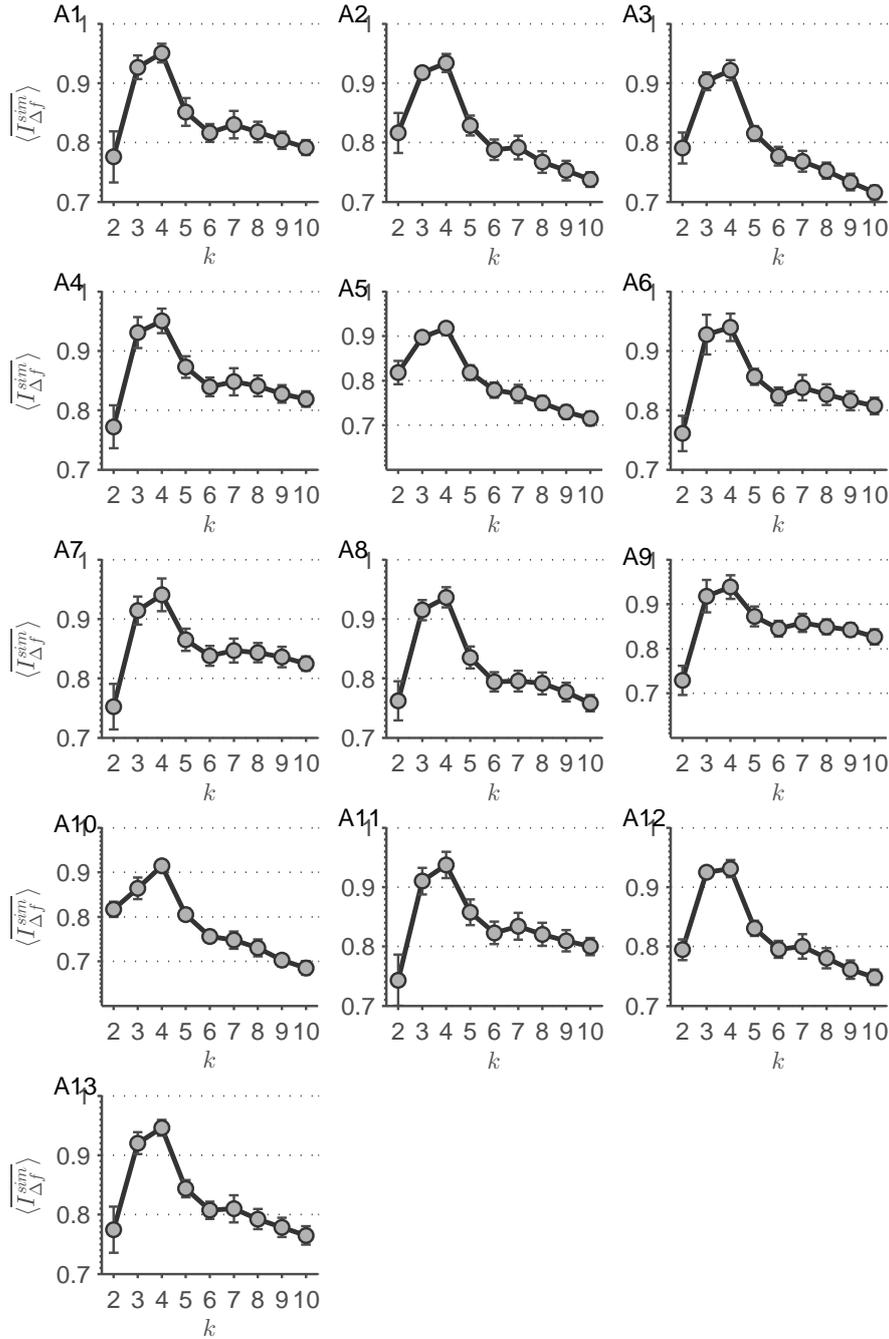


Figure 34: In this figure,  $\langle I_{\Delta f}^{sim} \rangle(k)$  is plotted for each simulation run for  $k_{inh} = 4$  data inherent connectivity-states. Error bars represent the standard deviation. Each panel represents a certain parameter combination: (A1)  $p_U = .6; a_U = .8; p_S = .8$ , (A2)  $p_U = 1; a_U = .8; p_S = .8$ , (A3)  $p_U = .8; a_U = 1; p_S = .8$ , (A4)  $p_U = .8; a_U = .6; p_S = .8$ , (A5)  $p_U = 1; a_U = .8; p_S = .6$ , (A6)  $p_U = 1; a_U = .6; p_S = .8$ , (A7)  $p_U = .6; a_U = .8; p_S = 1$ , (A8)  $p_U = .6; a_U = 1; p_S = .8$ , (A9)  $p_U = .8; a_U = .6; p_S = 1$ , (A10)  $p_U = .8; a_U = 1; p_S = .6$ , (A11)  $p_U = .8; a_U = .8; p_S = .1$ , (A12)  $p_U = .8; a_U = .8; p_S = .6$ , (A13)  $p_U = .8; a_U = .8; p_S = .8$ .

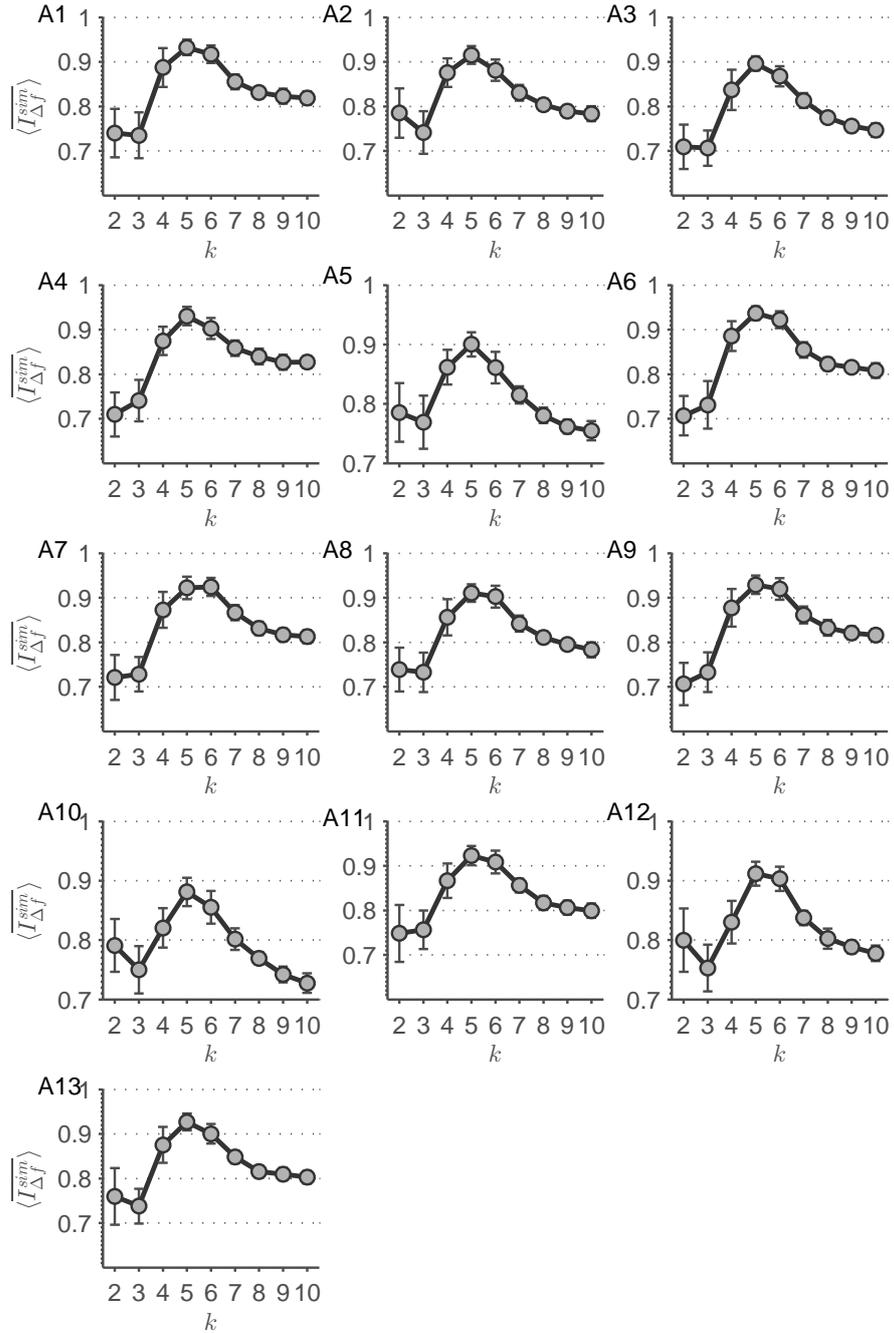


Figure 35: In this figure,  $\langle I_{\Delta f}^{sim} \rangle(k)$  is plotted for each simulation run for  $k_{inh} = 6$  data inherent connectivity-states. Error bars represent the standard deviation. Each panel represents a certain parameter combination: (A1)  $p_U = .6; a_U = .8; p_S = .8$ , (A2)  $p_U = 1; a_U = .8; p_S = .8$ , (A3)  $p_U = .8; a_U = 1; p_S = .8$ , (A4)  $p_U = .8; a_U = .6; p_S = .8$ , (A5)  $p_U = 1; a_U = .8; p_S = .6$ , (A6)  $p_U = 1; a_U = .6; p_S = .8$ , (A7)  $p_U = .6; a_U = .8; p_S = 1$ , (A8)  $p_U = .6; a_U = 1; p_S = .8$ , (A9)  $p_U = .8; a_U = .6; p_S = 1$ , (A10)  $p_U = .8; a_U = 1; p_S = .6$ , (A11)  $p_U = .8; a_U = .8; p_S = .1$ , (A12)  $p_U = .8; a_U = .8; p_S = .6$ , (A13)  $p_U = .8; a_U = .8; p_S = .8$ .

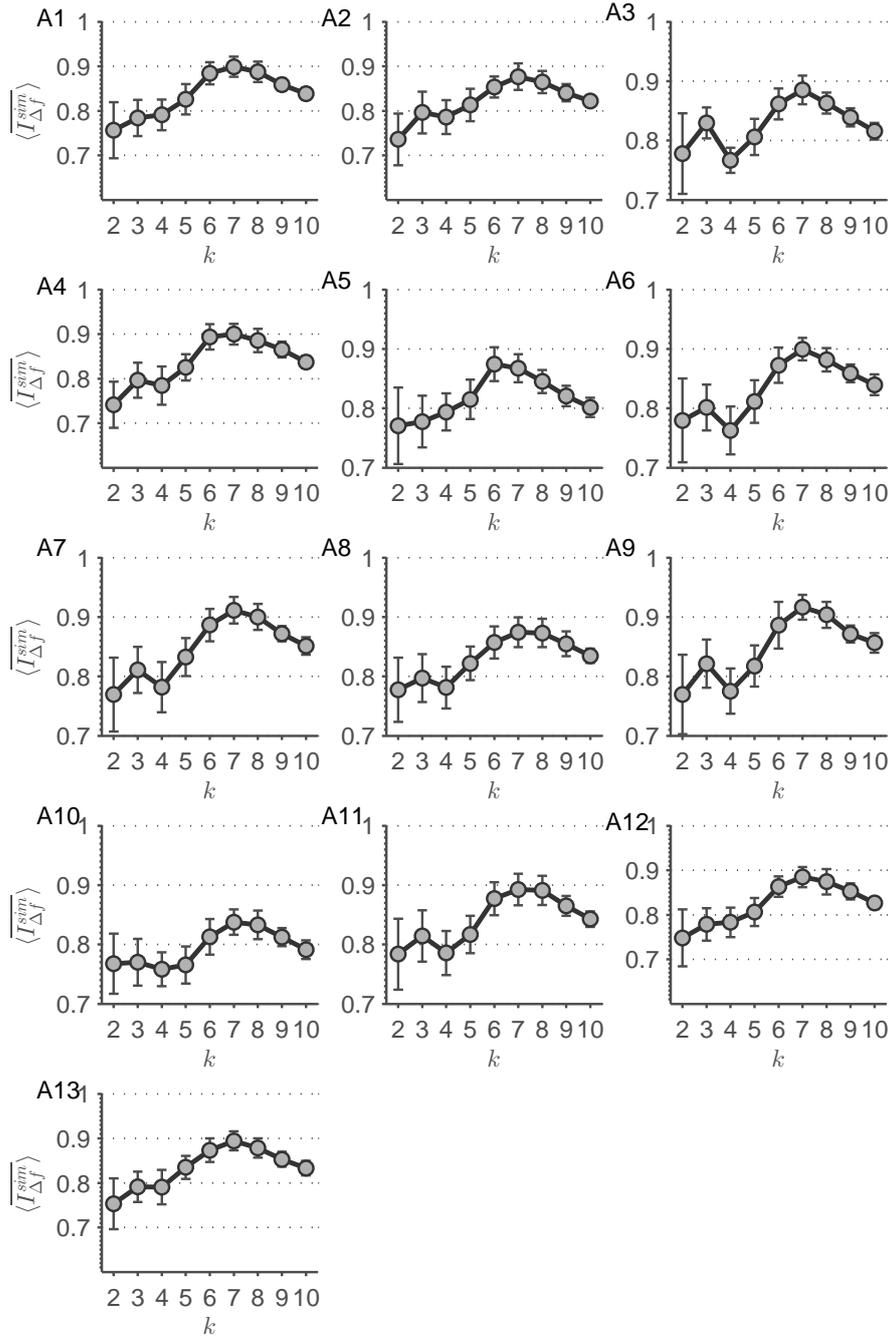


Figure 36: In this figure,  $\overline{\langle I_{\Delta f}^{sim} \rangle}(k)$  is plotted for each simulation run for  $k_{inh} = 8$  data inherent connectivity-states. Error bars represent the standard deviation. Each panel represents a certain parameter combination: (A1)  $p_U = .6; a_U = .8; p_S = .8$ , (A2)  $p_U = 1; a_U = .8; p_S = .8$ , (A3)  $p_U = .8; a_U = 1; p_S = .8$ , (A4)  $p_U = .8; a_U = .6; p_S = .8$ , (A5)  $p_U = 1; a_U = .8; p_S = .6$ , (A6)  $p_U = 1; a_U = .6; p_S = .8$ , (A7)  $p_U = .6; a_U = .8; p_S = 1$ , (A8)  $p_U = .6; a_U = 1; p_S = .8$ , (A9)  $p_U = .8; a_U = .6; p_S = 1$ , (A10)  $p_U = .8; a_U = 1; p_S = .6$ , (A11)  $p_U = .8; a_U = .8; p_S = .1$ , (A12)  $p_U = .8; a_U = .8; p_S = .6$ , (A13)  $p_U = .8; a_U = .8; p_S = .8$ .

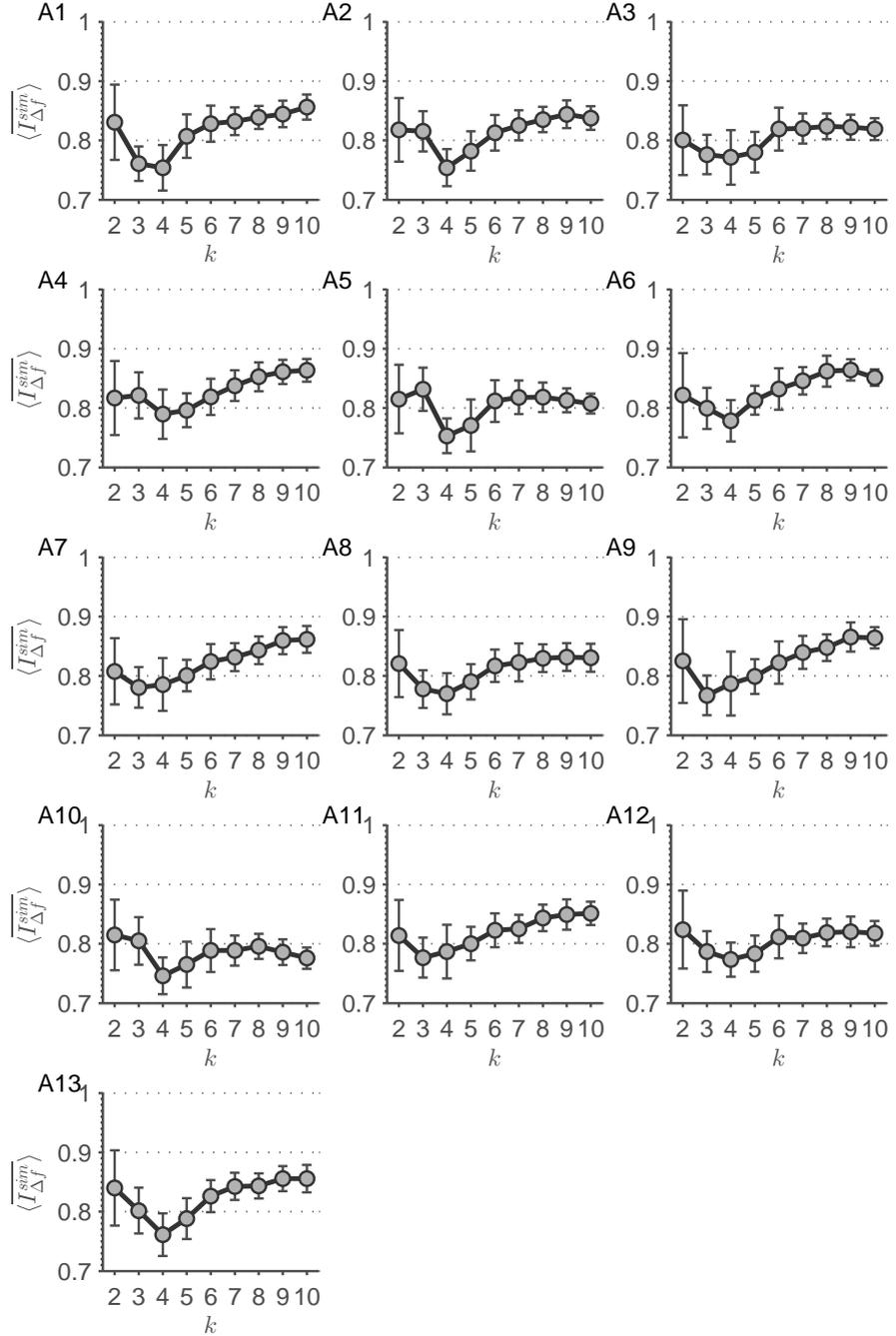


Figure 37: In this figure,  $\langle I_{\Delta f}^{sim} \rangle(k)$  is plotted for each simulation run for  $k_{inh} = 10$  data inherent connectivity-states. Error bars represent the standard deviation. Each panel represents a certain parameter combination: (A1)  $p_U = .6; a_U = .8; p_S = .8$ , (A2)  $p_U = 1; a_U = .8; p_S = .8$ , (A3)  $p_U = .8; a_U = 1; p_S = .8$ , (A4)  $p_U = .8; a_U = .6; p_S = .8$ , (A5)  $p_U = 1; a_U = .8; p_S = .6$ , (A6)  $p_U = 1; a_U = .6; p_S = .8$ , (A7)  $p_U = .6; a_U = .8; p_S = 1$ , (A8)  $p_U = .6; a_U = 1; p_S = .8$ , (A9)  $p_U = .8; a_U = .6; p_S = 1$ , (A10)  $p_U = .8; a_U = 1; p_S = .6$ , (A11)  $p_U = .8; a_U = .8; p_S = .1$ , (A12)  $p_U = .8; a_U = .8; p_S = .6$ , (A13)  $p_U = .8; a_U = .8; p_S = .8$ .

combinations found in panels 37A1,4,7,11. In panels 37A2,6,8,9,13,  $\langle I_{\Delta f}^{\text{sim}} \rangle(k)$  peaked at  $k = 9$ , in panel 37A3 at  $k = 8$ , in panel 37A5 at  $k = 3$ , and in panels 37A10, 12 at  $k = 2$ . As a weak criterion for detection of data inherent connectivity-states common features of the evolution of  $\langle I_{\Delta f}^{\text{sim}} \rangle(k)$  over all parameter combinations can be worked out. One characteristic of plots not meeting the hard criterion seems to be the peaking at  $k = k_{\text{inh}} - 1$ , which is dominant for  $k_{\text{inh}} = 6, 8$ . Additionally, the increase in  $\langle I_{\Delta f}^{\text{sim}} \rangle(k)$  for  $k < k_{\text{inh}} - 1$  or  $k < k_{\text{inh}}$  and its decrease for  $k > k_{\text{inh}} - 1$  or  $k > k_{\text{inh}}$  are also a common features. Those two features define the weak criterion. With this characteristics in mind, all simulations with  $k_{\text{inh}} = 6$  can be detected and with  $k_{\text{inh}} = 8$  all besides one. For  $k_{\text{inh}} = 10$ , the weak criterion yields the right detection in five parameter combinations, additional to the four detections by the hard criterion. Figure 32A1-5 depicts a selection of parameters sets, which result in good classification patterns.

In a next step, I apply the filter banks defined in section 4.3.2 to the time courses of  $k_{\text{inh}} = 4$  and parameter combination found in panel 34A4. This also results in time-series with  $F = 5, 8, 10, 12, 15$  frequency scales for the constant and adjusted order approach. Figure 38 shows the results of this analysis. The strong detection criterion can be met for all panels besides 38A1,2,6. But one has to keep in mind that the values of  $\langle I_{\Delta f}^{\text{sim}} \rangle(3)$  and  $\langle I_{\Delta f}^{\text{sim}} \rangle(4)$  are very similar, especially for the smaller numbers of bands. A remarkable aspect is that with increasing number of frequency scales the pattern identifying data inherent connectivity-states becomes clearer. Additionally, when using frequency scales extracted by filter banks, the scale stability measure  $\langle I_{\Delta f}^{\text{sim}} \rangle(k)$  shows values  $\approx 1$ , which represents almost perfect stability of connectivity-state structure over scales.

## 5.5 ON THE GLASSO PROCEDURE

The correlation matrices in my studies are estimated using the so-called GLASSO procedure. This method estimates the covariance matrix – or rather precision matrix – of data under the assumption that the variables can be approximated as multivariate Gaussian [36]. Since this assumption is most notably violated for time course segments from lower frequency scales, I compute correlation matrices from the MEMD derived scales by using the ordinary correlation. I compare the structure of the correlation matrices from the GLASSO approach to the ordinary ones by second-order correlation. Every connectivity matrix resulting from the GLASSO procedure is correlated with its pendent of the plain correlation. The structure of the correlation matrices is based on the Fisher transformed values, as these are the values entering the clustering procedure. I find that the structure is very similar for both approaches (fig. 39) and therefore the error concerning the structure of the correlation matrices introduced by employing the GLASSO procedure should be small concerning my results. This can be inferred, since the histograms in this figure show a distribution of the values of the second-order correlation mostly  $\rho > 0.9$  for all scales.

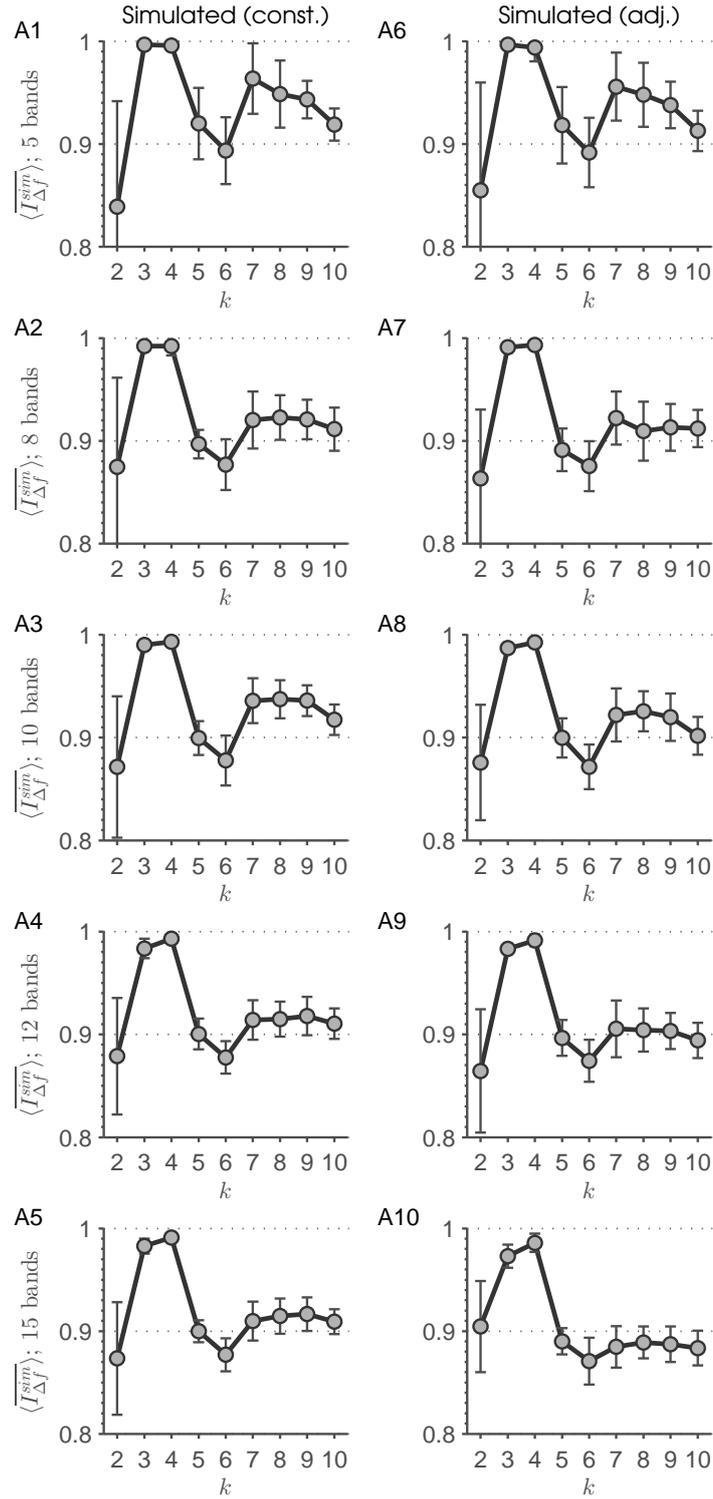


Figure 38: This figure shows evolutions of  $\langle I_{\Delta f}^{sim} \rangle(k)$  for filter bank data of simulated rs-fMRI time courses for  $k_{inh} = 4$  and parameter combination  $p_U = .8; a_U = .6; p_S = .8$ . (A1-5) represent results from the constant (A6-10) from the adjusted order approach. Error bars depict the standard deviation.

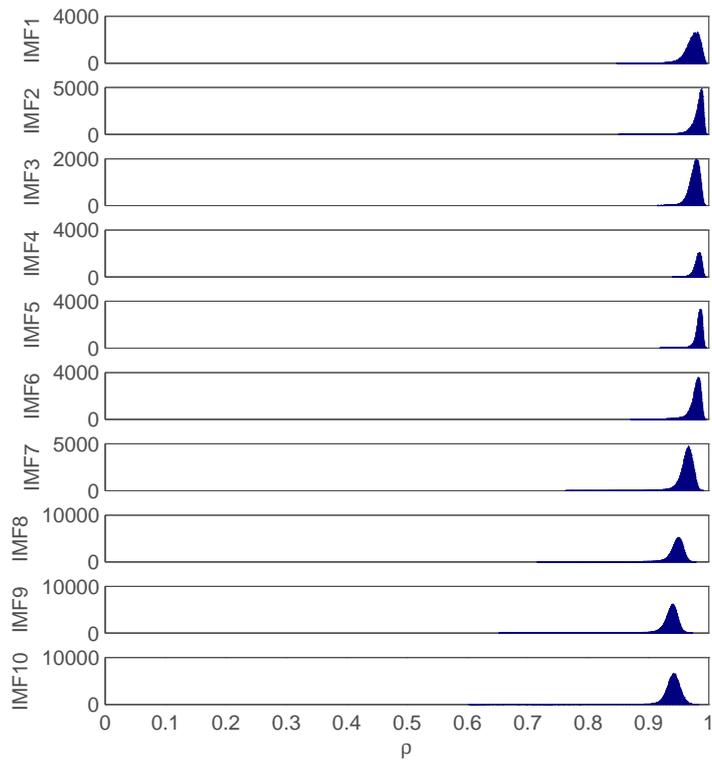


Figure 39: This figure depicts the distribution of the correlation coefficients of the second-order correlation between the connectivity matrices resulting from the GLASSO procedure and plain correlation for scales defined by the MEMD approach. Each panel represents an IMF index. For the histograms all subjects, sessions, and time points are used. The abscissa depicts the correlation values  $\rho$  and the ordinate the absolute frequency. It can be seen that the error introduced by GLASSO estimation is small. For this figure, the correlation matrices with the Fisher transformed values are used, since those are the values entering the clustering procedure.

On those correlation matrices derived with plain correlation coefficients instead of using the GLASSO procedure, I apply the same analyses, which have been conducted for figure 27A1. I find that using common correlation coefficients (fig. 40A1) changes the shape of the curve of  $\langle \overline{I_{\Delta f}^{sim}} \rangle(k)$  slightly losing the plateau for  $k \leq 4$ , but gaining a more  $\frac{k_{inh}}{k}$ -like behavior – assumed that  $k_{inh} = 4$  in the real data case – predicted by theoretical considerations (sec. 5.2.2). Additionally to this analysis, I also investigate the

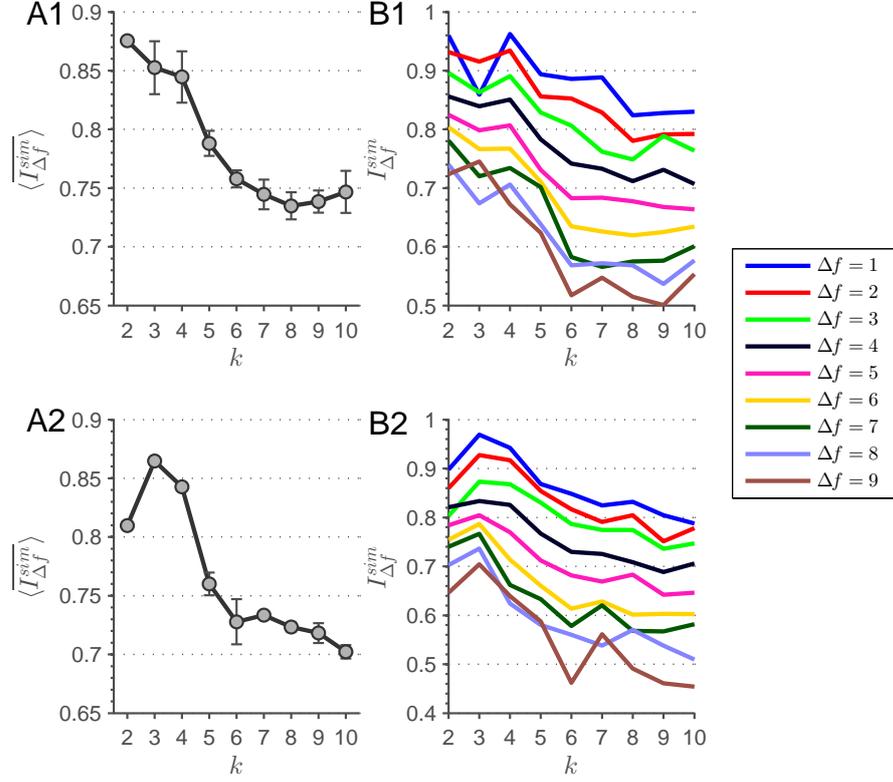


Figure 40: This figure depicts the evolution of  $\langle \overline{I_{\Delta f}^{sim}} \rangle(k)$  for the non-GLASSO, Fisher transformed (A1) and non-GLASSO, non-Fisher transformed (A2) case. Panels (B1) and (B2) show the evolution of  $\overline{I_{\Delta f}^{sim}}(k)$  correspondingly.

influence of using common correlation coefficients on the simulated results of figure 32. In this case, results (fig. 41) are almost identical with a better detection capability of  $\langle \overline{I_{\Delta f}^{sim}} \rangle(k)$  for at least  $k_{inh} = 4$  and  $k_{inh} = 8$ .

In a next step, I probe the influence of the Fisher transformation on the results derived by avoiding non-GLASSO estimation. Figure 40A2 shows the outcome of  $\langle \overline{I_{\Delta f}^{sim}} \rangle(k)$  for connectivity matrices that have not been Fisher transformed before entering the  $k$ -means clustering, and that have been calculated using plain correlations instead of the GLASSO method. It is intriguing that the drop from  $k = 4$  to  $k = 5$  increases again compared to figure 40A1, and that the shape of  $\langle \overline{I_{\Delta f}^{sim}} \rangle(k)$  is very similar to the evolution from the demeaned data case (fig. 27A2). Interestingly, when comparing the GLASSO estimated connectivity matrices and the non-GLASSO versions without Fisher transformation, then the error in structure almost vanishes

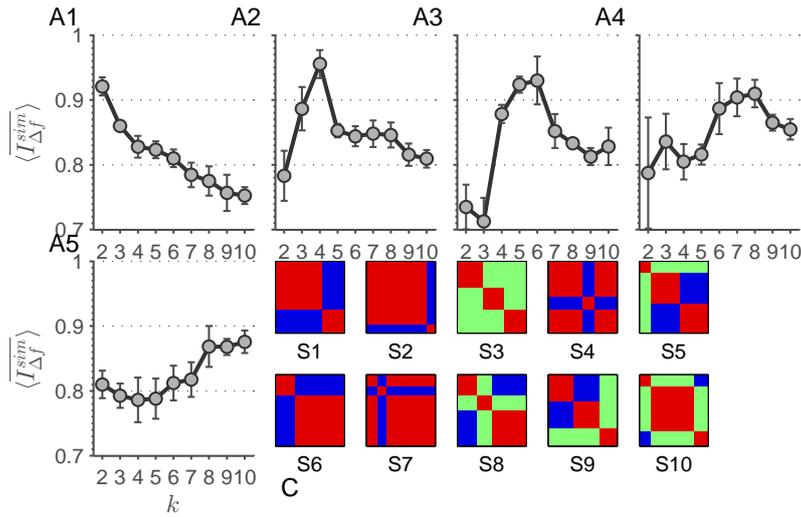


Figure 41: This figure depicts the results of  $\langle I_{\Delta_f}^{sim} \rangle(k)$  for the simulated data shown in figure 32. Please refer to this figure for details. The alteration in this figure is that correlation coefficients of the connectivity matrices entering the clustering procedure have not been calculated by the GLASSO method.

for the lower frequency scales (fig. 42). For all the simulated data in this section, only ten realizations of the k-means algorithm have been calculated.

And in a next step, I calculate  $\langle I_{\Delta_f}^{sim} \rangle(k)$  for the simulation results of figure 41, but with the alteration of also removing the Fisher transformation from the data. This yields slightly worse results compared to figures 32 and 41.

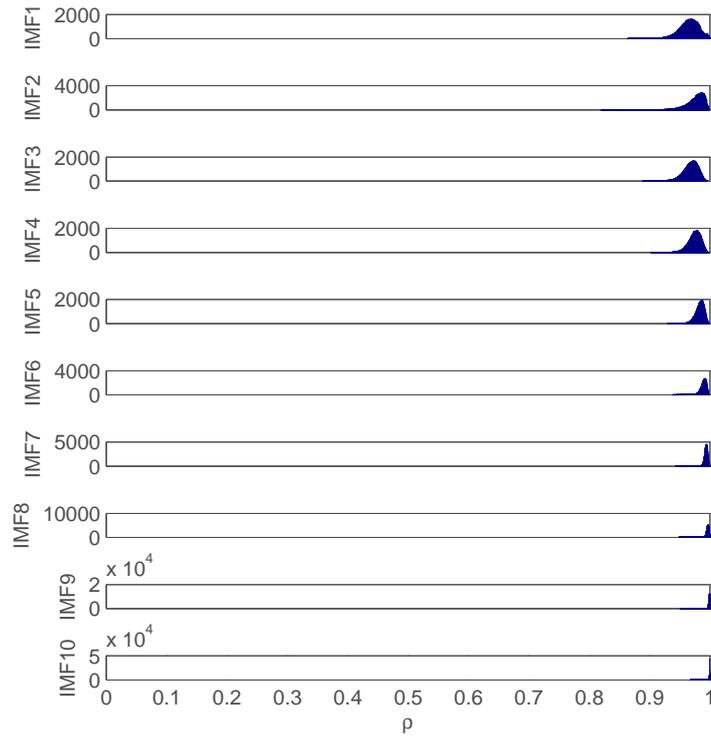


Figure 42: This figure depicts the second-order correlations derived in figure 39 with the alteration of using non-Fisher transformed correlation coefficients in the connectivity matrices.

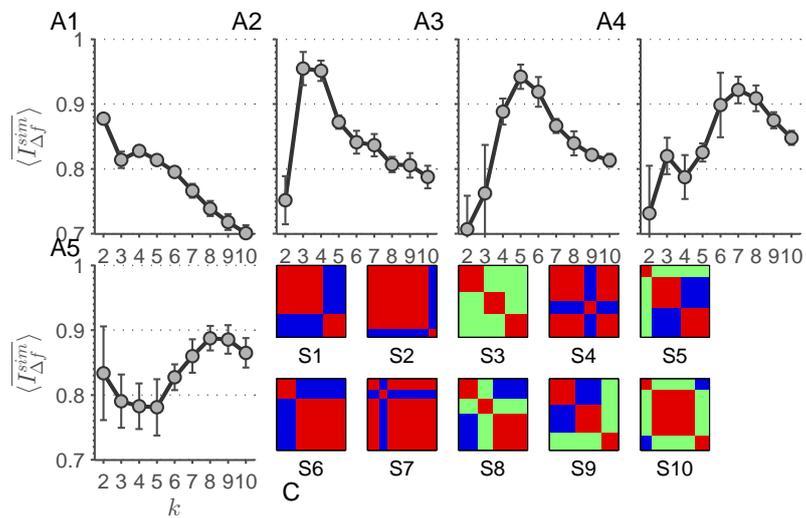


Figure 43: This figure depicts the results of  $\overline{\langle I_{\Delta f}^{sim} \rangle}(k)$  for the simulated data shown in figure 32. Please refer to this figure for details. The alteration in this figure is that correlation coefficients of the connectivity matrices entering the clustering procedure have not been calculated by the GLASSO method, and the Fisher transformation is removed from the data.

## Part III

### DISCUSSION, OUTLOOK, AND CONCLUSION

In this part, I sum up my results and discuss several aspects of my thesis in more detail (ch. 6). Furthermore, I elaborate on the outlook and limitations of my approaches in general, and in particular I focus on the project for my PostDoc phase. For the latter, I have created a proposal for my own funding in my PostDoc phase. A proposal to be admitted into the so-called 7. *Nachwuchsakademie Medizintechnik* has successfully passed the first selection step, but awaits final decision. In the course of this 7. *Nachwuchsakademie Medizintechnik*, I have submitted a proposal for a research fellowship directed to the German science foundation. In this research fellowship, I plan to join the group around Prof. Dr. Dr. Gustavo Deco in Barcelona, Spain. During this one year research fellowship, I am planning to investigate the ideas elaborated on in chapter 7. I finish by summing up my thesis (ch. 8).



*"The first principle is  
that you must not fool yourself –  
and you are the easiest person to fool."  
(Richard P. Feynman)*

In this thesis, I have investigated a frequency-resolved version of dFC. This has at first been done by MEMD as a data-driven method. The application of MEMD on time courses resulting from gICA has resulted in additional time-series on narrow banded frequency scales from IMFs. After the application of a sliding window procedure on those frequency-resolved time courses, clustering of the sets of correlation matrices has revealed scale stability as an inherent feature of connectivity-states, or rather cluster centroids. In *post hoc* analyses, null-model and filter bank investigations, as well as simulated data, my main findings have been corroborated. In the following, I discuss several aspects of my thesis in more detail – especially the comparison of scale stability of cluster centroids to cluster validity measures.

#### 6.1 DILEMMA OF OBJECTIVE CONNECTIVITY-STATE SELECTION

Connectivity-states, in their original definition [8], are centroids of k-means clusters (sec. 1.4). For this clustering approach the number of clusters to be extracted has to be predefined. To validate the most suitable number of clusters, an elbow criterion in a cluster validity plot can be employed. Those types of plots are also called scree plots. If we assume that  $k_{inh}$  is the data inherent number of clusters in our data set, then usually the used cluster validity measure shows large decrements for  $z < k_{inh}$  and small ones for  $z > k_{inh}$ . This describes the emergence of an elbow in the plot of the cluster validity measure. Ideally, the task of identifying this elbow can be done almost objectively given that it is very pronounced. Unfortunately, this is rarely the case. In their study, Allen et al. [8] only show their scree plot for simulated data in the supplementary information where they find a pronounced elbow. For the analyzed rs-fMRI data this plot is lacking. During my studies, I have also looked at scree plots for my data set and I have found that an objective selection of the estimated number of clusters inherent to the data is problematic. Figure 44 shows a scree plot by using the ratio of within cluster to between cluster distance

$$I_{val} = \frac{\sum_{t=1}^T \min_z (d(\mathbf{c}_z, \mathbf{x}_t))}{\sum_{t=1}^T \sum_{z \neq z_n} d(\mathbf{c}_z, \mathbf{x}_t)}, \quad (89)$$

which is also used by Allen et al. [8] for validation. Here,  $z_n$  is the nearest centroid of data point  $\mathbf{x}_t$ . One of my master students (T. Geigenfeind) has in-

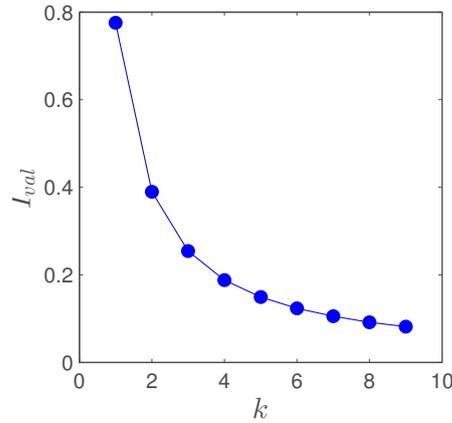


Figure 44: This figure depicts the evolution of the cluster validity measure  $I_{val}$  from equation 89 for  $k = 2, \dots, 10$ . It is plotted for dFC applied on the broad band signal of the rs-fMRI data from the 400 sessions used in this thesis. An objective way of selecting the number of data inherent clusters is not possible using this plot.

vestigated the clustering quality of dFC matrices in more detail in his thesis [37]. First, he investigates the mean distance of data points  $\mathbf{x}_t$  to their nearest centroids, which is an alternative version of the measure introduced in equation 89. His results for this criterion are shown in figure 45. He employs both cosine and city block distance metric, respectively. He also draws the conclusion from this plots that an objective selection of the data inherent number of clusters is not possible. He also looks at the so-called silhouette score [86]. He elaborates on this measure defined for every data point  $\mathbf{x}_t$  as

$$S(\mathbf{x}_t) = \frac{d_B - d_A}{\max(d_A, d_B)}, \quad (90)$$

where  $d_A$  is the average distance of data point  $\mathbf{x}_t$  to the data points belonging to its assigned cluster  $A$ , and  $d_B$  is the average distance to the data points belonging to the nearest cluster of cluster  $A$ . He elaborates further that the difference  $d_B - d_B$  is normalized by the maximum distance of these two groups of distances, which yields values between  $-1$  and  $1$ . He explains that  $-1$  stands for a completely failed clustering,  $0$  for a lack of cluster structure in the data, and  $1$  for a perfect clustering. He mentions that for toy data with uniformly distributed samples in two dimensions the average silhouette score over all points never drops below a value of  $0.32$  for  $k \leq 30$ . He compares this value to average silhouette scores of dFC matrices and finds a behavior shown in figure 46. Additionally, he looks at the so-called centroid silhouette score, where the distances  $d_A$  and  $d_B$  are replaced by the distance of every data point  $\mathbf{x}_t$  to centroids of clusters  $A$  or  $B$ , respectively. His results of the clustering of dFC matrices and this measures are shown in figure 47. With both silhouette score measures, he finds that clusterings

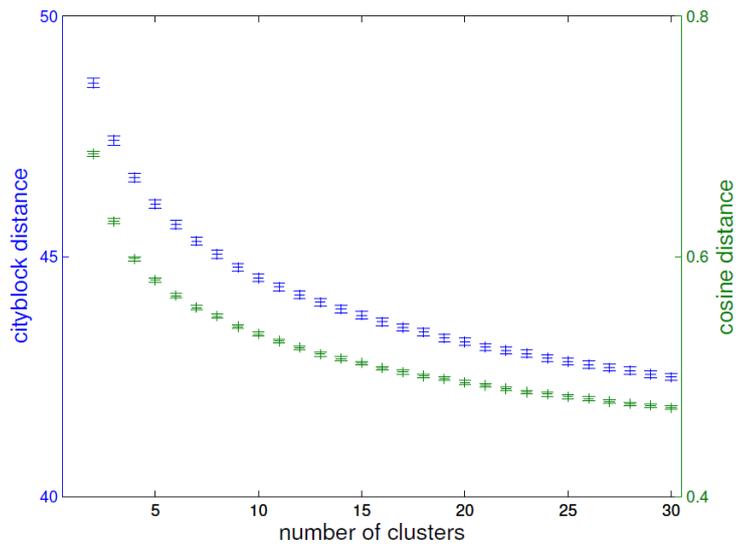


Figure 45: In this figure, the results of the mean distance between all data points  $x_t$  and their corresponding clusters is shown for  $k = 2, \dots, 30$  extracted clusters and two distance metrics – city block and cosine. An objective selection of the data inherent number of clusters is not possible. Error bars depict the standard error for several initializations of the clustering procedure. From [37].

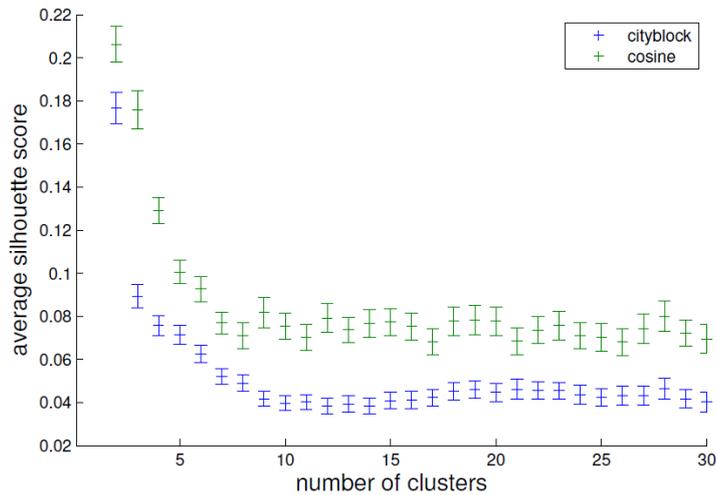


Figure 46: In this figure, the average silhouette score of clustered dFC matrices is shown for  $k = 2, \dots, 30$  for cosine and city block distance metric. Error bars depict the standard error. From [37].

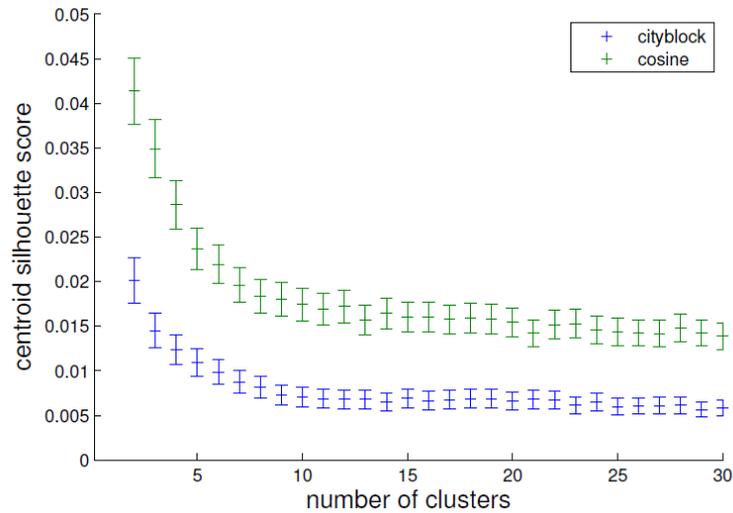


Figure 47: In this figure, the average centroid silhouette score of clustered dFC matrices is shown for  $k = 2, \dots, 30$  for cosine and city block distance metric. Error bars depict the standard error. From [37].

with  $k = 2$  are preferred, but the silhouette values are smaller compared to the clustering of random data points. He concludes that the data structure created by the dFC matrices is problematic. It has to be mentioned that the analyses of T. Geigenfeind are based on earlier versions of the data, in which a small interpolation error in the preprocessing of used time courses has been uncorrected. My have found that this error hardly influences the main results. Therefore, I suggest that his mentioned results in this thesis can be taken seriously.

Taking all those results into account, it is obvious that it is very hard to validate clustering of dFC matrices properly. Nevertheless, even with such problematic data sets my approach is still able to give a clear recommendation about the inherent cluster structure, which is also confirmed by simulated data.

#### 6.1.1 Generalizing the idea of selecting the number of clusters using frequency-resolved information

Since the idea of using correlation matrices as data points in a higher-dimensional space, which then serve as the substrate for a clustering procedure, is quite new but not restricted to the narrow field of neuroscience, my approach can also be extended and generalized to other fields. There are two ways of how clustering procedures could benefit from frequency-resolved data. First, if we are dealing with broad band signals like in the investigated case in this thesis, frequency resolution yields the possibility of employing scale stability of cluster centroids as a validity criterion. It seems that this approach vastly magnifies the capability of identifying the data inherent number of clusters compared to common validation procedures. Sec-

ond, if we are dealing with a monochromatic signal, then frequency resolution yields the possibility of revealing the important scales where the most information of the signal is contained. After frequency resolution, the band containing the signal can be extracted and then the common cluster validation procedures can be applied. This could yield a more pronounced elbow in the case of scree plots.

## 6.2 FILTER BANKS, MEMD, AND SELF-SIMILARITY

Investigating time courses that are suggested to have self-similar features by means of data-driven approaches is a natural way of approaching such type of data [31]. Many studies find that scale-freeness and self-similarity are inherent features of rs-fMRI data [25, 35, 60, 99] and my results confirm this circumstance. For rs-fMRI data the main finding holds for ten, 12, and 15 bands (fig. 31) and in case of simulated data it holds basically for all numbers of extracted bands (fig. 38). Additionally, I show that self-similarity can be exploited to yield deeper insights in broad band data and its cluster structure. My *post hoc* filter bank analyses strengthen the finding of broad band self-similarity. Furthermore, it is evident that for simulated data the increase in number of extracted scales by a filter bank yields a better carving out of the ground truth. In case of rs-fMRI, the same holds, if we assume that  $k = 4$  is the data inherent number of connectivity-states. Since in rs-fMRI data more sources of noise should be present than in simulated data, the finding that the number of data inherent connectivity-states can be identified easier than in the rs-fMRI case is not surprising. Comparing MEMD and frequency banks the former is able to find a proper number of frequency scales for a good classification of the number of data inherent clusters, whereas the latter needs adjustment at least in the number of extracted scales. Furthermore, the accurate mode alignment of IMFs resulting from MEMD ensures that there is a very similar number of periods within windows of different RSNs making those time course segments more comparable, which is beneficial for dFC analysis.

## 6.3 CONSTANT OR ADAPTED WINDOW SIZE

When entering frdFC, the sizes of windows used for applying a sliding window approach on different frequency scales have to be discussed. The advantage of a comparable number of periods within one window over different frequency scales is appealing. But when conducting the frdFC analysis in this way, the number of correlation matrices in the sets corresponding to different frequency scales differs vastly. Consider the extreme case of  $\alpha = 1$  in equation 88. With such a huge data set analyzed in this thesis, the number of correlation matrices for the scale of highest frequency is  $T_{f=1} = 462000$  and for the scale of lowest frequency  $T_{f=10} = 400$ . Such a discrepancy in the number of samples between the data sets questions the comparability of the results of the k-means clustering procedures *per se*. With increasing  $\alpha$

this discrepancy scales up. Looking at the results of figure 30 a continuous change for the scale stability measure from  $\alpha = 0.05$  to  $\alpha = 1$  is obvious. Considering a loss of comparability of the clustering procedures with increasing  $\alpha$ , the gradually vanishing drop from  $k = 4$  to  $k = 5$  can be interpreted as an artifact. Besides the comparability issue of clustering, there is another aspect favoring the constant window size approach of the sliding window procedure in the case of MEMD. This method decomposes the time courses of interest into IMFs, which result in the original time course when summed up. This means that each time point of the original time-series  $\mathbf{T}_i$  has its corresponding partner in each frequency-resolved time courses  $\mathbf{U}_i^f$ . In case of a constant window size, the time-series segments sum up to its original pendent. With increasing  $\alpha$ , more and more time points are taken into account in segments of scales of lower frequency that do not have corresponding partners in segments of previous scales. This could be interpreted as a decrease in signal-to-noise ratio and therefore renders another aspect in favor of a constant window size.

#### 6.4 LIMITATIONS AND FUTURE DIRECTIONS

We also have to look at the ranges of the frequency bands resulting from MEMD and filter banks. EEG studies show [46, 78] that infra-slow frequency (0.01 Hz – 0.1 Hz) fluctuations are associated with BOLD signal changes and hence foster a physiological connection to most of the frequency scales used in this thesis. At least for scales  $f > 7$  in case of MEMD, the physiological base can be questioned. Further studies have to show that such low frequency ranges have any physiological meaning. Note that even on the lowest frequency scales imprints of connectivity-states can be found. This means that connectivity-states can even be defined by their structure using just the lowest frequency information of the data.

MEMD is computationally a very costly method of decomposing time-series. Therefore, with the used computer infrastructure, it would not be feasible to increase the number of investigated RSNs to result in a complete parcellation of the cortex. Smith et al. [94] question the validity of doing connectomics with ICA analyses using a small number of components. They suggest that network analyses make sense, if the number of components is so large to result in a complete parcellation of the cortex. For future studies, it would be therefore interesting, whether this statement holds for my results with a significantly increased number of RSNs. I think that the simulation results are quite promising in this context, since the detection of the data inherent cluster structure is feasible even when using only 24 RSNs. With a more fine grained parcellation of the cortex, I would also expect that the cerebellar and subcortical ICs become less noisy and could therefore be included in my analysis.

Investigating brain graphs on a frequency-resolved basis is quite common in literature using wavelet analysis [1, 2, 11, 18]. To my knowledge, there is one study [113] investigating static brain graphs on different frequency

scales using MEMD as a decomposition tool. The method of frdFC could also be expanded to such frequency-resolved brain graph studies switching from static to dynamic investigations, where methods of time varying graph theory could be applied [76]. This extension to frequency-resolved dynamic brain graphs could result in higher diagnostic sensitivity when comparing groups of healthy and pathological subjects.

Furthermore, I suggest further studies to investigate the influence of the way of deriving the correlation coefficients for the dFC matrices in more depth. One question would be to look quantitatively at the extent of the error introduced by using the GLASSO procedure instead of calculating the ordinary correlation coefficients. Moreover, future studies should investigate the influence of the structure of simulated connectivity-states. This is a degree of freedom, which I have not varied in my investigations. It would also be interesting to look at the influence of different distance measures, when conducting the k-means clustering procedure, and to investigate the influence of using an averaging instead of median procedure, when calculating the cluster centroids.

I want to discuss results found in section 5.5 in more detail. The outcome of these simulations shows that the sensibility of my method concerning the detection of data inherent clusters is quite robust when it comes to simulated data. Only the non-GLASSO and non-Fisher transformed version shows a decrease in detection quality. For the real data case, results show a higher sensibility of my method concerning the GLASSO and Fisher transformation steps. Nevertheless, the overall picture and result of my method holds. Future studies should investigate the most robust and reliable parameter and (pre-)processing combinations for detecting data inherent cluster structures using my method.



THE HUMAN CONNECTOME AND ITS DYNAMICS AS A  
SUBSTRATE FOR ISING-LIKE BACKBONE SIMULATIONS  
IN HEALTH AND DISEASE

---

*"The honors are unreal to me.  
I don't believe in honors. [...]  
I can't stand it, it hurts me."  
(Richard P. Feynman)*

In this chapter, I elaborate on the ideas of my proposal to the German science foundation, which has been created during my PhD phase. This proposal passed the first selection step, but awaits final decision. It is based on the dFC knowledge, which I have gathered during my PhD.

### 7.1 INTRODUCTION

As mentioned briefly in section 3.1, connectomes can also be employed for simulations of dynamic systems like the Ising-model. In parallel to the switch from static functional connectivity to dFC, the comparison of Ising-like systems to connectomics emerged. One very recent aspect of this comparison is the development of so-called backbone simulations. In such studies, the connectivity matrix in focus is directly implemented into the dynamical system. The first study has been done by Marinazzo et al. [71], in which a structural connectome is implemented in the coupling constant  $J_{ij}$  of an Ising-like system with the Hamiltonian

$$H_i = -s_i \sum_j J_{ij} s_j \quad (91)$$

$$J_{ij} \propto X_{ij}^S, \quad (92)$$

where  $X_{ij}^S$  is the  $ij$ -th entry of the structural connectivity matrix. Marinazzo et al. [71] then simulates this Ising-like system on two structural connectomes with different spatial resolution. They find that there are corresponding peaks in the susceptibility  $\chi$ , the heat capacity  $C_v$ , and the transfer entropy at a certain temperature. They interpret this correspondence – especially for  $\chi$  and  $C_v$  – as evidence for a critical state of the system. Furthermore, they report that the peak of the transfer entropy also at this critical temperature suggests that the information flow is maximum at this point. A second study conducted by Hudetz et al. [52] investigates a spin-glass model

with implemented static functional connectivity matrix as backbone. Their model is defined as

$$H_i = -\frac{1}{2}s_i \sum_{j=1}^N J_{ij}s_j \quad (93)$$

$$J_{ij} \propto X_{ij}^{\text{SF}} \quad (94)$$

$$p_i \propto e^{-\beta \Delta E_i}, \quad (95)$$

where  $N$  is the number of vertex points on the lattice,  $X_{ij}^{\text{SF}}$  is the  $ij$ -th entry of the static functional connectivity matrix,  $p_i$  is the local flipping probability depending on the inverse temperature  $\beta$  and the energy difference of a flip  $\Delta E_i$ . Like it is common for spin-glass simulations, their system is annealed down to zero temperature 10000 times. They report that this results in varying end configurations, which are then hierarchically clustered to reveal the 40 most common patterns. Hudetz et al. [52] find that those patterns resemble known rs-fMRI configurations on the cortex.

In this proposal, I suggest to use dFC in such backbone simulations, which would represent a fluctuating metabolic activity instead of static functional connectivity used by Hudetz et al. [52]. First, I define the Ising-like models I use in this proposal. Then, I introduce the method for implementing dFC as a time-dependent backbone in Ising-like systems with various alterations. Afterwards, I show how to use backbone simulations as tools for the comparison of deliberately damaged and healthy systems. Finally, I introduce a variation of an Ising-like system to investigate the effect of neighborhood size on backbone simulations.

## 7.2 MODEL DEFINITIONS

In order to investigate the various below introduced backbone systems, I define the following Ising-like models. Each model is assigned to the approaches, where it is used.

### 7.2.1 Ising-like model without external magnetic field term

In this proposal, I plan to use the model from equation 92 without a magnetic field term introduced by Marinazzo et al. [71] in sections 7.3.1, 7.3.2, and 7.3.3 for simulations on a dynamic backbone. Furthermore, this model is used to replicate the findings of Marinazzo et al. [71].

### 7.2.2 Ising-like model with external magnetic field term

In addition to the model with just an interaction term in its Hamiltonian, I propose to use an Ising-like system with an external magnetic field term

$$H_i = - \sum_{\langle ij \rangle} J_{ij} s_i s_j - h_i s_i. \quad (96)$$

In this general case, the sum contains all nearest neighbors  $\langle ij \rangle$  of spin  $i$ . I adjust this Hamiltonian to be comparable to the model introduced by Marinazzo et al. [71]

$$H_i = -s_i \sum_j J_{ij} s_j - h_i s_i \quad (97)$$

$$J_{ij} \propto A_{ij}, \quad (98)$$

where the sum is not restricted to the nearest neighbors of  $i$  anymore, and  $A_{ij}$  is the  $ij$ -th entry in the connectome matrix of consideration. For each spin location  $s_i$  there is a magnetic field term  $h_i$  contributing to the Hamiltonian  $H_i$ . In section 7.3.4, this model is used to simulate backbone systems combining different modalities – structural and functional backbones.

### 7.2.3 Spin-glass model

The model of a spin-glass in equation 95 incorporating a static functional connectome introduced by Hudetz et al. [52] can be found in sections 7.3.1, 7.3.2, and 7.3.3 also to be simulated on a dynamic backbone. Additionally, this model is used to replicate the results of Hudetz et al. [52].

## 7.3 USING DYNAMIC FUNCTIONAL CONNECTIVITY AS A SUBSTRATE FOR ISING-LIKE BACKBONE SIMULATIONS

Consider the spatiotemporal set of correlation matrices resulting from dFC analysis  $\{\mathbf{X}_i^D(t)\}_{t=1, \dots, T-w}$  of the dimension  $C \times C \times T - w$ , where  $i$  is the session number,  $C$  the number of used loci in the brain, and  $T - w$  the number of time points, or rather windows. This set of correlation matrices is used as a dynamic backbone in the following simulations. I refer to this set as the dynamic connectome  $\mathbf{X}^D(t)$  in general. The main idea is to implement the dynamic connectome  $\mathbf{X}^D$  in the Ising-like system as a time-dependent coupling between spins  $i$  and  $j$

$$J_{ij}(t) \propto X_{ij}^D(t) \quad (99)$$

directly connecting to the approaches of Marinazzo et al. [71] and Hudetz et al. [52]. This yields a time-dependent coupling in Ising-like systems. I call the two major approaches of such an implementation quasistatic and dynamic.

### 7.3.1 Quasi static backbone

In this approach, I propose to simulate Ising-like systems on each of the correlation matrices from the set  $\{\mathbf{X}_i^D(t)\}_{t=1, \dots, T-w}$  independently. This means that for each time point  $t_k$  the coupling constant is chosen to be

$$J_{ij}^{t_k} \propto X_{ij}^D(t_k) \quad (100)$$

and the Ising-like system is simulated with  $U$  iteration steps. This is done for  $R$  realizations with varying random initial configurations resulting in an ensemble of micro state sets  $\{\{s_i(\mathbf{u})\}_{i,u=1,\dots,U}\}_{r=1,\dots,R}$  for each  $t_k$ . From these ensembles, physical parameters like susceptibility  $\chi$ , magnetization  $M$ , and heat capacity  $C_v$  can be calculated for each correlation matrix separately. This results in a temporal evolution of those physical parameters under the quasistatic constraint. Figure 48 illustrates this approach.

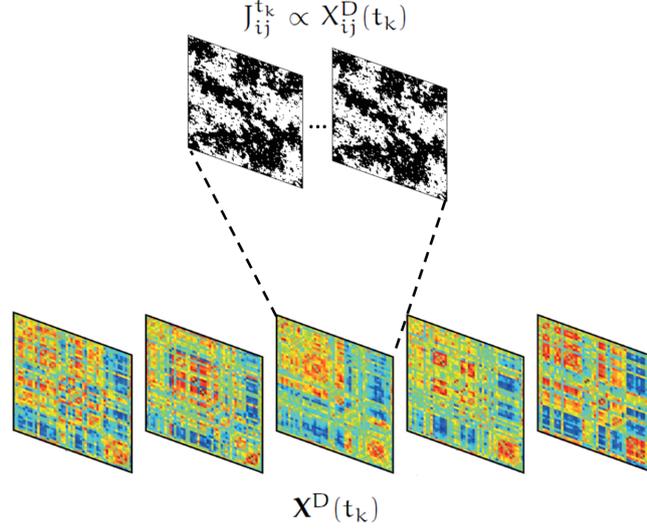


Figure 48: In this figure, the quasistatic approach of implementing dFC as a dynamic backbone is illustrated. On each correlation matrix resulting from dFC analysis, an Ising-like system is simulated with coupling  $J_{ij}^{t_k} \propto X_{ij}^D(t_k)$ . Assembled from [8, 60] and adjusted.

### 7.3.2 Dynamic backbone

In the dynamic approach, the dynamic backbone is implemented to yield an explicit time dependence in the Ising-like system. This is done by simulating the used models with coupling

$$J_{ij}(t) \propto X_{ij}^D(t), \quad (101)$$

which means that on each correlation matrix the Ising-like system is simulated for  $U$  iteration steps. But compared to the quasistatic approach, the  $U$ -th micro state  $\{s_i(\mathbf{U})\}_i$  for  $\mathbf{X}^D(t_i)$  is the initial configuration for  $\mathbf{X}^D(t_{i+1})$ . To generate an ensemble of  $R$  realizations, the Ising-like system is initialized with random configurations at the first iteration step of the first correlation matrix  $\mathbf{X}^D(t_1)$ . This approach results in spatiotemporal ensembles, from which the temporal evolution or averages of physical parameters can be estimated. The temporal resolution of this method is  $1 \leq U < \infty$  in terms of iteration steps.

With this approach, it is also possible to simulate more rapid evolving Ising-like systems. Consider adjacent correlation matrices  $\mathbf{X}^D(t_i)$  and

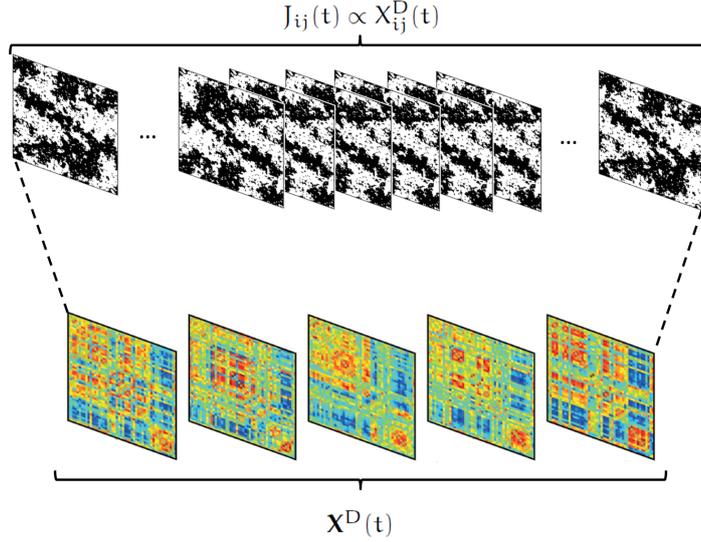


Figure 49: In this figure, the explicit time-dependent case of a backbone involving the dynamic connectome is depicted. The Ising-like system is simulated over the whole dynamic backbone  $\mathbf{X}^D(t)$  with coupling between spins  $i$  and  $j$  of  $J_{ij}(t) \propto X_{ij}^D(t)$ . Assembled from [8, 60] and adjusted.

$\mathbf{X}^D(t_{i+1})$ . By using spline interpolation it is possible to generate values for the correlation functions  $X_{ij}^D(t)$  between values  $X_{ij}(t_i)$  and  $X_{ij}(t_{i+1})$ . This means that for micro states of the Ising-like system, there is an infinite amount of backbone values available between those two empirical values. In other words, the temporal resolution in terms of iteration steps is extended by the range  $0 < \mathcal{U} < 1$ . The dynamic backbone approach is illustrated in figure 49.

### 7.3.3 Frequency-resolved dynamic backbone

The frequency decomposition of time courses  $\mathbf{T}_i$  of my above introduced approaches has yielded frequency-resolved dynamic connectomes  $\mathbf{X}_i^{I_f}(t)$  on frequency scales  $f = 1, \dots, F$  and session  $i$ . Those dynamic connectomes can be used as frequency-resolved dynamic backbones in Ising-like simulations. Instead of simulating on the full-band dynamic backbones, distinct frequency scales can be explored by implementing a time- and frequency-dependent coupling

$$J_{ij}^f(t) \propto X_{ij}^{I_f}(t). \quad (102)$$

The idea is that the simulation of Ising-like systems on frequency-resolved backbones could reveal patterns in the evolution of simulated physical parameters that would have been uncovered in the full-band approach.

### 7.3.4 Combination of backbones of different modalities

In a next step, I propose to combine backbones of different modalities in one Ising-like system. This is done by using an Ising-model with external magnetic field term  $h_i$

$$H_i = -s_i \sum_j s_j J_{ij} - s_i h_i, \quad (103)$$

where the dynamic connectome is still represented in the coupling  $J_{ij}(t) \propto X_{ij}^D(t)$ . The structural connectome  $\mathbf{X}^S$  is implemented in the external field by summing over all connections  $h_i \propto \sum_j X_{ij}^S$  reflecting the overall connectivity strength at brain area  $i$ . By uniformly varying the magnitude of the external field, or rather connectivity strength  $|h_i|$ , it is possible to investigate hysteresis behavior of the introduced Ising-like system. Furthermore, the dynamic connectome can be implemented in the external field making it time-dependent  $h_i(t) \propto \sum_j X_{ij}^D(t)$ , and the structural connectome can be represented by the coupling constant  $J_{ij} \propto X_{ij}^S$ .

Additionally, the Ising-like system with an external magnetic field term can also be used for simulations with the static functional connectome  $\mathbf{X}^F$ . This can be realized by combining it with the structural connectome, which would still be represented in the external magnetic field term, and the static functional connectome in the coupling  $J_{ij} \propto X_{ij}^F$ . Those results can then be compared to results from Hudetz et al. [52].

### 7.3.5 In detail considerations

In the case of a dynamic backbone, I suggest that dFC can also be seen as a slowly fluctuating background, on which the dynamic system is evolving. These low-frequency fluctuations of rs-fMRI data in the range of  $\omega < 0.15$  Hz can be linked to EEG low-frequency fluctuations [46, 59]. Therefore, backbone simulations could be used to emulate outcomes from a combined EEG/fMRI setup. One common term in EEG investigations are so-called micro states, which represent quasi-stable potential distributions on the human scalp [64]. The simulation of dynamic backbone systems yields a large number of Ising micro states. On those micro states, a clustering procedure can be applied to extract common patterns of the whole set, which in turn could resemble EEG micro states.

Furthermore, simulations of Ising-like systems on dynamic backbones could shed light on how dFC traverses between different connectivity-states. The transitions between connectivity-states are smooth, because of the low-frequency characteristics of the BOLD signal and the additional smoothing introduced by the sliding window procedure. This could prevent to uncover characteristics of a transition between connectivity-states. By simulating the dynamic system on this slowly changing dynamic backbone, which also undergoes these transitions between connectivity-states, characteristics in the simulated physical parameters could predict those transitions. E. g. a rise in

the simulated susceptibility  $\chi$ , which is a marker for a system undergoing a phase transition, could show up as a common feature for the dynamic connectome at a state transition.

#### 7.4 EFFECTS OF DEFECTS IN THE CONNECTOME BACKBONE

The introduced approaches with various backbone systems can also be used to simulate defects on the human cortex. Matrices resulting from anatomical or functional connectomics  $\mathbf{X}$  represent the corresponding connection strength on the cortex for  $C$  brain locations. These connectome matrices can be altered by deletion to reflect distinct damages on the cortex. By deleting certain rows and columns in the connectivity matrices of the used backbone models, the failure of complete brain areas can be simulated resulting in backbone matrices  $\mathbf{X}_r$  with  $C_r < C$ . Using the same parameters for the dynamic system evolving on both the deficient and 'healthy' backbone, the outcomes in simulated physical parameters can be compared.

I plan to rely on a collaboration with the university clinic of Regensburg to use rs-fMRI and diffusion MRI data from brain tumor patients to also compare the above mentioned simulations to empirical data. On a regular basis, the lab around Prof. Dr. Brawanski acquires 15 min rs-fMRI data and also structural scans from brain tumor patients. Those patients have distinct lesions introduced to the cortex, which can be identified by combining the acquired structural scans with different contrasts and cortical atlases like the AAL [101]. Additionally, neuropsychological measurements acquired by physicians of the university clinic of Regensburg can also be used to underpin the suggested damage of those cortices revealed by the structural scans. The above mentioned deletion of certain rows and columns in the connectome matrices can be adjusted to reflect those distinct lesions. By using the connectomes derived from brain tumor patients data for simulating Ising-like systems, the outcome of these simulations can be compared to the adjusted models with introduced artificial damage. This comparison could yield deeper insights, whether – or to what extent – the used Ising-like system is suitable in such simulations. If a lesion introduced to the cortex by a brain tumor can be almost doubtlessly verified by structural scans and neuropsychological investigations, then deviations in the simulated physical parameters between the artificially damaged and the pathological backbones could validate or falsify the used Ising-like system.

In addition to the local brain tumor expertise, the method of deliberately introducing lesions to cortices and comparing them to real data can also be applied to other brain diseases. On the one hand, deficits at the brain area level can be investigated, and on the other hand, defects at the connections level. The former is explained above by deleting certain columns and rows in the backbone matrices. The latter could be realized by deleting certain entries in the connectome matrices  $X_{ij}$ , representing failures in connections. There are several diseases, which can be characterized by connection deficits. Popular examples are schizophrenia and bipolar disorder. Calhoun et al. [20]

find alterations in static functional connectivity in schizophrenic (fig. 50A) and bipolar patients (fig. 50B). In schizophrenics, they find that there is sig-

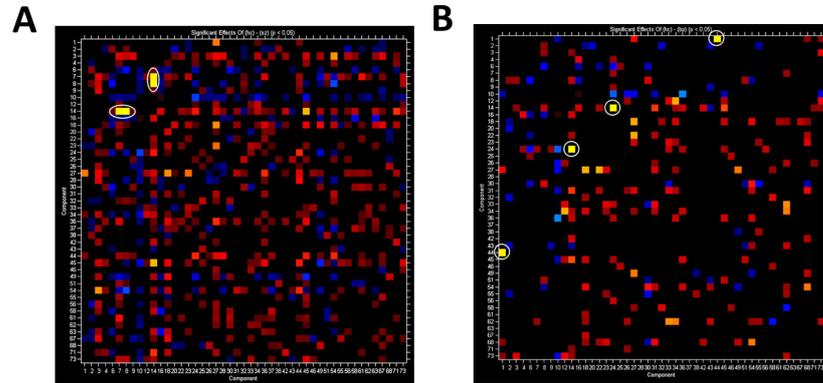


Figure 50: This figure depicts the group differences of the static functional connectome of schizophrenic (A) and bipolar patients (B). The circled connections show significantly larger correlations of the healthy connectomes compared to pathological ones. Assembled from [20] and adjusted.

nificantly lowered connectivity compared to healthy controls between the posterior cingulate and both the cerebellum and posterior temporal lobe. In patients with bipolar disorder, they find lowered correlation values compared to healthy controls, first, between posterior cingulate and temporo-occipital-parietal junction and, second, between anterior default mode regions and a posterior portion of the superior temporal lobe. By interpreting these connections as erroneous, the deletion of the corresponding entries in the connectome matrix  $X_{ij}$  yields an artificially damaged backbone, on which Ising-like systems can be simulated. For schizophrenics, I could rely on freely available data sets like the COBRE part of the NITRC<sup>11</sup>.

## 7.5 VARIATION OF THE ISING-MODEL

In the recent studies using Ising-like systems to do backbone simulations, the term Ising-model is used even though theoretically infinite-range interactions have been allowed in the Hamiltonian [71, 98], which is a typical characteristic of a certain spin-glass model – namely the Sherrington-Kirkpatrick model [91]. However, in the study of Hudetz et al. [52] a spin-glass model is used to simulate a backbone system. Both approaches reveal intriguing features of dynamic systems simulated in such a way. But there has not been suggested a resolution of whether an Ising-model or a spin-glass model is more suited for simulating backbone systems – or rather a model in between those two. I suggest a variation of the Ising-model to resolve this question in my proposal.

<sup>11</sup> [http://fcon\\_1000.projects.nitrc.org/indi/retro/cobre.html](http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html)

This is done by leaving the neighborhood variable in the Hamiltonian of an Ising-like system

$$H_i = -s_i \sum_{j=1}^n J_{ij} s_j, \quad (104)$$

where the connectivity matrix  $A_{ij}$  is implemented in the coupling  $J_{ij} \propto A_{ij}$ . The neighborhood  $\mathcal{N}$  can be varied between the two extreme cases of nearest neighbors  $n = |\mathcal{N}_0|$  and the whole lattice  $n = |\mathcal{N}_N|$ . By using such a variable neighborhood size, it is possible to sweep between a more Ising-like and more spin-glass like model. The realization of a variable neighborhood is problematic in the case of connectivity matrices as backbones. Consider the simulation of an Ising-like system on a connectome. The connectivity strength between brain areas  $i$  and  $j$  does not necessarily convey the distance information of those two areas. Therefore, an alternative way of defining a distance between two areas is suggested in my proposal. I propose to exploit the results and method introduced in a study of Misisic et al. [73]. They apply a **linear threshold model** on the human structural connectome to track the spreading of perturbations on this network (fig. 51). By placing a seed

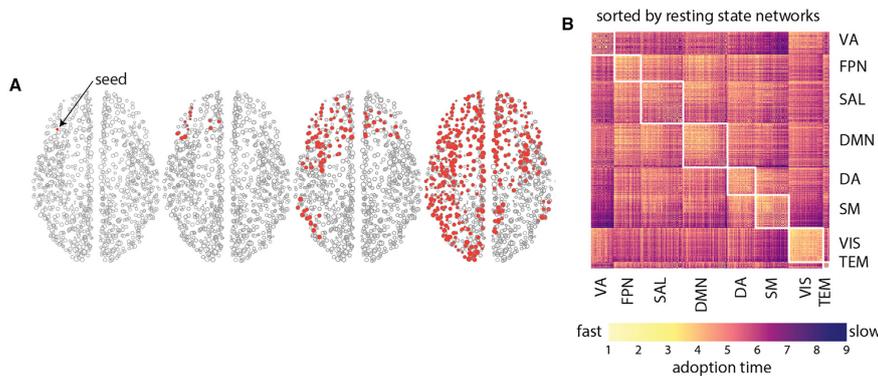


Figure 51: In this figure, the approach from Misisic et al. [73] is depicted. (A) shows the spreading of a perturbation (seed) on the human structural connectome by means of a linear threshold model. (B) The adoption time matrix represents how long it takes for a perturbation to travel from node  $i$  to node  $j$ . Taken from [73] and adjusted.

perturbation on each node  $i$  of the connectome and tracking its spreading behavior, they are able to simulate the time it takes for the perturbation to reach node  $j$ . Their results can be represented in a so-called adoption time matrix  $G$  shown in figure 51B. I propose to use this information to define neighborhoods of variable sizes. E. g. if the nodes included in the neighborhood of node  $i$  should have a distance of at most 5, then nodes with  $G_{ij} \leq 5$  time steps away are assigned to this neighborhood. This means that the entries in adoption time matrix  $G_{ij}$  serve as distance information to define a neighborhood of a certain node. Following this line of thoughts, nearest neighbors of node  $i$  would be nodes with adoption time  $G_{ij} = 1$ , and in general a neighborhood of distance  $n$  would be realized by nodes with  $G_{ij} \leq n$ .

*A linear threshold model describes the spreading of a perturbation through a network, where the activity of a node depends on the activity of its neighbors [73, 41].*

In order to choose, which neighborhood size is most suitable for Ising-like simulations, a proper model selection criterion has to be introduced. I propose to rely on the results found in the study of Marinazzo et al. [71]. They find that there are corresponding peaks in the simulated parameters susceptibility  $\chi$ , specific heat capacity  $C_v$ , and transfer entropy  $S_{Tr}$  (fig. 52) of their backbone model, which pointed to a system at a critical state at this particular temperature. Interpreting this tendency of corresponding peaks as an

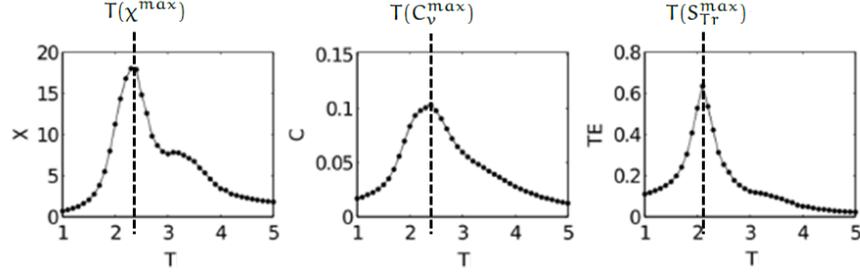


Figure 52: This figure depicts result from Marinazzo et al. [71]. They simulate an Ising-like system on the human structural connectome as a backbone and find that parameters like the susceptibility (left), specific heat capacity (middle), and transfer entropy (right) show a corresponding peak at a certain temperature. From [71].

inherent feature of Ising-like systems, a quality criterion can be constructed. I define a cost function

$$\psi_c \propto \Delta T(C_v^{\max}, \chi^{\max}) \Delta T(C_v^{\max}, S_{Tr}^{\max}) \Delta T(S_{Tr}^{\max}, \chi^{\max}), \quad (105)$$

with  $\Delta T(C_v^{\max}, \chi^{\max})$  being the difference of temperatures of the peak specific heat capacity  $T(C_v^{\max})$  and peak susceptibility  $T(\chi^{\max})$ . Analogously for the other two factors. This cost function  $\psi_c$  has a minimum, when all three peaks of the simulated parameters correspond to each other. For every simulated neighborhood size  $\mathcal{N}$ ,  $\psi_c$  could be evaluated on the static functional connectome as a backbone and the most suitable size could be selected.

## CONCLUSION

*"Und kurz umrühr'n – fertig. [...]  
Des schmeckt irgendwie – nach irgendwos."  
(Joe Waschl)*

In one part of my PhD thesis, I have investigated an extension of dFC to a frequency-resolved version of it by applying a sliding window approach on time courses from different frequency scales. The method of choice for the frequency decomposition has been MEMD applied on time-series resulting from a gICA of rs-fMRI data. Frequency scales have been defined by the resulting narrow band IMFs. Those frequency-resolved time courses have been segmented by using a box-car function slid over the whole time span with a step width of one TR. For each of those windows and within each frequency scale, correlation matrices have been calculated by cross-correlating the time course segments. This has resulted in sets of correlation matrices for each rs-fMRI session and scale. In a next step, on the correlation matrices from all sessions on each scale a k-means clustering procedure has been applied. This has resulted in frequency-resolved connectivity-states, which shows scale-stable behavior. I find a significant change of scale stability of connectivity-states from  $k = 4$  to  $k = 5$  extracted centroids. To discard findings by chance, I have investigated scale stability on shuffled and phase randomized null-data finding an opposite trend to the empirical results. Additionally, I have conducted *post hoc* filter bank investigations with a varying number of frequency scales confirming my results. In a next step, I have used simulated data traversing different numbers of predefined artificial connectivity-states. Investigating scale stability of this simulated data reveals the potential of my approach to detect data inherent cluster structure in a more objective way than commonly used measures. Furthermore, simulations of  $k_{inh} = 4$  connectivity-states have confirmed the hypothesis of  $k_{inh} = 4$  data inherent states in the rs-fMRI case. The behavior of my introduced scale stability measure is also supported by theoretical considerations. Since all those studies have been conducted using windows with a constant size over frequency scales, adaptation of window size on the varying frequency over scales is also investigated. Eventually, the course of this work has culminated in the development of a sensitive and potentially objective method for detecting data inherent cluster structures of specific types of data.

The other part of my thesis has dealt with the development of ideas for a proposal directed to the German science foundation for the first funding of my planned PostDoc phase. I have extended the recently emerging field of simulating Ising-like systems on static functional and structural connectomes as backbones to a dynamic version of it. I have proposed to use the dynamic connectome resulting from dFC analysis as a time varying back-

bone introducing a time dependency to the used Ising-like systems. I have suggested to do this in a quasistatic, dynamic, and frequency-resolved way. Additionally, I have proposed to combine different modalities in Ising-like systems employing an external magnetic field term. My ideas open up a way to investigate artificially introduced damages to the used connectomes. The dynamics of those damaged systems could then be compared to healthy ones and also to real lesioned cortices. Since it is not clear from literature what Ising-like system is to be preferred, I have proposed a variation of the Ising-model with a variable extent of the used neighborhood. By exploiting recent findings of spreading dynamics on the human connectome, I have suggested a way of defining variable neighborhoods. Literature findings on criticality in rs-fMRI data yield a way of model selection for the neighborhood of the introduced Ising-like system.

#### HIGHLIGHTING THE MAIN ACHIEVEMENTS DURING MY PHD PROCESS

- Developing a novel method for selecting data inherent clusters for specific types of data
- Presenting my ideas at the Human Brain Mapping conference in Honolulu, Hawaii as a talk and poster; achieving Merit abstract award
- Developing ideas for my PostDoc phase, becoming a member of the 7. Nachwuchsakademie Medizintechnik of the German science foundation, and creating my own proposal directed to the German science foundation for the first funding of my PostDoc phase; this proposal passed first selection step during my PhD phase, but awaits final decision

## ACKNOWLEDGEMENTS

---

I want to thank Prof. Dr. E. W. Lang for his very supportive, cooperative, constructive, friendly, and open long-term guidance and supervision of my work in his work group. It was a pleasure for me to sit together with him for many, many hours discussing and generating ideas. My acknowledgements are also directed to Prof. Dr. M. W. Greenlee, who always supported me during my work and gave me freedom in conducting my studies. I thank Prof. Dr. A. Tomé for her support and help during my PhD phase. I thank Dr. K. Rosengarth for her professional and emotional support – and for her friendship. I thank all my master students for the fruitful discussions and suggestions that also found their way into my work and this thesis. I thank the journal reviewers criticizing my work and suggesting changes and new analyses that I gratefully adopted for my work. Moreover, my thanks go to the open source community and especially the MATLAB file exchange platform, from which I used plenty of freely available code and toolboxes (e. g. `export_fig`, `nan-suite`, `tight_subplot`, `distinguishable_colors`, `lscatter`, `lscatterVarColor`, `suplabel`). And I also want to thank the 'Making pretty graphs' tutorial on the MATLAB blog<sup>12</sup>, for teaching me how to make pretty figures with MATLAB. I gratefully used the code snippets in this tutorial for my figures.

I thank my parents for their support and love. They always took the load off from me so that I could focus on my education. It would be so great, if you could see me getting my PhD, Mom. I also want to thank my girlfriend Stefanie for her love and emotional support. I thank G. Dumm for his long-term friendship and emotional support. Furthermore, I thank my crew and all my other friends for just being awesome and giving me feelings of security and belonging. Special thanks go to Dr. M. Böck for being as divine as he is.

---

<sup>12</sup> <http://blogs.mathworks.com/loren/2007/12/11/making-pretty-graphs/>



## APPENDIX

---

### A.1 VARIOUS MATHEMATICAL CONCEPTS

In this section, I introduce statistical concepts that are needed for understanding particular parts of my thesis, and which do not fit into the above chapters. It is referenced to each of the following sections in the text, when the reader can benefit from looking up one of the following particular concepts.

#### A.1.1 Pearson's correlation coefficient

The correlation coefficient of Pearson  $\rho_{ij}$  is defined as the covariance  $\text{cov}(X, Y)$  of two random variables  $X$  and  $Y$  normalized by the product of their standard deviations

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}, \quad (106)$$

with

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (x_i - \langle x \rangle)(y_i - \langle y \rangle)}{n - 1}, \quad (107)$$

where  $x_i$  and  $y_i$  for  $i = 1, \dots, n$  being the realizations of  $X$  and  $Y$ , respectively, and  $\langle \cdot \rangle$  denotes the mean over all realizations [15].

#### A.1.2 Partial correlation

Marrelec et al. [72] have been the first to introduce *partial correlation* as a measure for functional connectivity with the advantage of eliminating distracting information introduced by nodes not of concern. They mention that partial correlation represents "direct correlation" between two regions in the sense that it only considers correlation that cannot be accounted for by the other regions. They introduce partial correlation  $\tilde{\rho}_{ij}$  as a conditional coefficient

$$\tilde{\rho}_{ij} = \text{corr} [a_{v_i}(t), a_{v_j}(t) | \mathbf{a}_{\mathcal{R}}(t)], \quad (108)$$

where  $a_{v_i}(t)$  is the activity time course of node  $v_i$  and  $\mathbf{a}_{\mathcal{R}}(t)$  the remaining activity time courses of nodes  $\mathcal{R} = \mathcal{V} \setminus \{v_i, v_j\}$ , where  $\mathcal{V}$  is the set of all nodes. They describe the conditional dependence of  $\tilde{\rho}_{ij}$  with the projection of the two time courses  $a_{v_i}(t)$  and  $a_{v_j}(t)$  onto the subspace orthogonal to the time courses of the remaining nodes  $\mathbf{a}_{\mathcal{R}}(t)$ , and therefore correlation calculated in this subspace only refers to the variance of  $a_{v_i}(t)$  and  $a_{v_j}(t)$ .

### A.1.3 Mutual information

Given two random variables  $X$  and  $Y$  with a joint probability distribution  $P_{XY}(x_i, y_j)$  where  $x_i$  and  $y_j$  for  $i = 1, \dots, n$  reflect the realizations of the random variables, then *mutual information*  $I(X; Y)$  of the two random variables is defined as [63]

$$I(X; Y) = \sum_{i,j=1}^n P_{XY}(x_i, y_j) \log \frac{P_{XY}(x_i, y_j)}{P_X(x_i)P_Y(y_j)}, \quad (109)$$

where  $P_X(x_i)$  and  $P_Y(y_j)$  are the marginal probability distributions of the random variables  $X$  and  $Y$ , respectively. Latham and Roudi [63] show that mutual information can also be formulated by using Shannon's information entropy  $H(X) = -\sum_i P_X(x_i) \log P_X(x_i)$

$$I(X; Y) = H(X) - H(X|Y), \quad (110)$$

where  $H(X|Y)$  is the conditional entropy. This form of entropy is defined as [24]

$$H(X|Y) = -\sum_{i=1}^n P_Y(y_i) H(X|Y = y_i). \quad (111)$$

If entropy is interpreted as a measure of uncertainty of a random variable  $X$ , then mutual information can be seen as a reduction in the uncertainty of  $X$  by observing  $Y$  [63].

A.2 CONNECTIVITY-STATE ARRAYS OF  $k = 2, 3, 5, \dots, 10$

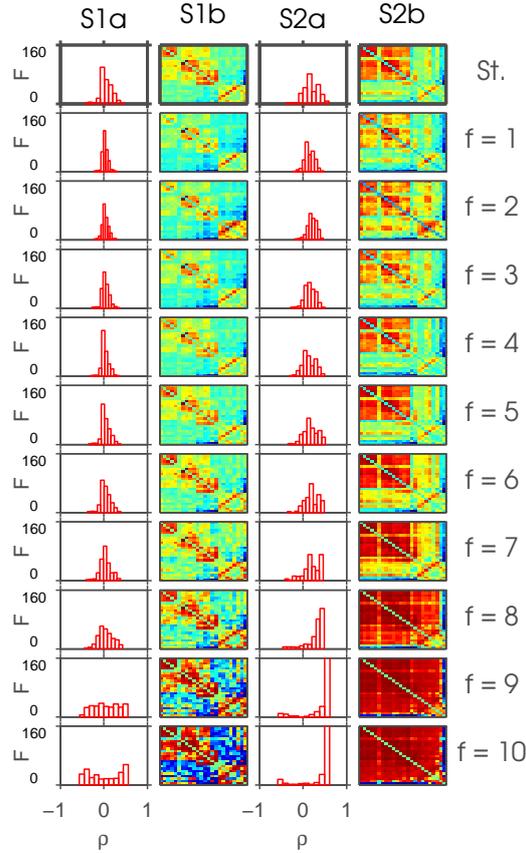


Figure 53: This figure depicts the result of the ordering algorithm 4 applied on connectivity-states resulting from frdFC procedure using  $\mathbf{U}_i^f$  extracted by MEMD. This is a realization of the ordering procedure without shuffling the frequency scales  $f$  and  $k = 2$  extracted connectivity-states on each frequency scale. In the top row, the connectivity-states resulting from the standard (St.) dFC procedure are shown. Below, each row shows the connectivity-states from frequency scales  $f = 1, \dots, 10$  with increasing frequency defined by the IMFs resulting from MEMD. Each column  $S^*b$  represents one connectivity-state, if it could be found robustly over frequency scales.. Color represents the range from minimum (blue) to maximum (red) value of correlation to highlight the common structure over scales. The columns  $S^*a$  show histograms of the absolute frequency of the correlation coefficients of the corresponding connectivity-states.

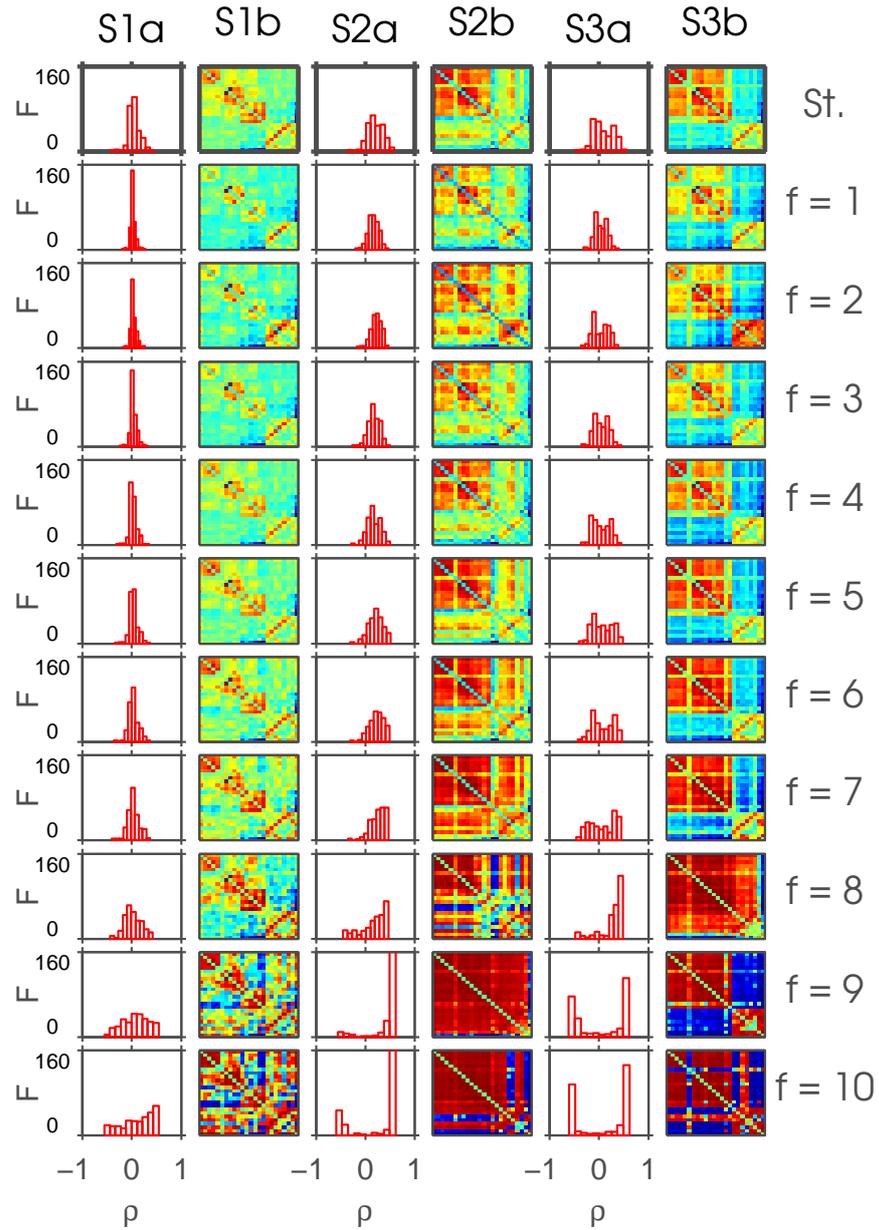


Figure 54: This figure depicts the result of the ordering algorithm 4 applied on connectivity-states resulting from frdFC procedure using  $U_i^f$  extracted by MEMD. This is a realization of the ordering procedure without shuffling the frequency scales  $f$  and  $k = 3$  extracted connectivity-states on each frequency scale. In the top row, the connectivity-states resulting from the standard (St.) dFC procedure are shown. Below, each row shows the connectivity-states from frequency scales  $f = 1, \dots, 10$  with increasing frequency defined by the IMFs resulting from MEMD. Each column  $S^*b$  represents one connectivity-state, if it could be found robustly over frequency scales. Color represents the range from minimum (blue) to maximum (red) value of correlation to highlight the common structure over scales. The columns  $S^*a$  show histograms of the absolute frequency of the correlation coefficients of the corresponding connectivity-states.

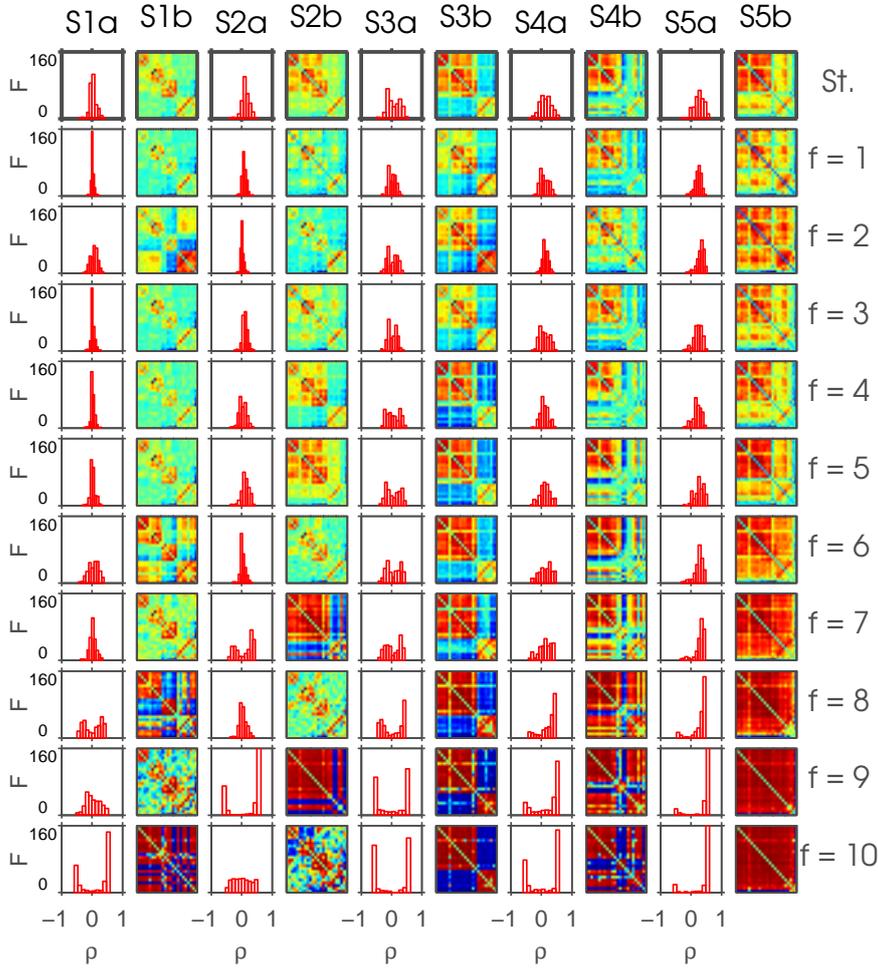


Figure 55: This figure depicts the result of the ordering algorithm 4 applied on connectivity-states resulting from frdFC procedure using  $\mathbf{U}_i^f$  extracted by MEMD. This is a realization of the ordering procedure without shuffling the frequency scales  $f$  and  $k = 5$  extracted connectivity-states on each frequency scale. In the top row, the connectivity-states resulting from the standard (St.) dFC procedure are shown. Below, each row shows the connectivity-states from frequency scales  $f = 1, \dots, 10$  with increasing frequency defined by the IMFs resulting from MEMD. Each column  $S^*b$  represents one connectivity-state, if it could be found robustly over frequency scales.. Color represents the range from minimum (blue) to maximum (red) value of correlation to highlight the common structure over scales. The columns  $S^*a$  show histograms of the absolute frequency of the correlation coefficients of the corresponding connectivity-states.

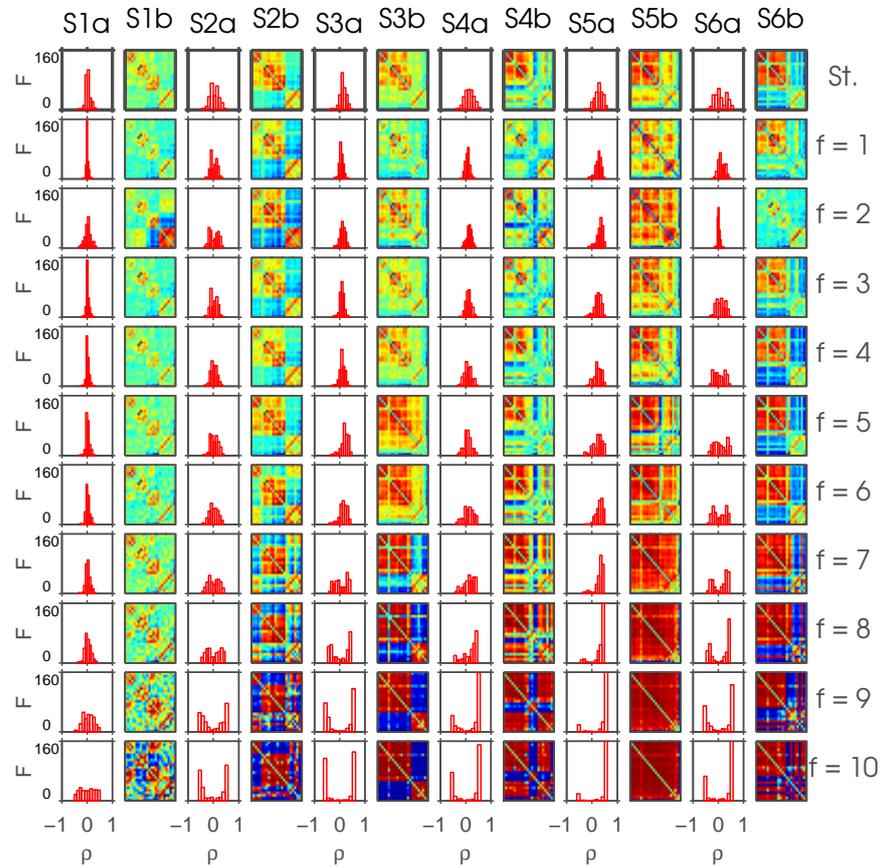


Figure 56: This figure depicts the result of the ordering algorithm 4 applied on connectivity-states resulting from  $\text{frdFC}$  procedure using  $\mathbf{U}_t^f$  extracted by MEMD. This is a realization of the ordering procedure without shuffling the frequency scales  $f$  and  $k = 6$  extracted connectivity-states on each frequency scale. In the top row, the connectivity-states resulting from the standard (St.) dFC procedure are shown. Below, each row shows the connectivity-states from frequency scales  $f = 1, \dots, 10$  with increasing frequency defined by the IMFs resulting from MEMD. Each column  $S^*b$  represents one connectivity-state, if it could be found robustly over frequency scales. Color represents the range from minimum (blue) to maximum (red) value of correlation to highlight the common structure over scales. The columns  $S^*a$  show histograms of the absolute frequency of the correlation coefficients of the corresponding connectivity-states.

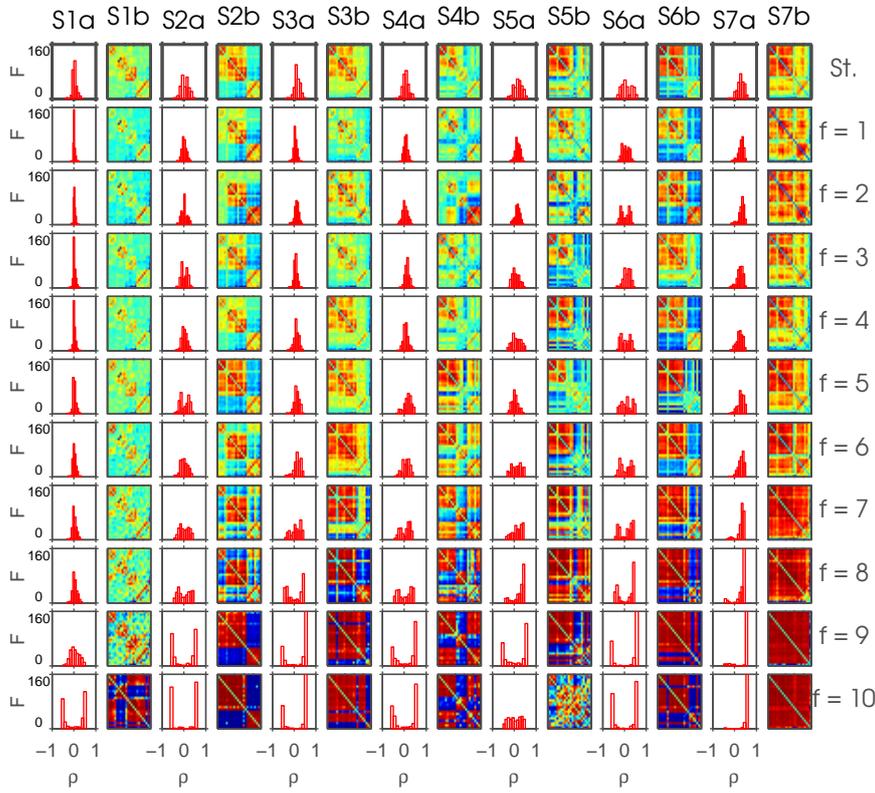


Figure 57: This figure depicts the result of the ordering algorithm 4 applied on connectivity-states resulting from frdFC procedure using  $U_i^f$  extracted by MEMD. This is a realization of the ordering procedure without shuffling the frequency scales  $f$  and  $k = 7$  extracted connectivity-states on each frequency scale. In the top row, the connectivity-states resulting from the standard (St.) dFC procedure are shown. Below, each row shows the connectivity-states from frequency scales  $f = 1, \dots, 10$  with increasing frequency defined by the IMFs resulting from MEMD. Each column  $S^*b$  represents one connectivity-state, if it could be found robustly over frequency scales.. Color represents the range from minimum (blue) to maximum (red) value of correlation to highlight the common structure over scales. The columns  $S^*a$  show histograms of the absolute frequency of the correlation coefficients of the corresponding connectivity-states.

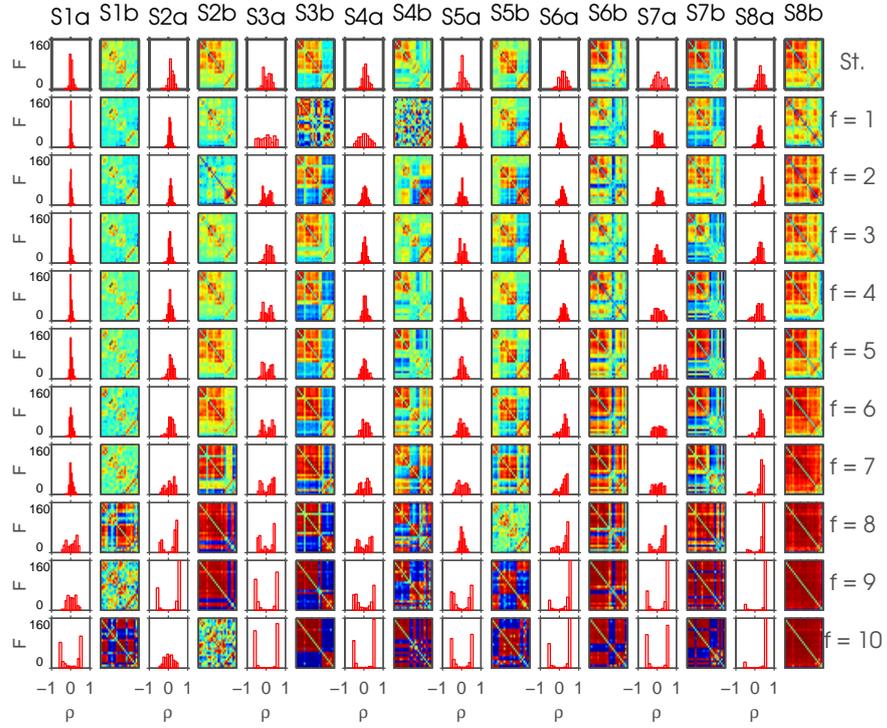


Figure 58: This figure depicts the result of the ordering algorithm 4 applied on connectivity-states resulting from frdFC procedure using  $U_i^f$  extracted by MEMD. This is a realization of the ordering procedure without shuffling the frequency scales  $f$  and  $k = 8$  extracted connectivity-states on each frequency scale. In the top row, the connectivity-states resulting from the standard (St.) dFC procedure are shown. Below, each row shows the connectivity-states from frequency scales  $f = 1, \dots, 10$  with increasing frequency defined by the IMFs resulting from MEMD. Each column  $S^*b$  represents one connectivity-state, if it could be found robustly over frequency scales. Color represents the range from minimum (blue) to maximum (red) value of correlation to highlight the common structure over scales. The columns  $S^*a$  show histograms of the absolute frequency of the correlation coefficients of the corresponding connectivity-states.

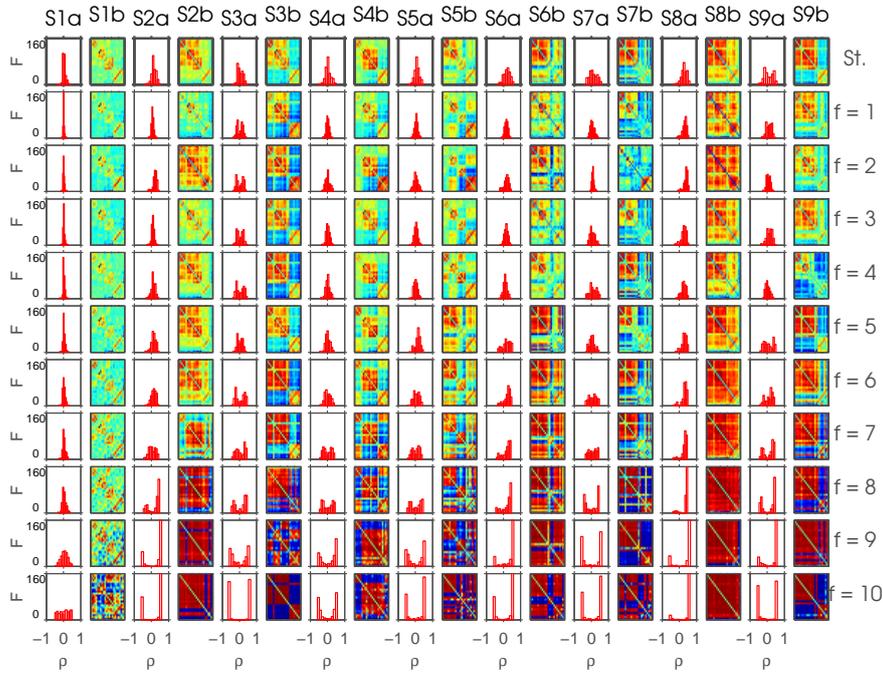


Figure 59: This figure depicts the result of the ordering algorithm 4 applied on connectivity-states resulting from frdFC procedure using  $\mathbf{U}_i^f$  extracted by MEMD. This is a realization of the ordering procedure without shuffling the frequency scales  $f$  and  $k = 9$  extracted connectivity-states on each frequency scale. In the top row, the connectivity-states resulting from the standard (St.) dFC procedure are shown. Below, each row shows the connectivity-states from frequency scales  $f = 1, \dots, 10$  with increasing frequency defined by the IMFs resulting from MEMD. Each column  $S^*b$  represents one connectivity-state, if it could be found robustly over frequency scales. Color represents the range from minimum (blue) to maximum (red) value of correlation to highlight the common structure over scales. The columns  $S^*a$  show histograms of the absolute frequency of the correlation coefficients of the corresponding connectivity-states.

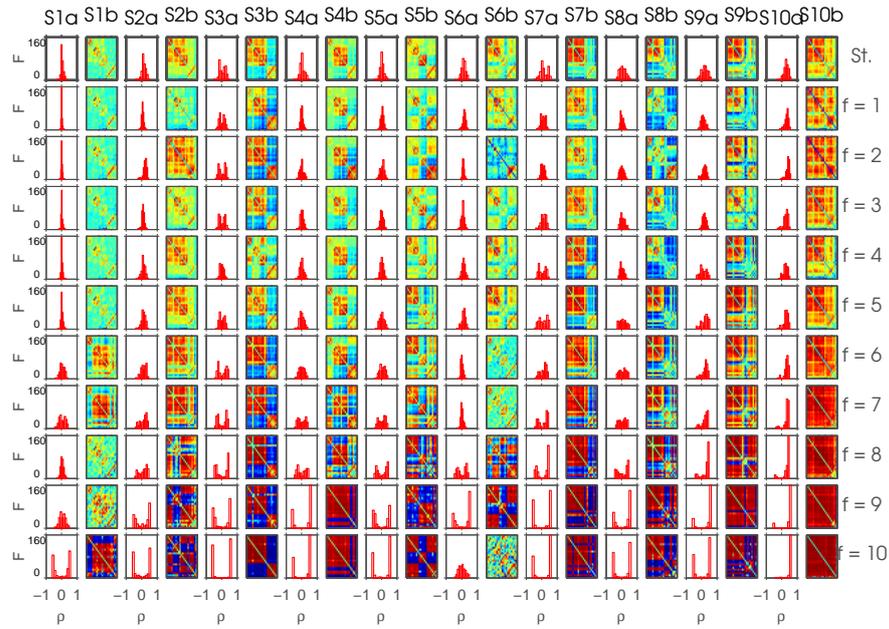


Figure 60: This figure depicts the result of the ordering algorithm 4 applied on connectivity-states resulting from  $\text{frdFC}$  procedure using  $U_i^f$  extracted by MEMD. This is a realization of the ordering procedure without shuffling the frequency scales  $f$  and  $k = 10$  extracted connectivity-states on each frequency scale. In the top row, the connectivity-states resulting from the standard (St.)  $\text{dFC}$  procedure are shown. Below, each row shows the connectivity-states from frequency scales  $f = 1, \dots, 10$  with increasing frequency defined by the IMFs resulting from MEMD. Each column  $S^*b$  represents one connectivity-state, if it could be found robustly over frequency scales. Color represents the range from minimum (blue) to maximum (red) value of correlation to highlight the common structure over scales. The columns  $S^*a$  show histograms of the absolute frequency of the correlation coefficients of the corresponding connectivity-states.

A.3 EXEMPLARY CONNECTIVITY-STATE ARRAY OF CLUSTERING WITH PRECEDING DEMEANING OF CORRELATION MATRICES

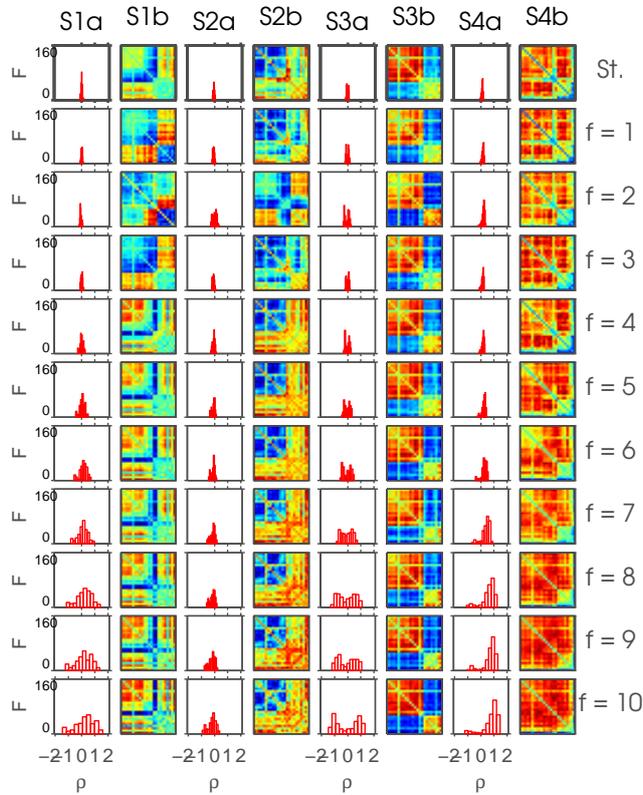


Figure 61: This figure depicts the result of the ordering algorithm 4 applied on connectivity-states resulting from frdFC procedure using  $\mathbf{U}_i^f$  extracted by MEMD. Before the clustering procedure was applied, the correlation matrices were demeaned. This is a realization of the ordering procedure without shuffling the frequency scales  $f$  and  $k = 4$  extracted connectivity-states on each frequency scale. In the top row, the connectivity-states resulting from the standard (St.) dFC procedure are shown. Below, each row shows the connectivity-states from frequency scales  $f = 1, \dots, 10$  with increasing frequency defined by the IMFs resulting from MEMD. Each column S\*b represents one connectivity-state, if it could be found robustly over frequency scales. Color represents the range from minimum (blue) to maximum (red) value of correlation to highlight the common structure over scales. The columns S\*a show histograms of the absolute frequency of the correlation coefficients of the corresponding connectivity-states. In this case, matrix entries show Fisher transformed, demeaned values represented by  $\rho$  in the histograms.

A.4 EXEMPLARY CONNECTIVITY-STATE ARRAYS OF SURROGATE TIME-SERIES

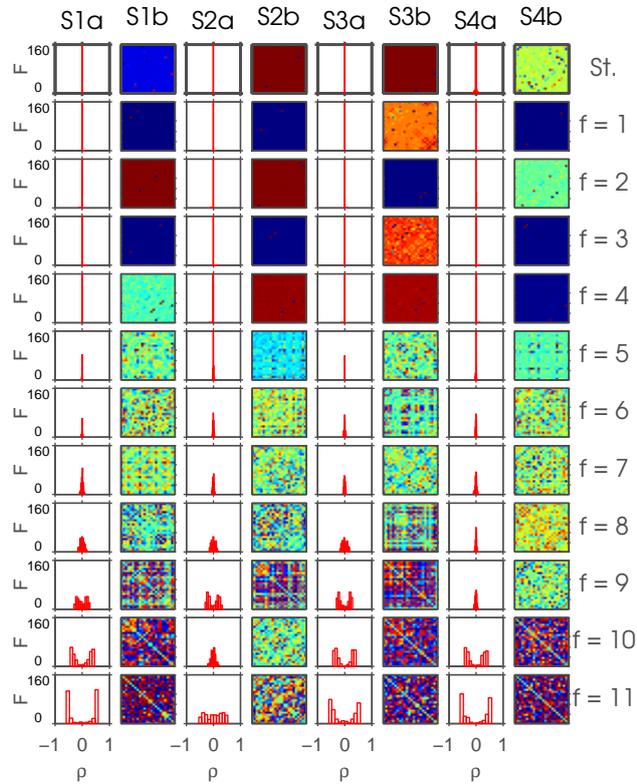


Figure 62: This figure depicts the result of the ordering algorithm 4 applied on connectivity-states resulting from frdFC procedure using  $U_i^f$  extracted by MEMD applied on shuffled time courses. In this case, shuffling was done in the time domain. This is a realization of the ordering procedure without shuffling the frequency scales  $f$  and  $k = 4$  extracted connectivity-states on each frequency scale. In the top row, the connectivity-states resulting from the standard (St.) dFC procedure are shown. Below, each row shows the connectivity-states from frequency scales  $f = 1, \dots, 10$  with increasing frequency defined by the IMFs resulting from MEMD. Color represents the range from minimum (blue) to maximum (red) value of correlation to highlight the common structure over scales. The columns S\*a show histograms of the absolute frequency of the correlation coefficients of the corresponding connectivity-states. All connectivity-states in this figure depict non-transformed correlation coefficients.

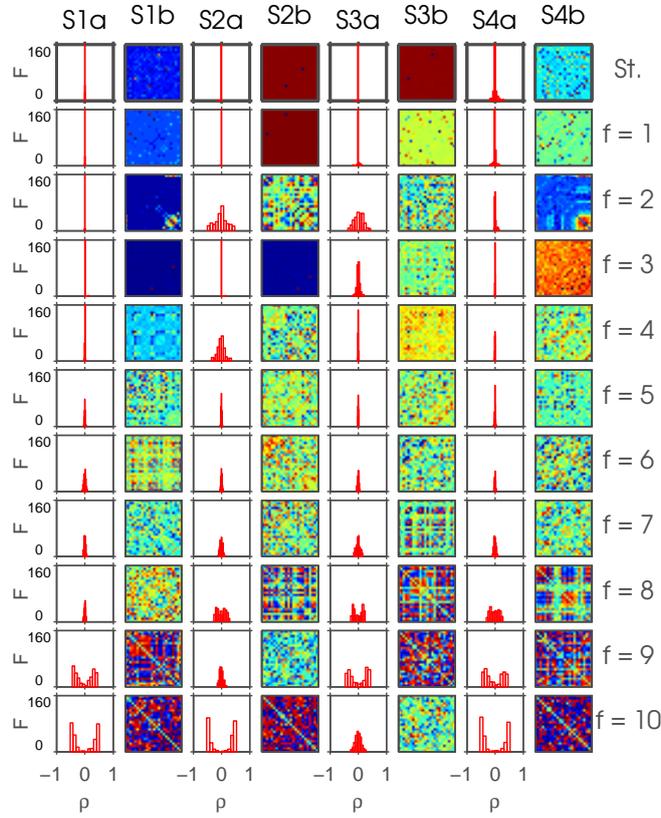


Figure 63: This figure depicts the result of the ordering algorithm 4 applied on connectivity-states resulting from frdFC procedure using  $\mathbf{U}_i^f$  extracted by MEMD applied on phase randomized time courses. This is a realization of the ordering procedure without shuffling the frequency scales  $f$  and  $k = 4$  extracted connectivity-states on each frequency scale. In the top row, the connectivity-states resulting from the standard (St.) dFC procedure are shown. Below, each row shows the connectivity-states from frequency scales  $f = 1, \dots, 10$  with increasing frequency defined by the IMFs resulting from MEMD. Color represents the range from minimum (blue) to maximum (red) value of correlation to highlight the common structure over scales. The columns  $S^*a$  show histograms of the absolute frequency of the correlation coefficients of the corresponding connectivity-states. All connectivity-states in this figure depict non-transformed correlation coefficients.



## BIBLIOGRAPHY

---

- [1] S. Achard. A Resilient, Low-Frequency, Small-World Human Brain Functional Network with Highly Connected Association Cortical Hubs. *Journal of Neuroscience*, 26(1):63–72, 2006. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.3874-05.2006. URL <http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.3874-05.2006>.
- [2] Sophie Achard and Ed Bullmore. Efficiency and cost of economical brain functional networks. *PLoS Computational Biology*, 3(2):0174–0183, 2007. ISSN 1553734X. doi: 10.1371/journal.pcbi.0030017.
- [3] M U Ahmed, N Rehman, D Looney, T M Rutkowski, P Kidmose, and D P Mandic. Multivariate entropy analysis with data-driven scales. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3901–3904. IEEE, mar 2012. ISBN 978-1-4673-0046-9. doi: 10.1109/ICASSP.2012.6288770. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6288770>.
- [4] Saad M. H. Al-Baddai. Correspondence, 2014.
- [5] Saad M H Al-Baddai, Karema S A Al-Subari, Ana Maria Tomé, Gregor Volberg, and Elmar W Lang. Combining EMD with ICA to analyze combined EEG-fMRI Data. pages 1–6, 2014.
- [6] K. Al-Subari, S. Al-Baddai, A.M. Tomé, M. Goldhacker, R. Faltermeier, and E.W. Lang. EMDLAB: A toolbox for analysis of single-trial EEG dynamics using empirical mode decomposition. *Journal of Neuroscience Methods*, jul 2015. ISSN 01650270. doi: 10.1016/j.jneumeth.2015.06.020. URL <http://linkinghub.elsevier.com/retrieve/pii/S0165027015002447>.
- [7] Elena A Allen, Erik B Erhardt, Eswar Damaraju, William Gruner, Judith M Segall, Rogers F Silva, Martin Havlicek, Srinivas Rachakonda, Jill Fries, Ravi Kalyanam, Andrew M Michael, Arvind Caprihan, Jessica A Turner, Tom Eichele, Steven Adelsheim, Angela D Bryan, Juan Bustillo, Vincent P Clark, Feldstein Ewing, Sarah W., Francesca Filbey, Corey C Ford, Kent Hutchison, Rex E Jung, Kent A Kiehl, Piyadasa Kodituwakku, Yuko M Komesu, Andrew R Mayer, Godfrey D Pearlson, John P Phillips, Joseph R Sadek, Michael Stevens, Ursina Teuscher, Robert J Thoma, and Vince D Calhoun. A Baseline for the Multivariate Comparison of Resting-State Networks. *Frontiers in Systems Neuroscience*, 5, 2011. ISSN 16625137. doi: 10.3389/fnsys.2011.00002.
- [8] Elena A Allen, Eswar Damaraju, Sergey M Plis, Erik B Erhardt, Tom Eichele, and Vince D Calhoun. Tracking whole-brain connectivity

- dynamics in the resting state. *Cerebral cortex (New York, N.Y. : 1991)*, 24(3):663–76, mar 2014. ISSN 1460-2199. doi: 10.1093/cercor/bhs352. URL <http://www.ncbi.nlm.nih.gov/pubmed/23146964>.
- [9] D. Arthur, D. Arthur, S. Vassilvitskii, and S. Vassilvitskii. k-means++: The advantages of careful seeding. *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, 8:1027–1035, 2007. ISSN 0898716241. doi: 10.1145/1283383.1283494. URL <http://portal.acm.org/citation.cfm?id=1283494>.
- [10] P A Bandettini, A Jesmanowicz, E C Wong, and J S Hyde. Processing strategies for time-course data sets in functional MRI of the human brain. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 30(2):161–173, 1993. ISSN 0740-3194. doi: 10.1002/mrm.1910300204.
- [11] Danielle Smith Bassett and Ed Bullmore. Small-world brain networks. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*, 12(6):512–23, dec 2006. ISSN 1073-8584. doi: 10.1177/1073858406293182. URL <http://www.ncbi.nlm.nih.gov/pubmed/17079517>.
- [12] Anthony J. Bell and Terrence J. Sejnowski. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation*, 7(6):1129–1159, nov 1995. ISSN 0899-7667. doi: 10.1162/neco.1995.7.6.1129. URL <http://www.mitpressjournals.org/doi/abs/10.1162/neco.1995.7.6.1129>.
- [13] Rasmus M. Birn, Jason B. Diamond, Monica A. Smith, and Peter A. Bandettini. Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI. *NeuroImage*, 31(4):1536–1548, 2006. ISSN 10538119. doi: 10.1016/j.neuroimage.2006.02.048.
- [14] Bharat Biswal, F Zerrin Yetkin, Victor M Houghton, and James S Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 34(4):537–41, 1995. ISSN 0740-3194. doi: 10.1002/mrm.1910340409. URL <http://www.ncbi.nlm.nih.gov/pubmed/8524021>.
- [15] Jürgen Bortz. *Statistik: Für Sozialwissenschaftler*. Springer-Verlag, 2010.
- [16] Korbinian Brodmann. *Vergleichende Lokalisationslehre der Großhirnrinde : in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. 1909.

- [17] Randy L Buckner, Jessica R Andrews-Hanna, and Daniel L Schacter. The brain's default network: anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, 1124:1–38, mar 2008. ISSN 0077-8923. doi: 10.1196/annals.1440.011. URL <http://www.ncbi.nlm.nih.gov/pubmed/18400922>.
- [18] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009. ISSN 1471-003X. doi: 10.1038/nrn2575.
- [19] V D Calhoun, T Adali, G D Pearlson, and J J Pekar. A method for making group inferences from functional MRI data using independent component analysis. *Human brain mapping*, 14(3):140–51, nov 2001. ISSN 1065-9471. doi: 10.1002/hbm.1048. URL <http://www.ncbi.nlm.nih.gov/pubmed/11559959>.
- [20] Vince D. Calhoun, Jing Sui, Kent Kiehl, Jessica Turner, Elena Allen, and Godfrey Pearlson. Exploring the psychosis functional connectome: Aberrant intrinsic networks in schizophrenia and bipolar disorder. *Frontiers in Psychiatry*, 2(January):1–13, 2012. ISSN 16640640. doi: 10.3389/fpsy.2011.00075.
- [21] D Cordes. Mapping functionally related regions of brain with functional connectivity MR imaging. *American Society of Neuroradiology*, 21 (October):1636–1644, 2000. ISSN 01956108.
- [22] D Cordes, V M Haughton, K Arfanakis, J D Carew, P a Turski, C H Moritz, M a Quigley, and M E Meyerand. Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data. *AJNR. American journal of neuroradiology*, 22(7):1326–33, aug 2001. ISSN 0195-6108. URL <http://www.ncbi.nlm.nih.gov/pubmed/11498421>.
- [23] Madalena Costa, Ary L. Goldberger, and C. K. Peng. Multiscale entropy analysis of biological signals. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 71(2):1–18, 2005. ISSN 15393755. doi: 10.1103/PhysRevE.71.021906.
- [24] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications. John Wiley & Sons, Inc., New York, USA, 2006.
- [25] Victor M. Eguíluz, Dante R. Chialvo, Guillermo a. Cecchi, Marwan Baliki, and a. Vania Apkarian. Scale-Free Brain Functional Networks. *Physical Review Letters*, 94(1):018102, jan 2005. ISSN 0031-9007. doi: 10.1103/PhysRevLett.94.018102. URL <http://link.aps.org/doi/10.1103/PhysRevLett.94.018102>.
- [26] Erik B. Erhardt, Elena a. Allen, Yonghua Wei, Tom Eichele, and Vince D. Calhoun. SimTB, a simulation toolbox for fMRI data under

- a model of spatiotemporal separability. *NeuroImage*, 59(4):4160–4167, 2012. ISSN 10538119. doi: 10.1016/j.neuroimage.2011.11.088.
- [27] Erik Barry Erhardt, Srinivas Rachakonda, Edward J Bedrick, Elena A Allen, Tülay Adali, and Vince D Calhoun. Comparison of multi-subject ICA methods for analysis of fMRI data. *Human brain mapping*, 32(12):2075–95, dec 2011. ISSN 1097-0193. doi: 10.1002/hbm.21170. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3117074&tool=pmcentrez&rendertype=abstract>.
- [28] Paul Expert, Renaud Lambiotte, Dante R Chialvo, Kim Christensen, Henrik Jeldtoft Jensen, David J Sharp, and Federico Turkheimer. Self-similar correlation function in brain resting-state functional magnetic resonance imaging. *Journal of the Royal Society, Interface / the Royal Society*, 8(September 2010):472–479, 2011. ISSN 1742-5662. doi: 10.1098/rsif.2010.0416.
- [29] David A Feinberg, Steen Moeller, Stephen M Smith, Edward Auerbach, Sudhir Ramanna, Matthias Gunther, Matt F Glasser, Karla L Miller, Kamil Ugurbil, and Essa Yacoub. Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. *PloS one*, 5(12):e15710, jan 2010. ISSN 1932-6203. doi: 10.1371/journal.pone.0015710. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3004955&tool=pmcentrez&rendertype=abstract>.
- [30] Bruce Fischl. FreeSurfer. *NeuroImage*, 62(2):774–81, aug 2012. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2012.01.021. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3685476&tool=pmcentrez&rendertype=abstract>.
- [31] Patrick Flandrin, Gabriel Rilling, and P. Goncalves. Empirical Mode Decomposition as a Filter Bank. *IEEE Signal Processing Letters*, 11(2):112–114, feb 2004. ISSN 1070-9908. doi: 10.1109/LSP.2003.821662. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1261951>.
- [32] Julien Fleureau, Amar Kachenoura, Laurent Albera, Jean-Claude Nunes, and Lotfi Senhadji. Multivariate empirical mode decomposition and application to multichannel filtering. *Signal Processing*, 91(12):2783–2792, dec 2011. ISSN 01651684. doi: 10.1016/j.sigpro.2011.01.018. URL <http://linkinghub.elsevier.com/retrieve/pii/S0165168411000338>.
- [33] M D Fox and M E Raichle. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat Rev Neurosci*, 8(9):700–711, 2007. ISSN 1471-003X. doi: nrn2201[pii]n10.1038/nrn2201. URL <http://www.ncbi>.

[nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=17704812](http://nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17704812).

- [34] Michael D Fox, Abraham Z Snyder, Justin L Vincent, Maurizio Corbetta, David C Van Essen, and Marcus E Raichle. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27):9673–8, 2005. ISSN 0027-8424. doi: 10.1073/pnas.0504136102. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1157105&tool=pmcentrez&rendertype=abstract>.
- [35] Daniel Fraiman, Pablo Balenzuela, Jennifer Foss, and Dante Chialvo. Ising-like dynamics in large-scale functional brain networks. *Physical Review E*, 79(6):061922, jun 2009. ISSN 1539-3755. doi: 10.1103/PhysRevE.79.061922. URL <http://link.aps.org/doi/10.1103/PhysRevE.79.061922>.
- [36] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics (Oxford, England)*, 9(3):432–41, jul 2008. ISSN 1468-4357. doi: 10.1093/biostatistics/kxm045. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3019769&tool=pmcentrez&rendertype=abstract>.
- [37] Thomas Geigenfeind. *Extracting Brain States with Multivariate Empirical Mode Decomposition*. PhD thesis, University of Regensburg, 2015.
- [38] Thomas Geigenfeind. Correspondence, 2015.
- [39] Matthew F Glasser, Stamatiou N Sotiropoulos, J Anthony Wilson, Timothy S Coalson, Bruce Fischl, Jesper L Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R Polimeni, David C Van Essen, and Mark Jenkinson. The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, 80:105–24, oct 2013. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2013.04.127. URL <http://www.ncbi.nlm.nih.gov/pubmed/23668970>.
- [40] Markus Goldhacker, Ana Maria Tomé, Mark W. Greenlee, and Elmar W. Lang. Frequency-resolved dynamic functional connectivity and scale-invariant connectivity-state behavior. <http://arxiv.org/abs/1511.00964>, 2015. URL <http://arxiv.org/abs/1511.00964>.
- [41] Mark Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978.
- [42] Ludovica Griffanti, Gholamreza Salimi-Khorshidi, Christian F Beckmann, Edward J Auerbach, Gwenaëlle Douaud, Claire E Sexton, Enikő Zsoldos, Klaus P Ebmeier, Nicola Filippini, Clare E Mackay, Steen

- Moeller, Junqian Xu, Essa Yacoub, Giuseppe Baselli, Kamil Ugurbil, Karla L Miller, and Stephen M Smith. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage*, 95:232–47, jul 2014. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2014.03.034. URL <http://www.ncbi.nlm.nih.gov/pubmed/24657355>.
- [43] Patric Hagmann, Leila Cammoun, Xavier Gigandet, Reto Meuli, Christopher J Honey, Van J Wedeen, and Olaf Sporns. Mapping the structural core of human cerebral cortex. *PLoS biology*, 6(7): e159, 2008. ISSN 1545-7885. doi: 10.1371/journal.pbio.0060159. URL <http://www.ncbi.nlm.nih.gov/pubmed/18597554>.
- [44] A. Hendrix. *Magnets, spins and resonances – an introduction to the basics of magnetic resonance*. SIEMENS AG, 2003.
- [45] Kristina Heyn. *Classification of skin tumor images with Bi-Dimensional Ensemble Empirical Mode Decomposition*. PhD thesis, University of Regensburg, 2014.
- [46] Tuija Hiltunen, Jussi Kantola, Ahmed Abou Elseoud, Pasi Lepola, Kalervo Suominen, Tuomo Starck, Juha Nikkinen, Jukka Remes, Osmo Tervonen, Satu Palva, Vesa Kiviniemi, and J Matias Palva. Infralow EEG fluctuations are correlated with resting-state network dynamics in fMRI. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 34(2):356–62, 2014. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.0276-13.2014. URL <http://www.ncbi.nlm.nih.gov/pubmed/24403137>.
- [47] Johan Himberg, Aapo Hyvärinen, and Fabrizio Esposito. Validating the independent components of neuroimaging time series via clustering and visualization. *NeuroImage*, 22(3):1214–22, jul 2004. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2004.03.027. URL <http://www.sciencedirect.com/science/article/pii/S1053811904001661><http://www.ncbi.nlm.nih.gov/pubmed/15219593>.
- [48] Joseph P. Hornak. *The Basics of MRI*. URL <https://www.cis.rit.edu/htbooks/mri/index.html>.
- [49] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 454(1971): 903–995, mar 1998. ISSN 1364-5021. doi: 10.1098/rspa.1998.0193. URL <http://rspa.royalsocietypublishing.org/cgi/doi/10.1098/rspa.1998.0193>.

- [50] N. E. Huang, M.-L. C. Wu, S. R. Long, S. S. P. Shen, W. Qu, P. Gloersen, and K. L. Fan. A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 459(2037):2317–2345, 2003. ISSN 1364-5021. doi: 10.1098/rspa.2003.1123.
- [51] Norden Huang, Zhaohua Wu, and Steven Long. Hilbert-Huang transform. *Scholarpedia*, 3(7):2544, 2008. ISSN 1941-6016. doi: 10.4249/scholarpedia.2544. URL [http://www.scholarpedia.org/article/Hilbert-Huang{}\\_transform](http://www.scholarpedia.org/article/Hilbert-Huang{}_transform).
- [52] Anthony G. Hudetz, Colin J. Humphries, and Jeffrey R. Binder. Spin-glass model predicts metastable brain states that diminish in anesthesia. *Frontiers in Systems Neuroscience*, 8(December): 1–9, 2014. ISSN 1662-5137. doi: 10.3389/fnsys.2014.00234. URL <http://journal.frontiersin.org/journal/10.3389/fnsys.2014.00234/abstract>.
- [53] Scott A. Huettel, Allen W. Song, and Gregory McCarthy. *Functional magnetic resonance imaging*. Sinauer Associates, Sunderland, vol. 1 edition, 2004.
- [54] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, Inc., 10 edition, 2001.
- [55] Mark Jenkinson, Peter Bannister, Michael Brady, and Stephen Smith. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, 17(2):825–841, oct 2002. ISSN 10538119. doi: 10.1006/nimg.2002.1132. URL <http://linkinghub.elsevier.com/retrieve/pii/S1053811902911328>.
- [56] Mark Jenkinson, Christian F Beckmann, Timothy E J Behrens, Mark W Woolrich, and Stephen M Smith. FSL. *NeuroImage*, 62(2):782–90, aug 2012. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2011.09.015. URL <http://www.ncbi.nlm.nih.gov/pubmed/21979382>.
- [57] J. R. Johnson. *Introduction to digital signal processing*. Prentice-Hall, Inc., New Jersey, 1989.
- [58] I Jolliffe. *Principal Component Analysis*. John Wiley & Sons, Inc., 2002.
- [59] Shella Dawn Keilholz. The Neural Basis of Time-Varying Resting-State Functional Connectivity. *Brain Connectivity*, 4(10):769–779, dec 2014. ISSN 2158-0014. doi: 10.1089/brain.2014.0250. URL <http://www.ncbi.nlm.nih.gov/pubmed/24975024><http://online.liebertpub.com/doi/abs/10.1089/brain.2014.0250>.
- [60] Manfred G Kitzbichler, Marie L Smith, Søren R Christensen, and Ed Bullmore. Broadband criticality of human brain network

- synchronization. *PLoS computational biology*, 5(3):e1000314, mar 2009. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1000314. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2647739&tool=pmcentrez&rendertype=abstract>.
- [61] Harold W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83-97, mar 1955. ISSN 00281441. doi: 10.1002/nav.3800020109. URL <http://doi.wiley.com/10.1002/nav.3800020109>.
- [62] Elmar W. Lang. Correspondence, 2015.
- [63] Peter Latham and Yasser Roudi. Mutual information. *Scholarpedia*, 4(1):1658, 2009. ISSN 1941-6016. doi: 10.4249/scholarpedia.1658. URL [http://www.scholarpedia.org/article/Mutual\\_information](http://www.scholarpedia.org/article/Mutual_information).
- [64] Dietrich Lehmann, Roberto Pascual-Marqui, and Christoph Michel. EEG microstates. *Scholarpedia*, 4(3):7632, 2009. ISSN 1941-6016. doi: 10.4249/scholarpedia.7632. URL [http://www.scholarpedia.org/article/EEG\\_microstates](http://www.scholarpedia.org/article/EEG_microstates).
- [65] Nora Leonardi, Jonas Richiardi, Markus Gschwind, Samanta Simioni, Jean-Marie Annoni, Myriam Schlupe, Patrik Vuilleumier, and Dimitri Van De Ville. Principal components of functional connectivity: a new approach to study dynamic brain connectivity during rest. *NeuroImage*, 83:937-50, dec 2013. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2013.07.019. URL <http://www.ncbi.nlm.nih.gov/pubmed/23872496>.
- [66] Nora Leonardi, William R Shirer, Michael D Greicius, and Dimitri Van De Ville. Disentangling dynamic networks: Separated and joint expressions of functional connectivity patterns in time. *Human Brain Mapping*, page n/a, jul 2014. ISSN 10659471. doi: 10.1002/hbm.22599. URL <http://www.ncbi.nlm.nih.gov/pubmed/25081921>.
- [67] Doron Levy. *Introduction to numerical analysis*. Department of Mathematics and Center for Scientific Computation and Mathematical Modeling, CSCAMM, University of Maryland, 2010.
- [68] David Looney, Apit Hemakom, and Danilo P Mandic. Intrinsic multi-scale analysis : a multi-variate empirical mode decomposition framework Subject Areas :. *Proc. R. Soc. A*, 471(20140709), 2015. doi: 10.1098/rspa.2014.0709.
- [69] D J C MacKay. An Example Inference Task: Clustering. In *Information Theory, Inference and Learning Algorithms*, chapter 20, page 640. Cambridge University Press, Cambridge, 2003. ISBN 0521642981.

- [70] Danilo P Mandic, Naveed Rehman, Zhaohua Wu, and Norden E Huang. Empirical Mode Decomposition-Based Time-Frequency Analysis of Multivariate Signals: The Power of Adaptive Data Analysis. *IEEE Signal Processing Magazine*, 30(6):74–86, nov 2013. ISSN 1053-5888. doi: 10.1109/MSP.2013.2267931. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6633076>.
- [71] Daniele Marinazzo, Mario Pellicoro, Guorong Wu, Leonardo Angelini, Jesús M. Cortés, and Sebastiano Stramaglia. Information transfer and criticality in the ising model on the human connectome. *PLoS ONE*, 9(4):1–7, 2014. ISSN 19326203. doi: 10.1371/journal.pone.0093616.
- [72] Guillaume Marrelec, Alexandre Krainik, Hugues Duffau, Mélanie Péligrini-Issac, Stéphane Lehericy, Julien Doyon, and Habib Benali. Partial correlation for functional brain interactivity investigation in functional MRI. *NeuroImage*, 32(1):228–237, aug 2006. ISSN 10538119. doi: 10.1016/j.neuroimage.2005.12.057. URL <http://linkinghub.elsevier.com/retrieve/pii/S1053811906000103>.
- [73] Bratislav Misic, Richard F. Betzel, Azadeh Nematzadeh, Joaquin Goñi, Alessandra Griffa, Patric Hagmann, Alessandro Flammini, Yong-Yeol Ahn, and Olaf Sporns. Cooperative and Competitive Spreading Dynamics on the Human Connectome. *Neuron*, 86(6):1518–1529, 2015. ISSN 08966273. doi: 10.1016/j.neuron.2015.05.035. URL <http://linkinghub.elsevier.com/retrieve/pii/S0896627315004742>.
- [74] Steen Moeller, Essa Yacoub, Cheryl A Olman, Edward Auerbach, John Strupp, Noam Harel, and Kâmil Uğurbil. Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 63(5):1144–53, may 2010. ISSN 1522-2594. doi: 10.1002/mrm.22361. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2906244&tool=pmcentrez&rendertype=abstract>.
- [75] James Munkres. Algorithms for the Assignment and Transportation Problems. *Journal of the Society for Industrial and Applied Mathematics*, 5(1):32–38, mar 1957. ISSN 0368-4245. doi: 10.1137/0105003. URL <http://epubs.siam.org/doi/abs/10.1137/0105003>.
- [76] Vincenzo Nicosia, John Tang, Cecilia Mascolo, Mirco Musolesi, Giovanni Russo, and Vito Latora. Graph Metrics for Temporal Networks. In *Temporal Networks*, pages 15–40. 2013. doi: 10.1007/978-3-642-36461-7\_2. URL [http://link.springer.com/10.1007/978-3-642-36461-7\\_2](http://link.springer.com/10.1007/978-3-642-36461-7_2).
- [77] A. V. Oppenheim and R. W. Schaffer. *Digital signal processing*. Prentice-Hall, Inc., London, 1975.

- [78] Wen-Ju Pan, Garth John Thompson, Matthew Evan Magnuson, Dieter Jaeger, and Shella Keilholz. Infralow LFP correlates to resting-state fMRI BOLD signals. *NeuroImage*, 74:288–297, 2013. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2013.02.035.
- [79] Jonathan D Power, Kelly A Barnes, Abraham Z Snyder, Bradley L Schlaggar, and Steven E Petersen. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, 59(3):2142–2154, 2012. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2011.10.018.
- [80] Dean Prichard and James Theiler. Generating surrogate data for time series with several simultaneously measured variables. 73(7): 4, 1994. ISSN 0031-9007. doi: 10.1103/PhysRevLett.73.951. URL <http://arxiv.org/abs/comp-gas/9405002>.
- [81] N. Rehman and D. P. Mandic. Multivariate empirical mode decomposition. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 466(2117):1291–1302, dec 2009. ISSN 1364-5021. doi: 10.1098/rspa.2009.0502. URL <http://rspa.royalsocietypublishing.org/cgi/doi/10.1098/rspa.2009.0502>.
- [82] Naveed Rehman and Danilo P Mandic. Filter Bank Property of Multivariate Empirical Mode Decomposition. *IEEE Transactions on Signal Processing*, 59(5):2421–2426, 2011. ISSN 1053-587X. doi: 10.1109/TSP.2011.2106779.
- [83] Naveed Rehman, Cheolsoo Park, Norden E Huang, and Danilo P Mandic. EMD via MEMD: multivariate noise-aided computation of standard EMD. *Advances in Adaptive Data Analysis*, 05(02):1350007, apr 2013. ISSN 1793-5369. doi: 10.1142/S1793536913500076. URL <http://www.worldscientific.com/doi/abs/10.1142/S1793536913500076>.
- [84] Simon Robinson, Gianpaolo Basso, Nicola Soldati, Uta Sailer, Jorge Jovicich, Lorenzo Bruzzone, Ilse Kryspin-Exner, Herbert Bauer, and Ewald Moser. A resting state network in the motor control circuit of the basal ganglia. *BMC neuroscience*, 10(Dm):137, jan 2009. ISSN 1471-2202. doi: 10.1186/1471-2202-10-137. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2785820&tool=pmcentrez&rendertype=abstract>.
- [85] Edmund T. Rolls, Marc Joliot, and Nathalie Tzourio-Mazoyer. Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *NeuroImage*, 122: 1–5, 2015. ISSN 10538119. doi: 10.1016/j.neuroimage.2015.07.075. URL <http://linkinghub.elsevier.com/retrieve/pii/S1053811915006953>.

- [86] Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(C):53–65, 1987. ISSN 03770427. doi: 10.1016/0377-0427(87)90125-7.
- [87] Sam Roweis. EM Algorithms for PCA and SPCA. *Advances in neural information processing systems*, pages 626–632, 1998.
- [88] Gholamreza Salimi-Khorshidi, Gwenaëlle Douaud, Christian F Beckmann, Matthew F Glasser, Ludovica Griffanti, and Stephen M Smith. Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, 90:449–68, apr 2014. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2013.11.046. URL <http://www.ncbi.nlm.nih.gov/pubmed/24389422>.
- [89] Raymond Salvador, John Suckling, Martin R. Coleman, John D. Pickard, David Menon, and Ed Bullmore. Neurophysiological architecture of functional magnetic resonance images of human brain. *Cerebral Cortex*, 15(9):1332–2342, 2005. ISSN 10473211. doi: 10.1093/cercor/bhio16.
- [90] Kawin Setsompop, Borjan A Gagoski, Jonathan R Polimeni, Thomas Witzel, Van J Wedeen, and Lawrence L Wald. Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 67(5):1210–24, may 2012. ISSN 1522-2594. doi: 10.1002/mrm.23097. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3323676&tool=pmcentrez&rendertype=abstract>.
- [91] David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. *Physical Review Letters*, 35(26):1792–1796, 1975. ISSN 00319007. doi: 10.1103/PhysRevLett.35.1792.
- [92] Amir Shmuel and David A. Leopold. Neuronal correlates of spontaneous fluctuations in fMRI signals in monkey visual cortex: Implications for functional connectivity at rest. *Human Brain Mapping*, 29(7):751–761, 2008. ISSN 10659471. doi: 10.1002/hbm.20580.
- [93] Karin Shmueli, Peter van Gelderen, Jacco A. de Zwart, Silvina G. Horovitz, Masaki Fukunaga, J. Martijn Jansma, and Jeff H. Duyn. Low-frequency fluctuations in the cardiac rate as a source of variance in the resting-state fMRI BOLD signal. *NeuroImage*, 38(2):306–320, 2007. ISSN 10538119. doi: 10.1016/j.neuroimage.2007.07.037.
- [94] Stephen M Smith, Karla L Miller, Gholamreza Salimi-Khorshidi, Matthew Webster, Christian F Beckmann, Thomas E Nichols, Joseph D Ramsey, and Mark W Woolrich. Network modelling methods for FMRI. *NeuroImage*, 54(2):875–91, jan 2011. ISSN 1095-9572. doi: 10.

- 1016/j.neuroimage.2010.08.063. URL <http://www.ncbi.nlm.nih.gov/pubmed/20817103>.
- [95] Stephen M Smith, Christian F Beckmann, Jesper Andersson, Edward J Auerbach, Janine Bijsterbosch, Gwenaëlle Douaud, Eugene Duff, David A Feinberg, Ludovica Griffanti, Michael P Harms, Michael Kelly, Timothy Laumann, Karla L Miller, Steen Moeller, Steve Petersen, Jonathan Power, Gholamreza Salimi-Khorshidi, Abraham Z Snyder, An T Vu, Mark W Woolrich, Junqian Xu, Essa Yacoub, Kamil Uğurbil, David C. Van Essen, and Matthew F Glasser. Resting-state fMRI in the Human Connectome Project. *NeuroImage*, 80:144–68, oct 2013. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2013.05.039. URL <http://www.ncbi.nlm.nih.gov/pubmed/23702415><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3720828&tool=pmcentrez&rendertype=abstract>.
- [96] Olaf Sporns. Brain connectivity. *Scholarpedia*, 2(10):4695, 2007. ISSN 1941-6016. doi: 10.4249/scholarpedia.4695. URL [http://www.scholarpedia.org/article/Brain{}\\_connectivity](http://www.scholarpedia.org/article/Brain{}_connectivity).
- [97] Bram Stieltjes, Romuald M. Brunner, Klaus H. Fritzsche, and Fredrik B. Laun. *Diffusion Tensor Imaging*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. ISBN 978-3-642-20455-5. doi: 10.1007/978-3-642-20456-2. URL <http://link.springer.com/10.1007/978-3-642-20456-2>.
- [98] Sebastiano Stramaglia, Mario Pellicoro, Leonardo Angelini, Enrico Amico, Hannelore Aerts, Jesus Cortés, Steven Laureys, and Daniele Marinazzo. Conserved Ising Model on the Human Connectome. <http://arxiv.org/abs/1509.02697>, pages 1–7, sep 2015. URL <http://arxiv.org/abs/1509.02697>.
- [99] Enzo Tagliazucchi, Pablo Balenzuela, Daniel Fraiman, and Dante R Chialvo. Criticality in large-scale brain FMRI dynamics unveiled by a novel point process analysis. *Frontiers in physiology*, 3:15, jan 2012. ISSN 1664-042X. doi: 10.3389/fphys.2012.00015. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3274757&tool=pmcentrez&rendertype=abstract>.
- [100] James Theiler, Stephen Eubank, André Longtin, Bryan Galdrikian, and J. Doyné Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, 58(1-4):77–94, 1992. ISSN 01672789. doi: 10.1016/0167-2789(92)90102-S.
- [101] N Tzourio-Mazoyer, B Landeau, D Papathanassiou, F Crivello, O Etard, N Delcroix, B Mazoyer, and M Joliot. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1):273–289, 2002. ISSN 1053-8119. doi: 10.1006/nimg.2001.0978.

- [102] Martijn P van den Heuvel and Hilleke E Hulshoff Pol. Exploring the brain network: a review on resting-state fMRI functional connectivity. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology*, 20(8):519–34, aug 2010. ISSN 1873-7862. doi: 10.1016/j.euroneuro.2010.03.008. URL <http://www.ncbi.nlm.nih.gov/pubmed/20471808>.
- [103] D C Van Essen, K Ugurbil, E Auerbach, D Barch, T E J Behrens, R Bucholz, A Chang, L Chen, M Corbetta, S W Curtiss, S Della Penna, D Feinberg, M F Glasser, N Harel, A C Heath, L Larson-Prior, D Marcus, G Michalareas, S Moeller, R Oostenveld, S E Petersen, F Prior, B L Schlaggar, S M Smith, A Z Snyder, J Xu, and E Yacoub. The Human Connectome Project: a data acquisition perspective. *NeuroImage*, 62(4):2222–31, oct 2012. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2012.02.018. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3606888&tool=pmcentrez&rendertype=abstract>.
- [104] Gaël Varoquaux, Alexandre Gramfort, Jean Baptiste Poline, and Bertrand Thirion. Brain covariance selection: better individual functional connectivity models using population prior. aug 2010. URL <http://arxiv.org/abs/1008.5071>.
- [105] Gang Wang, Xian-Yao Chen, Fang-Li Qiao, Zhaohua Wu, and Norden E. Huang. On intrinsic mode function. *Advances in Adaptive Data Analysis*, 02(03):277–293, jul 2010. ISSN 1793-5369. doi: 10.1142/S1793536910000549. URL <http://www.worldscientific.com/doi/abs/10.1142/S1793536910000549>.
- [106] Wikipedia. Low-discrepancy sequence. URL [https://en.wikipedia.org/wiki/Low-discrepancy\\_sequence](https://en.wikipedia.org/wiki/Low-discrepancy_sequence).
- [107] Tien-Tsin Wong, Wai-Shing Luk, and Pheng-Ann Heng. Sampling with Hammersley and Halton Points. *Journal of Graphics Tools*, 2(2):9–24, jan 1997. ISSN 1086-7651. doi: 10.1080/10867651.1997.10487471. URL <http://www.tandfonline.com/doi/abs/10.1080/10867651.1997.10487471>.
- [108] Z. Wu and N. E. Huang. A study of the characteristics of white noise using the empirical mode decomposition method. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 460(2046):1597–1611, jun 2004. ISSN 1364-5021. doi: 10.1098/rspa.2003.1221. URL <http://rspa.royalsocietypublishing.org/cgi/doi/10.1098/rspa.2003.1221>.
- [109] Zhaohua Wu and Norden E. Huang. Ensemble Empirical Mode Decomposition: a Noise-Assisted Data Analysis Method. *Advances in Adaptive Data Analysis*, 01(01):1–41, jan 2009. ISSN 1793-5369. doi: 10.1142/S1793536909000047. URL <http://www.worldscientific.com/doi/abs/10.1142/S1793536909000047>.

- [110] Zhaohua Wu, Norden E. Huang, and Xianyao Chen. The multi-dimensional ensemble empirical mode decomposition method. *Advances in Adaptive Data Analysis*, 1(3):339–372, jul 2009. ISSN 1793-5369. doi: 10.1142/S1793536909000187. URL <http://www.worldscientific.com/doi/abs/10.1142/S1793536909000187>.
- [111] J Xu, S Moeller, J Strupp, E J Auerbach, L Chen, D A Feinberg, K Ugurbil, and E Yacoub. Highly accelerated whole brain imaging using aligned-blipped-controlled-aliasing multiband EPI. *Proceedings of the 20th Annual Meeting of ISMRM*, page 2306, 2012.
- [112] Yan. DPARSF: a MATLAB toolbox for “pipeline” data analysis of resting-state fMRI. *Frontiers in System Neuroscience*, 4(May):1–7, 2010. ISSN 16625137. doi: 10.3389/fnsys.2010.00013. URL <http://journal.frontiersin.org/article/10.3389/fnsys.2010.00013/abstract>.
- [113] Tao Zhang, Peng Xu, Lanjin Guo, Rui Chen, Rui Zhang, Hui He, Qiankun Xie, Tiejun Liu, Cheng Luo, and Dezhong Yao. Multivariate empirical mode decomposition based sub-frequency bands analysis of the default mode network: a resting-state fMRI data study. *Applied Informatics*, 2(1), 2015. ISSN 2196-0089. doi: 10.1186/s40535-014-0005-z. URL <http://www.applied-informatics-j.com/content/2/1/2>.

## DECLARATION

---

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe des Literaturzitats gekennzeichnet. Weitere Personen waren an der inhaltlich-materiellen Herstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe eines Promotionsberaters oder anderer Personen in Anspruch genommen. Niemand hat von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

*Regensburg, 18. August 2016*

---

Markus Goldhacker, 18. August  
2016