VirtualBrainCloud

Personalized Recommendations for Neurodegenerative Disease

www.VirtualBrainCloud-2020.eu

Public deliverable report

D3.9: Evaluation of cross-model atlas-based compression for machine learning

Date	November 2020
Authors	Justin W.M. Domhof, Kyesam Jung, Oleksandr V. Popovych,& Simon B. Eickhoff Institute of Neuroscience and Medicine (Brain and Behaviour, INM-7) from Forschungszentrum Forschungzentrum Jülich
Dissemination level	public
Website	www.VirtualBrainCloud-2020.eu

This project has received funding from the European Union´s Horizon 2020 research and innovation programme under grant agreement No 826421

Table of content

Background

The current deliverable is developed in the framework of the VirtualBrainCloud project aimed at the leverage of the potential of big data and high-performance computing (HPC) for personalized prevention and treatment of neurodegenerative diseases (NDD). The suggested multi-disciplinary approach to this problem essentially involves the methods of enhanced data analytics and computational neuroscience by culminating them into a computational model that can contribute to the personalized diagnostics and treatments of NDD. On this way, an effective integration of an a priori knowledge on the brain organization and function, for example, as given by a brain atlas, into the data analysis and computational models is a crucial step. Compressing the big neuroimaging data into a network of only a few hundreds of brain regions parcellated according to a given brain atlas can strongly facilitate the analysis or, in the case of the whole-brain connectome of the whole-brain models investigated in this project, even make it feasible at all. Little is however known about the impact of different brain parcellations on the structural and functional empirical and simulated brain connectomes and their interrelations, which is evaluated in detail in this deliverable.

1. Introduction

Investigation of the brain dynamics and connectivity is inevitably linked with collection, processing and analysis of large empirical data, e.g., neuroimaging data of the magnetic resonance imaging (MRI) and a variety of its modalities. The complexity and size of the acquired brain data in many cases require a dimensionality reduction. This typically is accomplished by referencing to a brain atlas providing the respective brain parcellation (Eickhoff et al., 2018; Thirion et al., 2014), wherein several hundreds of thousands of voxels from high-resolution neuroimaging data are grouped into a few hundreds of brain regions. This approach also benefits the comparability of the results across subjects and the signal-to-noise ratio. Parcellating the brain into large regions is especially relevant for the modeling of the brain activity by whole-brain dynamical models (Honey et al., 2009), where voxel-based simulations become computationally intractable.

Currently, a great variety of possible techniques for brain parcellation and brain atlases exist (Eickhoff et al., 2018; Thirion et al., 2014), which makes the choice of a particular parcellation very difficult. There is no general consensus as to which parcellation would be most appropriate for a given analysis, and only recent investigations have reported initial empirical evidence for the effects of atlas selection on subject specificity, structure-function relationship and data-driven prediction of behavioral traits (Messe, 2019; Pervaiz et al., 2020; Zimmermann et al., 2019). The influence of parcellation granularity and local connectivity on slow and fast dynamics of coupled neuronal mass models was investigated in paper (Proix et al., 2016) for random splitting of the Desikan-Killiany atlas (Desikan et al., 2006) into smaller subregions. Still, the impact of one or another brain atlas on the results of the whole-brain modeling and properties of the empirical data used for the model derivation and validation remains practically unexplored.

This deliverable focuses on the evaluation of the cross-modal atlas-based compression of the neuroimaging data and its influence on the graph-theoretical and statistical properties of the compressed data and the results of the model validation used for the simulation of the resting-state brain dynamics. We considered in total 23 different cortical parcellations based on 10 brain atlases

provided by the literature and applied them for compression of the structural and functional brain connectomes. The latter were utilized to construct and validate the network models used to model brain activity. The results of the model fitting to empirical data were compared with each other across parcellations and individual subjects as well as with several data variables calculated from the empirical data. In such a way we evaluate the impact of the considered parcellations on the characteristics of the compressed empirical data, modeling results and their interdependencies when the type and spatial resolution of the brain atlases vary.

In the presentation below we first address the impact of the brain parcellations on the parameters of the processing of the diffusion-weighted MRI (dwMRI) data and extraction of the structural connectivity. We vary the number of streamlines of the whole-brain tractography (WTB) affecting its resolution. Here we considered the Schaefer atlas (Schaefer et al., 2018) and the Harvard-Oxford atlas (Desikan et al., 2006) representing two paradigmatically distinct parcellation approaches reflecting functional and anatomical brain properties, respectively. The consideration is then extended to more parcellations of the same atlases including different granularities for the Schaefer atlas and the maximal probability thresholds for the Harvard-Oxford atlas affecting the size of brain regions. We evaluate how the properties of the empirical data used for the model derivation and validation may influence the modeling results and show that the data variables may be classified into a few correlative types depending on their contribution to the model fitting at the level of individual subjects and brain parcellations. The results are finally generalized for more atlases and dynamical models as well as for the graph-theoretical metrics of the structural and functional connectomes confirming the great variability of the modeling results across parcellations and their relationship to empirical data.

2. Partners involved

This deliverable was prepared by the Institute of Neuroscience and Medicine (Brain and Behaviour, INM-7) from Forschungszentrum Jülich (FZJ). The computational resources were granted through JARA on the supercomputer JURECA at Forschungszentrum Jülich (Jülich Supercomputing Centre, 2018). The computational scripts used for the model simulations on CPU and GPU partitions of the supercomputer JURECA were optimized in collaboration with the Jülich Supercomputing Centre (JSC) of FZJ.

3. Description of work performed

As mentioned above, we systematically investigated the influence of brain parcellations on the structural and functional brain connectomes as well as on the goodness-of-fit of the models fit as measured by the Pearson correlation between the simulated functional connectivity (FC) generated by the large-scale whole-brain models and their empirical counterparts. First, we extracted the empirical structural connectivity (SC) and FC corresponding to a particular parcellation scheme from the dwMRI and the resting-state functional MRI (fMRI) data, respectively (Figure 1, green). The result of the SC reconstruction comprised two matrices: one with the number of streamlines and one with the average length of the streamlines between each pair of brain regions, which are referred to as the actual structural connectivity (SC), i.e., streamline count, and the path length (PL), respectively. The FC matrix contained Pearson cross-correlation coefficients between the regional BOLD response time series extracted from the fMRI data.

Subsequently, the SC matrices were fed to the model as prior knowledge, while the empirical FC matrix was compared with the simulated FC matrix produced by the model simulations (Figure 1, blue). Two models (a phase oscillator and a neural mass model) were used for the acquisition of simulation results, and we simulated the models for a broad range of global parameter settings to maximize the fit between empirical and simulated data. We also extracted some graph-theoretical and other statistical metrics from the empirical SC and FC matrices referred to as data variables (Figure 1, red).
Finally, we evaluated the correlations between the model simulation results and the extracted data variables using univariate and multivariate regression approaches combined with principal component analysis (PCA) (Figure 1, yellow). Below, we explain the utilized methodology in more detail.

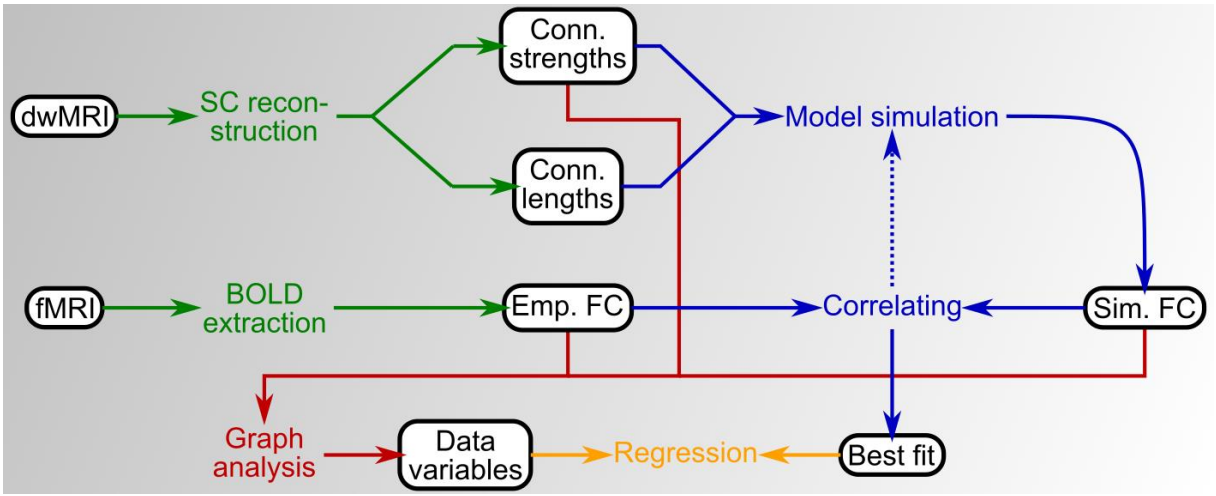

Figure 1: Summary of the methods used for this deliverable. Data extraction (green) comprises the construction of the empirical structural (SC) and functional connectivity (FC) from dwMRI and fMRI data, respectively. Both connectomes serve as input for the modelling stage (blue), where the model parameters are optimized to maximize the correlation between simulated and empirical data (dotted arrow). Data variables as given by some graph-theoretical and other statistical metrics were extracted from the empirical and simulated connectomes (red) and regressed with the model fitting results (orange).

3.1. Extraction of empirical connectomes

For the investigations presented in this deliverable, we used the neuroimaging data of 351 healthy, unrelated subjects from the Human Connectome Project (HCP) S1200 release dataset (http://www.humanconnectomeproject.org) (Van Essen et al., 2013, 2012).

3.1.1. SC extraction from dwMRI

For the extraction of the SC matrices from dwMRI data, we used a workflow developed in the framework of Deliverable 3.1, which consisted of four stages: (1) preprocessing of raw dwMRI images, (2) calculation of the whole-brain tractography (WBT), (3) transformation of the brain parcellation images into native space and (4) reconstruction of the SC. The workflow included functions from the ANTs (Tustison et al., 2010), FreeSurfer (Dale et al., 1999), FSL (Jenkinson et al., 2012) and MRtrix3 (Tournier et al., 2019) software packages. Computations were performed on the JURECA high-performance computing cluster (Jülich Supercomputing Centre, 2018).

(1) In the preprocessing stage, we used FreeSurfer functions to perform the following operations on the T1-weighted images: bias field correction, tissue segmentation, cortical (surface) reconstruction, volume-surface conversion and surface deformation. We also used FreeSurfer functions to correct the

dwMRI images with regard to head motions and eddy current distortions, while MRtrix3 functions were employed to denoise them and perform bias field correction. The dwMRI images were then co-registered to the T1-weighted images using the linear and nonlinear transformation functions included in FSL, and tissue segmentation was performed. (2) Subsequently, WBT was calculated using exclusively MRtrix3 functions. A multi-shell-multi-tissue constrained algorithm (Jeurissen et al., 2014) estimated the response functions for spherical deconvolution, which were subsequently used to determine the fiber-oriented distributions (FOD) from the dwMRI data. The WBT was then completed through a second-order integration over the FOD using a probabilistic algorithm (Tournier et al., 2010), where we extracted from 10M (used if not specified otherwise) to 10K WBT streamlines. (3) Next, the images of the brain parcellations used in this study (see below) were linearly and nonlinearly transformed from the standard MNI152 space (in which they were all originally sampled) to the native diffusion space using FSL. (4) Finally, we reconstructed the SC and PL matrices containing the streamline counts and path lengths between any two parcels included in a particular brain parcellation, respectively, by the MRtrix3 function tck2connectome.

3.1.2. FC extraction from fMRI

To construct the empirical FC matrix, BOLD signals were first extracted as the mean signals of the brain regions included in a given brain atlas from the resting-state fMRI data preprocessed using ICA-FIX as provided by the HCP repository (Griffanti et al., 2014). The FC matrix was calculated by taking the Pearson cross-correlation across the linearly detrended and z-scored BOLD time series for each pair of parcels. 4 resting-state fMRI sessions of 1200 volumes sampled with a repetition time of TR=720 ms were available for every subject (2 phase encoding directions scanned on 2 days). We also used the BOLD signals concatenated over all for sessions. We thus calculated 5 different empirical FCs per subject that were used for the validation of our models.

3.2. Graph-theoretical analysis of empirical connectomes

The SC and FC matrices were subjected to graph-theoretical and statistical analyses in order to extract data variables portraying the properties of the networks they represent. In these networks the brain regions were the nodes, and the individual elements of SC and FC matrices were undirected weighted edges between them (self-connections were removed). For the analyses presented below, different network properties were calculated including binarized and weighted node degree distribution, clustering coefficient, modularity, closeness and betweenness centrality, local and global efficiency (see Rubinov and Sporns (2010) for definitions). These graph metrics reflect a variety of network properties such as the extent of node centrality and network integration and segregation with different physical meaning when applied to SC, PL or FC matrices (Rubinov and Sporns, 2010).

To evaluate the impact of brain parcellations on the statistical properties of the resting-state brain dynamics and connectivity, other data variables were also calculated such as the standard deviation of time fluctuations of BOLD signals averaged over all parcels *aver*[*std*(BOLD)], an average extent of total functional connectivity (synchronization) in the brain *aver*[*aver*(FC)] or it's inter-region variability *std*[*aver*(FC)] for both functional (FC) and structural (SC, PL) connectomes. Also the structure-function correspondence *corr*(FC, SC) between empirical functional and structural connectivities was evaluated by Pearson correlation.

3.3. Model simulations

From the calculated empirical SC and PL matrices a system of coupled oscillators was derived to model the collective dynamics of the mean-field activities of the individual brain regions and to simulate the FC calculated from the simulated BOLD signals. The model was validated against the empirical data, where the similarity between the simulated and empirical FC matrices was quantified by the Pearson correlation. By exploring the model parameter space, the maximal similarity between the FC matrices could be found, which is henceforth also referred to as the *goodness-of-fit* of the model.

The local dynamics of the brain regions was modelled from different perspectives by two different models. The first model was the Kuramoto system of coupled phase oscillators (Kuramoto, 1984), and the other was an ensemble of Wilson-Cowan type neural mass models (Wilson and Cowan, 1972).

3.3.1. Kuramoto model

In the Kuramoto model (Kuramoto, 1984), the mean-field BOLD activity of brain region i is assumed to oscillate with a region-specific frequency f_i, and the dynamics of its phase are governed by the differential equation

$$\dot{\varphi}_i(t) = 2\pi f_i + \sum_{j=1}^{N} C_{ij} \sin\left(\varphi_j(t-\tau_{ij}) - \varphi_i(t)\right) + \sigma_p \nu_i(t). \quad (1)$$

Here $\nu_i(t)$ is independent white noise with zero mean and unit variance, and σ_p rad is the noise intensity. Furthermore, C_{ij} and τ_{ij} represent the individual coupling strength and delay values between brain regions, respectively, and are derived from the SC and PL matrices via

$$C_{ij} = \begin{cases} 0 & \text{if} \quad i = j \\ G \cdot \frac{\mathrm{SC}_{ij}}{N\langle \mathbf{SC} \rangle} & \text{otherwise} \end{cases} \quad \text{and} \quad \tau_{ij} = \begin{cases} 0 & \text{if} \quad i = j \\ \tau \cdot \frac{\mathrm{PL}_{ij}}{\langle \mathbf{PL} \rangle} & \text{otherwise} \end{cases}. \quad (2)$$

Here, the operator $\langle \cdot \rangle$ returns the mean over all elements of the matrix, and G and τ are scaling factors referred to as the global coupling and global delay, respectively. The latter two parameters were varied on a dense grid in the model parameter space in order to maximize the similarity between the simulated and empirical FCs as mentioned above. The frequencies f_i were derived from the power spectra of the empirical BOLD time series.

3.3.2. Neural mass model

The neural mass model used in this study was a Wilson-Cowan model (Wilson and Cowan, 1972) adapted from the paper (Deco et al., 2009). It models the interaction between the excitatory and inhibitory neuron ensembles of the i-th brain region, where their mean firing rates $E_i(t)$ and $I_i(t)$, i.e. the proportion of cells firing within a unit of time, respectively, are modelled via the coupled differential equations

$$\mu_E \dot{E}_i(t) = -E_i(t) + \kappa \mathcal{S}\left(\sum_{j=1}^{N} C_{ij} E_j(t-\tau_{ij}) - c_{EI} I_i(t) + I_b\right) + \sigma_n \nu_i(t) \quad (3)$$

and

$$\mu_I \dot{I}_i(t) = -I_i(t) + \kappa \mathcal{S}\left(c_{IE} E_i(t)\right) + \sigma_n \nu_i(t). \quad (4)$$

In these equations, μ_E and μ_I are the decay time constants of the excitatory and inhibitory activity, respectively. Both populations received the same zero-mean, independent Gaussian white noise of intensity σ_n. Parameters c_{EI} and c_{IE} regulate the inhibition of the excitatory cells by the inhibitory pool and the excitation of the inhibitory cells by the excitatory pool, respectively. $\mathcal{S}(x)$ is a sigmoid function defined by

$$\mathcal{S}(x) = \frac{1}{1+\exp(-\lambda(x-\gamma))} - \frac{1}{1+\exp(\lambda\gamma)}, \quad (5)$$

where λ and γ determine its width and the position of its inflexion point, respectively. Additionally, I_b is a constant external input to the excitatory neurons, and $\kappa = (1+\exp(\lambda\gamma))/\exp(\lambda\gamma)$ scales $\mathcal{S}(x)$ such that $\kappa\mathcal{S}(x) = 1$ as $x \to \infty$. Finally, C_{ij} and τ_{ij} have the same interpretations and similar associated expressions as with the Kuramoto model (Eq. 2):

$$C_{ij} = \begin{cases} c_{EE} & \text{if} \quad i = j \\ G \cdot \frac{\mathrm{SC}_{ij}}{N\langle \mathbf{SC} \rangle} & \text{otherwise} \end{cases} \quad \text{and} \quad \tau_{ij} = \begin{cases} 0 & \text{if} \quad i = j \\ \tau \cdot \frac{\mathrm{PL}_{ij}}{\langle \mathbf{PL} \rangle} & \text{otherwise} \end{cases}, \quad (6)$$

where c_{EE} is a parameter scaling the self-excitation of the excitatory pool.

We set the model parameters to the values listed in Table 1. The considered parameter configurations resulted in a low activity state of disconnected nodes ($G = 0$) and generation of limit-cycle oscillations with an alpha-band frequency when the individual regions were coupled ($G > 0$). The modelled alpha oscillations have been shown to be dominant in EEG of human resting-state brain activity (Spitoni et al., 2013) and to interact with BOLD responses (Mayhew et al., 2013). The simulated electrical activity was converted to the BOLD responses by the Balloon-Windkessel model (Friston et al., 2003), which was subsequently used for the calculation of simulated FC and model validation.

Table 1: Parameter settings of the neural mass model

Parameter	Value	Parameter	Value	Parameter	Value
μE	20 ms	γ	0.300	c_{EE}	1.000
μI	20 ms	I_b	0.100	c_{EI}	1.500
λ	20.000	σ_n	0.002	c_{IE}	0.600

3.3.3. Model implementation, simulation and parameter variation

The system of coupled phase oscillators (1) was used to simulate the BOLD and FC for 351 individual HCP subjects and for several conditions of SC extraction, where the number of WBT streamlines attained the values of 10M, 2M, 500K, 100K, 50K and 10K for two parcellations as given by the Schaefer atlas with 100-area parcellation and the Harvard-Oxford atlas with 96 cortical regions leading to 6 x 2 = 12 simulation conditions.

In another study the phase oscillators (1) were simulated for 272 individual HCP subjects for 10M WTB streamlines and for more parcellations. The latter were based on the Schaefer atlas of 100, 200 and 400 cortical regions and the Harvard-Oxford atlas with 96 non-overlapping cortical parcels with thresholds at 0%, 25%, 35%, and 45% of the maximal probability leading to 7 simulation conditions.

The results were generalized using both models (1) and (3-4), which were simulated for 200 individual HCP subjects and for 19 parcellations based on 10 atlases as listed in Table 2 leading to 2 x 19 = 38 simulation conditions. All model simulations were performed for each individual subject with respective individual empirical SC and FC on a dense grid of 48×64=3072 parameter points of the global delay and coupling, respectively. The optimal parameter values were selected, where the best correspondence between the simulated and empirical data was achieved. For the mentioned studies, subjects and conditions more than 40 million of model runs were performed.

Table 2: Overview of the used brain parcellation schemes with the index for reference in this study, the number of parcels (#parcels) and associated publications. Colors are used to highlight the parcellations in Figures 7 and 9.

Index	Name	#parcels	Refs.
1	MIST	31	(Urchs et al., 2019)
2		56	
3		103	
4		167	
5	Craddock	38	(Craddock et al., 2012)
6		56	
7		108	
8		160	
9	Shen 2013	79	(Shen et al., 2013)
10		156	
11	Schaefer (17 networks)	100	(Schaefer et al., 2018)
12		200	
13	Harvard-Oxford	48	(Desikan et al., 2006; Frazier et al., 2005; Goldstein et al., 2007; Makris et al., 2006)
14		96	
15	Desikan-Killiany	70	(Desikan et al., 2006)
16	von Economo-Koskinas	86	(Scholtens et al., 2018; von Economo and Koskinas, 1925)
17	AAL (version 2)	92	(Rolls et al., 2015; Tzourio-Mazoyer et al., 2002)
18	Destrieux	150	(Destrieux et al., 2010)
19	Brainnetome	210	(Fan et al., 2016)

4. Results

As mentioned in the Introduction, we present 3 studies that considered different number of parcellations and details of their investigation. In Sec. 4.1 only two parcellations were investigated based on the functional Schaefer atlas with 100 cortical parcels and anatomical Harvard-Oxford atlas with 25% probability threshold with 96 cortical parcels. They were investigated for 351 HCP subjects and for 6 additional simulation conditions reflecting the variation of the WBT resolution as given by the total number of streamlines. In Sec. 4.2 more parcellations were considered for the same two atlases of three granularities of 100, 200 and 400 parcels for the Schaefer atlas and four values of probability threshold for the Harvard-Oxford atlas leading to different 7 parcellations investigated for 272 HCP subjects. Finally, in Sec. 4.3 the results were generalized for 19 different parcellations of 10 brain atlases and for two whole-brain models based on the phase oscillators and neuronal mass model simulated for 200 individual HCP subjects.

4.1. Impact of parcellation on structural connectivity

System of coupled phase oscillators (Eq. 1) was simulated for varying global delay and global coupling, and the similarity between simulated FC (sFC) and empirical FC (eFC) was calculated for every scanned

parameter value (see Sec. 3.3). These calculations resulted in the model parameter planes shown in Figure 2A,B for the Schaefer and Harvard-Oxford atlases and two densities of WBT. The maximal goodness-of-fit between sFC and eFC was observed for small delays for both atlases, where the influence of the number of WBT streamlines is notable for the Schaefer atlas (Figure 2A), but not for the Harvard-Oxford atlas (Figure 2B). Moreover, the latter atlas could lead to a stronger fit between the sFC and eFC (Figure 2A,B).

The distributions of the maximal goodness-of-fit between sFC and eFC for all considered 351 HCP subjects are illustrated in Figure 2C,D for the 12 simulated conditions (6 WBT x 2 atlases). We found that for the Schaefer atlas the models with 10M and 2M WBTs performed better than the other WBTs, and the performance of the model decreased when the number of streamlines decreased (Figure 2C). On the other hand, the functional model fitting for the Harvard-Oxford atlas revealed the optimal condition at 50K or 100K WBT (Figure 2D).

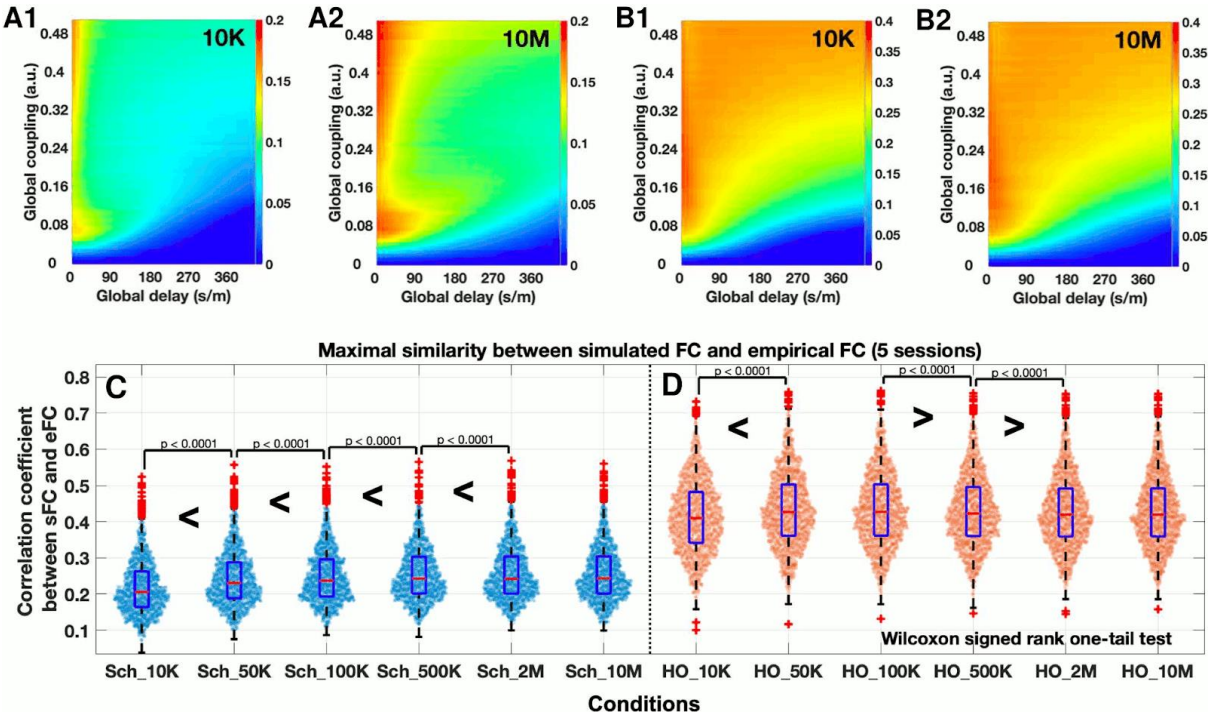

Figure 2: Results of model fitting for two parcellations and varying WBT resolution. (A-B) Parameter landscape of the correlation (indicated by color) between the empirical (eFC) and simulated (sFC) functional connectivity averaged over all 351 subjects and corresponding to (A) the Schaefer atlas with 100 parcels and 17 networks and (B) the Harvard-Oxford atlas with 96 parcels and 25% probability threshold for 10K and 10M of WBT streamlines as indicated in the plots. (C, D) Distributions of the maximal similarities between eFC and sFC for individual subjects, i.e., goodness-of-fit of the model for 6 different simulation conditions of 10K, 50K, 100K, 500K, 2M and 10M WBT streamlines and (C) the Schaefer atlas (Sch.) and (D) the Harvard-Oxford atlas (HO) as indicated on the horizontal axis. The results of the pairwise comparisons between the conditions (Wilcoxon signed rank one-tail test) are also indicated with the corresponding p-values in the cases of statistically significant differences (Bonferroni corrected $p < 0.05$).

To evaluate the impact of the brain parcellations on the architecture of the structural networks and its relation to modeling results, we calculated several main graph-theoretical network properties of the empirical SC and correlate them with the goodness-of-fit values of the model across 12 WTB conditions for every individual subject as illustrated in Figure 3. The two considered parcellations apparently demonstrate different distributions of the correlation coefficients for individual subjects. In particular, the relatively strong correspondence between network metrics and the modelling results for the

Schaefer atlas (Figure 3A) is less pronounced for the Harvard-Oxford atlas with stronger inter-individual variability of the relationship between goodness-of-fit and considered graph-theoretical metrics (Figure 3B).

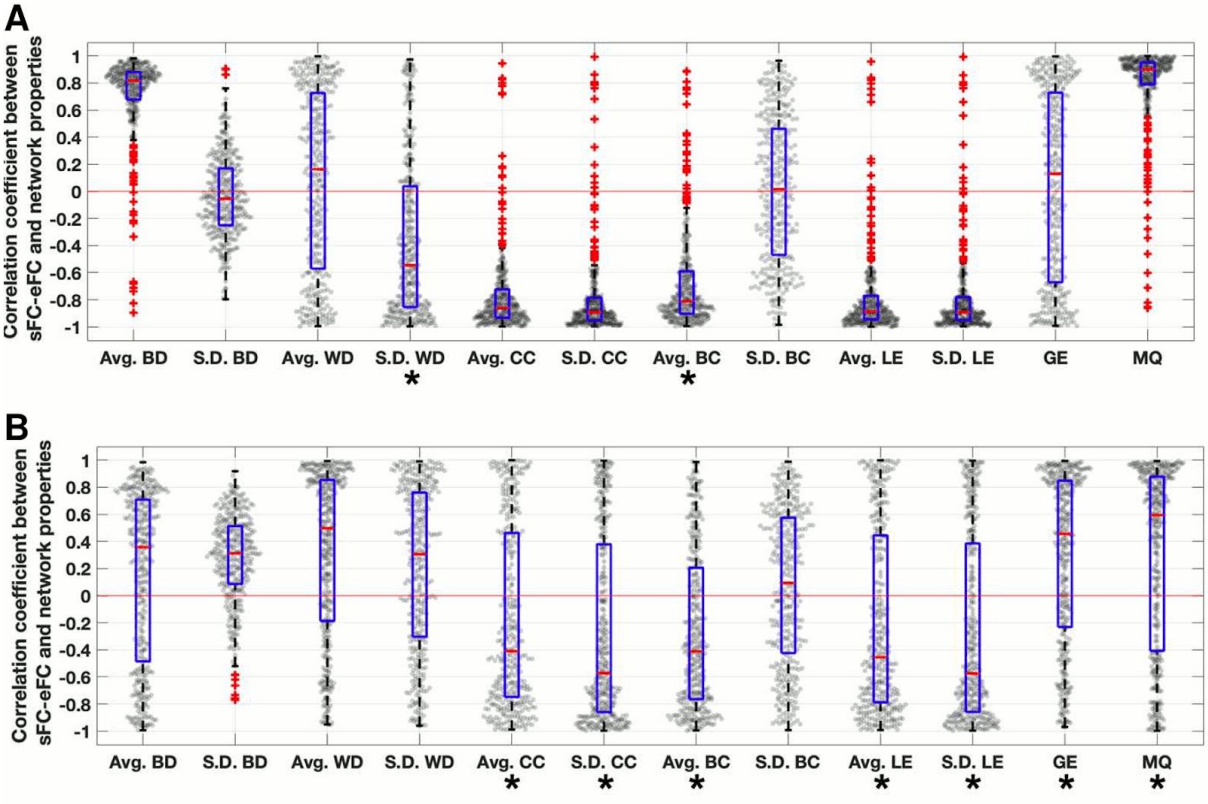

Figure 3: Correlation between the network properties indicated on the horizontal axes and the goodness-of-fit (maximal similarity between sFC and eFC) across 6 WBT conditions for all individual subjects for (A) the Schaefer atlas and (B) the Harvard-Oxford atlas. The gray dots represent the values for individual subjects, and the box plots illustrate the medians (red lines), the interquartile ranges (blue boxes) and the outliers (red pluses). The asterisks below the x-axes labels indicate statistically significant differences in the maximal goodness-of-fit values between the two subgroups of subjects with positive and negative correlations ($p < 0.05$ of two-sample one-tail t-test). Abbreviations: average (Avg.), standard deviation (S.D.), binarized node degree (BD), weighted node degree (WD), clustering coefficient (CC), betweenness centrality (BC), local efficiency (LE), global efficiency (GE), and modularity Q (MQ).

The splitting of the subjects between the two groups of positive and negative correlation in Figure 3 was used to stratify the individual subject into subgroups that can demonstrate different quality of the model validation and dependence on the WTB resolution. Two more stratification criteria were based on (1) a clustered distribution of the model parameters of the structure-function model fitting of the correspondence between sFC and eSC and (2) the dynamics of the maximal goodness-of-fit when the WTB resolution varies splitting the subjects with the best model performance for high or low WTB resolution. The stratification results are illustrated in Figure 4. As follows from the presented results, the optimal number of the WTB streamlines for the simulation with the Schaefer atlas should be considered large, for example, 2M or 10M (Figure 4A2). The model evaluation with the Harvard-Oxford atlas showed different optimal conditions, where the optimal streamline number may depend on the stratification subgroup to which the subject belongs (Figure 4B). For example, the optimal WTB resolution could range from 10M to 100K for the subjects from subgroup 1 in Figure 4B2 (solid red curve). On the other hand, for more than 20% of all subjects (group 3 in Figure 4B2, dashed blue curve), the optimal conditions are at ~50K WBT streamlines, and more streamlines may lead to the degradation

of the quality of the model validation. For the other 18% of the subjects (n = 66, group 2 in Figure 4B2, solid blue curve) a sparse WBT can also be a reasonable option.

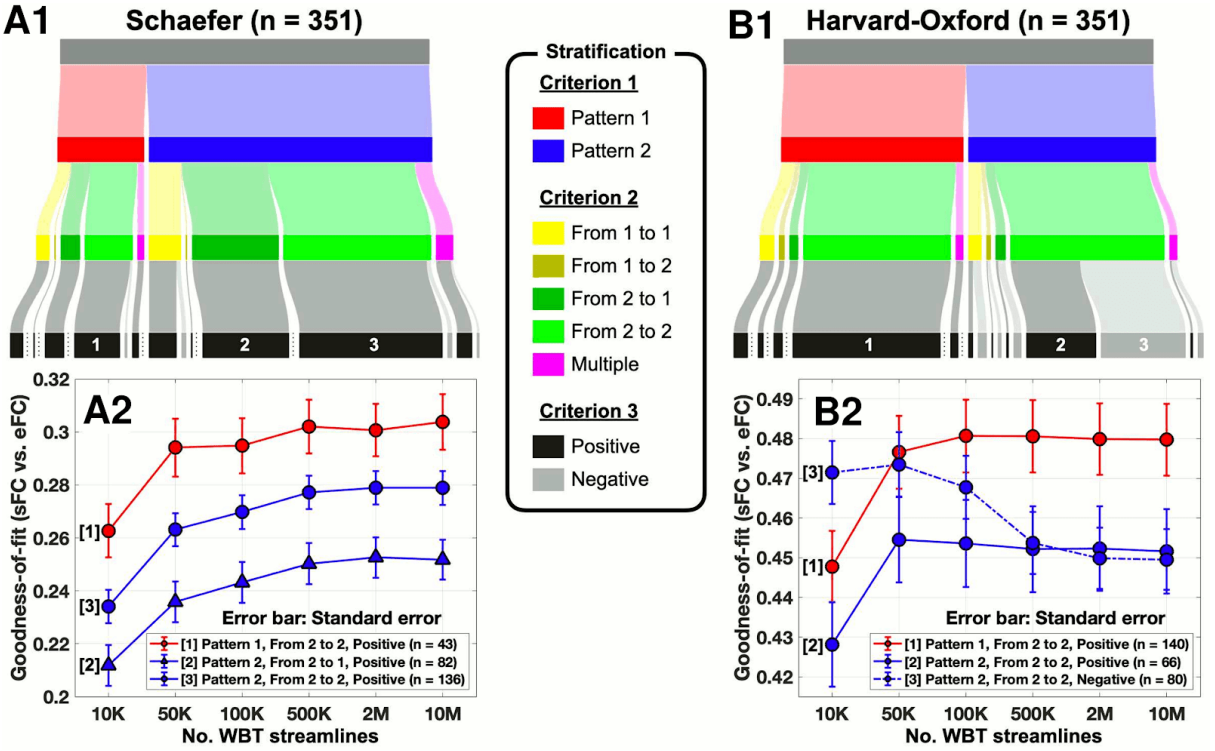

Figure 4: Stratification analysis for (A) the Schaefer atlas and (B) the Harvard-Oxford atlas. The alluvial plots (1) show how the subjects were stratified via several subgroups based on a few stratification criteria (see text for detail), and the bottom plots (2) show the goodness of fit as a function of the streamline condition for large stratified groups of subjects.

4.2. Data variables and their relation to modelling results

In the next study we investigated more parcellations as given by the Schaefer atlas of 100, 200 and 400 cortical regions and the Harvard-Oxford atlas with thresholds at 0%, 25%, 35%, and 45%. The effects of these parcellations on the statistical properties of the empirical data are illustrated in Figure 5. In particular, higher granularity for the Schaefer atlas and larger threshold for the Harvard-Oxford atlas led to smaller brain regions of the corresponding parcellations (Figure 5A). Increasing standard deviation of BOLD time fluctuations within regions and their variability across regions may indicate an enhanced intra-region homogeneity and inter-region heterogeneity when the region size decreases (Figure 5B), which also agrees with decreasing average inter-region synchronization in the brain as reflected by *aver*[*aver*(eFC)] in Figure 5C1. Other properties of eFC, eSC, and the structure-function relationships are also illustrated, where, for example, *corr*(eFC, eSC) decays with the region size in agreement with results of paper (Messe, 2019).

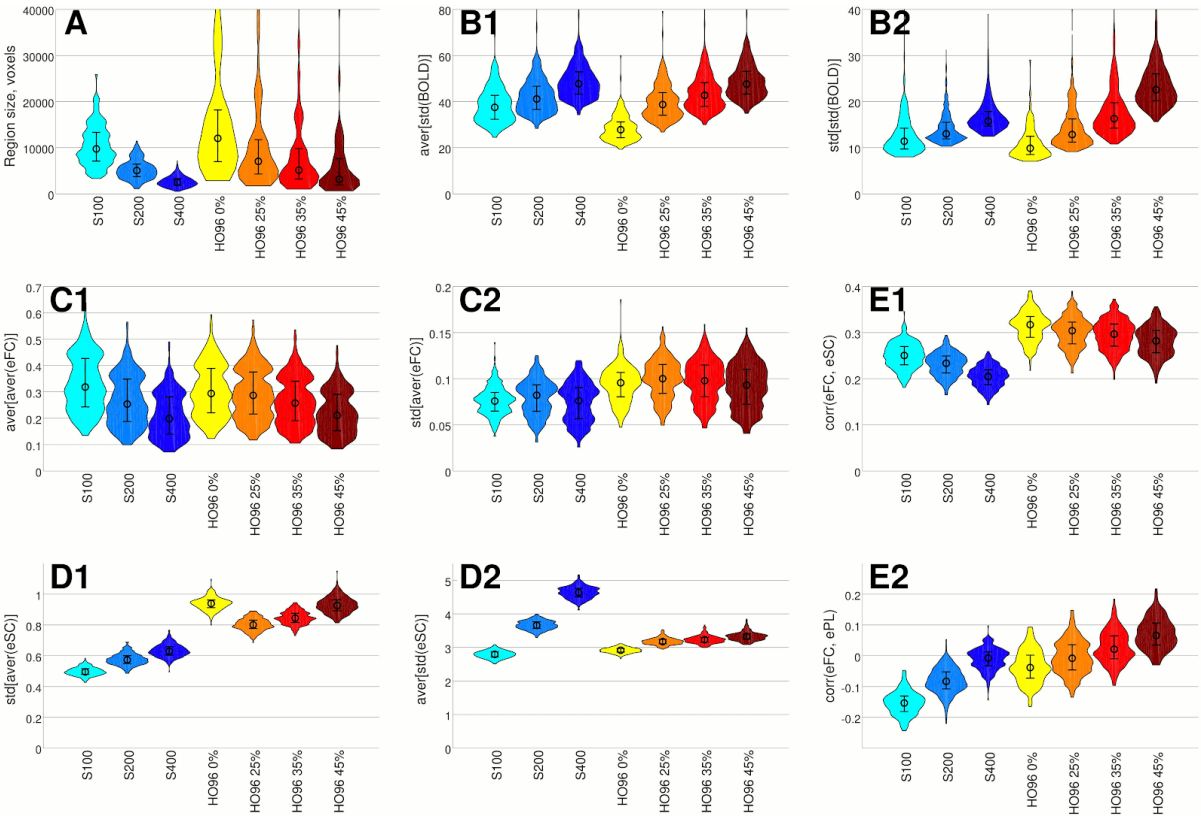

Figure 5: Distributions of (A) the region size (1 mm3 isotropic voxels) in each considered parcellations, (B-D) data variables reflecting some statistical properties of brain activity and connectivity and (E) their interrelation for a few parcellations indicated on the horizontal axes and based on the Schaefer atlas with 100, 200 and 400 cortical parcels (denoted as S100, S200 and S400, respectively), and Harvard-Oxford atlas with thresholds at 0%, 25%, 35%, and 45% of the maximal probability (denoted as HO96 0%, HO96 25%, HO96 35%, and HO96 45%, respectively). The data variables are calculated for individual subjects (the distributions for all subjects are shown in the plots) and indicated on the vertical axes, where "aver", "std" and "corr" stand for averaging, standard deviation and correlation, respectively.

We also investigate how the considered parcellations influence the results of the model validation and their relationship to the considered data variables. The distributions of the maximal similarity *Fit*(sFC, eFC) of the fitting of sFC to eFC are illustrated in Figure 6A for the considered brain atlases. The impact of the atlases is apparent for the much better model fitting for the Harvard-Oxford atlas (HHO96 xx%, yellow - dark red violins) than for the Schaefer atlas (S100-S400, blue violins), see also Figure 2. The best fitting for the Schaefer atlas was achieved for the parcellation S200 of intermediate granularity, although the differences are small (Figure 6A). The variation is much more pronounced among HO96 cases, where *Fit*(sFC, eFC) monotonically decays when the threshold for HO96 atlas increases (Figure 6A).

The correlations between the goodness-of-fit *Fit*(sFC, eFC) for every considered parcellation and the mentioned data variables across all considered subjects and scanning sessions are depicted in Figure 6B. Many of the data variables only weakly correlate with *Fit*(sFC, eFC) across subjects, for example, the empirical structure-function relationship *corr*(eFC, eSC) for both atlases, or *corr*(eFC, ePL) for the Schaefer atlas only. Some other data variables exhibit from moderate to relatively strong (anti-)correlation with *Fit*(sFC, eFC), for example, the spread of the natural frequencies *std*(f_i) and some properties of the BOLD signals and eFC (Figure 6B).

The joint correlation for the data merged over all parcellations revealed additional data variables exhibiting strong correlation with the goodness-of-fit, for example the data variables calculated from empirical SC, and also the empirical structure-function relationship *corr*(eFC, eSC) stated to play a role (Figure 6C, empty bars). These data variables can be used to account for the variation of the results of the model fitting across different parcellations. Such a variation can also be investigated by the multiple linear regression (MLR) model, where many data variables can participate in explaining the variance of the modeling results. The corresponding regression coefficients for merged data are illustrated in Figure 6D demonstrating which and how the considered data variables contribute to the model fitting results for the joint data. The corresponding scatter plot of the predicted Fit-values versus the simulated ones shows 72% of the explained variance (Figure 6E).

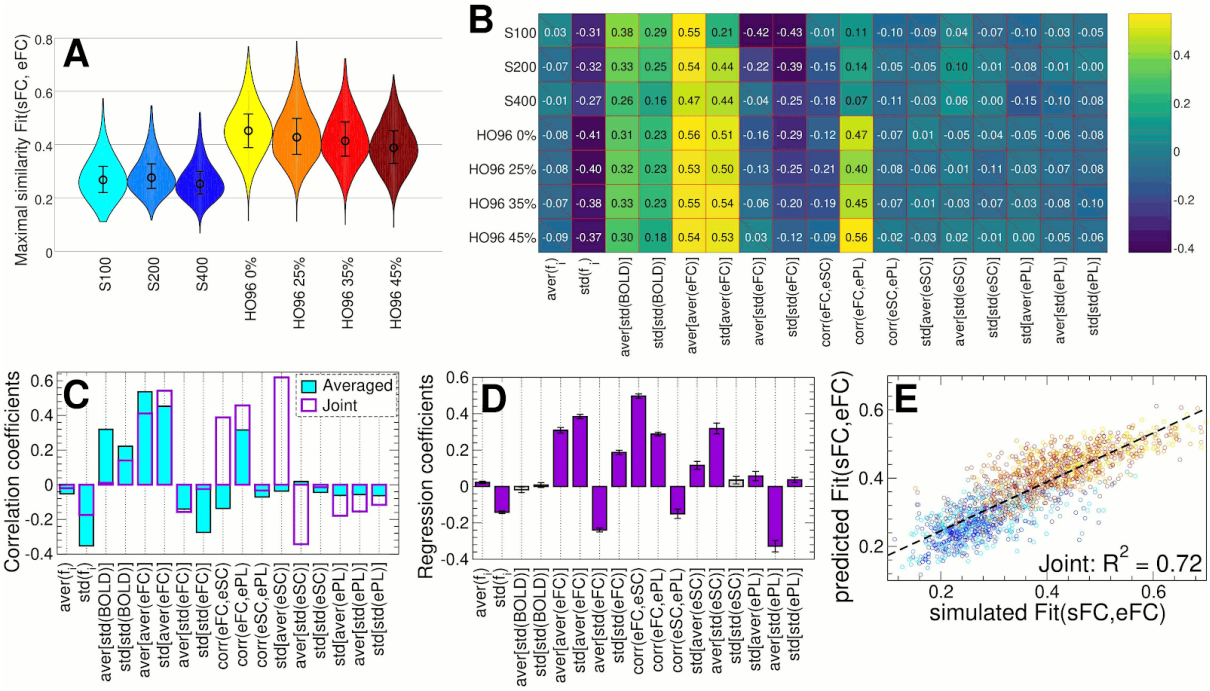

Figure 6: Results of the validation of the phase oscillators (Eq. 1) and their relationships with the data variables. (A) Distributions of the maximum similarity Fit(sFC, eFC) (goodness-of-fit) between eFC and sFC for the considered brain parcellations indicated on the horizontal axis. (B) Pearson correlation across subjects between the goodness-of-fit and several statistical properties (data variables) extracted from the empirical data as indicated on the horizontal axis for different parcellations (vertical axis). The crossed cells indicate the not significant correlations ($p>0.05$). (C) Joint correlation between goodness-of-fit and data variables for the data merged over all considered parcellations is depicted by empty bars. The filled bars represent the individual correlation from plot (B) averaged over all parcellations as indicated in the legend. (D-E) Results of the multiple linear regression (MLR) analysis of the joint goodness-of-fit Fit(sFC, eFC) merged over all considered parcellations with data variables as independent variables. (D) The corresponding regression coefficients with standard deviation, where gray bars indicate statistically non-significant coefficients, and (E) scatter plots with regression lines of the joint Fit(sFC, eFC) predicted by MLR, where the fraction of the explained variance R2 is also indicated.

4.3. Generalization for more models and parcellations

In this section the consideration is extended to more whole-brain models based on the coupled phase oscillators (Eq. 1) and neuronal mass models (Eq. 3, Eq. 4) as well as for more brain parcellations listed in Table 2. The results of the model simulations performed for 200 HCP subjects were compared with a few graph-theoretical metrics of the empirical connectomes reflecting the impact of the parcellations on both empirical and simulated data.

4.3.1. Graph-theoretical properties of empirical connectomes

We found a high variability in the graph-theoretical network properties of the structural connectome with respect to the different brain parcellations (Figure 7A-F). We observed that the modularity derived from the SC matrices (Q^{SC}), the degree distribution parameters α_{dd} and β_{dd} and the modularity calculated across subjects for the empirical FC matrices (Q^{FC}) show an increasing trend when the number of parcels grows (Figure 7C, G-I). The functional networks are however characterized by an enhanced inter-individual variability. An opposite decreasing trend seems to be demonstrated by the structure-functional relationship as given by the Pearson correlation coefficient $\rho_{SC,FC}$ between the empirical SC and FC matrices together with the global efficiency ε of the PL matrix as brain parcellations vary (Figure 7F,J). Such a simple relation was however not observed for the other considered statistics (Figure 7A-B,D).

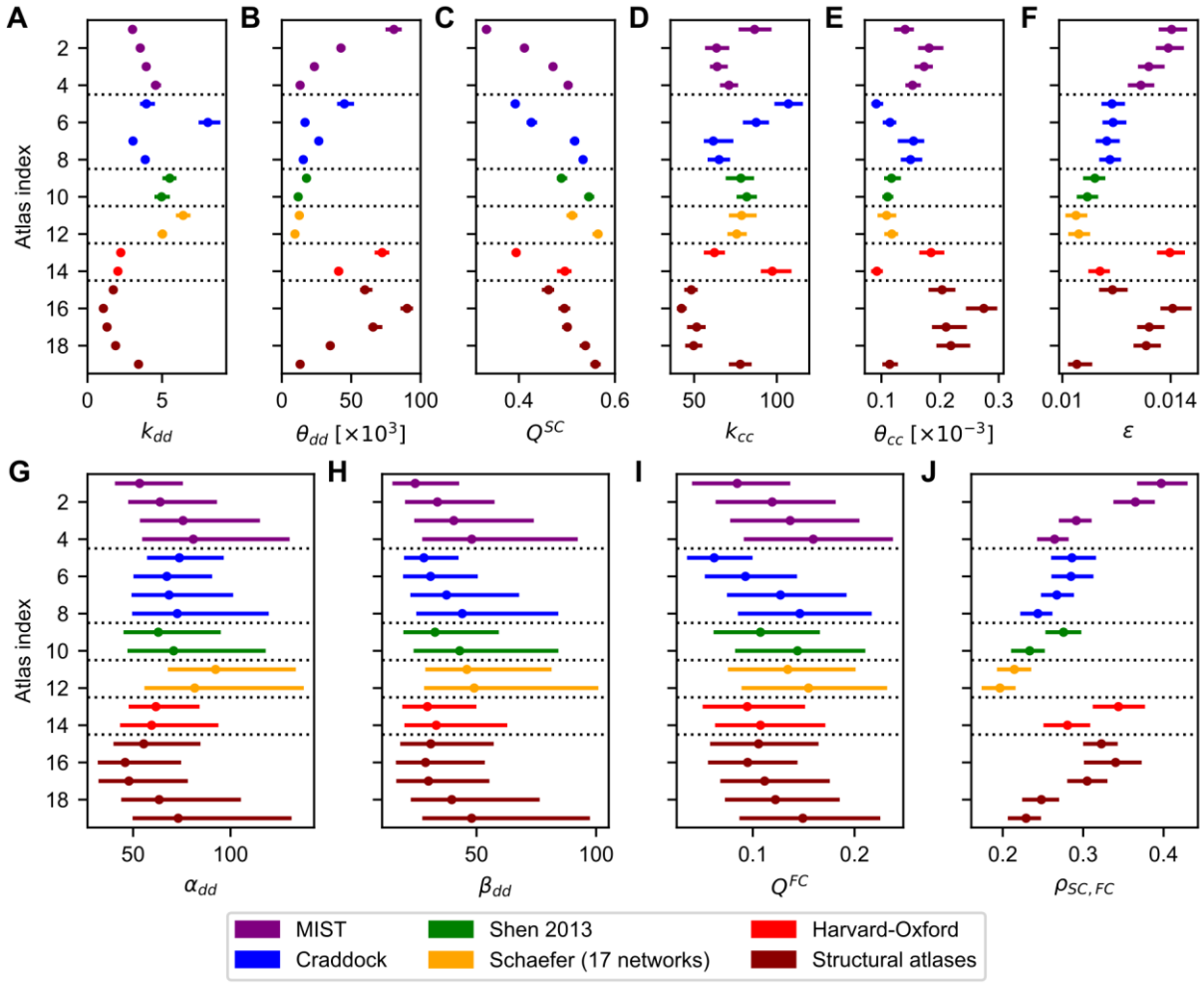

Figure 7: Heterogeneity of graph-theoretical properties of empirical structural and functional networks across parcellations. (A-C) Statistics extracted from the structural connectivity (SC) matrices, which are the shape (k_{dd}, A) and scale (θ_{dd}, B) parameters of the degree distributions and the modularities (Q^{SC}, C). (D-F) Statistics extracted from the path length (PL) matrices, which are the shape (k_{cc}, D) and scale (θ_{cc}, E) parameters of the closeness centrality distributions and the global efficiencies (ε, F). (G-I) Statistics extracted from the empirical functional connectivity (FC) matrices, which are the parameters of their degree distributions (α_{dd} and β_{dd} shown in panel G and H, respectively) and their modularities (Q^{FC}, I). (J) Structure-function relationship as given by Pearson correlation coefficient between

the empirical SC and FC matrices ($\rho_{SC,FC}$). Dots and lines depict the medians and interquartile ranges across subjects, respectively, and the colors and atlas indices on the vertical axes correspond to those in Table 2 and also indicated in the legend.

We also evaluated whether the inter-individual differences in the network statistics exhibited similar patterns between the brain parcellations used for the extraction of the connectomes. For each graph-theoretical measure, the vectors comprising the values for individual subjects were correlated with each other across subjects for any two of the considered parcellations, and the results are illustrated in Figure 8. We found that these correlations are highest when considering the FC matrices in spite of the enhanced inter individual variability: The Pearson correlation coefficient of their modularities, for instance, did not drop below 0.9, and it also was bounded by 0.6 from below for any combination of brain parcellations with respect to the α_{dd} parameter of the degree distributions (Figure 8C). Such correspondences are generally lower for the structural networks (Figure 8A-B), except for the global efficiencies ε of the PL matrices, where the similarity between parcellations can also be relatively high (Figure 8B, lower triangular matrix).

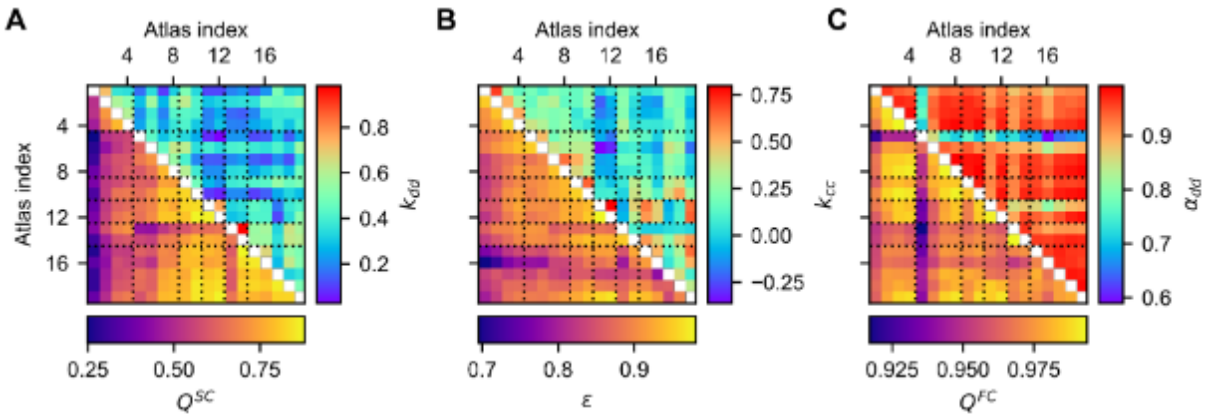

Figure 8: Cross-correlations across individual subjects of the network statistics derived from the empirical structural and functional connectomes between different parcellations. (A) Correlations between parcellations with respect to the shape parameter of the degree distributions (k_{dd}, upper triangular matrix) and the modularities (Q^{SC}, lower triangular matrix) of the structural connectivity matrices across subjects. (B) Same as panel A, but for the shape parameter of the closeness centrality distributions (k_{cc}, upper) and the global efficiencies (ε, lower) of the path length matrices. (C) Same as panel A, but for one of the parameters of the degree distributions (α_{dd}, upper) and the modularities (Q^{FC}, lower) of the functional connectivity matrices. The atlas indices correspond to those in Table 2.

4.3.2. Parcellation-induced variability of model fitting

In this section we present the results of the model simulations for all brain parcellations in Table 2 and the two considered whole-brain models of coupled phase oscillators (Eq. 1 and Eq. 2) and neuronal mass models (Eq. 3 - Eq. 6). For each combination of subject, brain parcellation and model, the optimal values of the global coupling and delay parameters were found by maximizing the Pearson correlation between the empirical and simulated FC matrices, which provides the goodness-of-fit of the model to empirical data (Figure 9A). The quality of the model fitting (goodness-of-fit values) exhibits a relatively high variability across subjects and parcellations. Nevertheless, we found that the patterns of the goodness-of-fit across individual subjects show relatively high correspondence with each other between any two of the considered parcellations and models as illustrated in Figure 9B. Somewhat lower correlations were observed for the Schaefer and also the Harvard-Oxford atlases, both within and across models (Figure 9B, atlas indices 11-14). We did not observe such clear, generally decreased agreement when

considering the correlations of the graph-theoretical statistics across brain parcellations (Figure 8). For the empirical FC matrices, the Craddock atlas with 38 parcels can however be distinguished in this respect (Figure 8C, atlas index 5), and only a slight indication of a lower correlation can be found for the node degree index α_{dd} for the Schaefer atlas with 100 parcels and Harvard-Oxford atlas with 48 parcels (Figure 8C, upper part, atlas indices 11 and 13).

The model fitting results, if restricted to the median values calculated over all subjects (Figure 9A, dots), can also strongly correlate with each other across parcellation between the two models and with those medians obtained for some of the tested graph-theoretical statistics (Figure 7, Figure 9C). In particular, θ_{dd} derived from the SC matrix and $\rho_{SC,FC}$ exhibit significant correlations, whereas the indices α_{dd} and β_{dd} of the degree distribution derived from the empirical FC matrix show significant anti-correlations with the fitting results for both models (Figure 9C). However, the inter-parcellation patterns of the modularities Q^{SC} and Q^{FC} are anti-correlated with those of the goodness-of-fit obtained for the neuronal mass model only (Figure 9C).

The network properties of the empirical connectomes (Figure 7) and the quality of the model validation as given by the values of the goodness-of-fit of the model to empirical data (Figure 9A) demonstrate a pronounced variation when the utilized brain parcellation is changing. To quantify their relationship we combined principal component analysis (PCA) with ordinary least squares linear regression. In this approach, we first built a dataset with the median values of the graph-theoretical statistics across subjects such that we obtain a 10x19 matrix in which each row is associated with one of the quantities (network metrics) depicted in Figure 7 and each column holds the values for those metrics corresponding to a particular brain parcellation from Table 2. Next, the first principal component (PC1) was extracted from the dataset, which was found to explain 59% of the variance in the data variables (network properties) across the brain parcellations. Subsequently, we regressed the PC1 scores with the medians of the goodness-of-fit calculated across subjects for every brain parcellation and found that this PC explains about 82% and 87% of the inter-parcellation variance in the modelling results for the phase oscillator and neural mass model, respectively (Figure 9E). With these results, we demonstrated that most of the inter-parcellation variation observed in the modelling results (Figure 9A) can be explained by the empirical data used to inform and validate the models.

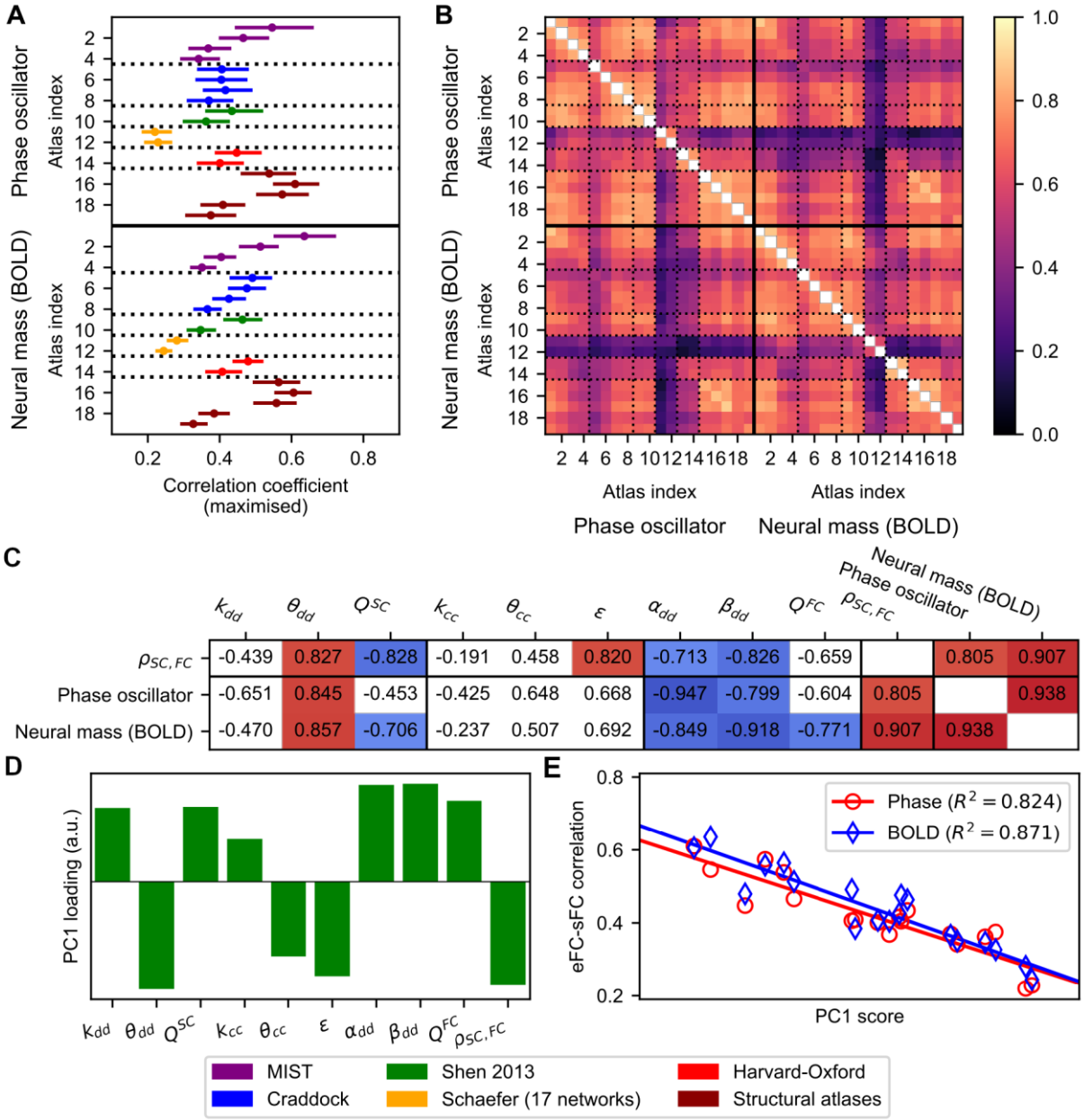

Figure 9: (A) Maximized correlations (goodness-of-fit) between the empirical and simulated FC matrices for the brain parcellation schemes and models investigated in this study as indicated on the vertical axis. Dots and lines depict the medians and interquartile ranges across subjects, respectively. (B) Correlations across subjects of the goodness-of-fit of the model between the considered parcellations and models. The color in plot A and the atlas indices on the axes correspond to those in Table 2 and also indicated in the legend at the bottom of the figure. (C) Cross-correlations between the graph-theoretical measures, empirical structure-function relationship $\rho_{SC,FC}$ and the goodness-of-fit of the models to empirical data. The correlation was calculated across parcellations between the median values over all subjects. Significant correlations are highlighted by colors ($p<0.05$, two sided, Bonferroni corrected). (D) Loadings of the first principal component (PC1) of the group-averaged graph-theoretical metrics, i.e. the contributions of the original empirical data variables (network properties depicted in Figure 7) to PC1. (E) Regressions of PC1 scores with the medians of the goodness-of-fit between empirical and simulated functional connectivity. The medians were calculated across subjects for each considered brain parcellation for the phase oscillator (red) and the neural mass model (blue) as indicated in the legend together with the fraction of the explained variance. The symbols stand for the individual parcellations.

5. Conclusion, next steps

The performed evaluation of the atlas-based compression of the neuroimaging data has involved a true cross-modal approach being applied to the empirical dwMRI data and structural connectome, resting-state fMRI data with BOLD time series and the functional connectome as well as data simulated by whole-brain dynamical models and results of the model fitting to the empirical data. The reported findings underlined an importance of investigation of the parcellation-induced changes of the graph-theoretical network metrics and other statistical properties of the empirical data as well as their impact on the model validation. The discussed inter-subject and inter-parcellation relationships between structural, functional and simulated brain connectomes can be used to explain the underlying mechanisms of the modeling results, e.g., the contribution of the empirical data to the simulated brain activity. It can also provide an insight into the optimal setups and parameters of the data processing, e.g., connectome resolution and model selection for simulation of the brain activity at the level of individual subjects.

The next steps will include the development of the used modeling approach involving a variety of brain parcellations for the subject-specific personalized simulations and model-based prediction of behavioral data for healthy and clinical cohorts with application of the methods of machine learning.

References

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. 2008, P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. NeuroImage 9, 179–194. https://doi.org/10.1006/nimg.1998.0395

Deco, G., Jirsa, V., McIntosh, A.R., Sporns, O., Kötter, R., 2009. Key role of coupling, delay, and noise in resting brain fluctuations. Proc. Natl. Acad. Sci. 106, 10302–10307. https://doi.org/10.1073/pnas.0901831106

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31, 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Eickhoff, S.B., Yeo, B.T.T., Genon, S., 2018. Imaging-based parcellations of the human brain. Nat Rev Neurosci 19, 672–686. https://doi.org/10.1038/s41583-018-0071-7

Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. Neuroimage 19, 1273–1302. https://doi.org/10.1016/S1053-8119(03)00202-7

Griffanti, L., Salimi-Khorshidi, G., Beckmann, C.F., Auerbach, E.J., Douaud, G., Sexton, C.E., Zsoldos, E., Ebmeier, K.P., Filippini, N., Mackay, C.E., Moeller, S., Xu, J., Yacoub, E., Baselli, G., Ugurbil, K., Miller, K.L., Smith, S.M., 2014. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. NeuroImage 95, 232–247. https://doi.org/10.1016/j.neuroimage.2014.03.034

Honey, C.J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.P., Meuli, R., Hagmann, P., 2009. Predicting human resting-state functional connectivity from structural connectivity. Proc Natl Acad Sci U A 106, 2035–2040. https://doi.org/10.1073/pnas.0811168106

Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. FSL. NeuroImage, 20 YEARS OF fMRI 62, 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Jeurissen, B., Tournier, J.-D., Dhollander, T., Connelly, A., Sijbers, J., 2014. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. NeuroImage 103, 411–426. https://doi.org/10.1016/j.neuroimage.2014.07.061

Jülich Supercomputing Centre, 2018. JURECA: Modular supercomputer at Jülich Supercomputing Centre. J. Large-Scale Res. Facil. 4, A132. https://doi.org/10.17815/jlsrf-4-121-1

Kuramoto, Y., 1984. Chemical oscillations, waves, and turbulence. Springer, Berlin.

Mayhew, S.D., Ostwald, D., Porcaro, C., Bagshaw, A.P., 2013. Spontaneous EEG alpha oscillation interacts with positive and negative BOLD responses in the visual–auditory cortices and default-mode network. NeuroImage 76, 362–372. https://doi.org/10.1016/j.neuroimage.2013.02.070

Messe, A., 2019. Parcellation influence on the connectivity-based structure-function relationship in the human brain. Hum Brain Mapp 1–14. https://doi.org/10.1002/hbm.24866

Pervaiz, U., Vidaurre, D., Woolrich, M.W., Smith, S.M., 2020. Optimising network modelling methods for fMRI. NeuroImage 211, 116604. https://doi.org/10.1016/j.neuroimage.2020.116604

Proix, T., Spiegler, A., Schirner, M., Rothmeier, S., Ritter, P., Jirsa, V.K., 2016. How do parcellation size and short-range connectivity affect dynamics in large-scale brain network models? Neuroimage 142, 135–149. https://doi.org/10.1016/j.neuroimage.2016.06.016

Rubinov, M., Sporns, O., 2011. Weight-conserving characterization of complex functional brain networks. NeuroImage 56, 2068–2079. https://doi.org/10.1016/j.neuroimage.2011.03.069

Rubinov, M., Sporns, O., 2010. Complex network measures of brain connectivity: Uses and interpretations. NeuroImage, Computational Models of the Brain 52, 1059–1069. https://doi.org/10.1016/j.neuroimage.2009.10.003

Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.-N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T.T., 2018. Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. Cereb. Cortex 28, 3095–3114. https://doi.org/10.1093/cercor/bhx179

Spitoni, G.F.P., Di Russo, F., Cimmino, R.L., Bozzacchi, C., Pizzamiglio, L., 2013. Modulation of spontaneous alpha brain rhythms using low-intensity transcranial direct-current stimulation. Front. Hum. Neurosci. 7. https://doi.org/10.3389/fnhum.2013.00529

Thirion, B., Varoquaux, G., Dohmatob, E., Poline, J.-B., 2014. Which fMRI clustering gives good brain parcellations? Front Neurosci 8, 167. https://doi.org/10.3389/fnins.2014.00167

Tournier, J.-D., Calamante, F., Connelly, A., 2010. Improved probabilistic streamlines tractography by 2nd order integration over fibre orientation distributions. Proc Intl Soc Mag Reson Med ISMRM 18.

Tournier, J.-D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C.-H., Connelly, A., 2019. MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. NeuroImage 202, 116137. https://doi.org/10.1016/j.neuroimage.2019.116137

Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: Improved N3 Bias Correction. IEEE Trans. Med. Imaging 29, 1310–1320. https://doi.org/10.1109/TMI.2010.2046908

Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K., 2013. The WU-Minn Human Connectome Project: An overview. NeuroImage, Mapping the Connectome 80, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041

Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T.E.J., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S.W., Della Penna, S., Feinberg, D., Glasser, M.F., Harel, N., Heath, A.C.,

Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S.E., Prior, F., Schlaggar, B.L., Smith, S.M., Snyder, A.Z., Xu, J., Yacoub, E., 2012. The Human Connectome Project: A data acquisition perspective. NeuroImage, Connectivity 62, 2222–2231. https://doi.org/10.1016/j.neuroimage.2012.02.018

Wilson, H.R., Cowan, J.D., 1972. Excitatory and Inhibitory Interactions in Localized Populations of Model Neurons. Biophys. J. 12, 1–24. https://doi.org/10.1016/S0006-3495(72)86068-5

Zimmermann, J., Griffiths, J., Schirner, M., Ritter, P., McIntosh, A.R., 2019. Subject specificity of the correlation between large-scale structural and functional connectivity. Netw. Neurosci. 3, 90–106. https://doi.org/10.1162/netn_a_00055