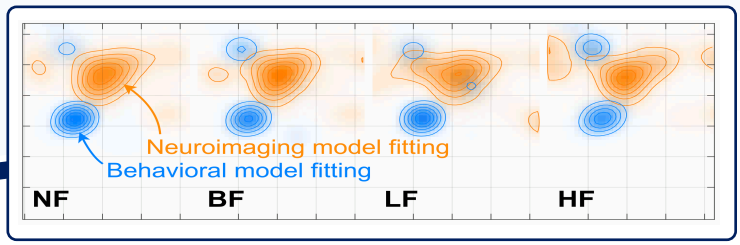
VirtualBrainCloud

Personalized Recommendations for Neurodegenerative Disease



www.VirtualBrainCloud-2020.eu

Public deliverable report

D3.3: Optimization and evaluation of pipeline for practical use concluded

Date	May 2023
Authors	Institute of Neuroscience and Medicine (INM-7; Brain and Behaviour) Forschungszentrum Jülich: Kyesam Jung, Shufei Zhang, Oleksandr V. Popovych, Simon B. Eickhoff

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**

Table of content

1. Introduction

The complexity of the neuroimaging data recorded by modern magnetic resonance imaging (MRI) requires a sophisticated data processing and signal extraction, which plays a crucial role for a successful data analysis and reliable results (Satterthwaite *et al.*, 2013; Ciric *et al.*, 2017; Maier-Hein *et al.*, 2017; Lindquist *et al.*, 2019; Botvinik-Nezer *et al.*, 2020). This is a complicated and challenging problem especially because of the lacking ground truth for empirical data and golden standards for the neuroimaging data processing. A variability in the latter can cause a variability of the derived results due to different data processing, which led to intense debates in the literature with respect to the best-practice approaches and parameters of the data processing delivering the most plausible results (Salimi-Khorshidi *et al.*, 2014; Qi *et al.*, 2015; Parkes *et al.*, 2018; Lindquist, 2020; Maffei *et al.*, 2022). Selecting an appropriate approach and parameters for the neuroimaging data processing is also crucial for modeling of the resting-state brain activity by the whole-brain dynamical models investigated in the framework of the VirtualBrainCloud project, where the empirical data are used for the model derivation and validation (Ghosh *et al.*, 2008; Honey *et al.*, 2009; Popovych *et al.*, 2019). Indeed, the modeling involves both structural (diffusion-weighted) and functional MRI data as well as brain parcellations for construction of the underlying model network and fitting of the model dynamics to the empirical brain dynamics. The modeling results appear to be sensitive to the effect of the varied data processing, selected parcellation and parameters of the brain tractography (Domhof *et al.*, 2021; Jung, Eickhoff and Popovych, 2021; Popovych *et al.*, 2021; Aquino *et al.*, 2022; Domhof, Eickhoff and Popovych, 2022). It is therefore important to verify the modeling output and derived conclusions with respect to a variety of reasonable choices of data processing parameters in order to evaluate their impact, identify potential convergence to the common output for robust conclusions and optimize the analysis procedure.

With the current deliverable we contribute to optimization of the data processing pipeline for the brain modeling by biophysically-motivated and data-driven dynamical models. We illustrate this on one example, where the modeling approach was successfully applied to classification of the clinical data by machine-learning approach (Jung *et al.*, 2022). The employed pipeline was developed in the framework of the VirtualBrainCloud project. The initial version of it was reported in the deliverable D3.1 “Initial version of full-scope containerized pipeline developed” (Jung *et al.*, 2020) that provided an individualized approach to the neuroimaging data processing. This is a full-scope pipeline for processing of the diffusion-weighted images (DWI) and the resting-state functional MRI (fMRI) human neuroimaging data that takes the raw MRI data as an input and delivers the whole-brain structural and functional connectomes necessary, in particular, for the whole-brain modeling. The pipeline has a modular structure, which contributes to flexibility of pipeline application appropriate for cross-platform usage on different hardware and software environments. In addition, the containerized scheme allows us to process MRI data with version-controlled software that can give consistent outcomes in different operating systems. It was tested for execution on single-core desktops and local clusters as well as on supercomputers (Jülich Supercomputing Centre, 2021), where the pipeline was optimized for parallel processing of several subjects on multi-thread computational nodes necessary for the handling of large datasets. The pipeline was developed to simplify the procedure of the high-quality and state-of-the-art data processing for the end user on the one hand, but also to provide enough freedom for selection of parameters and algorithms of the data cleaning and signal extraction on the other hand.

Further enhancements of the pipeline were implemented and reported in the deliverable D3.2 "Most appropriate RS-processing pipelines for TVB modelling finished" (Jung *et al.*, 2021), where the focus was on data personalization. In particular, the updated version of the pipeline was designed to provide the whole-brain tractography, blood oxygen level-dependent (BOLD) signals and the respective structural and functional connectomes in the corresponding native spaces of individual subjects. We avoided the complex nonlinear transformation of the individual brains to the standard MNI152 space (Evans *et al.*, 2012), which can especially be grave for the brains deviating from the template brain, e.g., in the case of brain atrophies or injuries frequently observed in clinical data and old subjects. Brain normalization may also suppress the inter-individual variability that can be an important feature for differentiation between individual subjects in health and disease. Such a variability can be implemented in personalized whole-brain models, where region- and subject-specific features extracted from the empirical data are incorporated in the model parameters (Deco *et al.*, 2017, 2019, 2021; Demirtas *et al.*, 2019; Wang *et al.*, 2019; Domhof *et al.*, 2021; Jung, Eickhoff and Popovych, 2021; Popovych *et al.*, 2021; Domhof, Eickhoff and Popovych, 2022). Personalized models can demonstrate better fit to empirical data, test-retest reliability and subject specificity and can be applied for modeling clinical data and treatments in the framework of personalized medicine (Falcon, Jirsa and Solodkin, 2016; Jirsa *et al.*, 2017, 2023; Deco *et al.*, 2021; Domhof, Eickhoff and Popovych, 2022). The developed pipeline may contribute to this topic by enhancing the model personalization already at the level of empirical data used for the model derivation and validation.

The structure of the developed pipeline is schematically illustrated in Fig. 1, see also the deliverables (Jung *et al.*, 2020, 2021) for more details. In the current deliverable we demonstrate how the personalized pipeline can be applied to processes the clinical neuroimaging data of Parkinson's disease that can be used for biophysical modeling and prediction based on machine learning. We vary the respective pipeline parameters and optimize them in order to appropriately prepare the processed empirical data for the modeling of the inter-individual variability in the clinical data. We in particular show that complementing the empirical connectomes by simulated ones obtained for the optimal pipeline parameters and the best neuroimaging and behavioral model fits can significantly improve the differentiation of the patients from healthy controls. Our results suggest that the personalized whole-brain models based on personalized data can serve as an additional source of information relevant for diagnosis of diseases and possibly for their treatment as well.

2. Partners involved

This deliverable was prepared by the Institute of Neuroscience and Medicine (Brain and Behaviour, INM-7) from the Forschungszentrum Jülich (FZJ). The computational resources were granted through JARA on the supercomputer JURECA (Jülich Supercomputing Centre, 2021) at Jülich Supercomputing Centre, Forschungszentrum Jülich.

3. Description of work performed

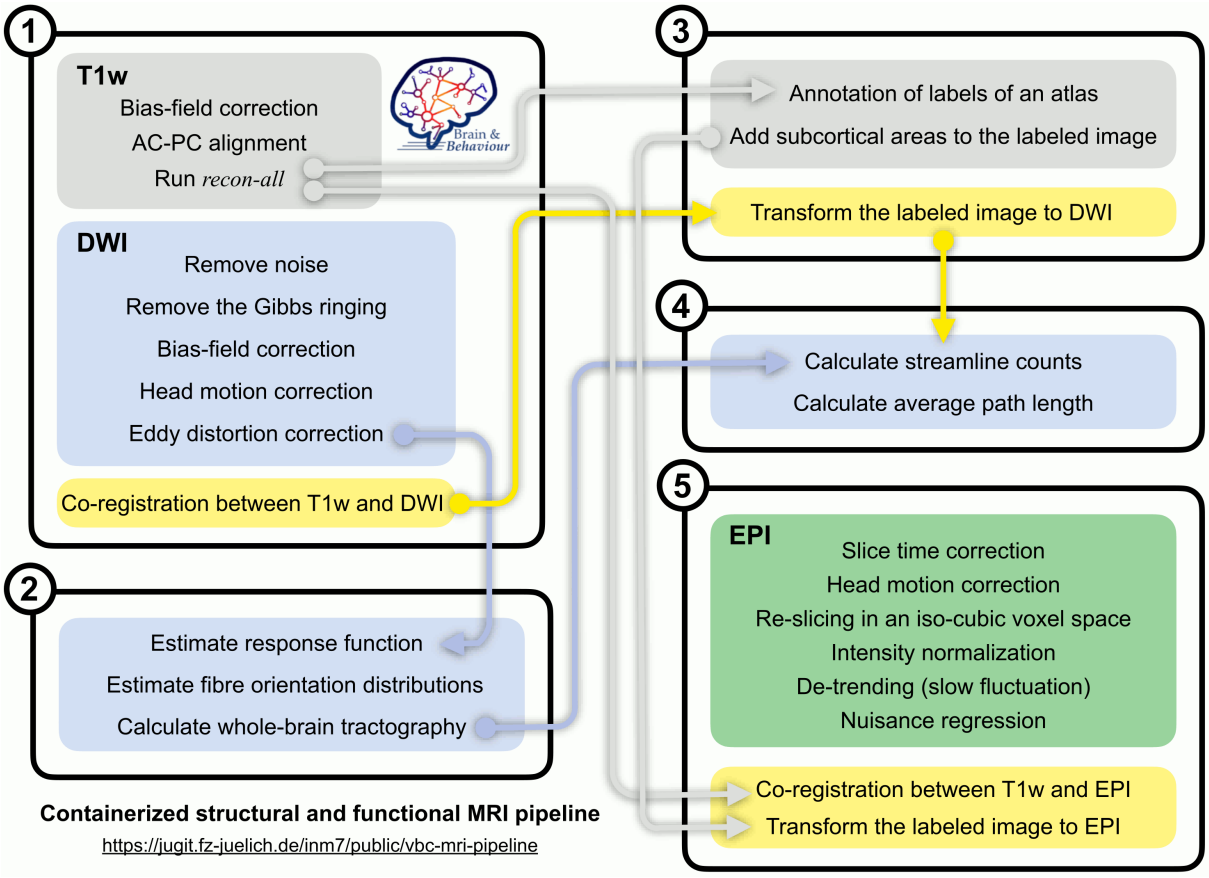


Figure 1: Schematic illustration of the pipeline workflow for personalized data processing and signal extraction. The independent modules are encircled in enumerated rectangular blocks with performed data processing steps indicated, where the operations shadowed in color are used to process T1w images (gray), dwMRI data (light blue) and fMRI data (green). The parts of the pipelines necessary for parcellations in the native spaces are indicated in yellow. The arrows depict the interdependence (required input) between the modules.

In this study, three whole-brain connectivities were calculated for 116 subjects consisting of 51 (30 males) healthy controls (HC) and 65 (45 males) patients diagnosed with Parkinson's disease (PD): empirical functional connectivity (eFC), empirical structural connectivity (SC), and simulated functional connectivity (sFC). Figure 2 schematically illustrates the data processing and simulation workflow. Four temporal filtering conditions were applied to empirical and simulated BOLD signals. Subsequently, we considered three types of connectivity relationships: correlation of eFC vs. SC, correlation of SC vs. sFC, and correlation of eFC vs. sFC. The relationships involving sFC vary depending on the two free parameters of global coupling and global delay of the computational model as illustrated by the eFC-sFC correlation landscape in the parameter space in Fig. 2 (the rightmost color plot). Such a landscape was used for the model validation, where the model parameters of the best fit (strongest similarity) between empirical and simulated data were considered as optimal.

Structural and functional MRI data were acquired using a 3T scanner (Siemens Trio). A structural brain image was acquired using a three-dimensional T1-weighted (T1w) sequence of the voxel dimension 240 × 256 × 160 with voxel size 1.0 × 1.0 × 1.1 mm3. Diffusion-weighted images (DWI) comprised a single non-weighted (B0) image and weighted (B = 1000 s/mm2) images with 64 directions and voxel dimension 90 × 90 × 55 with voxel size 2.4 × 2.4 × 2.4 mm3. Resting-state fMRI was obtained using an echo-planar imaging sequence during 663s with TR = 2.21 s, voxel dimension 64 × 64 × 36 with voxel

size 3.125 × 3.125 × 3.565 mm3, see (Jung et al., 2022) for more details. We performed quality control for all acquired MRI data, where the data of 73 subjects have been found to have no artifacts or abnormal brain shape. For the other 43 subjects included in the study, DWI volumes with artifacts were culled such that 3.9 impaired volumes were removed on average in the screened subjects, and 60.1 weighted volumes were on average kept for further analyses (Fig. 3). The maximal number of the removed volumes was 16 for one subject with 48 weighted images of good quality, and we therefore included all 116 subjects in the subsequent analyses after the data quality check and artifact cleaning.

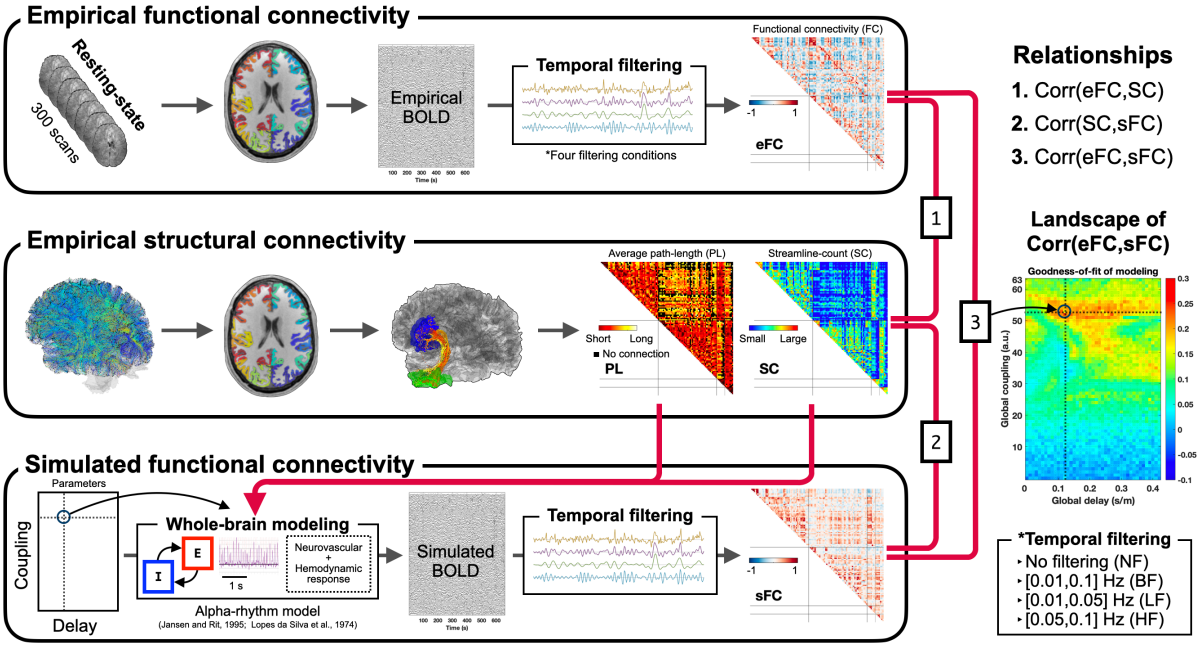


Figure 2: Data processing and simulation overview. First (upper box), brain parcellations in the native space of T1w were prepared and applied to the processed fMRI data, BOLD signals were extracted from the corresponding regions and filtered according to four temporal filtering conditions (see the right bottom box), and four empirical functional connectivities (eFC) were calculated. Second (middle box), the parcellations were also used for calculation of the structural connectivity by extracting streamlines from the whole-brain tractography reconstructed using diffusion-weighted images, where the number and length of streamlines connecting any two brain regions were collected into matrices of streamline count (SC) and the averaged path-length (PL). Third (lower box), the structural connectome (SC and PL) was used to build a brain network for the whole-brain modeling that simulates BOLD signals, which were filtered according to the considered filtering conditions (right bottom box) and used to calculate simulated FC (sFC). Subsequently, we calculated connectivity relationships (Pearson correlation) using those three connectivity matrices: (1) corr(eFC, SC), (2) corr(sFC, SC), and (3) corr(eFC, sFC). Model parameters of global coupling and global delay were varied to validate the model against empirical data. In particular, the correspondence (correlation) between empirical and simulated functional connectivities was calculated for each parameter point resulting in the parameter space landscape of relationship between eFC and sFC (rightmost color plot). The most pronounced correspondence (correlation) between eFC and sFC was selected together with the respective optimal model parameters as a result of neuroimaging model fitting for further analysis.

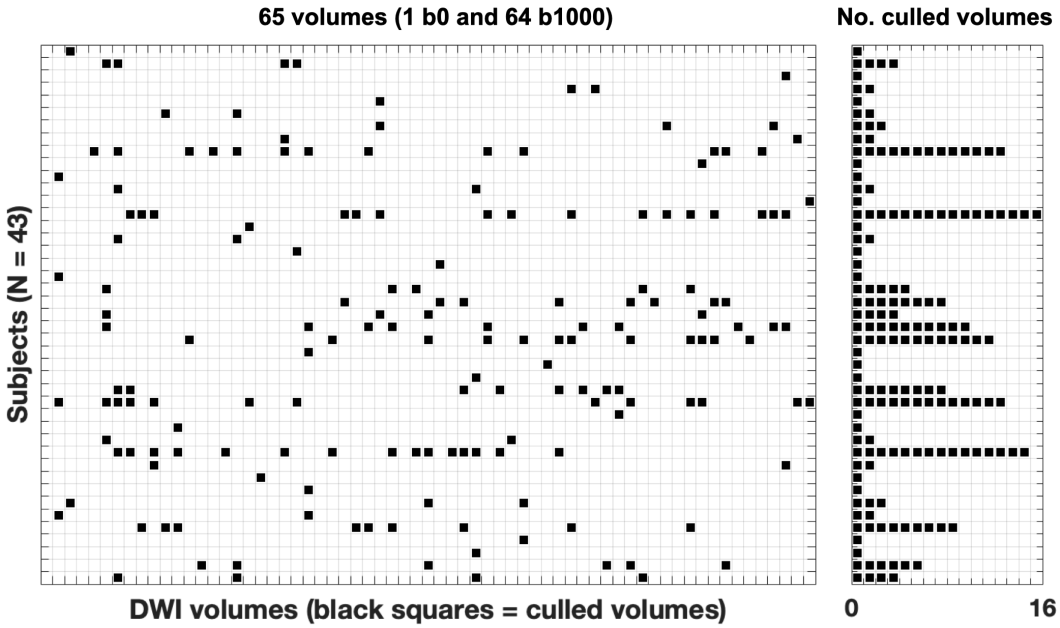


Figure 3: Culled volumes in diffusion-weighted images (DWI) after quality check.

For the personalized data processing, we used a containerized pipeline developed in the framework of the VirtualBrainCloud to process the structural and functional MRI in the corresponding native spaces (Jung *et al.*, 2020, 2021). The pipeline consists of five modules: preprocessing of structural MRI (T1w and DWI), whole-brain tractography (WBT) calculation, atlas transformation, reconstruction of structural connectivity (SC and PL), and preprocessing of fMRI (Fig. 1). The pipeline comprises Freesurfer (Dale, Fischl and Sereno, 1999), FSL (Smith *et al.*, 2004), ANTs (Tustison *et al.*, 2014), MRtrix3 (Tournier *et al.*, 2019), and AFNI (Cox, 1996). It is publicly available (https://jugit.fz-juelich.de/inm7/public/vbc-mri-pipeline).

3.1. Preprocessing of the structural MRI data

The preprocessing module of structural MRI performed the following steps: bias-field correction for T1w, alignment of anterior-posterior commissures (AC-PC) of T1w, recon-all by Freesurfer, removing the Gibbs ringing artifacts of DWIs, bias-field correction for DWIs, corrections of head motion, b-vector rotations and eddy distortion of DWIs, and co-registration between averaged DWI and T1w. This module segmented subcortical areas based on voxel intensities of the T1w images. It also prepared labeling annotations using a brain atlas for which a classifier was available. The latter can also be created from a training set by capturing region data either drawn by neuroanatomists or according to dedicated algorithms (Fischl *et al.*, 2004; Jung *et al.*, 2021). The WBT calculation module included only MRtrix3 functions. Fiber oriented distributions (FOD) were estimated from the DWIs using spherical deconvolution, and the WBT with 10 million total streamlines was created through the fiber tracking by the second-order integration over the FOD by a probabilistic algorithm (Tournier, Calamante and Connelly, 2010). The atlas transformation module annotated labels using a classifier to parcel cortical regions in the native T1w space by Freesurfer. In the present study, we applied two atlas classifiers for brain parcellations, the Schaefer atlas with 100 parcels (Schaefer *et al.*, 2018) and the Desikan-Killiany atlas with 68 parcels (Desikan *et al.*, 2006). The former is based on fMRI data, and the latter is labeled by gyral-based anatomical parcellation. Both atlases provide cortical parcellations, and additional 14 subcortical areas segmented by the preprocessing module were included and combined with the

labeled cortical parcels. The pipeline transformed the labeled image (cortical parcels and subcortical regions) from the T1w to DWI native spaces. Finally, the reconstruction module calculated the matrices of the streamline counts (SC) and the matrices of the average path lengths (PL) of the streamlines extracted between any two parcellated brain regions from the calculated WBT with the transformed labeled image in the native DWI space.

3.2. Preprocessing of the functional MRI data

The preprocessing module of the resting-state fMRI data performed slice time correction, head motion correction, re-slicing in a 2 mm iso-cubic voxel space, intensity normalization, de-trending with filtering very slow fluctuation out (high pass), co-registration to the T1w image, and calculation of regressors for the white-matter, cerebrospinal fluid (CSF), and brain global signals as well as for the head motion. The pipeline also transformed the labeled image of the brain parcellation generated in the native T1w space to the fMRI native space. Finally, we performed a nuisance regression with the prepared regressors (white-matter, CSF, and the brain global signals as well as head motions). After preprocessing of MRI, we extracted BOLD signals based on the annotated atlas labels and applied three temporal band-pass filtering conditions in the frequency ranges [0.01,0.1] Hz (broad band; BF), [0.01,0.05] Hz (low-frequency band; LF), and [0.05,0.1] Hz (high-frequency band; HF). Therefore, four filtering conditions were considered: no filtering (NF), BF, LF, and HF. The filtering was made by a script in the Python programming language (version 3.8, Python Software Foundation, https://www.python.org/) using the SciPy (version 1.5) signal processing module and NumPy (version 1.19) for the temporal band-pass filtering. We used the Butterworth digital filter of the order 6, *scipy.signal.butter*.

3.3. Model simulations

The whole-brain resting-state dynamics was simulated by a system of coupled neuronal models representing the mean activity of the brain regions. Each region contains two populations of excitatory and inhibitory neurons that interact with each other via post-synaptic potentials (PSP) (Lopes da Silva *et al.*, 1974). The considered convolution-based model was of the Jansen-Rit type (Jansen and Rit, 1995; Moran, Pinotsis and Friston, 2013) and simulated the PSP signals involving other brain regions interacting with time delay in coupling according to the calculated structural connectivity, i.e., SC and PL matrices, see Ref. (Jung *et al.*, 2022) for details. We considered two model fitting approaches that we refer to as *neuroimaging* and *behavioral* model fitting. The former fitting is well known from the literature and consists in validation of the model against neuroimaging empirical data. In this study, the Pearson correlation between eFC and sFC (upper triangle parts) was calculated and referred to as *goodness-of-fit (GoF)* values (Honey *et al.*, 2009; Deco *et al.*, 2017). We also considered the connectivity relationship between empricial SC (eSC) and sFC.

For the behavioral model fitting we calculated the effect size of the difference between the neuroimaging GoF values of the HC and PD subject groups for each model parameter point in the considered parameter space. We thus obtained the parameter landscape reflected the relation of the model to the behavioral data, which can thus be used for behavioral model fitting. The landscape of the effect size was verified with respect to statistical significance followed by correction for multiple comparisons, which resulted in a thresholded Z-score map for significant parameter areas (alpha < .05). Finally, we searched for the optimal model parameters in the significant areas corresponding to the

maximal effect size, i.e., maximal differences between HC and PD in the GoF value. We also considered two connectivity relationships (eFC vs. sFC and eSC vs. sFC) for the behavioral model fitting.

3.4. Cross-validation of prediction model

We used a regularized logistic regression with the least absolute shrinkage and selection operator (LASSO) for training and classification of the healthy and PD subjects (Tibshirani, 1996). To avoid an overfit, the training error included the deviance and an L1-penalty (Hastie, Tibshirani and Friedman, 2009). We used a cross-validation (CV) to train and test the regularized logistic regression, see (Jung *et al.*, 2022) for details. Since PD is a degenerative disease, the features for PD classification should be controlled by an age effect via confound regression. Due to a random sampling from the same cohort, we should prevent data leakage during the machine learning CV procedure, because the trained models can be biased throughout the confound regression and neuroimaging and behavioral fitting of the computational model. With this respect, we applied a cross-validated confound regression (More *et al.*, 2023) as well as a cross-validated fitting of the computational model to empirical data in order to prevent a data leakage during model training and remove the age effect from the trained model.

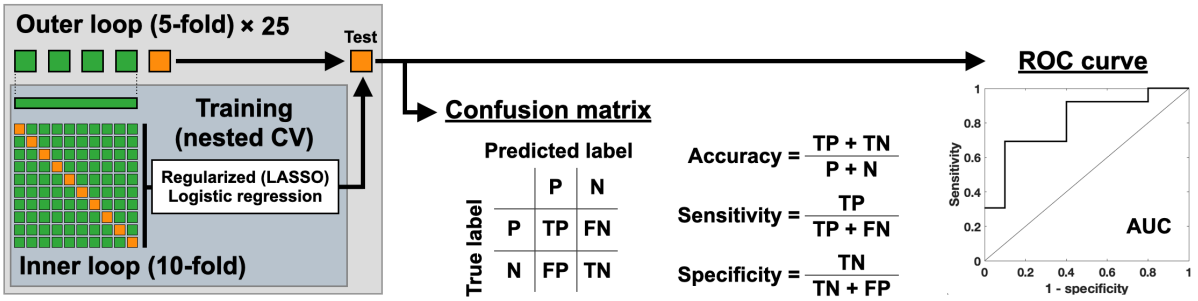


Figure 4: Schematic illustration of the nested cross-validation (CV) utilized in this study. The green boxes in the leftmost plot illustrate randomly split subject subgroups used for training of the model in the 5-fold outer loop and in the 10-fold inner loop. The orange box in the outer loop depicts the testing subject subgroup used for evaluation of the prediction performance of the trained model as given by accuracy, sensitivity, specificity and area under ROC curve. Abbreviations P: positive as patients, N: negative as controls, TP: true positive, FP: false positive, TN: true negative, FN: false negative, LASSO: least absolute shrinkage and selection operator, ROC: receiver operating characteristic, and AUC: area-under-curve.

We used a nested CV to train a regularized logistic regression for PD classification. Figure 4 illustrates the training and test processes via the nested CV. In an outer loop, we randomly split the subjects into five subsets with the same ratio of HC and PD within each subset around 0.8 (51:65 in the entire cohort). One subset of 20% of subjects was considered as a test set (unseen subjects, the orange box in the outer loop in Fig. 4), and the others (four subsets pulled together, the green boxes in the outer loop in Fig. 4) composed a training set. Then, the selected training set from the outer loop was split into ten subsets for the nested CV in the inner loop. The training process estimated the parameters of the logistic regression, i.e., the parameter of the L1-penalty which gives the minimal training error during the 10-fold repetitions. After that, the trained model was applied to predict the classes in the test set of the outer loop. We calculated a confusion matrix and plotted a receiver operating characteristic (ROC) curve (Fawcett, 2006) based on the prediction by shifting a hit criterion of a predicted probability from 0 to 1. Subsequently, we calculated the prediction accuracy and the area-under-curve (AUC) of the ROC curve. In the next step, we set one of the four training subsets in the outer loop as a test set and composed

the training set from the remaining subjects for the next iteration of the inner loop. Then we performed the nested CV (inner loop) again. Therefore, from one 5-fold random splitting of the subjects in the outer loop, five prediction results were obtained. We repeated this procedure 50 times leading to 5 × 50 = 250 prediction results.

4. Results

We performed the whole-brain simulations using the whole-brain dynamical model derived from and validated against MRI data from 116 subjects consisting of 51 HCs and 65 PD patients. We calculated sFC depending on the varied model parameters, brain parcellations and the filtering of the empirical and simulated BOLD signals in different frequency bands. The model was then validated based on the GoF values that were calculated as similarity (Pearson correlation) between the empirical and simulated connectomes and maximized by varying model parameters in the considered parameter space. As mentioned above, such a model validation was referred to as neuroimaging model fitting. Furthermore, we split subjects into HC and PD groups, compared the neuroimaging GoF values between the groups and found the optimal model parameters, where the difference between HC and PD subjects is maximal, i.e., the largest effect size of the difference, which was referred to as behavioral model fitting. Subsequently, we performed PD classification by using the connectivity relationships regarding the modality of model validation, i.e., neuroimaging and behavioral model fittings. Below we show that the differences between HC and PD subjects can be enhanced for the simulated connectomes as compared to empirical ones, and the performance of PD prediction by machine-learning approach can be improved when the empirical connectivity is complemented by simulation results.

4.1. Neuroimaging and behavioral model fitting

We calculated sFC using simulated BOLD signals for each parameter point on a dense grid of the 2-Dim model parameter space of the global delay and global coupling and obtained the GoF values (similarity, Pearson correlation) between eFC and sFC. Figure 5 shows the corresponding values in the parameter space averaged over all subjects, which define an average GoF parameter landscape. We calculated eFC and sFC for the frequency ranges of the corresponding BOLD signals provided by the employed temporal filtering, i.e., NF, BF, LF, and HF conditions (see Sec. 3.2 for details). This resulted in the GoF landscapes obtained for the corresponding filtering conditions (Fig. 5 A1 and B1). The profiles of the parameter landscapes were different between the considered brain atlases. The case of the Schaefer atlas showed the hill around maximal GoF values (the dashed circle in Fig. 3 A1), and the distributions of the optimal global delays for the maximal GoF were mainly in the biologically feasible range, from 0.06 to 0.25 s/m (Caminiti *et al.*, 2013), where we used the actual PL values in millimeters for the signal propagation in the model (Fig. 5 A4). On the other hand, the maximal GoF values for the Desikan-Killiany atlas exhibit a bi-modal distribution (the dashed circles in Fig. 5 B1) that are well separated along the parameter of global coupling (Fig. 5 B3). Moreover, stronger global coupling of the maximal GoF values is accompanied by a wide spread along with global delays (the upper dashed circle in Fig. 5 B1), and these optimal global delays may get out of the biologically feasible range as compared to the weaker global couplings (the lower dashed circle in Fig. 5 B1). We also note here that the strongest similarity between the simulated and empirical data appears to be observed for the non-filtered BOLD signals (NF) for both considered brain parcellations, where the relatively narrower frequency bands (LF and HF) resulted in significantly lower maximal GoF values than the cases of the broad (BF) or the entire frequency (NF) range of the resting-state BOLD signals.

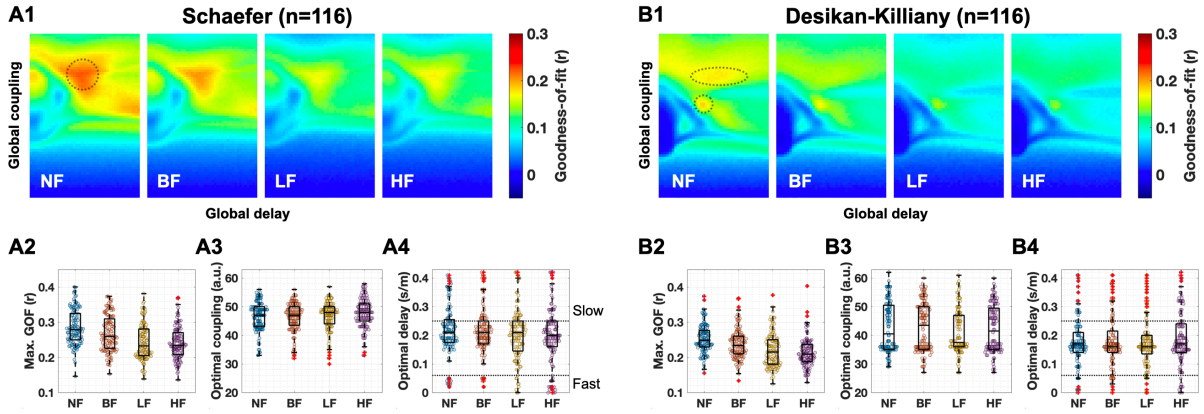
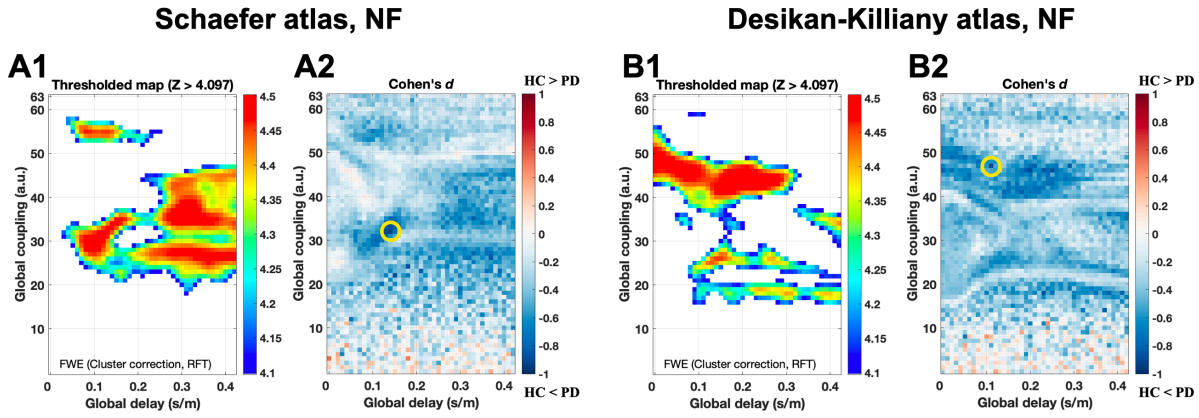


Figure 5: Results of the neuroimaging model fitting for (A) the Schaefer atlas and (B) the Desikan-Killiany atlas. (A1, B1) Parameter landscapes of the similarity (Pearson correlation) between empirical FC and simulated FC, i.e., goodness-of-fit (GoF) values averaged over the entire subject cohort. The landscapes are illustrated for each filtering condition (NF, BF, LF and HF, see Sec. 3.2 for details) indicated in the plots. The dashed circles delineate the hills around the maximal GoF values. Distributions of (A2, B2) the maximal GoF values, (A3, B3) optimal coupling parameters and (A4, B4) the respective optimal delays corresponding to the maximal GoF values for each filtering condition. The dashed horizontal lines in plots (A4, B4) indicate the biologically feasible delay range regarding the electrophysiological conduction speed (Caminiti et al., 2013). The middle lines in interquartile box plots indicate medians of distributions, and the red pluses are the outliers.



A3

Max. Cohen's *d*	NF	BF	LF	HF
Global coupling (a.u.)	32	32	32	32
Global delay (s/m)	0.14	0.14	0.14	0.14

B3

Max. Cohen's *d*	NF	BF	LF	HF
Global coupling (a.u.)	47	46	43	46
Global delay (s/m)	0.11	0.25	0.23	0.19

Figure 6: Behavioral model fitting based on a diagnostic criterion, i.e., on the differences between the HC and PD subjects for (A) the Schaefer atlas and (B) the Desikan-Killiany atlas. (A1, B1) Statistical maps of thresholded Z-scores for HC < PD relation between the neuroimaging GoF values in the model parameter plane of the global delay and global coupling obtained by the non-parametric Wilcoxon rank-sum one-tail test and corrected for multiple comparisons using random-field thresholding for cluster correction. The results are illustrated for NF condition. (A2, B2) Maps of the corresponding effect size between the groups. The maximal values are indicated by yellow circles. (A3, B3) Tables of the optimal parameters corresponding to the maximal effect size in the biologically feasible delay ranges from 0.06 to 0.25 s/m (Caminiti et al., 2013) for each temporal filtering condition.

For the behavioral model fitting, we considered the individual GoF values obtained during the neuroimaging model fitting (Fig. 5) and compared their distributions for the groups of HC and PD subjects for each parameter point in the model parameter plane of global delay and global coupling. This resulted in a statistical map of *p*-values of the Wilcoxon rank-sum one-tail (HC < PD) test in the model parameter space. We then corrected the *p*-values with respect to multiple comparison and applied the random-field thresholding (Worsley *et al.*, 1992) using a 2-dimensional Gaussian kernel with a 3-point full-width-half-maximum, i.e., by controlling the family-wise error by cluster correction. In the thresholded map (Z > 4.097 for alpha = .05), we observed a prominent disassociation between HC and PD groups in large parameter regions, where the neuroimaging GoF values of HC subjects are significantly smaller than those for the PD patients (Fig. 6 A1 and B1).

We also calculated the effect size between the GoF values of the HC and PD groups (Fig. 6 A2 and B2 for NF condition). The parameter maps of the effect size showed similar patterns with the thresholded statistical maps (Fig. 6 A1 and B1). Subsequently, for the parameter points of statistically significant GoF differences between HC and PD (Fig. 6 A1 and B1), we searched for the optimal parameter points corresponding to the maximal effect size (HC < PD, the yellow circles in Fig. 6 A2 and B2) for each filtering condition. These optimal model parameters and the corresponding effect size are considered as the results of the behavioral model fitting. Furthermore, we also performed bootstrapping analysis to see how stable the optimal parameters are when the subjects are randomly sampled. We found that the optimal parameter points for the maximal effect size were narrowly distributed across different subject samplings and filtering conditions (Fig. 7), which indicates their stability and robustness when the subject group configurations were varied. Interestingly, the optimal parameters for the maximal effect size were in the biologically plausible range of the model parameters (delay). Therefore, we selected these optimal parameter points in the biologically plausible range of the delay for the behavioral model fitting among a few local optima observed in the thresholded map (Figure 6 A3 and B3).

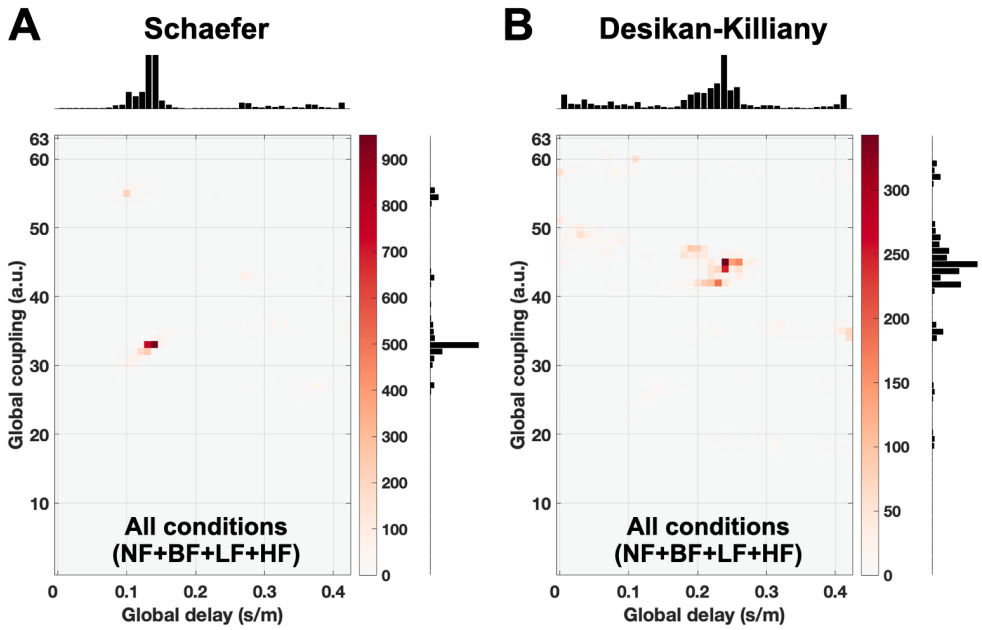


Figure 7: Results of the bootstrapping analysis (n=1000) for the optimal model parameters of the behavioral model fitting for (A) the Schaefer atlas and (B) the Desikan-Killiany atlas. The obtained distributions of the optimal model parameters are depicted in the model parameter plane and in the projections to the parameter subspaces, which

correspond to the maximal effect size of the difference between neuroimaging GoF values of the bootstrapped HC and PD subject groups.

4.2. Differentiation between HC and PD subjects

In this study we investigated three types of connectome relationships as mentioned in Fig. 2: eFC vs. eSC, eFC vs. sFC and eSC vs. sFC, to evaluate the group differences between HC and PD subjects. The empirical structure-functional relationship (eFC vs. eSC) has a single value per subject for each filtering condition and parcellation. Due to the varied model parameters of global delay and coupling, the other two connectome relationships attain as many values as the considered number of simulated parameter points of the model. For the two approaches of the model validation based on the neuroimaging and behavioral model fittings, the optimal model parameters and the corresponding sFC matrices were calculated based on the best model fit to neuroimaging or behavioral data, respectively. This approach is illustrated in Figs. 5-7 for the relation between simulated and empirical functional connectomes, eFC vs. sFC. The neuroimaging model fitting resulted in different optimal parameter points for different subjects corresponding to the different maximal GoF values in the parameter spaces of individual subjects (Fig.5 A2-A4, B2-B4). The behavioral fitting approach led to one optimal parameter point for the entire subject cohort and every filtering and parcellation condition, which corresponded to the maximal effect size of the difference of the neuroimaging GoF values between HC and PD subject groups (Fig. 6 A2-A3, B2-B3, and Fig. 7).

Similar procedure can be applied to the model fitting based on the structure-functional relationship between empirical SC and simulated FC, i.e., eSC vs. sFC, see Ref. (Jung *et al.*, 2022) for details. Accordingly, for every subject, filtering condition and brain parcellation, we calculated five optimal connectome relationships (Pearson correlation) for optimal model parameters obtained for the two model fitting approaches, which are corr(eFC, eSC) (empirical), corr(eFC, sFC) (neuroimaging), corr(eSC, sFC) (neuroimaging), corr(eFC, sFC) (behavioral), and corr(eSC, sFC) (behavioral). We used these five connectome relationships to evaluate how strongly they are differentiable between HC and PD. The empirical structure-function relationships corr(eFC, eSC) of HC and PD subjects were found to be from distributions with different medians (Fig. 8 A1 and B1). The case of the Schaefer atlas showed significantly different distributions between the groups for all filtering conditions (Fig. 8C, left, the first row). Here, HC subjects had a weaker correspondence between the brain structure and function. On the other hand, for the Desikan-Killiany atlas the differences in the distributions were significant for the LF filtering condition only (Fig. 8C, right, the first row).

The differentiation between PD and HC subject groups can be done based on the modeling results, where the optimal sFC is taken into account in the connectome relationships. The group differences obtained by involving sFC from the neuroimaging model fitting were small and non-significant for both atlases and all filtering conditions (Fig. 8, A2-A3, B2-B3, and C, the second and third rows). On the other hand, we observed significantly stronger agreements between the empirical and simulated connectomes for PD patients than HC subjects, when the optimal sFC from the behavioral model fitting was involved (Fig. 5, A4-A5, B4-B5 and C, the fourth and fifth rows). This holds for all considered parcellations, filtering conditions and functional and structure-functional connectome relationships. Therefore, involving the simulated connectomes in the analyses results in a stable differentiation between HC and PD, which is little influenced by the temporal filtering and parcellation scheme in contrast to the purely empirical data.

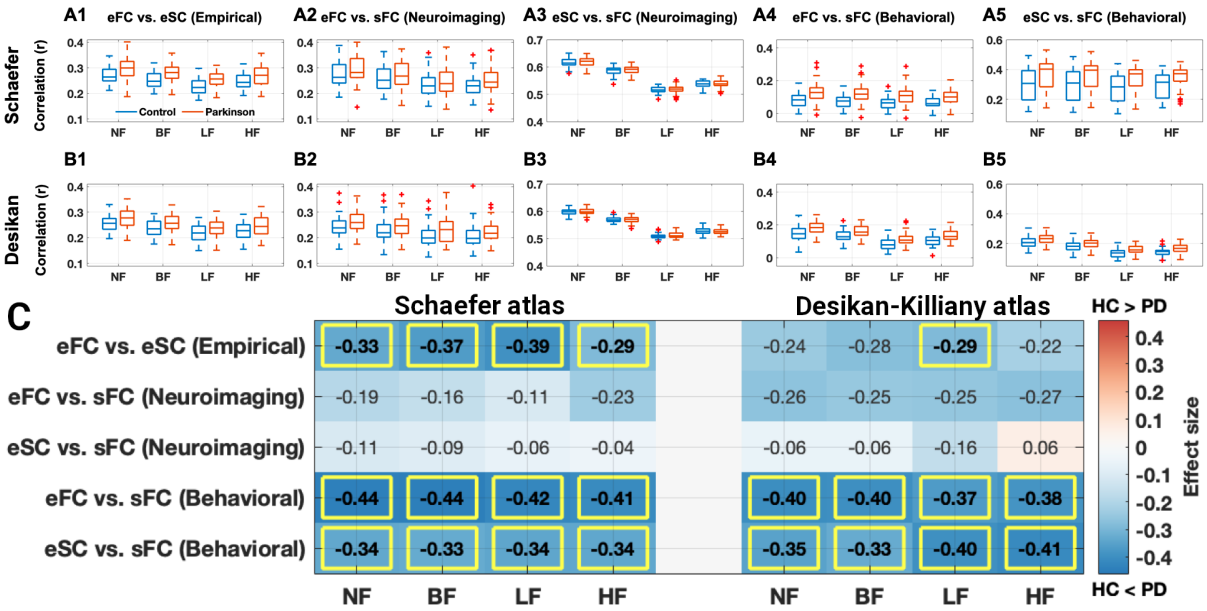


Figure 8: Differences between HC and PD subjects as reflected by the connectome relationships. (A, B) Box plots of five correspondences of the HC (blue) and PD (red) groups for (A) the Schaefer atlas and (B) the Desikan-Killiany atlas and for (1) the empirical structure-functional relationship corr(eFC, eSC), (2, 3) functional and structure-functional relationships for the simulated connectomes corr(eFC, sFC) and corr(eSC, sFC) for the neuroimaging model fitting, and (4, 5) connectome relationships corr(eFC, sFC) and corr(eSC,. sFC) for the behavioral model fitting. (C) Effect sizes between HC and PD subjects reflected by the empirical and simulated connectome relationships regarding both model fitting modalities. Summary tables of the effect sizes (blue for HC >PD and red for HC < PD) with indicated Bonferroni corrected p-values (yellow rectangles) of the Wilcoxon rank-sum two-tail test for (left) the Schaefer atlas and (right) the Desikan-Killiany atlas.

4.3. Classification of PD by machine learning

We employed the five whole-brain connectivity relationships for PD classification as the features of the machine-learning techniques, i.e., predictors in the logistic LASSO regression in our case. Three feature conditions were prepared by combining the ten features consisting of five connectome relationships mentioned above for two atlases: (i) empirical features, (ii) simulated features, and (iii) all features. Here, all ten features were included in the classification analysis, but some of them were randomly shuffled in accordance to the selected feature condition. The shuffling was performed by a random redistribution of the values of a given feature (connectome relationship) among subjects such that the correspondence of the feature to individual subjects (and to their behavioral characteristics, e.g., PD or HC) was destroyed. By focusing on some feature combination like conditions (i)-(iii) from above or brain parcellations, the other features were shuffled. For example, to focus on the empirical features for the Schaefer atlas, four simulated features (eFC vs. sFC and eSC vs. sFC for two model fitting modalities) of the Schaefer atlas and all five features (one empirical and four simulated features) of the Desikan–Killiany atlas were shuffled. For a given feature condition, the prediction model was trained according to the nested CV and applied to the unseen test subjects (Fig. 4). We then calculated a confusion matrix from the prediction results, plotted a ROC curve and evaluated the prediction performance (accuracy, sensitivity, specificity and balanced accuracy) and the area under a curve (AUC) of the ROC curve (Fig. 4).

Figure 9 shows the prediction performance of the trained models for each of the investigated conditional cases of brain parcellations, frequency bands of temporal filtering and feature conditions. The first important observation is that an involvement of the simulated connectomes can affect the classification of PD and HC, see Fig. 9 and compare blue bars (empirical features) to red bars (simulated features) and to yellow bars (all features). In particular, the prediction performance can be enhanced by complementing the empirical data by the simulated connectomes when comparing with the case of purely empirical features. Interestingly, the classification performance can further be improved if the features from both atlases were merged.

We can also address how the prediction performance varies depending on the filtering conditions (Fig. 9D). The effect of the temporal filtering was prominent for the empirical features for the Schaefer atlas, where the performance was significantly increased for the LF condition compared to the others (Fig. 9D, the Emp. column for the Schaefer atlas). On the other hand, the HF condition showed low performances, especially, for the specificity of the empirical features (Fig. 9B and D), where the LF filtering again seems to be a beneficial condition for PD prediction.

Summarizing, the temporal filtering conditions influenced the model performance, and the LF bandpass filtering resulted in the most effective PD prediction relying on the empirical and simulated connectome relationships. In other words, complementing empirical data with simulation results using LF filtering and involving multiparcellation (concatenating both atlases) is advisable for PD classification.

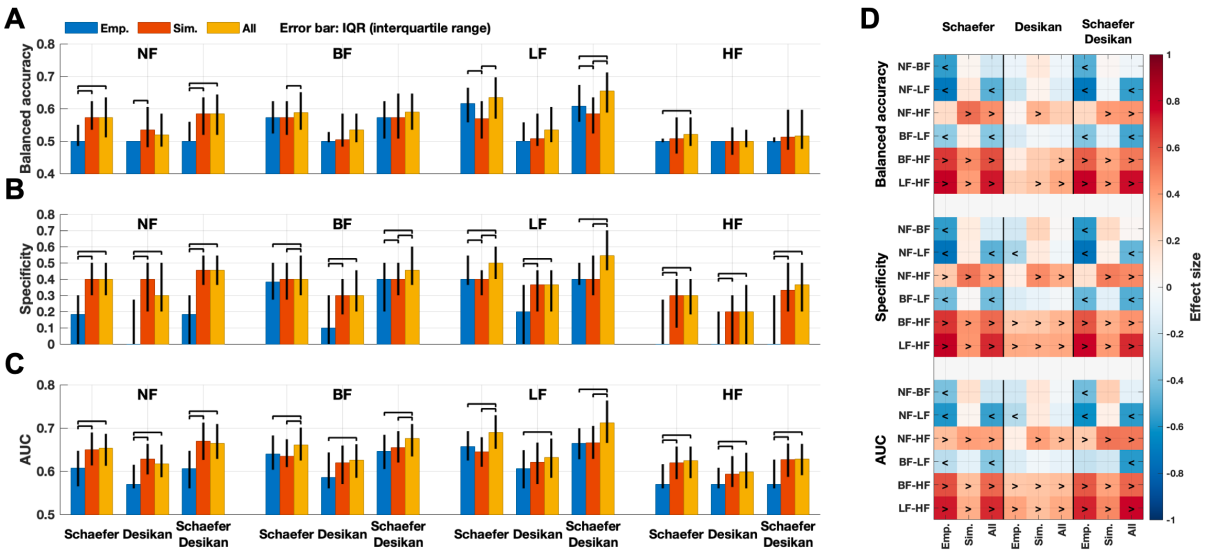


Figure 9: Summary of the performance of PD classification using the three different feature conditions: empirical features (blue bars), simulated features (red bars), and all features (yellow bars) after incorporating the cross-validated age controlling and the behavioral model fitting (Fig. 4). Median values of the balanced accuracy, specificity and area-under-curve (AUC) of the receiver operating characteristics (ROC) curves for all considered parcellations and filtering conditions are shown in each panel for (A) balanced accuracy and (B) specificity and (C) AUC. The error bars indicate the interquartile range across iterations of the outer loop of the nested cross-validation procedure (Fig. 4). The black lines connecting two conditions indicate significantly different performance between feature conditions. (D) Effect sizes between filtering conditions for each feature condition. The signs '<' and '>' indicate which condition is significantly larger than the other. For example, '<' sign for 'NF-LF' indicated on the vertical axes means NF < LF in the prediction performance for a given feature condition indicated on the horizontal axes. The Wilcoxon signed-rank two-tail test was used for comparisons across predictor sets and filtering conditions (Bonferroni corrected statistics).

5. Conclusion, next steps

We developed a pipeline that can be used for pre-processing of the neuroimaging MRI data in the native spaces of the brain images without involving a complex nonlinear transformation to the standard MNI152 space. This is supposed to reduce the complex manipulation with the data, make it applicable to processing of difficult clinical data and facilitate the investigation of the inter-individual variability. We applied this pipeline for processing of the clinical MRI data of PD patients used for the derivation and validation of the whole-brain dynamical model in particular. We showed that the obtained empirical and simulated data (structural and functional connectomes) can successfully be employed for differentiation (classification) of the patients from healthy subjects, where the performance can be improved when the empirical connectomes are complemented by simulation results. We also demonstrated that selected parameters of the data processing like temporal filtering or brain parcellation can strongly influence the obtained results, e.g., the prediction performance. It is therefore important to test the drawn conclusions for varying data processing conditions, and the developed pipeline is an appropriate toll for this and provides enough freedom for such investigations.

The pipeline was also used to calculate the empirical BOLD signals, eSC and eFC for a sample of subjects from the 1000BRAINS cohort (Caspers *et al.*, 2014), which we made available for neuroimaging and modeling communities via EBRAINS platform (Jung, Eickhoff and Popovych, 2022).

The pipeline is available in a dedicated project on GitHub:
https://jugit.fz-juelich.de/inm7/public/vbc-mri-pipeline.

The repository contains all code and documentation necessary to apply the pipeline to the pre-processing of the neuroimaging data and extraction of the signals including the whole-brain tractography, SC, BOLD signals and FC.

The next steps may include a systematic evaluation of the impact of personalized data processing and the extent of model personalization on the investigation of inter-individual variability and prediction performance of behavioral characteristics of individual subjects from the neuroimaging empirical and simulated data in health and disease.

6. References

Aquino, K.M. *et al.* (2022) 'On the intersection between data quality and dynamical modelling of large-scale fMRI signals', *NeuroImage*, 256, p. 119051. Available at: https://doi.org/10.1016/j.neuroimage.2022.119051.

Botvinik-Nezer, R. *et al.* (2020) 'Variability in the analysis of a single neuroimaging dataset by many teams', *Nature*, 582(7810), pp. 84–88.

Caminiti, R. *et al.* (2013) 'Diameter, Length, Speed, and Conduction Delay of Callosal Axons in Macaque Monkeys and Humans: Comparing Data from Histology and Magnetic Resonance Imaging Diffusion

Tractography', *Journal of Neuroscience*, 33(36), pp. 14501–14511. Available at: https://doi.org/10.1523/JNEUROSCI.0761-13.2013.

Caspers, S. *et al.* (2014) 'Studying variability in human brain aging in a population-based German cohort-rationale and design of 1000 BRAINS', *Front. Aging Neurosci.*, 6, p. 149. Available at: https://doi.org/10.3389/fnagi.2014.00149.

Ciric, R. *et al.* (2017) 'Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity', *Neuroimage*, 154, pp. 174–187. Available at: https://doi.org/10.1016/j.neuroimage.2017.03.020.

Cox, R.W. (1996) 'AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages', *Comput. Biomed. Res.*, 29(3), pp. 162–173. Available at: https://doi.org/10.1006/cbmr.1996.0014.

Dale, A.M., Fischl, B. and Sereno, M.I. (1999) 'Cortical surface-based analysis. I. Segmentation and surface reconstruction', *Neuroimage*, 9, pp. 179–194.

Deco, G. *et al.* (2017) 'The dynamics of resting fluctuations in the brain: metastability and its dynamical cortical core', *Sci. Rep.*, 7, p. 3095. Available at: https://doi.org/10.1038/s41598-017-03073-5.

Deco, G. *et al.* (2019) 'Awakening: Predicting external stimulation to force transitions between different brain states', *Proc. Natl. Acad. Sci. U. S. A.*, 116(36), pp. 18088–18097. Available at: https://doi.org/10.1073/pnas.1905534116.

Deco, G. *et al.* (2021) 'Dynamical consequences of regional heterogeneity in the brain's transcriptional landscape', *Science Advances*, 7(29), p. eabf4752. Available at: https://doi.org/10.1126/sciadv.abf4752.

Demirtas, M. *et al.* (2019) 'Hierarchical Heterogeneity across Human Cortex Shapes Large-Scale Neural Dynamics', *Neuron*, 101(6), pp. 1181–+. Available at: https://doi.org/10.1016/j.neuron.2019.01.017.

Desikan, R.S. *et al.* (2006) 'An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest', *Neuroimage*, 31(3), pp. 968–980. Available at: https://doi.org/10.1016/j.neuroimage.2006.01.021.

Domhof, J.W.M. *et al.* (2021) 'Parcellation-induced variation of empirical and simulated brain connectomes at group and subject levels', *Network Neuroscience*, 5(3), pp. 798–830. Available at: https://doi.org/10.1162/netn_a_00202.

Domhof, J.W.M., Eickhoff, S.B. and Popovych, O.V. (2022) 'Reliability and subject specificity of personalized whole-brain dynamical models', *NeuroImage*, 257, p. 119321. Available at: https://doi.org/10.1016/j.neuroimage.2022.119321.

Evans, A.C. *et al.* (2012) 'Brain templates and atlases', *NeuroImage*, 62(2), pp. 911–922. Available at: https://doi.org/10.1016/j.neuroimage.2012.01.024.

Falcon, M.I., Jirsa, V. and Solodkin, A. (2016) 'A new neuroinformatics approach to personalized medicine in neurology: The Virtual Brain', *Curr. Opin. Neurol.*, 29(4), pp. 429–436. Available at: https://doi.org/10.1097/WCO.0000000000000344.

Fawcett, T. (2006) 'An introduction to ROC analysis', *Pattern Recognition Letters*, 27(8), pp. 861–874. Available at: https://doi.org/10.1016/j.patrec.2005.10.010.

Fischl, B. *et al.* (2004) 'Automatically Parcellating the Human Cerebral Cortex', *Cereb. Cortex*, 14(1), pp. 11–22. Available at: https://doi.org/10.1093/cercor/bhg087.

Ghosh, A. *et al.* (2008) 'Noise during Rest Enables the Exploration of the Brain's Dynamic Repertoire', *PLoS Comput. Biol.*, 4(10), p. e1000196. Available at: https://doi.org/10.1371/journal.pcbi.1000196.

Hastie, T., Tibshirani, R. and Friedman, J. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics)*. New York, NY: Springer.

Honey, C.J. *et al.* (2009) 'Predicting human resting-state functional connectivity from structural connectivity', *Proc. Natl. Acad. Sci. U. S. A.*, 106(6), pp. 2035–2040. Available at: https://doi.org/10.1073/pnas.0811168106.

Jansen, B.H. and Rit, V.G. (1995) 'Electroencephalogram and Visual-evoked Potential Generation In A Mathematical-model of Coupled Cortical Columns', *Biol. Cybern.*, 73(4), pp. 357–366. Available at: https://doi.org/10.1007/BF00199471.

Jirsa, V. *et al.* (2023) 'Personalised virtual brain models in epilepsy', *The Lancet Neurology* [Preprint]. Available at: https://doi.org/10.1016/S1474-4422(23)00008-X.

Jirsa, V.K. *et al.* (2017) 'The Virtual Epileptic Patient: Individualized whole-brain models of epilepsy spread', *Neuroimage*, 145, pp. 377–388. Available at: https://doi.org/10.1016/j.neuroimage.2016.04.049.

Jülich Supercomputing Centre (2021) 'JURECA: Data Centric and Booster Modules implementing the Modular Supercomputing Architecture at Jülich Supercomputing Centre', *Journal of large-scale research facilities*, 7, p. A182. Available at: https://doi.org/10.17815/jlsrf-7-182.

Jung, K. *et al.* (2020) 'Initial version of full-scope containerized pipeline developed', *VirtualBrainCloud, ID 82642, H2020*, Deliverable 3.1.

Jung, K. *et al.* (2021) 'Most appropriate RS-processing pipelines for TVB modelling finished', *VirtualBrainCloud, ID 82642, H2020*, Deliverable 3.2.

Jung, K. *et al.* (2022) 'Whole-brain dynamical modelling for classification of Parkinson's disease', *Brain Communications*, 5(1), p. fcac331. Available at: https://doi.org/10.1093/braincomms/fcac331.

Jung, K., Eickhoff, S.B. and Popovych, O.V. (2021) 'Tractography density affects whole-brain structural architecture and resting-state dynamical modeling', *NeuroImage*, 237, p. 118176. Available at: https://doi.org/10.1016/j.neuroimage.2021.118176.

Jung, K., Eickhoff, S.B. and Popovych, O.V. (2022) 'Parcellation-based structural and resting-state functional whole-brain connectomes of 1000BRAINS cohort (v1.1) [Data set]', *EBRAINS* [Preprint]. Available at: https://doi.org/10.25493/8XY5-BH7.

Lindquist, M. (2020) 'Pipeline choices alter neuroimaging findings', *Nature*, 582(7810), pp. 36–37. Available at: https://doi.org/10.1038/d41586-020-01282-z.

Lindquist, M.A. *et al.* (2019) 'Modular preprocessing pipelines can reintroduce artifacts into fMRI data', *Hum. Brain Mapp.*, 40(8), pp. 2358–2376. Available at: https://doi.org/10.1002/hbm.24528.

Lopes da Silva, F.H. *et al.* (1974) 'Model of brain rhythmic activity', *Kybernetik*, 15(1), pp. 27–37.

Maffei, C. *et al.* (2022) 'Insights from the IronTract challenge: Optimal methods for mapping brain pathways from multi-shell diffusion MRI', *NeuroImage*, 257, p. 119327. Available at: https://doi.org/10.1016/j.neuroimage.2022.119327.

Maier-Hein, K.H. *et al.* (2017) 'The challenge of mapping the human connectome based on diffusion tractography', *Nat. Commun.*, 8(1), p. 1349.

Moran, R., Pinotsis, D.A. and Friston, K.J. (2013) 'Neural masses and fields in dynamic causal modeling', *Front. Comput. Neurosci.*, 7, p. 57. Available at: https://doi.org/10.3389/fncom.2013.00057.

More, S. *et al.* (2023) 'Brain-age prediction: A systematic comparison of machine learning workflows', *NeuroImage*, 270, p. 119947. Available at: https://doi.org/10.1016/j.neuroimage.2023.119947.

Parkes, L. *et al.* (2018) 'An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI', *NeuroImage*, 171, pp. 415–436. Available at: https://doi.org/10.1016/j.neuroimage.2017.12.073.

Popovych, O.V. *et al.* (2019) 'What Can Computational Models Contribute to Neuroimaging Data Analytics?', *Front. Syst. Neurosci.*, 12, p. 68. Available at: https://doi.org/10.3389/fnsys.2018.00068.

Popovych, O.V. *et al.* (2021) 'Inter-subject and inter-parcellation variability of resting-state whole-brain dynamical modeling', *NeuroImage*, 236, p. 118201. Available at: https://doi.org/10.1016/j.neuroimage.2021.118201.

Qi, S.L. *et al.* (2015) 'The influence of construction methodology on structural brain network measures: A review', *J. Neurosci. Methods*, 253, pp. 170–182. Available at: https://doi.org/10.1016/j.jneumeth.2015.06.016.

Salimi-Khorshidi, G. *et al.* (2014) 'Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers', *Neuroimage*, 90, pp. 449–468. Available at: https://doi.org/10.1016/j.neuroimage.2013.11.046.

Satterthwaite, T.D. *et al.* (2013) 'An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data', *Neuroimage*, 64, pp. 240–256. Available at: https://doi.org/10.1016/j.neuroimage.2012.08.052.

Schaefer, A. *et al.* (2018) 'Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI', *Cereb. Cortex*, 28(9), pp. 3095–3114. Available at: https://doi.org/10.1093/cercor/bhx179.

Smith, S.M. *et al.* (2004) 'Advances in functional and structural MR image analysis and implementation as FSL', *Neuroimage*, 23, pp. S208–S219. Available at: https://doi.org/10.1016/j.neuroimage.2004.07.051.

Tibshirani, R. (1996) 'Regression shrinkage and selection via the Lasso', *Journal of the Royal Statistical Society Series B-methodological*, 58(1), pp. 267–288. Available at: https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

Tournier, J.D. *et al.* (2019) 'MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation', *Neuroimage*, 202, p. UNSP 116137. Available at: https://doi.org/10.1016/j.neuroimage.2019.116137.

Tournier, J.-D., Calamante, F. and Connelly, A. (2010) 'Improved probabilistic streamlines tractography by 2nd order integration over fibre orientation distributions', *Proc. Intl. Soc. Mag. Reson. Med.*, 18, p. 1670.

Tustison, N.J. *et al.* (2014) 'Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements', *Neuroimage*, 99, pp. 166–179. Available at: https://doi.org/10.1016/j.neuroimage.2014.05.044.

Wang, P. *et al.* (2019) 'Inversion of a large-scale circuit model reveals a cortical hierarchy in the dynamic resting human brain', *Science Advances*, 5(1). Available at: https://doi.org/10.1126/sciadv.aat7854.

Worsley, K.J. *et al.* (1992) 'A 3-dimensional Statistical-analysis For CBF Activation Studies In Human Brain', *J. Cereb. Blood Flow Metab.*, 12(6), pp. 900–918. Available at: https://doi.org/10.1038/jcbfm.1992.127.