Effective workflow from multi-modal MRI data to model-based prediction

Kyesam Jung[1,2], Kevin J. Wischnewski[1,2,3], Simon B. Eickhoff[1,2], Oleksandr V. Popovych[1,2*]

[1]Institute of Neurosciences and Medicine - Brain and Behaviour (INM-7), Research Centre Jülich, Jülich, Germany

[2]Institute of Systems Neuroscience, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

[3]Institute of Mathematics, Faculty of Mathematics and Natural Sciences, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

***Corresponding author:**
Oleksandr V. Popovych
o.popovych@fz-juelich.de

Keywords: brain MRI, whole-brain modeling, parameter optimization, machine learning, classification, prediction.

Abstract

Predicting human behavior from neuroimaging data remains a complex challenge in neuroscience. To address this, we propose a systematic and multi-faceted framework that incorporates a model-based workflow using dynamical brain models. This approach utilizes multi-modal MRI data for brain modeling and applies the optimized modeling outcome to machine learning. We demonstrate the performance of such an approach through several examples such as sex classification and prediction of cognition or personality traits. We in particular show that incorporating the simulated data into machine learning can significantly improve the prediction performance compared to using empirical features alone. These results suggest considering the output of the dynamical brain models as an additional neuroimaging data modality that complements empirical data by capturing brain features that are difficult to measure directly. The discussed model-based workflow can offer a promising avenue for investigating and understanding inter-individual variability in brain-behavior relationships and enhancing prediction performance in neuroimaging research.

1 Introduction

Since the concept of the human connectome (Sporns et al., 2005) was proposed almost two decades ago, whole-brain connectivity derived from neuroimaging data has been employed to address questions across various topics including cognitive functions (Sporns, 2014) and brain disorders (Fornito et al., 2015). An important characteristic of magnetic resonance imaging (MRI) data is their multi-modality that has enabled the researchers to view the brain connectivity from multiple perspectives of structural and functional connections between brain regions (Park and Friston, 2013). For instance, diffusion-weighted MRI (dwMRI) can be used to investigate the microstructure of white matter as well as to estimate axonal fibers connecting brain regions via tracking streamlines. The latter are interpreted as anatomical connectivity and also referred to as structural connectivity (SC) (Wiegell et al., 2000). On the other hand, resting-state functional MRI (rsfMRI) provides a way to obtain the degree of similarity of activity patterns between brain regions over time, representing

functional connectivity (FC) (Biswal et al., 1995). These two connectivities (SC and FC), constructed in different ways, evidently have different meanings and interpretations and, accordingly, can be utilized in several ways. For example, temporal changes of brain activity will be represented in FC (Schaefer et al., 2018; Tavor et al., 2016), while anatomical white matter changes in long-term periods can be revealed through SC (Damoiseaux, 2017; Zhao et al., 2015). Furthermore, comparing these connectomes and calculating their similarity led to the notion of the brain structure-function relationship as a possible methodological approach to explore the interdependence between structure and function of the human brain (Suarez et al., 2020). However, the strength of the structure-function relationship is usually relatively low, might depend on many factors including brain parcellation into separate regions, and its mechanism is still unclear (Batista-Garcia-Ramo and Fernandez-Verdecia, 2018; Messe, 2020).

Integration of model-based approaches into whole-brain connectome research can expand the scope of investigation to understand the brain. The models can, for example, be used to generate simulated FC that together with the fitted model parameters can serve as an additional data modality. This approach provides further attributes that characterize brain dynamics in great detail (Popovych et al., 2019). In the framework of the whole-brain dynamical modeling, the models were suggested as a possible mediator between brain structure and function, where the empirical SC and FC are used for the model derivation and validation (Honey et al., 2009). A natural output of such models is the relationship between simulated and empirical connectomes, which can in particular be used for investigation of the brain structure-function relationship. One of the main advantages of a model-based approach is a great freedom of considering many *in silico* models, ranges of their parameters and the respective brain activity that may be hidden in a few *in vivo* measurements (Pathak et al., 2022). The modeling results may thus contain the information going well beyond that of empirical data and can also validate the biophysical properties of the brain that have been discovered so far or even provide new insights (Havlicek et al., 2015). In addition, with increased power of high-performance computational clusters, a variety of experimental and data-processing conditions can be simulated including modeling of virtual brain interventions in order to identify and test the optimal conditions and parameters, which is hardly possible *in vivo* (Jirsa et al., 2023; Jung et al., 2024).

In this report we suggest a framework that advances the applicability of the model-based approach for neuroimaging research and outline an effective workflow for applying simulated data to machine-learning analysis (Figure 1). With the suggested framework, we illustrate a few examples of model-based machine learning applied to classification and prediction by employing simulated data that is beneficial for the performance compared to using solely empirical neuroimaging data. We consider connectome relationships as features for predictions. Purely empirical connectome relationship (empirical SC vs. empirical FC) is used as empirical feature and simulated connectome relationship (empirical FC vs. simulated FC) is used as simulated feature which involves simulated data. We then compare the cases of using empirical features, simulated features, and their combination. Such an enhancement of model applicability might be of relevance, for example, in medical research, where the classification of subjects into patients and healthy individuals might be well assisted by models (Jung et al., 2023).

The simulated and empirical connectome relationships exhibit weak similarity between each other with low or even negative correlations across individuals (Popovych et al., 2021). This indicates that the simulated data showing stronger relationships might contain additional and possibly useful information for the machine-learning prediction analysis if included as features. Along this line, we recently reported that model-based simulated connectomes show higher correlation with clinical

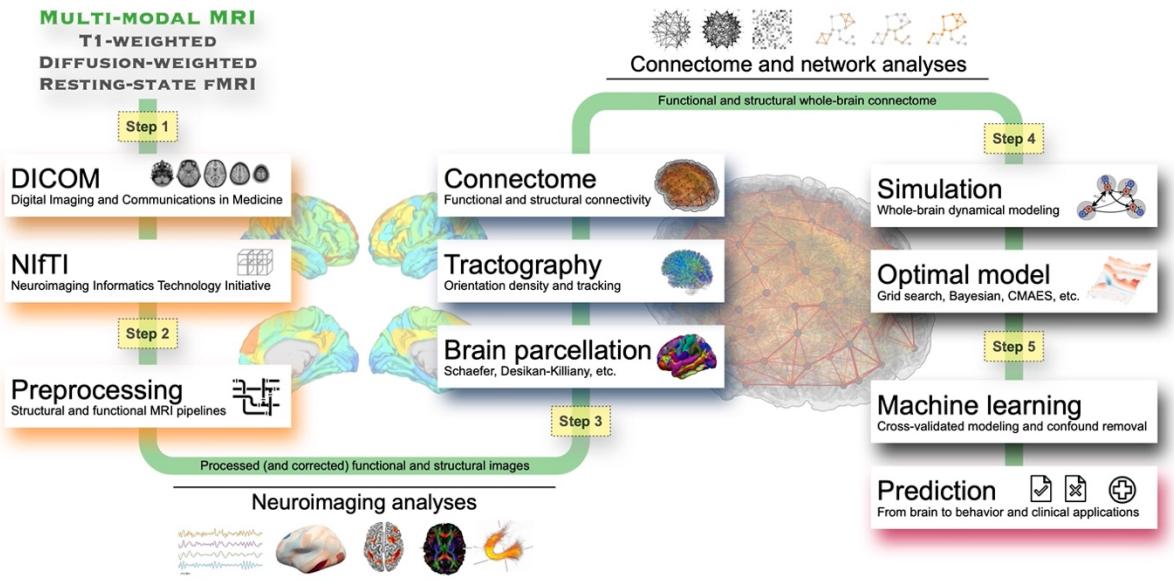

Figure 1 A workflow for model-based prediction research. It can be divided into five steps. The first step is the acquisition of multi-modal MRI data (T1-weighted, diffusion-weighted, and resting-state functional MRI). The second step is preprocessing the acquired MRI data, which can be used for neuroimaging analysis. The third step is to calculate whole-brain tractography and apply a brain parcellation to reconstruct the whole-brain structural and functional connectomes. The fourth step involves whole-brain dynamical modeling including parameter optimization, where the optimal whole-brain model is identified and used to simulate and investigate the brain dynamics *in silico*. The final step is applying the simulated data for machine-learning analyses.

scores than that of empirical connectomes, thereby outperforming the latter in this respect (Jung et al., 2024).

Including simulated data as an additional data modality in the mentioned studies was motivated by several previous results demonstrating distinct properties of simulated and empirical data in spite of the fact that the models were fitted to the latter. One of the important issues in brain MRI research is the low reliability of findings. This problem has particularly been brought up in the resting-state functional imaging of the whole-brain connectome (Andellini et al., 2015). However, model-based connectome relationships can offer relatively good reliability and improved subject specificity compared to a fair reliability and low specificity of empirical functional data (Domhof et al., 2022a). Enhanced data reliability might also be important for the prediction analysis (Chen et al., 2023). Therefore, applying model-based simulated connectome features, which exhibit distinct patterns along with enhanced reliability and inter-subject variability, to machine learning could lead to consistent results and potentially improved prediction performance as we illustrate in a few examples in this study.

2 Methods

In the suggested workflow (Figure 1), the first step of the model-based approach required multi-modal MRI data, including T1-weighted, dwMRI, and rsfMRI scans. The second step involved processing the MRI data, which included inhomogeneous field/motion corrections, tissue segmentation, cortical rendering, and image registration. In the next step, we applied brain parcellation schemes and computed the whole-brain connectome, including both SC and FC. The fourth step consisted of selecting a dynamical model for the research objectives and optimizing

model parameters by fitting simulated data to empirical data. Finally, machine learning was performed using features derived from both the measured and model-based data. We utilized empirical human connectomes, *i.e.*, SC derived from the white-matter fiber tracking and FC calculated by Pearson's correlation between resting-state Blood Oxygenation Level-Dependent (BOLD) signals of parcellated brain regions. Subsequently, simulated BOLD signals were generated via the considered whole-brain model informed by empirical neuroimaging data and validated by parameter optimization, where the model showed the highest similarity, *i.e.*, Goodness-of-Fit (GoF) between simulated and empirical FCs, and GoF is considered as simulated features which involved simulated data. Then the connectome relationships between empirical and simulated brain connectomes were calculated by Pearson's correlation between empirical SC (eSC), empirical FC (eFC) and simulated FC (sFC). These connectome relationships were considered as brain features and utilized by machine-learning techniques for prediction of behavioral characteristics of individual subjects, for instance, sex classification or prediction of cognitive scores and five personality traits. The subsections below describe details of each step in the workflow.

2.1 Multi-modal MRI data: Step 1

The current study used the Human Connectome Project (HCP) S1200 young adult dataset (Van Essen et al., 2013) including 270 unrelated subjects of 142 females and 128 males with ages in 28.5 ± 3.5 (mean ± standard deviation) years. HCP data were acquired using MRI protocols approved by the Washington University institutional review board (IRB #20124036). Informed consent was obtained from all subjects. Anonymized data are publicly available (https://db.humanconnectome.org). Multi-modal MRI data including T1-weighted MRI (T1w), rsfMRI, and dwMRI were used in the current workflow.

2.2 MRI processing: Step 2

A pipeline of MRI processing that consists of structural and functional modules was applied to the multi-modal MRI data, *i.e.*, T1w, rsfMRI, and dwMRI. The pipeline is available via a public repository (https://jugit.fz-juelich.de/inm7/public/vbc-mri-pipeline). The pipeline uses functions in AFNI (Cox, 1996), ANTs (Tustison et al., 2010), FreeSurfer (Dale et al., 1999), FSL (Smith et al., 2004), MRtrix3 (Tournier et al., 2019), and Connectome Workbench (Marcus et al., 2011). The entire MRI pipeline was aiming at obtaining the whole-brain human connectome. The Schaefer atlas with 100 parcels (Schaefer et al., 2018) and the Harvard-Oxford atlas with 96 parcels (Desikan et al., 2006) were utilized in this study for brain parcellation in the MNI space.

Resting-state BOLD signals were extracted from the rsfMRI processed with FMRIB's ICA-based X-noiseifier (ICA-FIX) provided by a pipeline of the HCP repository (Griffanti et al., 2014). There were four rsfMRI sessions (1200 volumes, TR = 720 ms) conducted over two different days and consisting of two different phase-encoding directions on each day. In order to obtain the mean regional BOLD signals, the brain was parcellated according to a given brain atlas, and the voxel-wise BOLD signals in every brain region were averaged over all voxels of the region at each time point. A concatenated BOLD signal was then generated by combining all four z-scored BOLD signals from the four rsfMRI sessions.

For the whole-brain tractography (WBT) calculation, response functions were estimated for spherical deconvolution using the constrained deconvolution algorithm (Tournier et al., 2007). Fiber oriented distributions (FODs) were estimated from the dwMRI using spherical deconvolution (Jeurissen et al., 2014), and WBT including 10 million streamlines was created through the fiber tracking by second-order integration over the FOD by a probabilistic algorithm (Tournier et al., 2010).

2.3 Whole-brain connectome: Step 3

For eFC, Pearson's correlation coefficients between the concatenated regional BOLD signals of each pair of brain regions of the considered brain parcellation were calculated, resulting in the whole-brain resting-state FC. For eSC, the atlases were transformed from the MNI space to the native space of dwMRI. Following the transformation, labeled voxels masked within gray matter were selected for seed and target regions and applied to the WBT. Subsequently, streamlines connecting the seed and target regions were selected for each pair of brain regions, and we ultimately obtained the whole-brain SC matrices including streamline counts and average path lengths of them. With eFC and eSC, we can apply connectome and graph-theoretical network properties for further analyses.

2.4 Mathematical whole-brain model and model fitting: Step 4

We simulated a whole-brain dynamical model of N coupled phase oscillators (Kuramoto, 1984; Yeung and Strogatz, 1999). Their temporal dynamics can be described by the following set of differential equations:

$$\dot{\varphi}_i(t) = 2\pi f_i + \frac{C}{N}\sum_{j=1}^{N} k_{ij} \sin\left(\varphi_j\left(t-\tau_{ij}\right) - \varphi_i(t)\right) + \sigma\eta_i,\ i = 1,2,\cdots,N. \quad (1)$$

The number of oscillators N corresponds to the number of brain regions as defined by a given brain atlas, where $\varphi_i(t)$ models the phase of the mean BOLD signal of the corresponding region, and the simulated BOLD was calculated as $\sin\varphi_i(t)$. C is a global coupling which scales the level of couplings of the whole-brain network. η_i is an independent noise perturbing oscillator i, which is sampled from a random uniform distribution from the interval [-1,1]. $\sigma = 0.3$ denotes the noise intensity. The natural frequencies f_i were estimated from the empirical data as frequencies of the maximal spectral peaks (restricted to the frequency range from 0.01 Hz to 0.1 Hz) of the empirical BOLD signals of the corresponding brain regions. k_{ij} stands for the coupling strength between oscillators i and j, and τ_{ij} approximates the time delay of the signal propagation between oscillators i and j. They were calculated from the streamline counts and average path-length matrices and determined by the following equations:

$$k_{ij} = \frac{w_{ij}}{\langle W\rangle}, \quad (2)$$

where w_{ij} is the number of streamlines between the i^{th} and j^{th} parceled region and $\langle W\rangle$ is an average number of streamlines over all connections except self-connections. The delays we calculated as

$$\tau_{ij} = \frac{L_{ij}}{\langle V\rangle} = \tau L_{ij}, \quad (3)$$

where L_{ij} is the average path length of the selected streamlines connecting the i^{th} and j^{th} region, and τ is a global delay, which is a reciprocal of an average speed of signal propagation $\langle V\rangle$ through the whole-brain network. The time step of the numerical integration of Eq. 1 by the stochastic Heun method was fixed to 0.04 s, and the simulated signals were generated for 3500 seconds after skipping 500 seconds of the initial transient. The simulated BOLD signals and the corresponding sFC matrices were calculated from the phases down-sampled to TR = 0.72 s, which is the repetition time of the current rsfMRI acquisition.

The considered mathematical model (Eq. 1) has two global parameters: global coupling C and global delay τ. These were optimized within the ranges $C \in [0,1]$ and $\tau \in [0,100]$ with the aim to maximize Pearson's correlation between eFC and sFC. We will refer to this setting with two free parameters as the low-dimensional parameter optimization. Further, we also considered the model fitting in high-dimensional spaces of model parameters, where the noise intensity σ and additional local (regional) parameters of natural frequencies f_i (see Eq. 1) of the brain regions were included in the optimization process. For both scenarios, we applied the Covariance Matrix Adaptation Evolution Strategy (CMAES) for parameter optimization (Hansen and Ostermeier, 1996; Hansen and Ostermeier, 2001). At the parameter optimization by CMAES, the number of particles sampled per generation was chosen as $\lambda = 24$ based on the previous study (Wischnewski et al., 2022). To account for a possible result variability of such a parameter optimization, we performed CMAES 30 times for every subject with different initial conditions and selected the optimal model parameters corresponding to the largest GoF for further analyses.

2.5 Machine learning for model-based prediction: Step 5

To illustrate the benefits in machine learning via including simulated data into the features, we used the empirical connectome relationship (Pearson's correlation between eFC and eSC) and the simulated connectome relationship (Pearson's correlation between eFC and sFC, that is the best GoF of the model to eFC). The empirical and simulated connectome relationships were used for sex classification (n=270) as well as the prediction of cognitive composite scores (n=268, 2 subjects had no cognitive scores) and personality traits (n=269, 1 subject had no data on personality traits) by using machine learning. We also merged the two features (empirical and simulated) and used them for the same machine-learning approach for the classification and prediction analyses. Afterward, we compared the performances with feature conditions of empirical only, simulated only, and merged features.

For the sex classification, we used a nested 5-fold cross-validation (CV) scheme, where every outer CV loop (k=5) included the embedded 5 inner loops as a nested CV (inner 5-fold CV) for training the prediction model using hyperparameter optimization. The training procedure started with a random splitting of the entire subject sample into 5 equally-sized subgroups while maintaining the ratio of female/male in each subgroup. Subsequently, in every outer loop, one subgroup was selected after another as a testing set, and the other 4 subgroups were united into a training set. In the inner loop with the training set, we performed a confound removal (CR) to remove the effect of brain volumes on the sex classification from the features, *i.e.*, connectome relationships. For this we used the univariate linear regression with the brain volumes (sum of cortical, subcortical and white matter volumes), estimated the parameters of the linear model, and z-scored the obtained residuals across subjects in the training set. Finally, we used a logistic regression with an L2 penalty for the training in the nested CV, and the regularizing parameter was optimized by the Limited memory Broyden-Fletcher-Goldfarb-Shanno algorithm (L-BFGS). After the training in the nested CV, the best model was selected and applied to the testing set to classify the unseen subjects as females or males. Here, the respective CR and z-scoring with parameters obtained for the training set were applied beforehand. Such a CV-CR scheme prevents a data leakage, where no information from the testing set was used during the training (More et al., 2021). We repeated this prediction process 100 times for different random subject splits into 5 subgroups. Finally, we calculated a balanced accuracy using predicted probability and target variables (female or male).

The CV-CR scheme (5-fold nested CV and CR with brain volumes and ages) was used for predicting the total cognitive function composite score (*CogTotalComp_Unadj*) as general intelligence acquired

in the NIH toolbox (https://www.nihtoolbox.org) and also the Five-Factor Model (McCrae and Costa, 2004) known as the big five personality traits including openness, conscientiousness, extraversion, agreeableness and neuroticism. The entire group was split into training and testing sets as before while keeping the shape of scores' distributions for the training and testing sets for an efficient and reliable CV performance (Budka and Gabrys, 2013). Here, the training and testing sets were created by stratifying the subjects among 7 subgroups balanced within 7 intervals of each target score (cognitive and personality traits) in order to mimic the distribution of the entire cohort. We applied a ridge regression with L2 penalty for training the prediction model, and the optimal regularizing parameter values were selected among several discrete values of 10^{-6}, 10^{-5}, ..., 10^{5}, and 10^{6}. The model with the optimal regularizing parameter was selected, which demonstrated the highest Pearson's correlation coefficient between predicted values and the target scores across subjects in the training set. Consequently, the best model trained through the nested CV on the training set was applied to the testing set to predict the target scores of unseen subjects. We repeated this CV-CR prediction 100 times with each iteration having different stratified subject splits. Finally, we calculated Pearson's correlation between predicted and measured scores for prediction performance.

For the machine-learning approach, we used Python version 3.11 with modules including Scikit-learn version 1.3.0 (Pedregosa et al., 2011), NumPy version 1.24.4 (Harris et al., 2020), and SciPy version 1.11.1 (Virtanen et al., 2020).

2.6 Statistical analysis

Effect sizes of the difference between prediction performance of feature conditions were calculated by the Rosenthal formula (Rosenthal et al., 1994) which used z-statistics also utilized for calculation of the *p*-values of Wilcoxon rank-sum two-tail test. Bonferroni correction was applied for corrected *p*-values in multiple comparisons. Principal component analysis (PCA) was performed for features, and loadings of each principal component were estimated. All statistical tests and data visualizations were performed in MATLAB (R2024a; MathWorks).

3 Results

By leveraging empirical whole-brain connectomes for the whole-brain dynamical modeling, we successfully generated sFC that can be used alongside the eFC. This allows us to characterize whole-brain dynamics through connectome relationships, highlighting inter-individual variability. Both empirical and simulated connectome relationships can be considered as individual features of whole-brain dynamics and used to classify subjects into different categories or predict their behavioral characteristics using machine-learning approaches. Here, simulated data can complement empirical neuroimaging data or serve as stand-alone features, which can improve the prediction performance. As an example of the proposed framework, we demonstrate that the modeling results can effectively classify subjects by sex (male vs. female) and predict their general intelligence as well as personality traits, showing improved performance compared to using empirical features alone.

3.1 Model-based connectome relationships as leveraged feature information

To calculate brain connectomes and their relationships, we utilized two brain parcellation schemes. One is the Schaefer atlas with 100 regions (Schaefer et al., 2018), where the cortical surface was divided based on functional characteristics of the brain. The other one is the Harvard-Oxford atlas with 96 regions (Desikan et al., 2006), where structural brain characteristics were used for cortical parcellation. The connectome relationships as given by the Pearson's correlation between the respective connectivity matrices were calculated for every subject leading to distributions of their

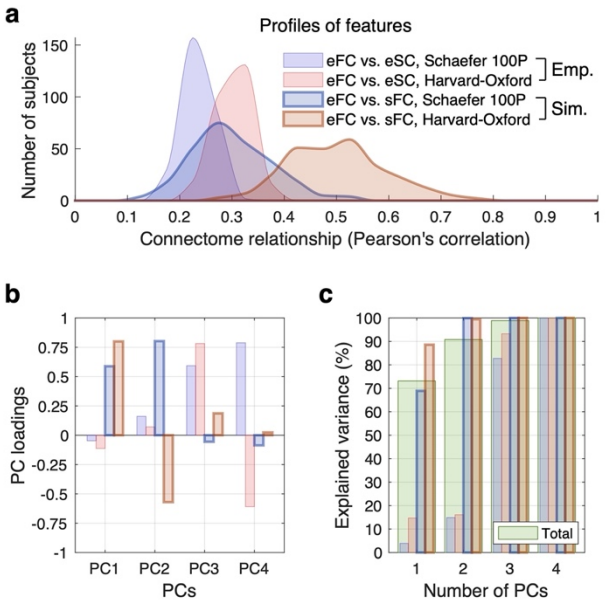

Figure 2 Empirical and simulated features (connectome relationship) for machine learning. **(a)** Feature distributions across individual subjects for two brain parcellations given by the Schaefer atlas (100 regions) and the Harvard-Oxford atlas (96 regions) as indicated in the legend. **(b-c)** Principal component analysis (PCA) of the feature variability across 4 feature conditions: empirical, simulated and the two considered brain parcellations. The loadings and the fractions of the explained variance by different principal components are illustrated in plots (b) and (c), respectively. The color and line schemes are as in plot (a). The cumulative explained variance across all conditions is depicted in plot (c) by bars in light green color. Abbreviations: eFC: empirical functional connectivity, eSC: empirical structural connectivity, sFC: simulated FC, PC: principal component.

values for a given subject cohort (Figure 2a). We observed that the two considered parcellation schemes yielded different distributions for the empirical structure-function connectome relationships corr(eFC, eSC) (Figure 2a, Emp.). In particular, the Harvard-Oxford atlas supported a somewhat stronger structure-function relationship as compared to the Schaefer atlas. Similarly, the simulated connectome relationships corr(eFC, sFC) also produced different ranges of values depending on the parcellation scheme applied (Figure 2a, Sim.). The considered functional connectome relationship involving the simulated data sFC is important in the brain modeling and frequently used in the literature as a measure of the best fitted model to empirical functional data (Honey et al., 2009). We observed that the simulated connectome relationships (eFC vs. sFC) exhibited a much broader spread as compared to the empirical connectome relationships (eFC vs. eSC) including an enhanced inter-individual variability when the simulated data were involved (interquartile ranges, empirical vs. simulated features: 0.033 vs. 0.104 for the Schaefer atlas, and 0.044 vs. 0.136 for the Harvard-Oxford atlas). The effect size of the difference between the atlases was similarly large for both empirical and simulated data (effect size: 1.088 vs. 1.077 for empirical data and simulated data, respectively). Furthermore, the difference between the mean values of each atlas is larger for the simulated data (effect size: 0.679 vs. 1.189 for the Schaefer and Harvard-Oxford atlas, respectively, see Figure 2a).

The illustrated empirical and simulated connectome relationships can be considered as features for the machine-learning prediction approaches, where the enhanced inter-subject variability of the simulated features (larger spread of the feature distributions) might be a good indication for involving the simulated data in the analyses. To examine the extent of overlap and difference in the feature information under the considered four conditions (2-by-2) shown in the legend in Figure 2a, we

performed PCA using empirical and simulated features of the connectome relationships. Interestingly, we found that the first two principal components (PC1 and PC2), which deliver the largest fraction of the explained variance of all connectome relationships, primarily related to the simulated features (Figure 2b-c), while the next two components (PC3 and PC4) explained the empirical features. Furthermore, PC1 and PC3 represented common contributing factors in the simulated and empirical connectome relationships, respectively, of the two parcellation schemes (Figure 2b). In contrast, PC2 and PC4 distinguished the parcellation schemes in the respective simulated and empirical feature conditions (Figure 2b). The two first PCs with the loading by simulated data cumulatively explained up to 90% of the variance of all features (Figure 2c). The observed segregation of the empirical and simulated features into different PCs as well as the leading role of the latter features in PCA further support the expectations of a positive contribution of the simulated features to prediction results, which can be used either as stand-alone features or as a complement to empirical ones.

3.2 Classification and prediction performance

Since the empirical and simulated connectome relationships exhibit distinct variabilities across individuals (Figure 2), these two types of connectome relationships might contribute differently to a machine-learning prediction process. To investigate this, we prepared three distinct feature sets:

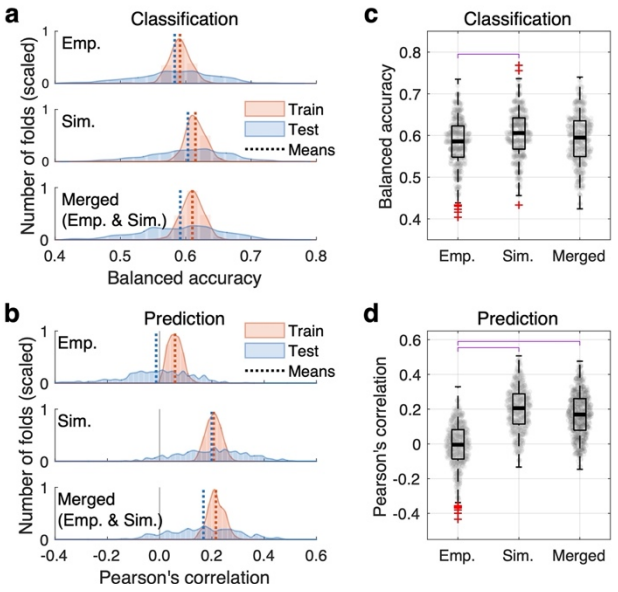

Figure 3. Machine-learning performances in sex classification and prediction of general intelligence. **(a)** Accuracy of sex classification as given by the fraction of correctly classified subjects for the training and testing sets as indicated in the legend. Distributions of balanced accuracy across cross-validation (CV) folds are shown. The mean values of the distributions are indicated by vertical dashed lines. The three plots illustrate the cases of (from top to bottom) the empirical features (Emp.), the simulated features (Sim.), and combination of the empirical and simulated features (Emp. & Sim.). The features of the two parcellations were merged in each condition. **(b)** Prediction of the total cognitive function composite scores (Pearson's correlation between predicted and empirical scores) for training and testing sets with the same scheme of (a). **(c)** Comparison of the sex classification and **(d)** prediction performance of the three feature conditions. The magenta bars depict statistically significant differences (with $p < 0.05$ of the Wilcoxon rank-sum two-tail test Bonferroni-corrected for multiple comparisons) in the distribution of prediction between the feature conditions indicated on the horizontal axis.

empirical (Emp.), simulated (Sim.), and combined empirical and simulated (Emp. & Sim.) features. Here, the first feature set (Emp.) includes the empirical structure-function relationships (Pearson's correlation between eSC and eFC), the second feature set (Sim.) includes the relationships between eFC and sFC (GoF values), and the third feature set (Emp. & Sim.) includes both the empirical and simulated features. We then performed two machine-learning analyses using these features to (*i*) classify the individual subjects as females or males and (*ii*) predict a continuous behavioral score as given by the general intelligence based on the total cognitive function composite score (Akshoomoff et al., 2013). For both cases and under each feature condition, we calculated prediction performance on the training set and after applying the model to the testing set of unseen subjects (Figure 3a, b). Sex classification on the test subject sets shows that the balanced accuracy was significantly enhanced, when the simulated features were employed in the classification analysis as compared to the case of the empirical features (Bonferroni-corrected $p < 0.05$) (Figure 3c, compare "Emp." to " Sim."). The machine-learning analysis applied to predict the general intelligence also exhibited improved performance with features that contain the simulated data. This was confirmed by statistical tests demonstrating a significant improvement of the prediction performance for the simulated features (Sim.) as well as for a combination of the empirical and simulated features (Emp. & Sim.) compared to the case of the empirical features (Emp.) (Figure 3d).

3.3 Enhanced performance with high-dimensional parameter optimization

We also fitted the model to empirical data in high-dimensional parameter spaces, where around 100 model parameters were simultaneously optimized by the CMAES algorithm. In such a way we obtained an increased GoF, where the simulated FCs closely approached the empirical FCs of individual subjects leading to a higher model personalization. For example, the mean GoF = 0.607

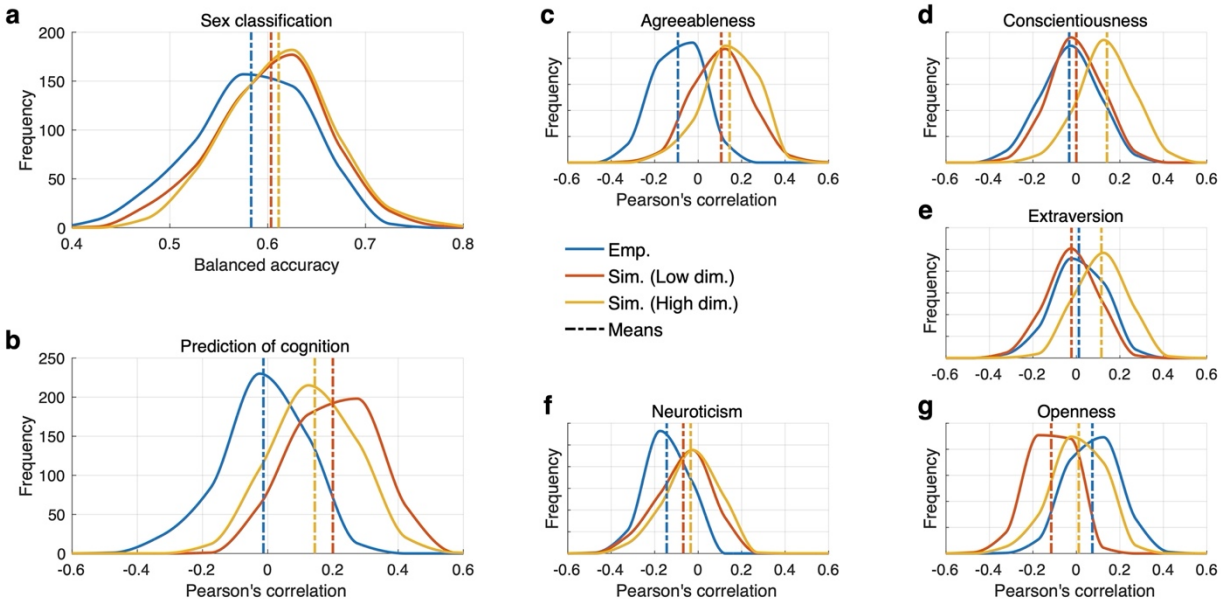

Figure 4. Prediction results with different feature conditions based on empirical features: Emp., simulated features with low-dimensional parameter optimization: Sim. (Low dim.) and high-dimensional parameter optimization: Sim. (High dim.). **(a)** Results of sex classification with five different feature conditions shown in the figure legend in the center. Dashed vertical lines indicate mean values of performance in each feature condition. **(b-g)** Results of prediction for cognition (b: general intelligence) and five personality traits (c-g: agreeableness, conscientiousness, extraversion, neuroticism, and openness).

and 0.724 for high-dimensional model fitting of the Schaefer and Harvard-Oxford atlases, respectively, may be compared to the respective GoF = 0.299 and 0.501 for the low-dimensional model fitting (Figure 2a). We then applied the simulated connectome relationships of the high-dimensional model fitting as features to machine learning. Interestingly, the results showed that involving the simulated features obtained through high-dimensional optimization yielded the best outcome in the sex classification (Figure 4a). Additionally, in the prediction of personality traits (Figure 4c-g), the simulated features showed the best results in four out of five traits, except for the openness, where the empirical features demonstrated the best performance (Figure 4g). These findings indicate that whole-brain dynamical modeling can enhance the performance of machine learning. This is especially evident in predictions of cognitive ability and personality traits, where the empirical features mainly showed correlations near zero, whereas the simulated features demonstrated clearly improved results.

Finally, when looking at the overall concatenated results of predicting all five personality traits, the simulated features obtained through the high-dimensional parameter optimization showed the highest prediction correlation (Figure 5), and the difference from the results based on the empirical features was particularly large (effect size is 0.836).

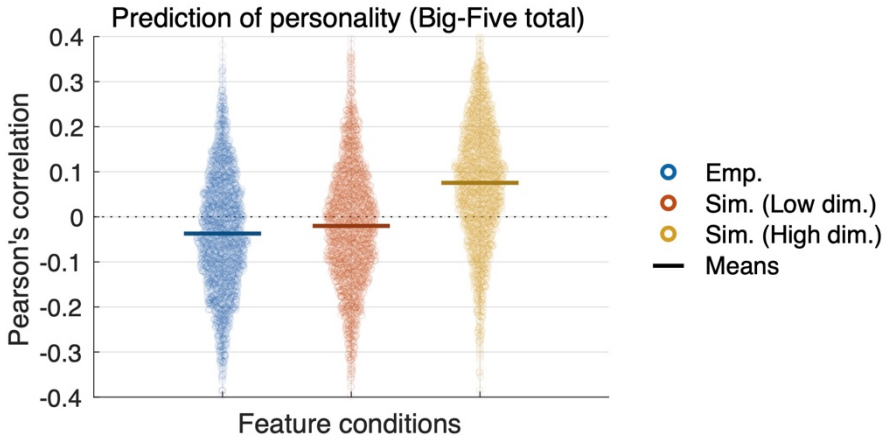

Figure 5. Overall concatenated prediction of personality traits. Each distribution includes Pearson's correlation coefficients between predicted and measured personality of all five traits in machine learning with 100 iterations and 5 folds cross-validation scheme, leading to 2500 points.

4 Discussion

In this report, we demonstrated that connectome relationships derived from the whole-brain dynamical modeling can represent individual variability of brain dynamics in a distinct way compared to empirical connectome relationships. We also showed that involving simulated connectomes in the machine-learning prediction analysis can enhance its prediction performance. Furthermore, machine learning using simulated and empirical features in a complementary way exhibited comparable or even improved performance in relation to a separate utilization of these feature configurations. Our results suggest that incorporating model-based features alongside empirical ones can enhance the extent of information extracted from the features provided by neuroimaging data. Building on these findings, it is important to delve into the specific advantages offered by the model-based approach.

The framework for the effective workflow proposed in this study consists of five steps, and the necessary procedures and possible approaches for each step are as follows:

- **Step 1:** For the whole-brain dynamical modeling, three types of MRI data are required: T1w, dwMRI, and resting-state fMRI. For neuroimaging research, raw data, *i.e.*, Digital Imaging and Communications in Medicine (DICOM) (Mildenberger et al., 2002) can be converted to a standard format such as Neuroimaging Informatics Technology Initiative (NIfTI) (Li et al., 2016). In addition, the data can be organized according to a consensus data organization called Brain Imaging Data Structure (BIDS) (Gorgolewski et al., 2016). Nowadays, many published datasets exist, which provide brain MRI necessary for the workflow such as OpenNeuro (https://openneuro.org) or other data collections, *e.g.*, ADNI (https://adni.loni.usc.edu), AOMIC (Snoek et al., 2021), PPMI (Marek et al., 2011) and research projects like 1000BRAINS (Caspers et al., 2014), HCP (Van Essen et al., 2013), MOUS (Schoffelen et al., 2019), PNC (Satterthwaite et al., 2014), etc.
- **Step 2:** This step of the workflow involves completing the preprocessing of MRI data and signal extractions through a pipeline. In this step, a careful selection of the data processing parameters with high quality control is necessary to check for errors or missing information in the acquired data. Small differences at the early stages can have a significant impact in the final stage of data modeling (Aquino et al., 2022; Domhof et al., 2022a; Domhof et al., 2021; Jung et al., 2021; Jung et al., 2023; Popovych et al., 2021; Zhang et al., 2024). The pipeline for processing MRI provided in this study (https://jugit.fz-juelich.de/inm7/public/vbc-mri-pipeline) can be configured in various ways depending on the purpose of the study. Alternatively, public pipelines such as fMRIPrep (Esteban et al., 2019), MRtrix3 (Tournier et al., 2019), QSIPrep (Cieslak et al., 2021), SPM (Friston et al., 1994), and FreeSurfer (Fischl, 2012) can be used. The processed data can be utilized to study the functional or structural characteristics of the brain through imaging analysis, as well as for modeling.
- **Step 3:** This step involves parcellating the brain into multiple regions according to a given brain atlas considering various schemes (Amunts et al., 2020; Desikan et al., 2006; Glasser et al., 2016; Pijnenburg et al., 2021; Schaefer et al., 2018) and calculating functional and structural connectivity of each pair of regions in order to construct the human brain connectome (Sporns et al., 2005). At this stage, the data necessary for the modeling (step 4) will finally be prepared. A few datasets of BOLD signals, SC and FC calculated for many brain parcellations are available on the EBRAINS (https://www.ebrains.eu) platform ready for analysis and modeling (Domhof et al., 2022b, c; Jung et al., 2022). Furthermore, since the structural and functional connectivities between brain regions can be interpreted as underlying structures of the information flow and its processing within the brain networks (Rubinov and Sporns, 2010), studies can be conducted to explore the relationships between network characteristics of SC and FC and behavioral, cognitive and clinical scores (Fornito et al., 2015; Sporns, 2014).
- **Step 4:** A whole-brain dynamical model can be constructed based on the empirical whole-brain connectomes and used to simulate brain dynamics such as electrical neuronal activity and BOLD signals. By varying the model parameters, one can analyze the simulated brain dynamics in comparisons with empirical data using BOLD signals (Friston et al., 2003), FC (Honey et al., 2009), dynamic FC that captures evolution of FC over time (Glomb et al., 2017), SC (Popovych et al., 2021), metastability (Deco and Kringelbach, 2016), behavioral or clinical scores (Jung et al., 2024; Jung et al., 2023), etc. This allows us to find optimal model parameters, where the model best replicates empirical brain dynamics and behavior depending on the study objectives. Several software packages are available for the modeling of neuronal brain dynamics, for example, The Virtual Brain (Sanz-Leon et al., 2015), NEST (Gewaltig and Diesmann, 2007) and DCM (Friston et al., 2003) to mention a few. Furthermore, by employing dedicated parameter optimization algorithms (Wischnewski et al., 2022), we can obtain fine-tuned models for an improved replication of empirical data. Such a whole-brain dynamical modeling approach provides

personalized optimal model parameters after model fitting toward specific target neuroimaging or behavioral scores of individual subjects, thereby showing the strongest relationship between simulated results of optimal models and clinical characteristics (Jirsa et al., 2023; Jung et al., 2024; Jung et al., 2023) or cognitive functions as demonstrated in the present study.

- **Step 5:** This stage involves conducting machine-learning prediction analysis using model-based data obtained from the previous steps. In this step, the cross-validated model-based scheme (Jung et al., 2023) extracts effective simulated features derived from personalized optimal models, and their predictive performances are evaluated using machine-learning techniques. This approach allows us to incorporate additional model-based features into the machine-learning process while keeping the established protocols of conventional machine-learning methodologies based on neuroimaging empirical data. The cross-validated model-based machine-learning approach has demonstrated improved prediction performance, as evidenced by medical data (Jung et al., 2023) and this study.

The discussed model-based approach can effectively be used for testing a variety of experimental and data-processing conditions applicable to many topics of brain research (Jirsa et al., 2023; Popovych et al., 2019). This approach has several advantages including enhanced reliability and flexibility as well as cost efficiency as it eliminates the burden to repeatedly acquire whole-brain dynamics from participants under different experimental conditions in the scanner. Additionally, given the diversity of approaches for the whole-brain modeling (Cabral et al., 2017; Deco et al., 2021; Jirsa et al., 2017; Moran et al., 2013), researchers can select and utilize models that best align with their research objectives, thereby facilitating model-based connectome investigation.

A critical aspect of this modeling process is the selection of data processing pipelines, including brain parcellation schemes and other parameters, which can significantly influence the modeling outcomes (Jung, 2023). More than 20 brain parcellation schemes have been employed in neuroimaging research, contributing to the diversity of the simulated brain dynamics (Domhof et al., 2021; Popovych et al., 2021). Moreover, variations in neuroimaging processing pipelines can substantially affect research outcomes (Aquino et al., 2022; Jung et al., 2021; Zhang et al., 2024), and multiple strategies of model-fitting methods can be applied to optimizing whole-brain models in different ways (Deco et al., 2017; Jung et al., 2024; Jung et al., 2023; Wischnewski et al., 2022). The variability of simulated connectomes across subjects can also provide more personalized data across a broader range of perspectives compared to analyses based solely on empirical results (Domhof et al., 2022a).

By incorporating model-based features alongside empirical data, we can extensively explore brain connectomes and their relationships, offering enhanced performance and other benefits. At the same time, researchers can gain a deeper understanding of the brain dynamics. Given the recent advancements in digital brain research, integrating and expanding brain models (Amunts et al., 2024), the systematic model-based approach proposed in this report represents a promising method for advancing brain models and their applications. Consequently, this approach underscores the potential for leveraging integrated data to provide comprehensive insights and improved predictive capabilities in neuroimaging research.

5 Conflict of Interest

The authors report no competing interests, and the research was conducted in the absence of any commercial or financial relationships.

6 Author Contributions

KJ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing - original draft, Writing - review & editing; **KJW**: Data curation, Methodology, Software, Validation, Writing - review & editing; **SBE**: Conceptualization, Funding acquisition, Resources, Project administration, Writing - review & editing; **OVP**: Conceptualization, Funding acquisition, Methodology, Resources, Project administration, Software, Supervision, Validation, Writing - review & editing.

7 Funding

This work was supported by the Portfolio Theme Supercomputing and Modeling for the Human Brain by the Helmholtz association, the Human Brain Project and the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreements 785907 (HBP SGA2), 945539 (HBP SGA3) and 826421 (VirtualBrainCloud). Open-access publication was funded by the Deutsche Forschungsgemeinschaft (German Research Foundation) - 491111487. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

8 Acknowledgments

The authors gratefully acknowledge computing time on the supercomputer JURECA (Jülich Supercomputing Centre, 2021) at Forschungszentrum Jülich under grant ‘cjinm71’.

9 Supplementary Material

No supplementary Material.

10 Data Availability Statement

The features for machine learning in this study can be found in the GitHub repository including Python scripts for training prediction models and a MATLAB script to analyze results and generate figures illustrated in this study (https://github.com/kyesam-jung/model-based-prediction).

References

Akshoomoff, N., Beaumont, J.L., Bauer, P.J., Dikmen, S.S., Gershon, R.C., Mungas, D., Slotkin, J., Tulsky, D., Weintraub, S., Zelazo, P.D., Heaton, R.K., 2013. NIH Toolbox Cognition Battery (CB): composite scores of crystallized, fluid, and overall cognition. Monogr Soc Res Child Dev 78, 119-132.

Amunts, K., Axer, M., Banerjee, S., Bitsch, L., Bjaalie, J.G., Brauner, P., Brovelli, A., Calarco, N., Carrere, M., Caspers, S., Charvet, C.J., Cichon, S., Cools, R., Costantini, I., D'Angelo, E.U., De Bonis, G., Deco, G., DeFelipe, J., Destexhe, A., Dickscheid, T., Diesmann, M., Düzel, E., Eickhoff, S.B., Einevoll, G., Eke, D., Engel, A.K., Evans, A.C., Evers, K., Fedorchenko, N., Forkel, S.J., Fousek, J., Friederici, A.D., Friston, K., Furber, S., Geris, L., Goebel, R., Güntürkün, O., Hamid, A.I.A., Herold, C., Hilgetag, C.C., Hölter, S.M., Ioannidis, Y., Jirsa, V., Kashyap, S., Kasper, B.S., d’Exaerde, A.d.K., Kooijmans, R., Koren, I., Kotaleski, J.H., Kiar, G., Klijn, W., Klüver, L., Knoll, A.C., Krsnik, Z., Kämpfer, J., Larkum, M.E., Linne, M.-L., Lippert, T., Abdullah, J.M., Maio, P.D., Magielse, N., Maquet, P., Mascaro, A.L.A., Marinazzo, D., Mejias, J., Meyer-Lindenberg, A., Migliore, M., Michael, J., Morel, Y., Morin, F.O., Muckli, L., Nagels, G., Oden, L., Palomero-

Gallagher, N., Panagiotaropoulos, F., Paolucci, P.S., Pennartz, C., Peeters, L.M., Petkoski, S., Petkov, N., Petro, L.S., Petrovici, M.A., Pezzulo, G., Roelfsema, P., Ris, L., Ritter, P., Rockland, K., Rotter, S., Rowald, A., Ruland, S., Ryvlin, P., Salles, A., Sanchez-Vives, M.V., Schemmel, J., Senn, W., de Sousa, A.A., Ströckens, F., Thirion, B., Uludağ, K., Vanni, S., van Albada, S.J., Vanduffel, W., Vezoli, J., Vincenz-Donnelly, L., Walter, F., Zaborszky, L., 2024. The coming decade of digital brain research: A vision for neuroscience at the intersection of technology and computing. Imaging Neuroscience.

Amunts, K., Mohlberg, H., Bludau, S., Zilles, K., 2020. Julich-Brain: A 3D probabilistic atlas of the human brain's cytoarchitecture. Science 369, 988-992.

Andellini, M., Cannata, V., Gazzellini, S., Bernardi, B., Napolitano, A., 2015. Test-retest reliability of graph metrics of resting state MRI functional brain networks: A review. J Neurosci Methods 253, 183-192.

Aquino, K.M., Fulcher, B., Oldham, S., Parkes, L., Gollo, L., Deco, G., Fornito, A., 2022. On the intersection between data quality and dynamical modelling of large-scale fMRI signals. Neuroimage 256, 119051.

Batista-Garcia-Ramo, K., Fernandez-Verdecia, C.I., 2018. What We Know About the Brain Structure-Function Relationship. Behav Sci (Basel) 8.

Biswal, B., Yetkin, F.Z., Haughton, V.M., Hyde, J.S., 1995. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn Reson Med 34, 537-541.

Budka, M., Gabrys, B., 2013. Density-preserving sampling: robust and efficient alternative to cross-validation for error estimation. IEEE Trans Neural Netw Learn Syst 24, 22-34.

Cabral, J., Kringelbach, M.L., Deco, G., 2017. Functional connectivity dynamically evolves on multiple time-scales over a static structural connectome: Models and mechanisms. Neuroimage 160, 84-96.

Caspers, S., Moebus, S., Lux, S., Pundt, N., Schutz, H., Muhleisen, T.W., Gras, V., Eickhoff, S.B., Romanzetti, S., Stocker, T., Stirnberg, R., Kirlangic, M.E., Minnerop, M., Pieperhoff, P., Modder, U., Das, S., Evans, A.C., Jockel, K.H., Erbel, R., Cichon, S., Nothen, M.M., Sturma, D., Bauer, A., Jon Shah, N., Zilles, K., Amunts, K., 2014. Studying variability in human brain aging in a population-based German cohort-rationale and design of 1000BRAINS. Front Aging Neurosci 6, 149.

Chen, J., Ooi, L.Q.R., Tan, T.W.K., Zhang, S., Li, J., Asplund, C.L., Eickhoff, S.B., Bzdok, D., Holmes, A.J., Yeo, B.T.T., 2023. Relationship between prediction accuracy and feature importance reliability: An empirical and theoretical study. Neuroimage 274, 120115.

Cieslak, M., Cook, P.A., He, X., Yeh, F.C., Dhollander, T., Adebimpe, A., Aguirre, G.K., Bassett, D.S., Betzel, R.F., Bourque, J., Cabral, L.M., Davatzikos, C., Detre, J.A., Earl, E., Elliott, M.A., Fadnavis, S., Fair, D.A., Foran, W., Fotiadis, P., Garyfallidis, E., Giesbrecht, B., Gur, R.C., Gur, R.E., Kelz, M.B., Keshavan, A., Larsen, B.S., Luna, B., Mackey, A.P., Milham, M.P., Oathes, D.J., Perrone, A., Pines, A.R., Roalf, D.R., Richie-Halford, A., Rokem, A., Sydnor, V.J., Tapera, T.M., Tooley, U.A., Vettel, J.M., Yeatman, J.D., Grafton, S.T., Satterthwaite, T.D., 2021. QSIPrep: an integrative platform for preprocessing and reconstructing diffusion MRI data. Nat Methods 18, 775-778.

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29, 162-173.

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9, 179-194.

Damoiseaux, J.S., 2017. Effects of aging on functional and structural brain connectivity. Neuroimage 160, 32-40.

Deco, G., Kringelbach, M.L., 2016. Metastability and Coherence: Extending the Communication through Coherence Hypothesis Using A Whole-Brain Computational Perspective. Trends Neurosci 39, 125-135.

Deco, G., Kringelbach, M.L., Arnatkeviciute, A., Oldham, S., Sabaroedin, K., Rogasch, N.C., Aquino, K.M., Fornito, A., 2021. Dynamical consequences of regional heterogeneity in the brain's transcriptional landscape. Sci Adv 7.

Deco, G., Kringelbach, M.L., Jirsa, V.K., Ritter, P., 2017. The dynamics of resting fluctuations in the brain: metastability and its dynamical cortical core. Sci Rep 7, 3095.

Desikan, R.S., Segonne, 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.

Domhof, J.W.M., Eickhoff, S.B., Popovych, O.V., 2022a. Reliability and subject specificity of personalized whole-brain dynamical models. Neuroimage 257, 119321.

Domhof, J.W.M., Jung, K., Eickhoff, S.B., Popovych, O.V., 2021. Parcellation-induced variation of empirical and simulated brain connectomes at group and subject levels. Netw Neurosci 5, 798-830.

Domhof, J.W.M., Jung, K., Eickhoff, S.B., Popovych, O.V., 2022b. Parcellation-based resting-state blood-oxygen-level-dependent (BOLD) signals of a healthy cohort (v1.0). EBRAINS.

Domhof, J.W.M., Jung, K., Eickhoff, S.B., Popovych, O.V., 2022c. Parcellation-based structural and resting-state functional brain connectomes of a healthy cohort (v1.1). EBRAINS.

Esteban, O., Markiewicz, C.J., Blair, R.W., Moodie, C.A., Isik, A.I., Erramuzpe, A., Kent, J.D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S.S., Wright, J., Durnez, J., Poldrack, R.A., Gorgolewski, K.J., 2019. fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat Methods 16, 111-116.

Fischl, B., 2012. FreeSurfer. Neuroimage 62, 774-781.

Fornito, A., Zalesky, A., Breakspear, M., 2015. The connectomics of brain disorders. Nat Rev Neurosci 16, 159-172.

Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. Neuroimage 19, 1273-1302.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., Frackowiak, R.S.J., 1994. Statistical parametric maps in functional imaging: A general linear approach. Human Brain Mapping 2, 189-210.

Gewaltig, M.-O., Diesmann, M., 2007. Nest (neural simulation tool). Scholarpedia 2, 1430.

Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., Smith, S.M., Van Essen, D.C., 2016. A multi-modal parcellation of human cerebral cortex. Nature 536, 171-178.

Glomb, K., Ponce-Alvarez, A., Gilson, M., Ritter, P., Deco, G., 2017. Resting state networks in empirical and simulated dynamic functional connectivity. Neuroimage 159, 388-402.

Gorgolewski, K.J., Auer, T., Calhoun, V.D., Craddock, R.C., Das, S., Duff, E.P., Flandin, G., Ghosh, S.S., Glatard, T., Halchenko, Y.O., Handwerker, D.A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B.N., Nichols, T.E., Pellman, J., Poline, J.B., Rokem, A., Schaefer, G., Sochat, V., Triplett, W., Turner, J.A., Varoquaux, G., Poldrack, R.A., 2016. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci Data 3, 160044.

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.

Hansen, N., Ostermeier, A., 1996. Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation. Proceedings of IEEE international conference on evolutionary computation. IEEE, pp. 312-317.

Hansen, N., Ostermeier, A., 2001. Completely derandomized self-adaptation in evolution strategies. Evol Comput 9, 159-195.

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., Del Rio, J.F., Wiebe, M., Peterson, P., Gerard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with NumPy. Nature 585, 357-362.

Havlicek, M., Roebroeck, A., Friston, K., Gardumi, A., Ivanov, D., Uludag, K., 2015. Physiologically informed dynamic causal modeling of fMRI data. Neuroimage 122, 355-372.

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 S A 106, 2035-2040.

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.

Jirsa, V., Wang, H., Triebkorn, P., Hashemi, M., Jha, J., Gonzalez-Martinez, J., Guye, M., Makhalova, J., Bartolomei, F., 2023. Personalised virtual brain models in epilepsy. Lancet Neurol 22, 443-454.

Jirsa, V.K., Proix, T., Perdikis, D., Woodman, M.M., Wang, H., Gonzalez-Martinez, J., Bernard, C., Benar, C., Guye, M., Chauvel, P., Bartolomei, F., 2017. The Virtual Epileptic Patient: Individualized whole-brain models of epilepsy spread. Neuroimage 145, 377-388.

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 JLSRF 7.

Jung, K., 2023. Impact of data processing parameters on whole-brain dynamical models. Heinrich-Heine-Universität Düsseldorf, pp. 1-111.

Jung, K., Eickhoff, S.B., Caspers, J., UKD-PD team, Popovych, O.V., 2024. Simulated brain networks reflecting progression of Parkinson's disease. Network Neuroscience, 1-31.

Jung, K., Eickhoff, S.B., Popovych, O.V., 2021. Tractography density affects whole-brain structural architecture and resting-state dynamical modeling. Neuroimage 237, 118176.

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

Jung, K., Florin, E., Patil, K.R., Caspers, J., Rubbert, C., Eickhoff, S.B., Popovych, O.V., 2023. Whole-brain dynamical modelling for classification of Parkinson's disease. Brain Commun 5, fcac331.

Kuramoto, Y., 1984. Chemical Oscillations, Waves, and Turbulence, 1 ed. Springer Berlin, Heidelberg.

Li, X., Morgan, P.S., Ashburner, J., Smith, J., Rorden, C., 2016. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. J Neurosci Methods 264, 47-56.

Marcus, D.S., Harwell, J., Olsen, T., Hodge, M., Glasser, M.F., Prior, F., Jenkinson, M., Laumann, T., Curtiss, S.W., Van Essen, D.C., 2011. Informatics and data mining tools and strategies for the human connectome project. Front Neuroinform 5, 4.

Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., Coffey, C., Kieburtz, K., Flagg, E., Chowdhury, S., 2011. The Parkinson progression marker initiative (PPMI). Progress in neurobiology 95, 629-635.

McCrae, R.R., Costa, P.T., 2004. A contemplated revision of the NEO Five-Factor Inventory. Personality and Individual Differences 36, 587-596.

Messe, A., 2020. Parcellation influence on the connectivity-based structure-function relationship in the human brain. Hum Brain Mapp 41, 1167-1180.

Mildenberger, P., Eichelberg, M., Martin, E., 2002. Introduction to the DICOM standard. Eur Radiol 12, 920-927.

Moran, R., Pinotsis, D.A., Friston, K., 2013. Neural masses and fields in dynamic causal modeling. Front Comput Neurosci 7, 57.

More, S., Eickhoff, S.B., Caspers, J., Patil, K.R., 2021. Confound Removal and Normalization in Practice: A Neuroimaging Based Sex Prediction Case Study. Springer International Publishing, Cham, pp. 3-18.

Park, H.J., Friston, K., 2013. Structural and functional brain networks: from connections to cognition. Science 342, 1238411.

Pathak, A., Roy, D., Banerjee, A., 2022. Whole-Brain Network Models: From Physics to Bedside. Frontiers in Computational Neuroscience 16.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine learning in Python. the Journal of machine Learning research 12, 2825-2830.

Pijnenburg, R., Scholtens, L.H., Ardesch, D.J., de Lange, S.C., Wei, Y., van den Heuvel, M.P., 2021. Myelo- and cytoarchitectonic microstructural and functional human cortical atlases reconstructed in common MRI space. Neuroimage 239, 118274.

Popovych, O.V., Jung, K., Manos, T., Diaz-Pier, S., Hoffstaedter, F., Schreiber, J., Yeo, B.T.T., Eickhoff, S.B., 2021. Inter-subject and inter-parcellation variability of resting-state whole-brain dynamical modeling. Neuroimage 236, 118201.

Popovych, O.V., Manos, T., Hoffstaedter, F., Eickhoff, S.B., 2019. What Can Computational Models Contribute to Neuroimaging Data Analytics? Front Syst Neurosci 12, 68.

Rosenthal, R., Cooper, H., Hedges, L., 1994. Parametric measures of effect size. The handbook of research synthesis 621, 231-244.

Rubinov, M., Sporns, O., 2010. Complex network measures of brain connectivity: uses and interpretations. Neuroimage 52, 1059-1069.

Sanz-Leon, P., Knock, S.A., Spiegler, A., Jirsa, V.K., 2015. Mathematical framework for large-scale brain network modeling in The Virtual Brain. Neuroimage 111, 385-430.

Satterthwaite, T.D., Elliott, M.A., Ruparel, K., Loughead, J., Prabhakaran, K., Calkins, M.E., Hopson, R., Jackson, C., Keefe, J., Riley, M., Mentch, F.D., Sleiman, P., Verma, R., Davatzikos, C., Hakonarson, H., Gur, R.C., Gur, R.E., 2014. Neuroimaging of the Philadelphia neurodevelopmental cohort. Neuroimage 86, 544-553.

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.

Schoffelen, J.M., Oostenveld, R., Lam, N.H.L., Udden, J., Hulten, A., Hagoort, P., 2019. A 204-subject multimodal neuroimaging dataset to study language processing. Sci Data 6, 17.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 Suppl 1, S208-219.

Snoek, L., van der Miesen, M.M., Beemsterboer, T., van der Leij, A., Eigenhuis, A., Steven Scholte, H., 2021. The Amsterdam Open MRI Collection, a set of multimodal MRI datasets for individual difference analyses. Sci Data 8, 85.

Sporns, O., 2014. Contributions and challenges for network models in cognitive neuroscience. Nat Neurosci 17, 652-660.

Sporns, O., Tononi, G., Kotter, R., 2005. The human connectome: A structural description of the human brain. PLoS Comput Biol 1, e42.

Suarez, L.E., Markello, R.D., Betzel, R.F., Misic, B., 2020. Linking Structure and Function in Macroscale Brain Networks. Trends Cogn Sci 24, 302-315.

Tavor, I., Jones, O.P., Mars, R.B., Smith, S.M., Behrens, T.E., Jbabdi, S., 2016. Task-free MRI predicts individual differences in brain activity during task performance. Science 352, 216-220.

Tournier, J.D., Calamante, F., Connelly, A., 2007. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. Neuroimage 35, 1459-1472.

Tournier, J.D., Calamante, F., Connelly, A., 2010. Improved probabilistic streamlines tractography by 2nd order integration over fibre orientation distributions. Proceedings of the International Society for Magnetic Resonance in Medicine, 1670.

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.

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.

Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.U.-M.H., 2013. The WU-Minn Human Connectome Project: an overview. Neuroimage 80, 62-79.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, I., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy, C., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 17, 261-272.

Wiegell, M.R., Larsson, H.B., Wedeen, V.J., 2000. Fiber crossing in human brain depicted with diffusion tensor MR imaging. Radiology 217, 897-903.

Wischnewski, K.J., Eickhoff, S.B., Jirsa, V.K., Popovych, O.V., 2022. Towards an efficient validation of dynamical whole-brain models. Sci Rep 12, 4331.

Yeung, M.K.S., Strogatz, S.H., 1999. Time Delay in the Kuramoto Model of Coupled Oscillators. Physical Review Letters 82, 648-651.

Zhang, S., Jung, K., Langner, R., Florin, E., Eickhoff, S.B., Popovych, O.V., 2024. Impact of data processing varieties on DCM estimates of effective connectivity from task-fMRI. Hum Brain Mapp 45, e26751.

Zhao, T., Cao, M., Niu, H., Zuo, X.N., Evans, A., He, Y., Dong, Q., Shu, N., 2015. Age-related changes in the topological organization of the white matter structural connectome across the human lifespan. Hum Brain Mapp 36, 3777-3792.