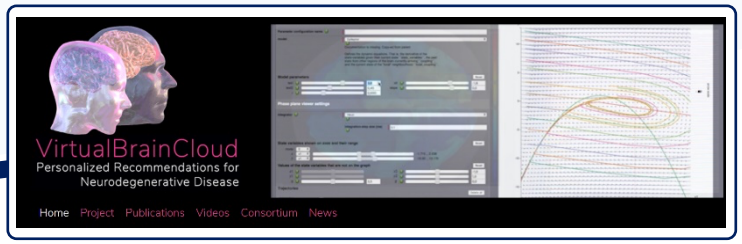
VirtualBrainCloud

Personalized Recommendations for Neurodegenerative Disease



www.VirtualBrainCloud-2020.eu

Public deliverable report

Deliverable 3.8: Seamless integration of electrophysiological data workflows with TVB modeling.

Date	May 2023
Authors	Ehtasham Javed, Satu Palva, Matias Palva (UH) Isabel Suárez, Gianluca Susi, Fernando Maestú (UCM)

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

Summary

Integrating electrophysiological data to TVB computational models is a key milestone in achieving the overarching goal of The VirtualBrainCloud (TVB-Cloud), that is, personalized prevention and treatment of dementia. In this context, we extracted individual dynamic measures including local dynamics (long-range temporal correlation (LRTC), functional excitation-inhibition ratio (fE/I)), and inter-areal dynamics (functional connectomes of phase (i.e., phase locking value (PLV), corrected imaginary phase-locking value (ciPLV), weighted-phase lag index (wPLI)) from the MEG dataset of the Madrid cohort. We show that both local and inter-areal dynamics progressively change and their spectral profiles reliably dissociate NC, SCD, and MCI individuals and are critical for computational model fitting and validation. The description of workflow, computation of dynamic measures, their importance in characterizing early stages of Alzheimer's disease, and their integration with the computational models are described in this document for public use.

2. Partners involved

UNIVERSITY OF HELSINKI (UH)
UNIVERSIDAD COMPLUTENSE DE MADRID (UCM)

3. Description of the work performed

3.1 MEG database

The MEG full-cohort database was introduced in The Virtual Brain Cloud (TVB-Cloud) public Deliverable 3.5[1]. This shared database included MEG data from older adults classified into four diagnostic categories: healthy control (HC), subjective cognitive decline (SCD), mild cognitive impairment (MCI), and Alzheimer's disease (AD). This shared database included a cross-sectional dataset of 364 participants, as well as a longitudinal dataset of 40 participants who were followed-up every six months to investigate AD conversion (Pusil et al., 2018). Besides MEG, MRI, clinical, neuropsychological, and genetic data were also provided. All participants were right-handed, native Spanish speakers, aged 65 to 80. Data collection was carried out after obtaining written consent from each participant. Beforehand, they were provided with all the necessary information to ensure that the collection and processing of personal data was conducted in a fair, lawful, and transparent manner (see also the Annex to Deliverable 3.5). In the context of the TVB-Cloud project, the MEG data was processed at the Complutense University of Madrid (UCM) and at the University of Helsinki (UH) according to the workflows described in Section – Workflows. In the public Deliverable 3.7[2], the database was completed with additional individual dynamic measures, including local dynamics (i.e., long-range temporal correlations (LRTCs) and functional excitation-inhibition (fE/I)) and inter-areal dynamics (i.e., functional phase-connectomes, amplitude coupling, cross-frequency synchrony (CFS),

[1] https://virtualbraincloud-2020.eu/files/tvb/documents/public%20deliverables/period%202/TVB-Cloud_Del3.5.pdf
[2] https://virtualbraincloud-2020.eu/files/tvb/documents/public%20deliverables/period%202/TVB-Cloud_D3.7_cohort

and phase-amplitude coupling (PAC)). The primary goal of this analysis array was to equip the consortium with robust tools to support personalized The Virtual Brain (TVB) parameter validation, while concurrently generating diagnostic and prognostic biomarkers for AD (see Section – Results).

Table 1. Summary description for the shared MEG dataset.

	HC	SCD	MCI	AD
N	119	88	142	15
Sex (F)	79	70	93	7
Age (years)	70.29 ± 4.38	72.34 ± 5.21	73.40 ± 5.44	76.73 ± 5.20
MEG	119	88	142	15
T1-MRI	119	88	142	15
dw-MRI	107	80	122	11

3.1.1 Sample

The shared database includes 119 HC (79 females; 70.29 ± 4.38 years), 88 participants with SCD (70 females; 72.34 ± 5.21 years), 142 participants with MCI (93 females; 73.40 ± 5.44 years), and 15 participants with AD (7 females; 76.73 ± 5.20 years). Following the recommendations made by the SCD Initiative Working Group, the participants eligible for the SCD group: (1) reported self-experienced persistent cognitive concerns (mainly associated with memory) in an interview with an expert clinician; (2) performed within the normal range on standardized cognitive tests that discriminate MCI and prodromal AD; (3) felt that their cognitive decline affected their daily activities; (4) had requested medical consultation; and (5) were ≥ 60 years at the onset of SCD, having it occurred within the last five years. MCI diagnosis was established following the recommendations of the National Institute on Aging–Alzheimer's Association (NIA–AA) criteria: (1) self- or informant-reported cognitive complaints; (2) objective evidence of impairment in one or more cognitive domains; (3) preserved independence in functional abilities; and (4) not demented. All participants classified as having AD fulfilled the National Institute of Neurological and Communicative Disorders and Stroke–Alzheimer's Disease and Related Disorders Association (NINCDS–ADRDA) criteria for probable AD. This requires patients to meet the clinical criteria for all-cause dementia, including (1) insidious onset; (2) clear history of worsening of cognition by report or observation; and (3) prominent cognitive deficits that include amnestic presentation and/or deficits in language presentation, visuospatial presentation, and executive function. The inclusion and exclusion criteria were detailed in D10.1: H – Requirement No. 1.

3.1.2 MEG data

MEG data were acquired using a 306-channel (102 magnetometers, 204 planar gradiometers) Vectorview MEG system (Elekta Neuromag Oy, Helsinki, Finland) installed inside a magnetically shielded room (VacuumSchmelze GmbH, Hanau, Germany) at the Center for Biomedical Technology (Madrid, Spain). During the data acquisition, the participants rested awake with their eyes closed, and a total of five minutes of MEG data were recorded. Before the data acquisition, the participants' head shapes were digitized using a three-dimensional Fastrak digitizer (Polhemus, Colchester, Vermont). Specifically, three fiducial points were registered (nasion, right preauricular, and left preauricular), as well as an outline of approximately 400 scalp points. Four head position indicator (HPI) coils were applied to the participant's scalp. To record eye blinks and ocular movements, two electrooculogram (EOG) electrodes were placed above and below the left eye. To record the cardiac activity, two electrocardiogram (EKG) electrodes were placed across the chest, forming a diagonal in a bipolar setup. MEG raw data were acquired using a sampling rate of 1,000 Hz and an in-line anti-aliasing band-pass filter between 0.1 and 330 Hz. Afterward, the MEG data were processed offline with the temporospatial filtering algorithm (tSSS) as implemented in MAXFILTER (Elekta Neuromag Oy, Helsinki, Finland) with a correlation limit of 0.9 and a window length of 10 s.

3.1.3 MRI data

Three-dimensional T1-weighted MRI images (T1-MRI) were acquired for each participant within a month after the MEG session, using a 1.5 T MRI scanner (GE Healthcare, Chicago, Illinois) with a high-resolution antenna and a homogenization PURE filter (fast spoiled gradient-echo sequence, with parameters: repetition time/echo time/inversion time: 11.2/4.2/450 ms; flip angle: 12°; slice thickness: 1 mm, 256×256 matrix, and field of view: 256 mm).
Three-dimensional diffusion-weighted MRI images (dw-MRI) were acquired with a single-shot echo-planar imaging sequence (with parameters: echo time/repetition time: 96.1/12,000 ms; NEX 3 for increasing the SNR; slice thickness: 2.4 mm, 128×128 matrix, and field of view: 30.7 cm), yielding an isotropic voxel of 2.4 mm; 1 image with no diffusion sensitization (i.e., T2-weighted b0 image); and 25 dw-MRI (b = 900 s/mm^2).

3.1.4 Neuropsychological data

The following domains were exhaustively assessed for each participant: (1) memory, (2) language, (3) executive function, (4) cognitive status, (5) subjective memory, and (6) functional capacity and mood. Memory was assessed with the Digit Span Test (forward and backward) of the Wechsler Memory Scale-III-R (Spanish version), and the Texts of Verbal Memory and the Word List of the Wechsler Memory Scale-III. Language function was assessed with the Boston Naming Test and the Phonemic and Semantic Fluency Tests (Controlled Oral Word Association Test). Executive function was assessed with the Trail Making Test parts A and B. General cognitive status was assessed with the Mini-Mental State Examination. Memory in everyday life was assessed with the Rivermead Behavioral Memory Test, and mood was assessed with the Geriatric Depression Scale-Short Form.

3.1.5 Database implementation

The full-cohort MEG database is available upon request at the URL https://vbc.ucm.es. Following the Brain Imaging Data Structure (BIDS) standard (Niso et al., 2017), the data was structured in a root

folder and a *derivatives* subfolder comprising the derived data generated by the different processing pipelines (see Section – Workflows). In addition, an extra folder (*derivatives/TVB*) includes a set of TVB-compliant structural data, the time-series associated with the ROIs, and the dynamic features provided for TVB-model validation.

3.2 Workflows

3.2.1 MEG

The developed pipelines encompass the three primary components of MEG analysis, namely: (1) the pre-processing of raw MEG recordings (*pre-processing* pipeline); (2) the MEG source reconstruction (*source-reconstruction* pipeline); and (3) the extraction of the individual dynamic features (*analysis* pipeline). All the scripts are executed within the MATLAB environment (Mathworks, Inc.) and incorporate functions from:

- **Fieldtrip** software toolbox for MEG analysis (Oostenveld et al., 2011); https://www.fieldtriptoolbox.org/.
- **FreeSurfer** software (Dale, Fischl, & Sereno, 1999); https://surfer.nmr.mgh.harvard.edu/.
- **SPM12** software; https://www.fil.ion.ucl.ac.uk/spm/software/spm12/.
- **Neurophysiological Biomarker** toolbox; https://www.nitrc.org/projects/nbt/.

Each pipeline comprises several modules designed for sequential execution. The source-reconstruction and analysis pipelines run automatically and require minimal input from the end user. The pre-processing pipeline necessitates a higher degree of user input and decision-making, particularly in relation to quality control. The individual outputs are finally converted to BIDS formatting. The UCM and UH pipelines present slight variations that are described in the text. To access either version email requests can be sent to fmaestuu@ucm.es (UCM version) and matias.palva@helsinki.fi (UH version).

The pre-processing pipeline aims to achieve automatic artifact extraction, artifact review, independent component (IC) extraction and labeling, quality control, and data segmentation. This pipeline is specifically designed to perform the following tasks (see Table 2 for a summary):

1. Automatic artifact extraction (*s1_artifactExtraction*): Fieldtrip's code is used to create a temporal definition for three types of artifacts (ocular, muscular, and system artifacts).
2. Artifact review (*s2_artifactReview*): The user corrects for potential misidentifications. If available, EOG data is displayed alongside MEG data to facilitate the detection of ocular artifacts.
3. ICs extraction and labeling (*s3_extractIC*, *s4_selectArtifactsIC*, and *s6_markEKGlead*): A blind source separation algorithm based on second-order statistics is used to extract the mixing matrix. Then, the ICs are visually inspected to identify EOG and EKG activity. EOG artifacts are highlighted to facilitate EOG-IC identification. The topographic distribution of the ICs is also displayed in the GUI.
4. Quality control: Optional quality-control modules are used to, e.g., recuperate data segments previously labeled as EOG artifacts after EOG-ICs identification.
5. Data segmentation or data interpolation (*s7_saveSegments*): (UCM) The clean data is segmented into 4-second epochs after discarding the artifacts and saved. (UH) The artifacts are interpolated to preserve the original-length of the recordings and the clean data is saved.

Table 2. Main modules of the data-preprocessing pipeline, objectives, data inputs and outputs, and user requirements.

Module	Objective	Data input	Data output	User input
s1_artifactExtraction	Performs automatic artifact detection.	MEG raw/tSSS data	Auto artifact definition	Auto
s2_artifactReview	Reviews the automatic artifact detection. Quality control.	MEG raw/tSSS data Auto artifact definition	User-corrected artifact definition	Yes
s3_extractIC	Performs Independent Component Analysis (ICA).	MEG raw/tSSS data User-corrected artifact definition	Mixing/unmixing matrices	Auto
s4_selectArtifactsIC	Identifies EOG and EKG independent components (ICs).	MEG raw/tSSS data Mixing/unmixing matrices	ICs definition	Yes
s5_extractSketch	Builds the trial definition. Identifies clean trials and clean ICs.	MEG raw/tSSS data User-corrected artifact definition ICs definition	Trial definition Trial classification (clean/noisy) ICs classification (clean/noisy/EOG/EKG)	Auto
s6_markEKGlead	Processes the EKG components.	Trial definition Trial classification (clean/noisy) ICs classification (clean/noisy/EOG/EKG)	EKG lead IC	Yes
s7_saveSegments	Saves the clean sensor-space data. Removes/interpolates the artifacted trials and deals with the noisy/EOG/EKG ICs.	Trial definition Trial classification (clean/noisy) ICs classification (clean/noisy/EOG/EKG) EKG lead IC.	Clean sensor-space data	Auto

The source-reconstruction pipeline aims to achieve MEG source-reconstruction. This pipeline is specifically designed to perform the following tasks (see Table 3. for a summary):

1. Fiducial identification, segmentation, and mask generation *(s1_loadMRI*, *s2_segmentMRI*, and *s3_createMasks*): Three fiducial points (nasion, right preauricular, left preauricular) and three SPM landmarks are identified in the T1-weighted MRI. The software SPM12 is used to segment the MRI data into probability maps for the different brain tissues (white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), and bone). Following the tissue segmentation, we obtain masks for the brain, the skull, and the scalp.
2. Head model (*s4_createMeshGrid* and *s5_createHeadModel*): This task builds a single-shell head model with a unique boundary defined by the inner skull.
3. Source model (*s4_createMeshGrid*): (UCM) A source model of 2,459 sources placed in a homogeneous grid of 1 cm in a Montreal Neurological Institute (MNI) template is converted to subject space by an affine transformation. (UH) The white-matter surface is reconstructed using FreeSurfer. The dipoles are defined at the vertices of the white-matter surface with around 7mm inter-dipole distance.
4. Co-registration (*s6_realignMRI*): Between the MEG and the T1-MRI.

5. Forward problem *(s7_createLeadfield)*: Leadfield matrix calculation.
6. Inverse problem (*s8_getSources*): (UCM) Source reconstruction is performed using beamforming. The beamformer filters are obtained for each classical frequency band (θ [4 - 8] Hz, α [8 - 12] Hz, β [12 - 30] Hz, γ [30 - 45] Hz, and broadband [2 - 45] Hz) using the previously computed leadfield, the epoch-averaged covariance matrix, and a 20% regularization factor. (UH) Source reconstruction is performed using Fieldtrip's minimum norm estimates (MNE) algorithm. The inverse operator is obtained using the previously computed leadfield, the covariance matrix, and a 0.11 regularization factor. Next, using the forward and inverse operators alongside simulated data, fidelity weights for each dipole (source) are estimated to avoid the erroneous estimation of source time-series. Then, the fidelity-weighted source time-series are parcellated into the Schaefer's 400 parcels time-series.

Table 3. Main modules of the source-reconstruction pipeline, objectives, data inputs and outputs, and user requirements.

Module	Objective	Data input	Data output	User input
s1_loadMRI	Loads the raw MRI file. Saves the fiducial markers and the SPM landmarks coordinates (user input).	Raw MRI file	Formatted MRI file Fiducial markers and SPM landmark coordinates	Yes
s2_segmentMRI	Obtains the probability maps for the different brain tissues (WM, GM, CSF, and bone). Uses SPM12.	Formatted MRI file SPM landmark coordinates	Probability maps for the different brain tissues	Auto
s3_createMasks	Uses the probability maps to create masks (logicals) for the different brain tissues.	Formatted MRI file Probability maps for the different brain tissues	Scalp, brain, and skull masks	Auto
s4_createMeshGrid	Builds the mesh that delimits the brain surface using the single-shell method (Nolte, 2003) and 10,000 vertices. Warps the MNI source template to the individual MRI.	Formatted MRI file Scalp, brain, and skull masks	Single-shell mesh Warped source model	Auto
s5_createHeadModel	Creates the head model.	Single-shell mesh	Head model	Auto
s6_realignMRI	Uses the fiducial landmarks to realign the head model relative to the sensor positions.	Head model Head digitization Fiducial markers	Realigned head model	Yes
s7_createLeadfield	Calculates the leadfield matrix.	Head model Source model	Leadfield matrix	Auto
s8_getSources	Solves the inverse problem and calculates the source-space time-series (using beamforming (UCM) or minimum-norm estimates (MNE) (UH)).	Clean sensor-space data Leadfield matrix	Beamformer filters (UCM) MNE solution (UH)	Auto

The analysis pipeline provides an analysis array aimed to equip partners with the relevant features to support personalized TVB parameter validation. A complete description of the analysis array can be found in Deliverable 3.7. The main modules of the analysis pipeline used in the present deliverable are summarized in Table 4.

Table 4. Main modules of the analysis pipeline, objectives, data inputs and outputs, and user requirements.

Module	Objective	Data input	Data output	User input
s1_getConnectivityPLV	Calculates the sources×sources and the ROIs×ROIs FC connectivity matrix using the hypothesis of phase synchrony and the phase-locking value (PLV).	Clean sensor-space data Beamformer filters	PLV matrix.	Auto Input parameters
s2_getConnectivityCIPLV	Calculates the sources×sources and the ROIs×ROIs FC connectivity matrix using the hypothesis of phase synchrony and the corrected imaginary phase-locking value (ciPLV).	Clean sensor-space data Beamformer filters	ciPLV matrix.	Auto Input parameters
s3_getDFA	Calculates the DFA exponents.	Source-space data	Fluctuation function DFA exponents	Auto Input parameters
S4_getfEI	Calculates the fE/I.	Source-space data	fE/I	Auto Input parameters

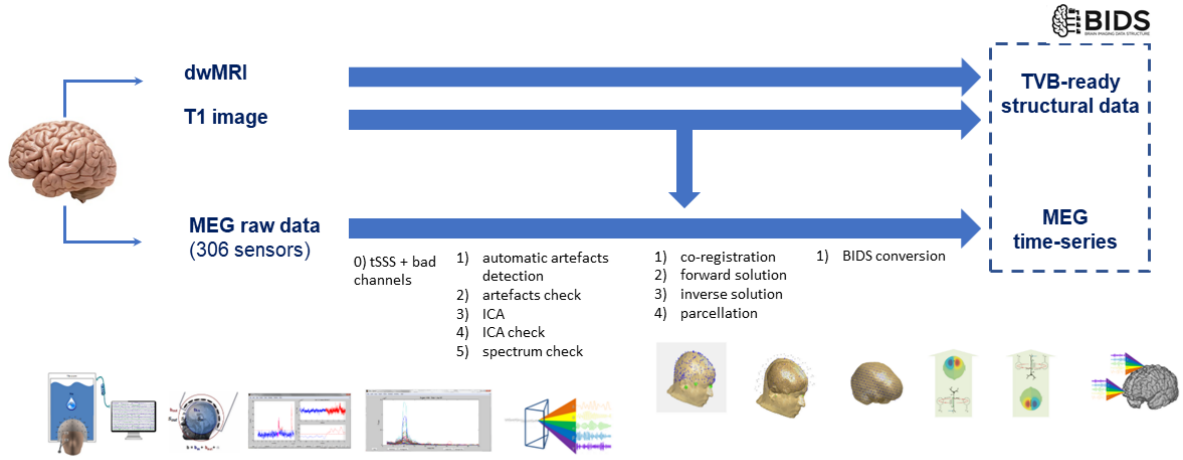


Figure 1. Summary of the MEG workflows.

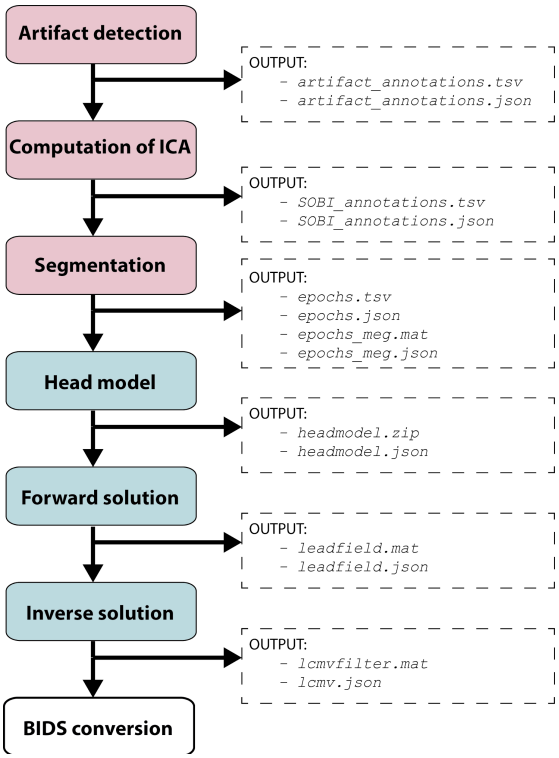


Figure 2. Workflow for the pre-processing and source-reconstruction pipelines and BIDS outputs that can be found in the full-cohort MEG database.

3.2.2 MRI

The T1-weighted images were processed using FreeSurfer v.6.0 *recon-all* function that includes motion correction, intensity non-uniformity correction, intensity normalization, segmentation of the different brain tissues, and constructs a cortical surface mesh for each image. This cortical surface mesh is inflated to a sphere and registered to a common surface-space. An anatomical atlas was used to assign neuroanatomical labels to each native brain voxel. Lastly, the T1-space cortical atlas was registered to each subject's dw-MRI space using FSL *flirt* with 7 degrees of freedom.
dw-MRI data was processed using the MRtrix3 software (version 3.0.2) and included the following steps: dw-MRI denoising, Gibbs ringing artifacts removal, eddy current and movement correction, dw-MRI bias field correction, generation of a tissue-type segmented image for anatomically constrained tractography, and the estimation of the WM, GM, and CSF response functions. The single-Shell 3-Tissue CSD (SS3T-CSD) method was applied to obtain WM-like fiber orientation distributions as well as GM-like and CSF-like compartments in all voxels using the MRtrix3Tissue fork (https://3Tissue.github.io). Finally, we performed multi-tissue informed log-domain intensity normalization, and the generation of the tractogram (25 million streamlines, maximum tract length = 250, FA cutoff = 0.06, dynamical seeding).

3.3 MEG Derivatives

3.3.1 Computation of functional connectivity

Functional connectivity (FC) was estimated under the hypothesis of phase synchronization using the phase-locking value (PLV) algorithm (Lachaux et al. 1999), which demonstrates high reliability across MEG recordings (Garcés et al. 2016). The PLV values between each pair of sources were calculated using the source-reconstructed time-series as the input for all the classical frequency bands. To reduce the dimensionality of the FC matrix, the PLV values were averaged following the Automated Anatomical Labeling/Harvard-Oxford/Schaefer parcellation to obtain, for each parcellation scheme, a single PLV value between each pair of regions.

Despite being widely popular, the PLV encounters a significant limitation due to its susceptibility to volume conduction and source-leakage effects. By considering the strong association between PLV and coherency, it is possible to derive a PLV-based FC measure that is insensitive to volume conduction (given that the imaginary part of coherency discards zero-lag connectivity and thus is insensitive to volume conduction). This particular formulation is referred to as corrected imaginary PLV (ciPLV), and its definition can be found in the work of Bruña et al. (2018). ciPLV was estimated in the classical frequency bands using the source-reconstructed time-series as the input and averaged following the above-mentioned parcellation schemes.

3.3.2 Computation of criticality markers

3.3.2.1 Long-range temporal correlations (LRTCs)

Linear *detrended fluctuation analysis* (DFA) has been widely used as a quantitative measure of LRTCs in narrow-band time-series irrespective of whether they arise from stationary processes or not (Peng, Havlin, Stanley, & Goldberger, 1995). The resulting scaling exponent α is a measure of the autocorrelation properties of a signal, i.e., $\alpha < 0.5$: anti-correlated, $\alpha \approx 0.5$: uncorrelated, and $0.5 < \alpha < 1$: correlated.

The DFA exponents were calculated as in (J. M. Palva et al., 2013; Zhigalov et al., 2015). For each participant, scaling exponents were estimated from parcel time-series, which were narrow-band filtered using a bank of 32 Morlet wavelets (width parameter m = 5) at log-equidistant spacing within a range of 2–90 Hz. The fitting interval included window sizes from 2–25 seconds (as opposed to 0.08–120 seconds) to avoid the influence of filtering artifacts on the scaling exponent estimation. The measures have been extracted for each wavelet frequency and parcel.

3.3.2.2 Functional excitation-inhibition ratio (fE/I)

Custom MATLAB scripts were implemented to compute fE/I as in (Bruining et al., 2020) for each parcel time-series: (i) parcel time-series were wavelet-filtered into 32 narrow-band signals as described above and their amplitude envelopes were extracted; (ii) the cumulative sum of each (demeaned) signal was calculated (the signal profile); (iii) the signal profile was divided by its mean amplitude in fixed windows of 40 cycles, to remove the effect of the original signal magnitude; (iv) each normalized signal profile window was linearly detrended; (v) the standard deviation was calculated for each window to get the windowed-normalized fluctuation function w_nF(t); (vi) mean windowed amplitudes (wAmp) were estimated (as in iii); (vii) the Pearson correlation between w_nF(t) and wAmp was calculated, that is, corr(w_nF(t),wAmp); and (viii) fE/I values were quantified as fE/I=1-corr(w_nF(t),wAmp), so that fE/I = 1 indicates balanced E/I, fE/I < 1 indicates dominant inhibition, and fE/I > 1 indicates dominant excitation.

3.4 Results

3.4.1 Attenuation of LRTCs dissociates early SCD and MCI stages

Mean DFA exponents were obtained by averaging the individual DFA exponents across all parcels separately for the NC, SCD, and MCI groups and showed a broad peak from alpha to beta-frequencies (7–30 Hz) (Fig. 3a). We tested for significant group differences in mean DFA exponents (Fig. 3b) with non-parametric statistics. Both SCD and MCI participants showed lower DFA exponents relative to the NC group from alpha to low-gamma frequencies (7–40 Hz) (Fig. 3a, Kruskal-Wallis test, $p < 0.05$; False Discovery Rate (FDR) corrected across frequencies). Additional post hoc tests revealed a salient progressive attenuation of LRTCs with disease development, with significant differences in DFA exponents between the NC and SCD groups (hereafter, NC-SCD) in the alpha-band (7–12 Hz), and between the SCD and MCI groups (SCD-MCI) in the beta-band (12–22 Hz) (Fig. 3b-c). We next examined the regional specificity of this measure by testing

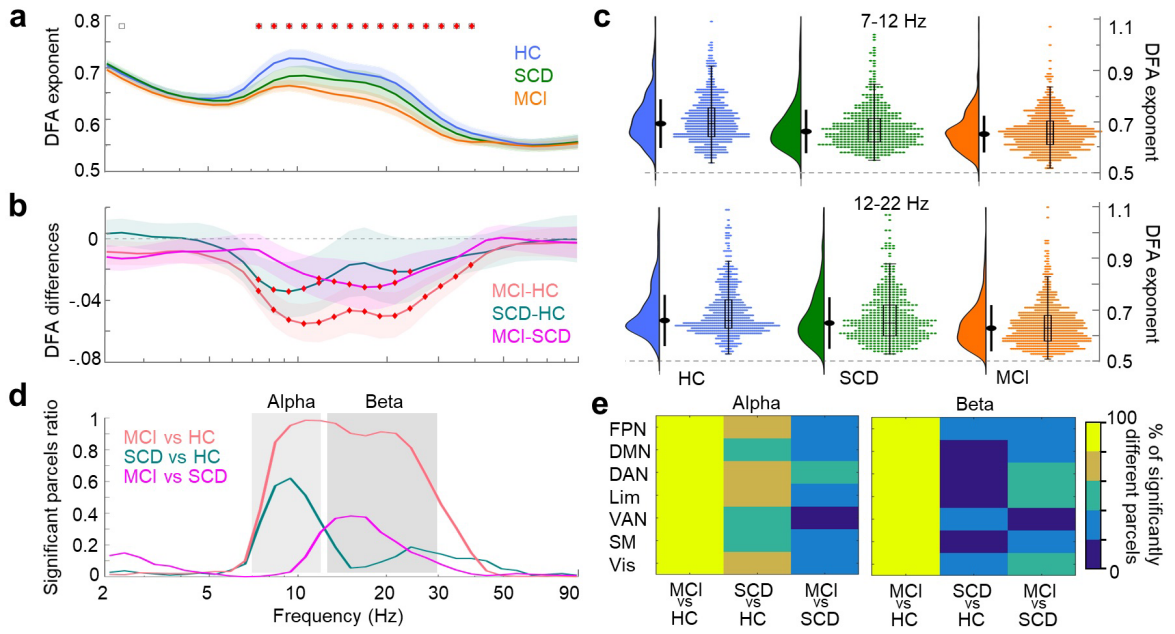


Figure 3. LRTCs dissociate NC, SCD and MCI cohorts. a, Mean DFA exponent, averaged across parcels and within cohorts. Shaded areas represent bootstrapped (n=10,000) 95% confidence intervals. Red diamonds highlight the frequencies with significant differences (Kruskal-Wallis test, p<0.05, FDR corrected). **b,** Pairwise differences between cohorts in averaged DFA exponents. Red diamonds highlight significant differences. **c,** Density plots (left) for DFA exponents averaged within alpha (7-12 Hz) and beta (12-22 Hz) bands where SCD-NC and SCD-MCI differences were found within cohorts, respectively. The black-filled dot denotes median, and the line length represents the standard deviation. Individual participants' DFA exponents (right) with an overlaid boxplot denoting the first, median and third quartiles, and the whisker lengths representing 1.5 times the interquartile range. **d,** Percentage of parcels showing statistically significant differences (Wilcoxon rank-sum test, p<0.05, FDR (Q=20%) corrected) in DFA exponents between NC-SCD, NC-MCI, and SCD-MCI. **e,** Differences between cohorts at the brain functional networks in the alpha and beta bands.

pairwise group differences in DFA exponents at the parcel level at an anatomical resolution of 400 parcels in total. In line with the whole-brain results, differences in DFA exponents between the NC and MCI groups (hereafter, NC-MCI) were extensive and found in nearly all parcels in the alpha and beta bands (Fig. 3d). The NC-SCD and SCD-MCI differences showed expansion of pathological dynamics across disease progression, so that the early NC-SCD deterioration in LRTCs occurred only in the alpha-band, while the progressing SCD-MCI differences expanded to the beta band, together overlapping with the observed NC-MCI differences (Fig. 3d). Region-wise, NC-MCI differences were widespread and observed for all the brain functional systems (Fig. 3e), while NC-SCD and SCD-MCI

differences showed regional specificity: NC-SCD alpha-band differences were visible in the frontoparietal network (FPN), dorsal attentional network (DAN), limbic network (Lim), and visual system (Vis), whereas SCD-MCI beta-band differences were localized within the DAN, Lim, and Vis networks (Fig. 3e).

3.4.2 Elevated excitation characterizes disease progression

Alterations in the E/I ratio have been proposed to be a key driving factor for brain network dysfunction in AD (Busche & Konnerth, 2016). Also, theoretical models of critical brain dynamics suggest that maximum (as opposed to dampened) neural fluctuations occur in the critical (as opposed to the sub- or supercritical) regime

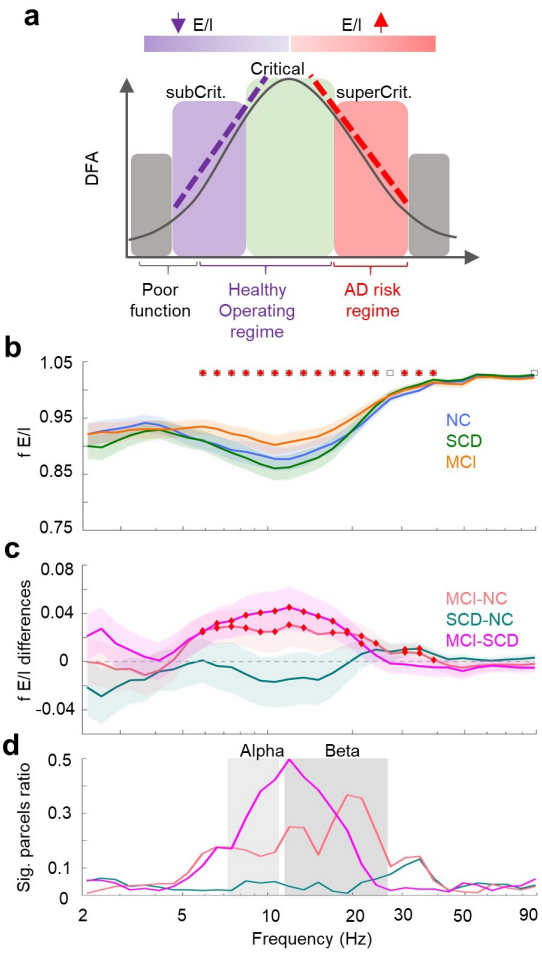


Fig. 4. fE/I balance change across disease trajectory a, Hypothesised dependence of DFA exponents on the brain critical state and the E/I balance in the classical brain criticality framework. Brain operating point is regulated by the E/I balance. DFA exponents peak at the critical transition point (inset: critical-like; green dots, no-linear dependence) at balanced E/I. The subcritical side (inset: purple dots, inhibition- dominant, positive-linear dependence) is characterized by stronger inhibition and the supercritical side (inset: red dots, excitation-dominant, negative-linear dependence) by stronger excitation. Quadratic dependence appears for all points across regimes **b,** Averaged mean fE/I values for each cohort (blue =NC, green = SCD, and orange = MCI) with shaded areas describing 95% confidence interval calculated using bootstrapping (n=10,000) method. Significance as in (Fig. 1b). **c,** Averaged pairwise differences between cohorts in the mean fE/I with 95% confidence intervals in bright and shaded colors, respectively. The diamonds mark the frequencies with significant ($p<0.05$; Kruskal-Wallis, FDR corrected) differences. **d,** Percentage of parcels showing significant differences between cohorts in fE/I. Shaded grey areas highlight frequencies in the alpha and beta bands.

(Cocchi et al., 2017; S. Palva & Palva, 2018). Hence, scale-free behavior and LRTCs should peak in the critical regime, characterized by conditions of balanced E/I, and decrease in the subcritical and supercritical phases, associated with excessive net inhibition and excitation, respectively (Fig. 4a).

To elucidate whether altered E/I balance would characterize SCD and MCI using fE/I measure, we found significant (Kruskal-Wallis test, $p < 0.05$; FDR corrected) group differences in mean fE/I in alpha-to-gamma frequencies (Fig. 4b), similar to our findings in DFA exponents (Fig. 3b). Additional post hoc tests showed that these differences were mainly due to a significant increase of fE/I in the MCI relative to the NC and SCD stages. At the parcel level, we found large NC-MCI and SCD-MCI differences, although not as robust as in DFA analysis. SCD-MCI differences were more widespread, whereas NC-SCD differences were overall weak (Fig. 4e), supporting the idea that E/I alterations echo

disease progression and might represent a pathway of ongoing network disruption beginning at the SCD stage.

4. Computational models

4.1 Hierarchical Kuramoto model

In the classical Kuramoto model a population is defined as multiple coupled all-to-all oscillators mediated by a sinusoidal interaction function with a given frequency for each oscillator

$$\theta_i = \omega_i + \frac{1}{N}\sum_{j=1}^{N} K_{ij}\, sin(\theta_j - \theta_i)$$

Where *N* is total number of oscillators, θ_i is a phase, ω_i is a frequency of i-th oscillator and K_{ij} is the coupling parameter between *i*-th and *j*-th oscillators.

However, the model represents the basic dynamics and for the final analysis data is often aggregated across all oscillators e.g., by computing a model order defined as absolute value of average oscillator phase. This approach makes comparison with the real data a challenging task. To overcome this, we introduced a hierarchical extension of the Kuramoto model with multiple nodes where each node may correspond to a recording from a single brain area or electrode.
Each node is made of multiple oscillators and behavior of each oscillator could be explained in three components:

$$\theta_i = Natural + Internal + External + Noise$$

where $Natural = \omega_i$ or the central frequency of an oscillator,

$$Internal = \frac{K}{N}\sum_{j=1}^{N} sin(\phi_i - \phi_j)$$

where the phase of each oscillator is shifted towards the value, which is calculated as the average of phase differences of oscillators in the same node,

$$External = L_{ij}\sum_{j=1}^{k} W_{ij}\, sin(\phi_i - \Phi_j)$$

where the phase of each oscillator is compared to an average phase of other nodes. Thus K_i is an internal coupling coefficient (local control parameter) between *i*-th node and L_{ij} is a coupling coefficient between *i*-th and *j*-th nodes (global control parameter), and Noise which is a random Gaussian phase distortion with mean of 0 and given scale.

Later, we aggregated signals for each node by averaging phase of all oscillators from the same node thus one is able to match a model's node to a single brain area or any other source of LFP recordings.

4.2 Model statistics

One of the basic statistics of a model is its order - measurement of oscillators synchrony inside each node defined as:

$$R = \left|\frac{1}{N}\sum_{i=1}^{N} \theta_i\right|$$

To estimate inter-areal phase interactions, we computed synchrony metrics for each pair of nodes, we used the phase locking value (PLV), obtained as the absolute value of complex PLV

$$cPLV = \frac{1}{N}\sum_{t=1}^{N}\frac{x(t) * y(t)}{|x(t)|\,|y(t)|}$$

LRTCs were quantified using DFA, which is computed using a timeseries of a model's order for simulated data and, for corresponding real data, the amplitude of a narrow-band filtered MEG data. To speed up the analysis we computed DFA in the Fourier domain (Nolte el al., 2019). The fluctuations were fitted with a robust linear regression with a bisquare weight function to obtain the DFA exponents, all with negligible fit error.
For all simulations in the paper, we used a model with 200 nodes to match the Schaefer parcellation of functional brain areas (Schaefer et al., 2018). We used structural connectome to derive coupling coefficients between different nodes of the model and central frequencies of oscillators were set as random value from the Gaussian distribution with mean of 10Hz and variance of 1.

4.3 Model fitting

Two types of metrics have been used for model fitting: FC (i.e., the PLV) and LRTC. The former, FC, grows monotonically as a function of coupling strength between nodes and therefore shows the gradient direction. The latter, LRTCs behave as a unimodal function with peak in the critical regime and we can approximate it using the quadratic function and optimize it as a function of the local control parameter.
Thus, fitting of coupling strength between nodes is based on PLV loss between target metric and model metric and the sign of difference shows the direction of gradient as well. Equation shows the gradient calculation:

$$grad = \frac{PLV' * loss}{N_{nodes}}$$

where *PLV'* is a derivative of *PLV*

The dependency of DFA over *K* was assumed as an inverted parabola where the critical state is an extremum of the DFA graph over. Thus, the derivative of a parabola is used for fitting. To get the direction of the gradient, the sign of the PLV difference was used. Equation shows the gradient calculation:

$$grad = \frac{-2 * loss * direction * K}{N_{nodes}}$$

The learning rate was optimized with *RMSprop* dividing coefficient by the root of a squared gradient, where the squared gradient is:

$$squared\ grad = 0.9 * squared\ grad + 0.1 * grad^2$$

where 0.9 is the default value for the moving average parameter.
Thus the equation for the parameter update:

$$P = P - \frac{learning\ rate}{\sqrt{squared\ grad}} * grad$$

4.4 Integration of the brain dynamics measures into the database for model fitting

Following the BIDS standard, the data is structured in a main folder (VBC_Madrid_BIDS) and a derivatives subfolder comprising the derived data generated by the different processing pipelines (derivatives/pipeline preprocessing and derivatives/pipeline sources). In addition, an extra folder (derivatives/TVB) has been added to the database for computational modeling purposes (see deliverable 3.5). The latter, so far composed of a set of TVB-compliant structural data and the time-series associated with the ROIs, has been updated with brain dynamics measures to serve as a compact tool for model fitting (Fig.5).

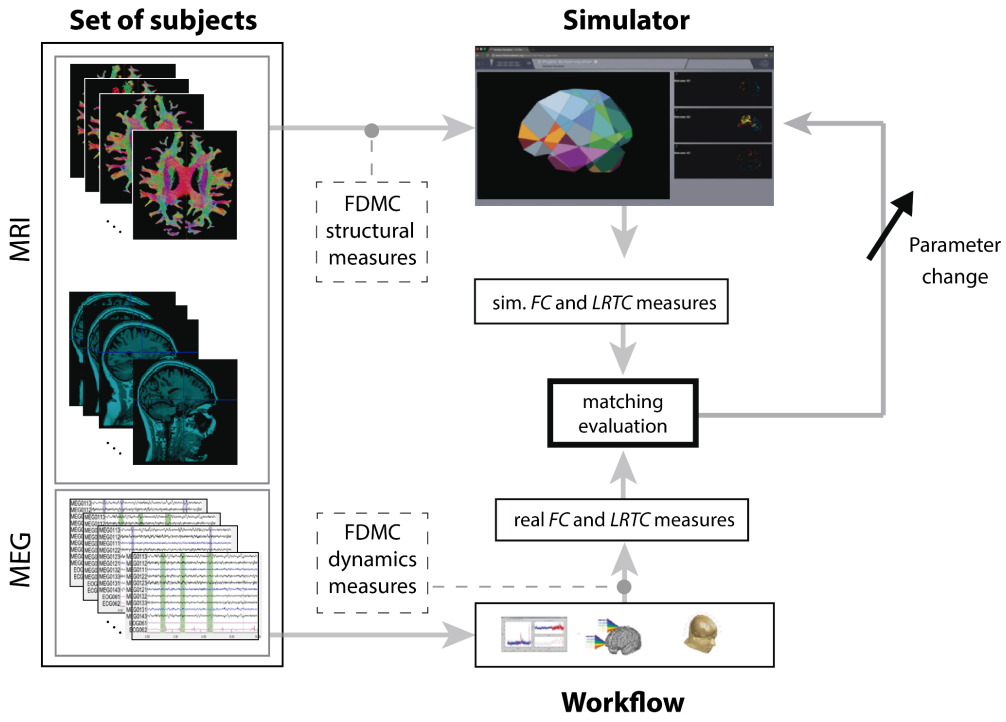


Fig. 5. Scheme of the simulation procedure and model fitting.

4.5 Results

Firstly, we investigated how both parameters affect model dynamics. We varied *K* and *L* parameters and computed model standard deviation, LRTC and synchrony between nodes for each combination of control parameters.

We found that increasing both global and local parameters lead to increase in model synchrony between nodes where low control parameters cause almost no inter and intra synchrony while high values lead the model to almost full synchronization. DFA and standard deviation shows a diagonal critical ridge indicating the model order transition period.

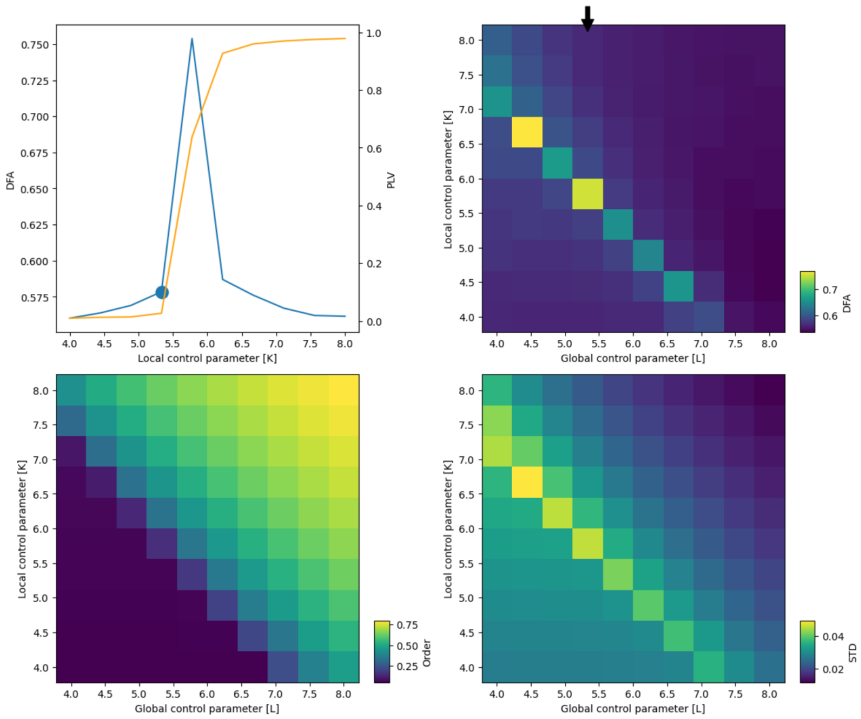


Fig. 6. Model metrics. The dependency of PLV and DFA metrics over local control parameter K. The distribution of DFA, PLV, and STD over the combination of control parameters.

4.5.1 Fitting to a predefined phenotype

At first, we fit a model to the artificial data with known properties, simulated ground truth. For that purpose, we generated several phenotypes by changing baseline parameters such as structural connectome or vector of *K* values, simulated time series using phenotype model parameters and fit the model using starting parameters to it.

We found that after those alterations we are able to reproduce the model activity with high similarity (Pearson's correlation coefficient ~ 0.8) and to restore the phenotype parameters using both FC and LRTC as the target variable.

Figure 7 shows an example of the fitting *Vis* phenotype *Le* matrix based on the PLV metric. *Le* matrix alternate weights of structural connectome according to target functional areas (Visual network). The ground truth (GT) is simulated using connectome multiplied by phenotype target weights (Figure 7). There are 25 runs of the fitting thus the plots of fitted PLV and fitted weights show the average value. The correlation between target and fitted metrics raised to 0.8 during fitting steps on average. We can notice that the changes in fitted weights are similar to the target weights.

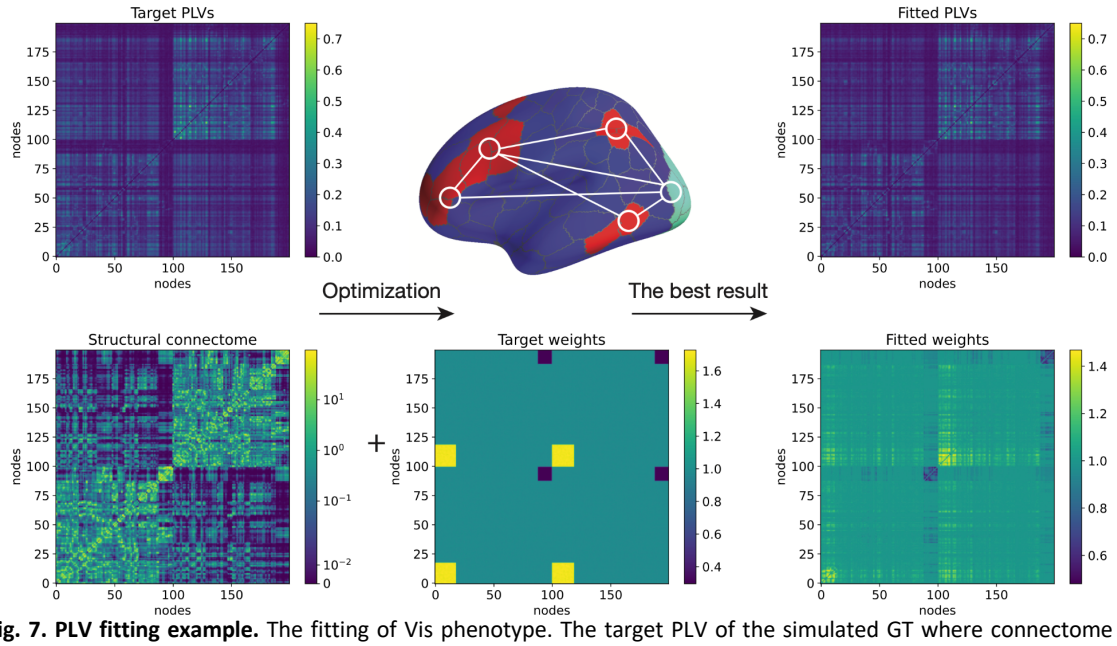


Fig. 7. PLV fitting example. The fitting of Vis phenotype. The target PLV of the simulated GT where connectome was multiplied by phenotype target weights. The fitted PLV and fitted weights show the average value of 25 fitting runs.

We used the detrended fluctuation analysis (DFA) which analyzes time-series of a signal for individual *K* fitting. Figure 8 shows an example of the fitting Default phenotype. For the nodes in the target area, the coupling coefficient of n^{th} node (K_n) was higher, and the mean of all nodes ($K_n s$) was 1. Each fitting run consists of 25 steps, and the correlation between target DFA and fitted DFA raised from 0.63 to 0.83 during these steps on average. After fitting, individual $K_n s$ demonstrate a similar increase in target areas (Figure 8).

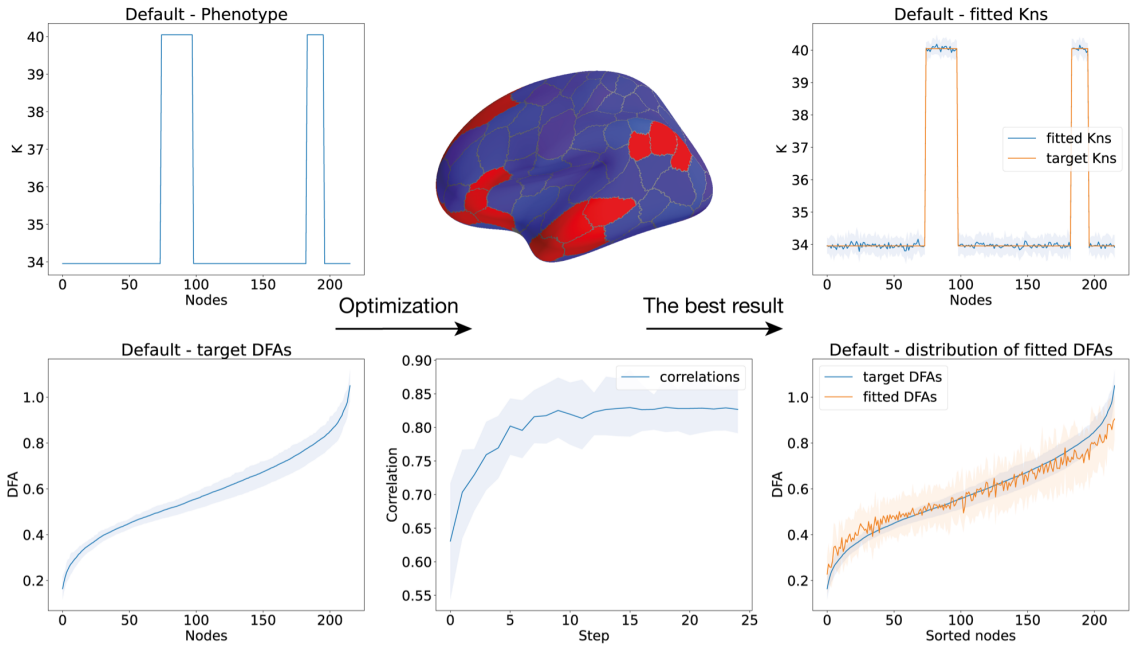


Fig. 8. DFA fitting example. The fitting of Default phenotype. Kns increase in the target nodes of Default network, mean of Kns for all nodes is 1. Fitted Kns demonstrate increase at the same nodes. There are 25 runs of the fitting thus the plots show the mean line and the spread is a standard deviation. The fitted DFA is aligned with the target DFA. The Pearson correlation coefficient rises up to 0.83 at average.

4.5.2 Fitting to a real data

As the next step we derived digital twins - a set of models made to reproduce dynamics of the real data. We filtered MEG data into equally log-spaced frequencies from 2Hz to 100Hz, computed wPLI, DFA and fit the model for the 10Hz frequency.

We found that after fitting the model is able to reproduce dynamics with fairly high similarity (intraclass correlation (ICC) = 0.47 for PLV and ICC = 0.4 for DFA, Figure 9). For the real data fitting, we use dual-gradient-descent model fitting to combine both gradients for global and local parameters. In Figure 9, the fitted DFA demonstrates a similar increase in the occipital lobe to the real DFA with ICC 0.4. The fitted PLV shows the increase in the synchronization between nodes similarly to the real wPLI after weight fitting and shows ICC 0.47.

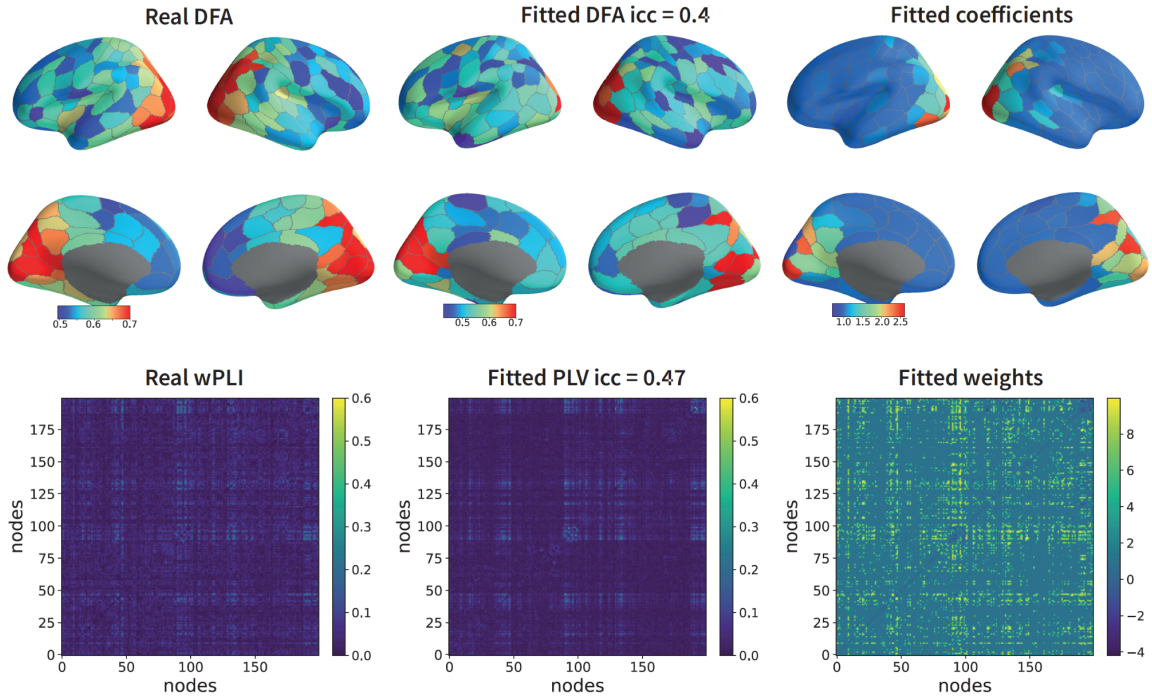


Fig. 9. Fitting to the real data example. The distribution of the real DFA values over brain areas. The fitted DFA demonstrates intraclass correlation (ICC) equal to 0.4 for DFA. The distribution of the fitted coefficients over brain areas. The real wPLI and the fitted PLV that demonstrates ICC equal to 0.47. The fitted weights for the model to demonstrate similar PLV.

5. Conclusions

In this document we detail the workflow adopted for the final computation of brain dynamics measures of the FDMC dataset, and discuss their importance in characterizing early stages of Alzheimer's disease.

The FDMC dataset, arranged using the BIDS standard to foster interoperability and to address the heterogeneity of data organization, has been importantly extended to allow the easy integration with computational models of large-scale brain dynamics. Specifically, an extra folder (derivatives/TVB), which has been previously inserted in the database for computational modeling purposes, has now been updated with brain dynamics measures to serve as a compact tool for model fitting.

We show how the attenuation of LRTCs (assessed with DFA) is able to dissociate early SCD and MCI stages, and how elevated excitation (assessed with fEI) characterizes disease progression. Further, with the aim of computational modeling, we showed that we are able to reproduce the model activity with high similarity and to restore the phenotype parameters using both FC and LRTC as the target variable.

Importantly, the subset of TVB-compliant data enables personalized simulations, and the FDMC on the whole can be used as a test bench for computational neuroscience methods and machine learning within the TVB-Cloud project.

6 Glossary

AD Alzheimer's disease

BIDS Brain Imaging Data Structure

CFS cross-frequency synchrony

CSF cerebrospinal fluid

DAN dorsal attentional network

DFA detrended fluctuation analysis

dw-MRI diffusion-weighted magnetic resonance imaging

EOG electrooculogram

EKG electrocardiogram

FC functional connectivity

FDR false discovery rate

FPN frontoparietal network

fE/I functional excitation-inhibition ratio

GM grey matter

GT ground truth

HC healthy control

HPI head position indicator

IC independent component

ICA independent component analysis

Lim limbic network

LFP local field potential

LNCyC Laboratory of Cognitive and Computational Neuroscience

LRTCs long-range temporal correlations

MCI mild cognitive impairment

MEG magnetoencephalography

MNE minimum norm estimates

MNI Montreal Neurological Institute

MRI magnetic resonance imaging

NIA–AA National Institute on Aging–Alzheimer's Association

NINCDS–ADRDA National Institute of Neurological and Communicative Disorders and Stroke–Alzheimer's Disease and Related Disorders Association

PAC phase-ampitude coupling

PLV phase-locking value

ciPLV corrected imaginary PLV

ROI region of interest

SCD subjective cognitive decline

SNR signal-to-noise ratio

tSSS temporal signal space separation

TVB The Virtual Brain

UCM Universidad Complutense de Madrid

UH University of Helsinki

Vis visual system

WM white matter

References

Bruining, H., Hardstone, R., Juarez-Martinez, E. L., Sprengers, J., Avramiea, A. E., Simpraga, S., ... & Linkenkaer-Hansen, K. (2020). Measurement of excitation-inhibition ratio in autism spectrum disorder using critical brain dynamics. *Scientific Reports*, *10*(1), 9195.

Bruña, R., Maestú, F., & Pereda, E. (2018). Phase locking value revisited: teaching new tricks to an old dog. *Journal of Neural Engineering*, *15*(5), 056011.

Bruining, H., Hardstone, R., Juarez-Martinez, E. L., Sprengers, J., Avramiea, A. E., Simpraga, S., ... & Linkenkaer-Hansen, K. (2020). Measurement of excitation-inhibition ratio in autism spectrum disorder using critical brain dynamics. *Scientific Reports*, *10*(1), 9195.

Cocchi, L., Gollo, L. L., Zalesky, A., & Breakspear, M. (2017). Criticality in the brain: A synthesis of neurobiology, models and cognition. *Progress in Neurobiology*, *158*, 132-152.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage*, *9*(2), 179-194.

Garcés, P., Martín-Buro, M. C., & Maestú, F. (2016). Quantifying the test-retest reliability of magnetoencephalography resting-state functional connectivity. *Brain Connectivity*, *6*(6), 448-460.

Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, *8*(4), 194-208.

Niso, G., Gorgolewski, K. J., Bock, E., Brooks, T. L., Flandin, G., Gramfort, A., ... & Baillet, S. (2018). MEG-BIDS, the brain imaging data structure extended to magnetoencephalography. *Scientific Data*, *5*(1), 1-5.

Nolte, G. (2003). The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Physics in Medicine & Biology*, *48*(22), 3637.

Nolte, G., Aburidi, M., & Engel, A. K. (2019). Robust calculation of slopes in detrended fluctuation analysis and its application to envelopes of human alpha rhythms. *Scientific Reports*, *9*(1), 1-16.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 1-9.

Buldyrev, S. V., Goldberger, A. L., Havlin, S., Mantegna, R. N., Matsa, M. E., Peng, C. K., ... & Stanley, H. E. (1995). Long-range correlation properties of coding and noncoding DNA sequences: GenBank analysis. *Physical Review E*, *51*(5), 5084.

Palva, J. M., Zhigalov, A., Hirvonen, J., Korhonen, O., Linkenkaer-Hansen, K., & Palva, S. (2013). Neuronal long-range temporal correlations and avalanche dynamics are correlated with behavioral scaling laws. *Proceedings of the National Academy of Sciences*, *110*(9), 3585-3590.

Palva, S., & Palva, J. M. (2018). Roles of brain criticality and multiscale oscillations in temporal predictions for sensorimotor processing. *Trends in Neurosciences*, *41*(10), 729-743.

Pusil, S., López, M. E., Cuesta, P., Bruna, R., Pereda, E., & Maestú, F. (2019). Hypersynchronization in mild cognitive impairment: the 'X' model. *Brain*, *142*(12), 3936-3950.

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X. N., Holmes, A. J., ... & Yeo, B. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral Cortex*, *28*(9), 3095-3114.