Neural Networks 170 (2024) 72–93

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Full Length Article

A recurrent Hopfield network for estimating meso-scale effective connectivity in MEG

Giorgio Gosti [a,b,c,*], Edoardo Milanetti [d,a], Viola Folli [a,e], Francesco de Pasquale [f], Marco Leonetti [a,b,e], Maurizio Corbetta [g,h,i], Giancarlo Ruocco [a,d,1], Stefania Della Penna [j,1]

[a] Center for Life Nano- & Neuro-Science, Istituto Italiano di Tecnologia, Viale Regina Elena, 291, 00161, Rome, Italy
[b] Soft and Living Matter Laboratory, Institute of Nanotechnology, Consiglio Nazionale delle Ricerche, Piazzale Aldo Moro, 5, 00185, Rome, Italy
[c] Istituto di Scienze del Patrimonio Culturale, Sede di Roma, Consiglio Nazionale delle Ricerche, CNR-ISPC, Via Salaria km, 34900 Rome, Italy
[d] Department of Physics, Sapienza University of Rome, Piazzale Aldo Moro, 5, 00185, Rome, Italy
[e] D-TAILS srl, Via di Torre Rossa, 66, 00165, Rome, Italy
[f] Faculty of Veterinary Medicine, University of Teramo, 64100 Piano D'Accio, Teramo, Italy
[g] Department of Neuroscience, University of Padova, Via Belzoni, 160, 35121, Padova, Italy
[h] Padova Neuroscience Center (PNC), University of Padova, Via Orus, 2/B, 35129, Padova, Italy
[i] Veneto Institute of Molecular Medicine (VIMM), Via Orus, 2, 35129, Padova, Italy
[j] Department of Neuroscience, Imaging and Clinical Sciences, and Institute for Advanced Biomedical Technologies, "G. d'Annunzio" University of Chieti-Pescara, Via Luigi Polacchi, 11, 66100 Chieti, Italy

ARTICLE INFO

Keywords:
Recurrent neural networks
Hopfield networks
Magnetoencephalography
Effective connectivity
Generative models

ABSTRACT

The architecture of communication within the brain, represented by the human connectome, has gained a paramount role in the neuroscience community. Several features of this communication, *e.g.*, the frequency content, spatial topology, and temporal dynamics are currently well established. However, identifying generative models providing the underlying patterns of inhibition/excitation is very challenging. To address this issue, we present a novel generative model to estimate large-scale effective connectivity from MEG. The dynamic evolution of this model is determined by a recurrent Hopfield neural network with asymmetric connections, and thus denoted Recurrent Hopfield Mass Model (RHoMM). Since RHoMM must be applied to binary neurons, it is suitable for analyzing Band Limited Power (BLP) dynamics following a binarization process.

We trained RHoMM to predict the MEG dynamics through a gradient descent minimization and we validated it in two steps. First, we showed a significant agreement between the similarity of the effective connectivity patterns and that of the interregional BLP correlation, demonstrating RHoMM's ability to capture individual variability of BLP dynamics. Second, we showed that the simulated BLP correlation connectomes, obtained from RHoMM evolutions of BLP, preserved some important topological features, *e.g*, the centrality of the real data, assuring the reliability of RHoMM.

Compared to other biophysical models, RHoMM is based on recurrent Hopfield neural networks, thus, it has the advantage of being data-driven, less demanding in terms of hyperparameters and scalable to encompass large-scale system interactions. These features are promising for investigating the dynamics of inhibition/excitation at different spatial scales.

1. Introduction

The functional organization of the human brain at rest plays a fundamental role in cognition. Network neuroscience and graph theory (Bassett & Sporns, 2017) show that a healthy connectome has a specific topology characterized by the presence of functional and structural modules connected through hubs with a specific temporal dynamic (Carbo et al., 2017; de Pasquale, Corbetta, Betti, & Della Penna, 2018; de Pasquale et al., 2012; de Pasquale, Della Penna, Sporns, Romani, & Corbetta, 2016; de Pasquale, Spadone, et al., 2021; Della Penna, Corbetta, Wens, & de Pasquale, 2019; Gu et al., 2015; Tang

* Corresponding author at: Center for Life Nano- & Neuro-Science, Istituto Italiano di Tecnologia, Viale Regina Elena, 291, 00161, Rome, Italy.
E-mail addresses: giorgio.gosti@iit.it (G. Gosti), edoardo.milanetti@uniroma1.it (E. Milanetti), Viola.Folli@iit.it (V. Folli), fdepasquale@unite.it (F. de Pasquale), marco.leonetti@iit.it (M. Leonetti), maurizio.corbetta@unipd.it (M. Corbetta), giancarlo.ruocco@roma1.infn.it (G. Ruocco), stefania.dellapenna@itab.unich.it (S. Della Penna).
[1] Authors contributed equally.

https://doi.org/10.1016/j.neunet.2023.11.027
Received 17 February 2023; Received in revised form 7 November 2023; Accepted 9 November 2023
Available online 11 November 2023

et al., 2017). This topology balances segregated and integrated brain functions (Bullmore & Sporns, 2012; Sporns, 2013), and it represents a trade-off between the cost of such properties and their functionality. However, the complexity of this architecture makes it very challenging to unravel its principles of organization (Betzel & Bassett, 2017). This is the aim of network generative modeling, in which wiring rules are identified and algorithmically implemented to produce synthetic network architectures with the same properties, i.e., summary statistics and topology, of the collected data (Magrans de Abril, Yoshimoto, & Doya, 2018; Sporns & Kötter, 2004). Once trained, a generative model can predict unobserved and out-of-sample data. Moreover, a generative model can reveal mechanisms guiding the formation of a system and its hierarchical structure. This feature is particularly important for the ultimate and ambitious aim of neuroscience consisting of perturbing the brain connectome in a targeted and predictable way (Gu et al., 2015) (but see Tu, Rocha, Corbetta, Zampieri, Zorzi, & Suweis, 2018).

To identify a generative model, a set of neural models of communication must be trained, *e.g.*, by exploiting the available knowledge on the structural/functional connectivity or on the dynamics of activity. Different approaches can be found in the literature: assumption-based (Friston, Li, Daunizeau, & Stephan, 2011; Vidaurre, Smith, & Woolrich, 2017) or data-driven (Singh, Braver, Cole, & Ching, 2020) generative models; certain generative models form synthetic networks by combining geometrical rules and topological properties (Betzel et al., 2016); other micro-scale nonlinear models of spontaneous neuronal activity produce large spatial scale interactions (at the level of cerebral networks) spanning different temporal scales (Honey, Kötter, Breakspear, & Sporns, 2007). More recently, Graph Neural Networks (GNN), *i.e.*, deep neural networks that operate in the graph domain Zhou et al. (2020), were proposed to model neural dynamics. Notably, the edges in a GNN would have to capture implicit knowledge of the data and affect the dynamics in a non-trivial way as discussed in Zang and Wang (2020) where the dynamical equation is modeled by a GNN. This approach may be very successful, but we may not be able to interpret all the parameters provided by the GNN in addition to the connectivity weights, as in the case of Monti, Fiorentino, Milanetti, Gosti, and Tartaglia (2022).

Among these approaches, Neural Mass Models (NMMs), in their standard implementation, use an approximate model for small groups of neurons to provide the activity dynamics within small portions of the cortex or a limited number of portions, allowing the description of the local architecture. They are extremely powerful since they constitute the basic elements for modeling mesoscale brain networks based on their interactions (Breakspear, 2017). Different NMMs have been developed to explain multiple aspects of Resting State Networks (RSNs) in fMRI data (Cabral, Kringelbach, & Deco, 2017; Deco, Jirsa, & McIntosh, 2011). Specifically, among the NMMs, a variant of the Kuramoto model with coupled and delayed local oscillators was able to reproduce both fMRI RSN patterns (Cabral, Hugues, Sporns, & Deco, 2011), and the spectral content observed by MEG connectivity (Cabral, Kringelbach, & Deco, 2014). Another popular class of generative models is Dynamic Causal Modeling (DCM). In this case, NMMs still represent basic elements producing models of observable responses through the estimation of effective connectivity (EC) within a Bayesian framework (Friston et al., 2011). An effective network is defined as one of the networks associated with one or more minimal neural circuit models reproducing the observed time course dynamics, and quantifying how the connected regions influence the activity dynamics of each other (Friston, 2011). DCM is widely used for fMRI (Friston et al., 2019) and EEG/MEG (Kiebel, Garrido, Moran, & Friston, 2008). Notably, both classes of models rely on a complex biophysical model of the NMMs and have expensive computational burdens, especially if applied to MEG data. Thus, these models are mostly used for small networks. Moreover, in most cases, a set of simplifying assumptions is applied to reduce their computational complexity and to apply them to larger networks. However, these assumptions make the model less realistic.

Even though the above NMMs are suited for a limited number of nodes, they are widely used to model mesoscale networks, while a popular model for small-scale neuronal networks is the recurrent Hopfield neural network (Hopfield, 1982). It is a network based on binary McCulloch-Pitts neurons (Amit, Gutfreund, & Sompolinsky, 1985; Brunel, 2016; Folli, Gosti, Leonetti, & Ruocco, 2018; Gosti, Folli, Leonetti, & Ruocco, 2019; Hillar, Chan, Taubman, & Rolnick, 2021; Hillar & Tran, 2018; Leonetti, Folli, Milanetti, Ruocco, & Gosti, 2020), in which every processing unit is connected to all the other ones through a set of weights. Most of the literature considers fully connected and symmetric Hopfield networks since this simplifies the mathematical analysis of the model, although recently the investigation of diluted networks (Brunel, 2016; Kim, Park, & Kahng, 2017) and asymmetric networks (Folli et al., 2018; Leonetti et al., 2020) has been introduced. In most cases, a recurrent Hopfield neural network is typically trained by the Hebbian learning principle. This states that the coactivation of two nodes, *i.e.*, sharing the same state at the same time, will strengthen their connection. This is inspired by biological neural networks where the simultaneous activation of neurons leads to increments in synaptic strength (Hebb, 1949). This training allows the network to store many patterns or memories mimicking the capacity of an associative network (Amit et al., 1985; Folli, Leonetti, & Ruocco, 2017). The Hebbian learning model is widely used to explain local mechanisms of memory storage (Lansner, 2009). Although this model was introduced some years ago, direct in-vivo applications of a recurrent Hopfield neural network to identify learning rules in a network of firing neurons were only recently reported (Pereira & Brunel, 2018). Another approach for storing neural patterns using recurrent Hopfield neural networks is based on Perceptron learning and the minimization of an energy function (Brunel, 2016). This approach has often been used to store stationary patterns. Further, it has never been systematically studied to store dynamic sequences, although this property was theoretically recognized.

In this work, we developed a novel coarse-grained generative model, which is simpler than the above-discussed approaches to estimate the effective connectivity that encodes the temporal dynamics of MEG BLP. Specifically, we designed a binary mass model with dynamics determined by a recurrent Hopfield network with asymmetric connections. Thus, we denote this model as the Recurrent Hopfield Mass Model (RHoMM). This approach is data-driven since it does not rely on the adoption of a specific biophysical model or priors, *e.g.*, structural connectivity (Cabral, Luckhoo, et al., 2014; Frässle et al., 2017; Friston, 2011), and it is based on a minimal number of assumptions that reflect specific features of the data. These assumptions are: (i) the BLP data could be binarized preserving their natural nonlinearity and the associated functional connectivity (FC) patterns estimated through BLP correlation; (ii) the activity at a certain step is predicted only by the previous one. Moreover, to make RHoMM a generative model that is able to predict the BLP dynamics, we applied a strategy based on prediction error minimization.

We applied RHoMM to BLP time courses from ten healthy subjects, during fixation. Each unit was a node representing the activity of a brain parcel belonging to one of a set of RSNs covering the entire cortex. Finally, to validate RHoMM generative performance, we compared the BLP correlation connectomes obtained from theoretical evolutions produced by RHoMM with the experimental ones.

2. Materials and methods

The overall concept of RHoMM is described in Fig. 1. RHoMM is a binary mass model applied to activities (binarized BLP time series) generated by mesoscale regions (voxels) in the brain (Fig. 1 bottom middle) to generate the output at time sample t based on the previous sample $(t-1)$. It implements a recurrent network with Hopfield-like dynamics (Fig. 1, top left), where the network neurons are McCulloch-Pitts all-to-all asymmetrically connected. At each time sample, the state

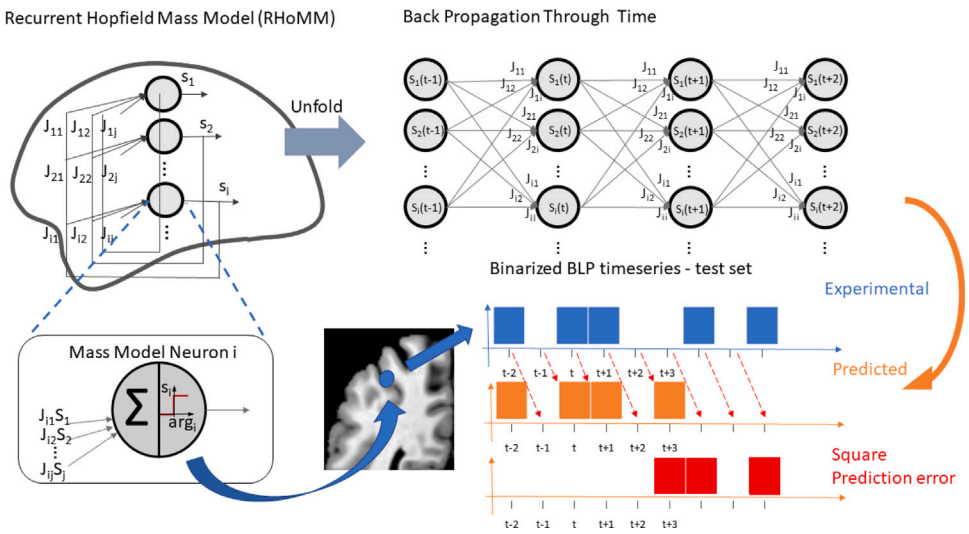

Fig. 1. Conceptual layout of RHoMM. RHoMM is a recurrent network (top left) trained as a feedforward network on time sample pairs (with the sample (t-1) as input and the sample t as output) (top right). The network neurons are associated with regions in the brain (bottom middle) and are regulated by a weighted sum of the inputs and an activation function (bottom left). The predicted binary time series are compared to the experimental binarized BLP sequences to obtain the square predicted error which is used to compute the number of incorrectly predicted items and the training cost function (bottom right).

of the network neurons (active-1 or rest-0) is set by a binary activation function on the base of a weighted sum of the input connections (Fig. 1 bottom left). The model network is trained as a feedforward network on the single transition (from the sample (t-1) to the sample t) using backpropagation through time (Fig. 1 top right). The predicted and experimental binary BLP sequences are compared to retrieve the model parameters in the training set. Specifically, in this training phase, we initialized RHoMM with random parameters. Then, we gradually tuned the network until it recovered the observed brain dynamics with as few errors as possible. The obtained EC matrix is applied to the test set to produce predicted binary BLP sequences (Fig. 1 bottom right). Details are provided in the following subsections.

The entire pipeline applied in this study is illustrated in Fig. 2 and can be summarized as follows. First, MEG data were preprocessed, at the source level, to obtain the Band Limited Power (MEG BLP) time courses. Second, BLP time courses were binarized (0-resting, 1-active) to be assumed as the activity of nodes of the network model. The binarized data were divided into two chunks: the training and the test set. Third, to train RHoMM, we minimized the prediction error on the training set, and in this way, we estimated the model parameters. Fourth, we tested the trained network on the remaining part of the data, the test set. In this way, we estimated the effective connectivity among the considered nodes. This estimated network was used to generate BLP dynamics that should replicate the functional properties of the acquired data. Thus, to validate the proposed approach, we derived BLP correlation connectomes from both experimental and estimated time series. We compared them in terms of some topological features typically reported in literature, *e.g.*, the global architecture and betweenness centrality.

2.1. MEG data acquisition and preprocessing

In this study, a subset (10 subjects) of the data published in de Pasquale et al. (2010) was used. Each subject underwent two or three resting-state runs (lasting 3 or 5 min) during which the subjects maintained fixation on a small visual target. In total, we analyzed 24 runs. Neuro-magnetic signals (filter settings 0.16–250 Hz, 1025 Hz sampling rate) were recorded using the 153-magnetometer MEG system installed in a 4-layer magnetically shielded room at the University of Chieti (Della Penna et al., 2000). Two ECG and two EOG channels were recorded simultaneously with the MEG signals and were used for offline artifact rejection. After each run, the subject's head position relative to the MEG sensor was estimated from the field produced by five coils placed on the scalp, whose positions were digitized together with anatomical landmarks (nasion, left and right preauricular, vertex) using a 3D digitizer (3Space Fastrak; Polhemus). A full description of the preprocessing steps to obtain Band Limited Power (BLP) time series can be found in Mantini et al. (2011) and it was extensively applied to resting-state (de Pasquale et al., 2010, 2012; Marzetti et al., 2013), and event-related (Favaretto et al., 2021; Larson-Prior et al., 2013; Sebastiani et al., 2014) studies. Here, we summarize the basic steps. First, environmental and physiological (*e.g.*, cardiac, ocular) artifacts were identified from the sensor space MEG time series using an ICA-based approach, and only ICs that were not artifacts were retained for further analysis. Then, the sensor maps of the non-artifactual ICs maps were projected on a 3D Cartesian grid (64 mm^3 cubic voxels) using a Weighted Minimum Norm Least Square (WMNLS) estimation. For every voxel in the 3D grid, the MEG activity was represented by a current dipole with 3 orthogonal directions. The individual 3D grid was then projected into the MNI 152 atlas space so that every voxel centroid was assigned to a set of MNI coordinates. The 3 directions of the MEG activity were obtained at each voxel from the linear combination of IC time courses weighted by the corresponding source maps. Source-space signals were filtered in the alpha [7-14 Hz] and beta [14–25 Hz] bands, as in previous studies (Brookes et al., 2011; de Pasquale et al., 2018, 2012, 2016; Hipp, Hawellek, Corbetta, Siegel, & Engel, 2012) it was shown that the strongest interactions within/across RSNs occur in these bands. Source-space power time series were estimated over 200 ms windows sliding every 20 ms to cover the entire resting-state run. In this study, the BLP was extracted on a parcellation scheme composed of 45 nodes belonging to 8 Resting State Networks (RSNs): Dorsal and Ventral Attention (D/VAN), Visual (VIS), Somatomotor (SMN), Default Mode (DMN), Fronto Parietal (FPN), Salience (SAL) and Executive Control (EXE). Based on Yeo 17-network parcels (Thomas Yeo et al., 2011), we selected seed voxels representing the geometrical centroid of the selected parcels listed in Table 1. The selected voxels covered the whole brain, as shown in Fig. A.1. The seed BLP time courses constituted an n-dimensional vector $x(t)$ which defined a brain's state at time t, to be used to estimate the EC matrix.

2.2. Data binarization

The fundamental computational unit of our coarse-grained neural mass model is the McCulloch-Pitts neuron, which was associated with

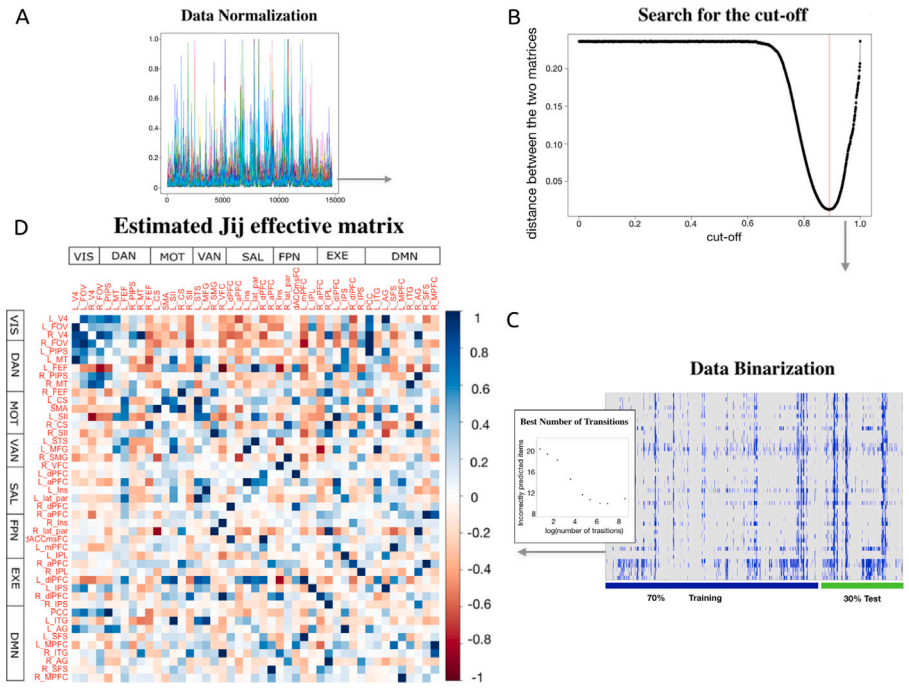

Fig. 2. Analysis pipeline for effective matrix estimation and statistical analysis. (A) For a representative subject, BLP Time courses for each node. (B) Distance between the individual correlation matrix for the continuous data C_{con} and the one estimated from the binary data $C_{bin}(\tau)$ as a function of the binarization threshold τ. The red vertical line illustrates the optimal threshold value τ_{opt} for which the distance between C_{con} and $C_{bin}(\tau)$ is minimal. (C) Binarized signal, obtained setting all experimental values acquired, at any time, to 0 or 1, compliant with the McCulloch-Pitts neuron model. In the insertion, it is reported the average of the incorrectly predicted items as obtained by the application of the estimated EC matrices on corresponding experimental data as a function of the number of transitions (or time samples) used for training the matrices. This allows us to identify the optimal number of samples that allows the effective matrix estimation. The average is calculated considering all runs for each subject. (D) Average estimated effective matrix for a single subject computed on multiple subsets of data.

each seed voxel. This can assume only two discrete values: zero when at rest and one when active. Thus, we had to transform the BLP time series into binary series. This was possible since, as demonstrated by Tagliazucchi, Balenzuela, Fraiman, and Chialvo (2012) with fMRI data, the BLP time course binarization most likely preserves the functional connectivity dynamics. To determine the optimal threshold for the binarization in each run, we proceeded as follows. First, for each seed voxel i, we divided its BLP time series $x_i(t)$ by its maximum value, so that all the time courses ranged between 0 and 1. Second, we computed the binarized time course at a fixed threshold value τ. Third, we varied the threshold values to minimize the information loss associated with the binarization of the BLP data. To this end, we calculated the Euclidean distance between the BLP and the binarized Pearson correlation matrices at each threshold value (Fig. 2B). The optimal threshold value was associated with the minimal distance:

$$\tau_{opt} = \underset{\tau\in[0,1]}{\operatorname{argmin}} \sum_{ij} (C_{con} - C_{bin}(\tau))^2, \tag{1}$$

where C_{con} and $C_{bin}(\tau)$ are respectively the Pearson correlation matrices of the BLP time series and the binarized time series. To train the RHoMM, we used the optimally binarized BLP timeseries, $s_i(t)$.

2.3. The recurrent Hopfield mass model

We denote with J_{ij} the elements of the effective matrix indicating the influence that node j has on node i. This EC matrix, to be estimated by RHoMM, should be able to reproduce the observed binarized time course.

In this coarse-grained model (see Fig. 1 top left), at time step t, the evolution of the brain state $s_i(t)$ in each McCulloch Pitts neuron (see Fig. 1 bottom left) is described by a non-linear dynamic equation determined by the recurrent Hopfield neural network

$$s_i(t) = \theta\left[\sum_{j=1}^{n} J_{ij} s_j(t-1)\right], \tag{2}$$

where n is the number of neurons and $\theta(x)$, the activation function, is the Heaviside step function ($\theta(x) = 1$ for $x \geq 0$, and $\theta(x) = 0$ otherwise). If the summation of the inputs on neuron i is above zero, the neuron is active, otherwise, it is silent.

2.4. Estimation of the effective network

We estimated the effective network by training RHoMM with Back-propagation Through Time (BPTT) (see Fig. 1 top right) to minimize a given cost function (Werbos, 1990). BPTT is a training method based on a gradient descent approach that trains a given recurrent network as a feed-forward neural network, with a database composed of the observed system transitions $(\boldsymbol{s}(t-1), \boldsymbol{s}(t))$, where $\boldsymbol{s}(t)$ is a vector of all the single neuron states $s_i(t)$. Thus, given an initial state $\boldsymbol{s}(t-1)$, the trained RHoMM will propose a prediction of the next system state $\bar{s}_i(t)$ as prescribed in the dynamical model, Eq. (2). More specifically, BPTT aims at minimizing a cost function that measures the rate of incorrect predictions. In our case, the cost function was the mean square error:

$$E(s_i(t), \bar{s}_i(t)) = \frac{1}{m} \sum_{t=2}^{m} \sum_{i=1}^{n} (s_i(t) - \bar{s}_i(t))^2, \tag{3}$$

where $s_i(t)$ and $\bar{s}_i(t)$ are respectively the observed and predicted states of the brain node i at the time sample t, n is the number of neurons (*i.e.*, nodes) in the network, and m represents the time samples in the time series. Given that $s_i(t)$ and $\bar{s}_i(t)$ are binary, the mean square error (Eq. (3)) is equivalent to the average number of incorrectly predicted items at each time sample, because $(s_i(t) - \bar{s}_i(t))^2 = 1$ only when $s_i(t) \neq \bar{s}_i(t)$, and it is zero otherwise. At each step of the iterative BPTT, the prediction $\bar{s}_i(t)$ is produced by the estimated $\bar{J}_{ij}$ by only considering the previous state $s_i(t-1)$. According to the standard procedure for gradient descent algorithms, to minimize the cost function $E(s_i(t), \bar{s}_i(t))$ we perturbed the estimated $\bar{J}_{ij}$ with the cost function gradient:

$$\bar{J}'_{ij} = \bar{J}_{ij} - \alpha \frac{dE}{d\bar{J}_{ij}}, \tag{4}$$

Table 1
Selected seed voxels representing the parcel's geometrical centroid and the network it belongs to.

Seed area	Seed label	Network	MNI coordinates (x,y,z)
Left primary visual cortex V1	L_VI	VIS	−24 −79 −7
Left Fovea	L_FOV	VIS	−8 −79 18
Right primary visual cortex V1	R_VI	VIS	27 −77 −4
Right Fovea	R_FOV	VIS	13 −75 22
Left Posterior Intraparietal Sulcus	L_PIPS	DAN	−22 −68 52
Left Middle Temporal visual area	L_MT	DAN	−45 −59 −8
Left Frontal Eye Field	L_FEF	DAN	−27 −6 55
Right Posterior Intraparietal Sulcus	R_PIPS	DAN	24 −64 53
Right Middle Temporal visual area	R_MT	DAN	49 −52 −11
Right Frontal Eye Field	R_FEF	DAN	25 −7 57
Left Central Sulcus	L_CS	MOT	−23 −27 63
Supplementary Motor Area	SMA	MOT	−1 −25 63
Left Secondary somatosensory area	L_SII	MOT	−51 −16 16
Right Central Sulcus	R_CS	MOT	23 −24 63
Right Secondary Motor Area	R_SII	MOT	50 −12 15
Left Superior Temporal Sulcus	L_STS	VAN	−59 −32 27
Left Middle Frontal Gyrus	L_MFG	VAN	−44 0 1
Right Supramarginal Gyrus	R_SMG	VAN	61 −30 28
Right Ventral Inferior Frontal Gyrus	R_vIFG	VAN	44 17 3
Left dorsal Prefrontal Cortex	L_dPFC	SAL	2 24 33
Left anterior prefrontal cortex	L_aPFC	SAL	−43 38 12
Left Insula	L_Ins	SAL	−34 21 −1
Right Insula	R_Ins	SAL	39 23 1
Left Lateral Part of the prefrontal cortex	L_lat_part	SAL	−62 −42 35
Right dorsal Prefrontal Cortex	R_dPFC	SAL	−6 23 33
Right anterior Prefrontal Cortex	R_aPFC	SAL	32 46 28
Right Lateral Part of the prefrontal cortex	R_lat_par	SAL	−6 −51 31
Left Medial Prefrontalcortex	L_medial	FPN	−3 28 46
Left anterior Prefrontal Cortex	L_ant_PFC	FPN	−43 23 40
Right Superior Part of the prefrontal cortex	R_sup_par	FPN	53 −51 47
Left Superior Part of the prefrontal cortex	L_sup_par	FPN	−48 −54 50
Left dorsal Prefrontal Cortex	L_dPFC	EXE	−43 13 28
Left Intra Parietal Sulcus	L_IPS	EXE/CON	−38 −51 44
Right Intra Parietal Sulcus	R_IPS	EXE/CON	62 −38 35
Right dorso-lateral Prefrontal Cortex	R_dlPFC	EXE/CON	45 21 23
Right Intraparietal Sulcus	R_IPS	EXE/CON	42 −49 45
Left Posterior Cingulate	L_PCC	DMN	−57 −12 −18
Left Inferior Temporal Gyrus	L_ITG	DMN	−45 −68 38
Left Angular Gyrus	L_AG	DMN	−22 28 46
Left Superior Frontal Sulcus	L_SFS	DMN	−8 51 5
Left medial Prefrontal Cortex	L_MPFC	DMN	62 −6 −18
Right Inferior Temporal Gyrus	R_ITG	DMN	52 −57 29
Right Angular Gyrus	R_AG	DMN	24 36 44
Right Superior Frontal Sulcus	R_SFS	DMN	8 50 5
Right Medial prefrontal cortex	R_MPFC	DMN	−24 −79 −7

To explicitly compute the cost function gradient, we had to regularize the activation function, which in RHoMM is non-derivable. Consequently, we computed the gradient $dE/d\bar{J}_{ij}$ through the pseudo-derivative (see Appendix A), obtaining an equation analogous to the Perceptron Learning algorithm, which is a well-known feed-forward neural network (Carpenter, 1989; Widrow & Hoff, 1960):

$$\bar{J}'_{ij} = \bar{J}_{ij} + \alpha \sum_{t=2}^{m} \left[s_i(t) - \bar{s}_i(t) \right] s_j(t-1), \tag{5}$$

where α is the learning rate of the algorithm. We used Eq. (5) to run the BPTT using the training data, and for each training set and specific hyper-parameterization, we initialized the $\bar{J}_{ij}$ matrix elements with random values from a uniform distribution between −1 and +1 and we set the diagonal elements to zero. Consequently, for each training set we estimated a single matrix $\bar{J}_{ij}$. In this way, at each step, the BPTT updated the $\bar{J}_{ij}$ matrix locally minimizing the cost function by using all the transitions (Eq. (3)), and we checked the gradual improvement of the cost function until it reached a plateau corresponding to the obtained minima. More precisely, given a learning rate α, we ran the BPTT algorithm for a certain number of steps (N_{steps}) until the cost function reached a plateau in which it steadily fluctuated around a given value (Bottou, 2012). This value, due to the approximations of the model and the inherent noise of the data, might not be zero, which would correspond with the perfect prediction of the data. Furthermore, to correctly estimate the $\bar{J}_{ij}$ matrix we ensured that the cost function minimum obtained with a certain α was a good local minimum, implying that α assumed an optimal value and that the BPTT was run for a sufficiently large number of N_{steps}, which in our case was 1000. If α is too large the system may converge to a suboptimal solution, and if α is too small the system may not reach the good local minimum in the given number of N_{steps}. Consequently, we started from a large α and decreased its value. If the BPTT, after N_{steps}, reached a plateau that corresponded to a suboptimal value of the cost function, this value would decrease when decreasing the learning rate α. Otherwise, when we decreased α, the BPTT converged to a plateau corresponding to the same optimal value. Thus, we continued reducing α until the value of the minimized cost function corresponding to the plateau did not decrease any-more.

Notably, RHoMM is invariant under the multiplication of each J_{ij}-row by an arbitrary positive value a_i, because it does not change the deterministic transitions between the states that produce the dynamics of the system. This implies that the non-linear dynamics equation (Eq. (2)) produces the same transition if we substitute J_{ij} with $a_i J_{ij}$ as long as $a_i > 0$,

$$s_i(t) = \theta \left[\sum_{j=1}^{n} J_{ij} s_j(t-1) \right] = \theta \left[\sum_{j=1}^{n} a_i J_{ij} s_j(t-1) \right], \tag{6}$$

This invariance suggests that there exists a partition of the space of all J_{ij} matrices determined by subsets of matrices which lead to the same dynamics. Thus, in our protocol, at each gradient step, we normalize the rows of the estimated matrix $\bar{J}_{ij}$, with the equation

$$\bar{J}'_{ij} = \frac{\bar{J}_{ij}}{\sum_j |\bar{J}_{ij}|}, \tag{7}$$

This normalization prevents the network weights from growing too large, and allows us to use gradient descent also with larger learning rates (Zhang, Jiang, Wei, & Dai, 2015).

2.5. Performance evaluation

Once the MEG time-series data was binarized, each run was divided into two parts. The first part, composed of 70% of the data, was used as the training set. The second part, consisting of 30% of the data, was used as the test set. The training set was used for the estimation of the model parameters, thus, to compute the estimated network weights $\bar{J}_{ij}$, which were then validated with the test set. In our implementation, instead of using the whole training set to provide a single estimate of $\bar{J}_{ij}$, we split the training set in N_h contiguous temporal subsequences, which were used to estimate J^h_{ij} matrices with the BPTT method (h denotes the subsequence index). Each matrix J^h_{ij} was normalized by the square norm. To estimate each matrix J^h_{ij} we used a learning rate $\alpha = 8 \times 10^{-5}$ and $N_{steps} = 1000$. These values were chosen after an initial hyperparameter search on the test set subsequences (we used grid search hyperparameter tuning Bergstra & Bengio, 2012), and they were large enough to ensure the gradient descent convergence together with the identification of a good local minimum in the cost function. We then provided a robust estimation of the effective matrix as the average of the N_h matrices, as a compromise between the predictive power and the variability of the estimation (Fig. 2D). It is important to consider that for a small number N_h of time-series subsequences, we may get an improved predictive ability of the single sequence network model. However, the standard errors of the final averaged matrix $\bar{J}_{ij}$ are less statistically reliable. To determine the optimal subsequence size, and thus N_h, we studied the average number of incorrectly predicted items, using the cost function in Eq. (3) for different numbers of contiguous subsequences N_h. Thus, we selected the optimal number N_h which corresponded to the minimum number of subsequences where this function reached a plateau, as shown in Fig. 3A.

2.6. Statistical analysis

From the 10 subjects in our data set, we estimated a total of 24 $\bar{J}_{ij}$ matrices from the alpha-band binarized BLP, and 24 $\bar{J}_{ij}$ matrices from the beta-band. First, to assess the significance of the estimated $\bar{J}_{ij}$ matrices, we compared the distributions of the incorrectly predicted items obtained from $\bar{J}_{ij}$ with the ones obtained from random EC matrices R_{ij}, testing whether these distributions were statistically different. The R_{ij} random subsequences were drawn from a normal distribution with the same mean value as each $\bar{J}_{ij}$. Then, for a given binarized state, we used the $\bar{J}_{ij}$ and R_{ij} matrices to predict the next time sample. We then established the performance of $\bar{J}_{ij}$ and R_{ij} as the number of incorrectly predicted items for the training set (first 70% of the signal) and the test set (remaining 30%). Second, we compared the matrices in the alpha and beta band (alpha-alpha and beta-beta comparison) within and across subjects, through Pearson's correlation coefficient. We expected a larger consistency for the within-subject comparison. As a control, we ran the same analysis on the interregional BLP correlation matrices. Then, we analyzed the relationship between the across-run variability of the estimated $\bar{J}_{ij}$ matrices and the variability of the interregional BLP correlation matrices. To further examine the $\bar{J}_{ij}$ patterns, we compared the alpha and beta $\bar{J}_{ij}$ matrices across runs, through Network Based Statistics Directed (NBSdirected) (Zalesky, Fornito, & Bullmore, 2010). Finally, for each band, we estimated the EC matrices at the RSN-level. These represent the relative RSN weights for each within-RSN or across-RSN effective interaction. They were computed as follows. For each band and matrix row, we averaged the elements of the group-level $\bar{J}_{ij}$ matrix across the columns corresponding to the same RSN. Then, we counted separately the number of excitatory and inhibitory averaged links over the rows, and we normalized them by the total number of links (in each RSN or for pairwise across-RSN interactions).

2.7. Ground-truth validation of RHoMM

We first validated RHoMM through a ground-truth simulation. We generated synthetic binary sequences for the set of nodes used in this paper, applying the Hopfield model on randomly generated EC matrices J_{ij} (our "target" matrices) and then applied RHoMM to retrieve the related estimated $\bar{J}_{ij}$ matrices. As ground-truth target matrices, we selected three random graph architectures. In each target $\bar{J}_{ij}$, the edge weights were randomly drawn from a uniform distribution $[-1, 1]$, and values smaller than a threshold, one for each random matrix architecture, were set to zero. We selected the three thresholds 0.04, 0.2, 0.4. Then, starting from different first guesses, we produced $N_h = 29$ subsequences, similarly to the experimental data. For each random network, we repeated this procedure 5 times, starting from different seeds. To show that the parameters we selected were not suboptimal, we inspected error measures as a function of the number of time samples.

2.8. Comparison between interregional BLP correlation patterns obtained from experimental and simulated data

To further validate the method, for each band, we compared the correlation matrix estimated from the binarized experimental signals (C_{obs}) with the one obtained from simulated runs using RHoMM (C_{sim}). For the generation of the simulated data, we applied the following procedure. For each run, we applied the associated RoHMM to an initial random state, consisting of 45 binary elements. The initial state evolved until it reached an attractor or a stationary state producing a binary time sequence with variable duration. This procedure was repeated 1000 times, and the resulting time sequences were concatenated to produce a simulated run. We then estimated C_{sim} for each run. To evaluate the statistical significance of the comparison between simulated and experimental connectivity matrices, we applied the Bramila Mantel test (Mantel, 1967) on the average z-score of connectivity, obtained as the difference between the correlation value in each element and the mean over the correlation matrix, normalized by its standard deviation. Further, to check if C_{sim} preserved the fundamental topological features of C_{obs}, we associated a weighted graph to each group-averaged z-score connectivity matrix obtained from the experimental and simulated signals. We based our comparison on three graph metrics: the Degree, the Betweenness Centrality, and the Clustering Coefficient (Rubinov & Sporns, 2010). Specifically, for the Degree we adopted the undirected-weighted Degree definition (Rubinov & Sporns, 2010). For the Clustering Coefficient, we adopted Constantini and Perugini's generalization of the Zhang and Horvath formula (see Costantini & Perugini, 2014; Onnela, Saramäki, Kertész, & Kaski, 2005; Zhang & Horvath, 2005). This allows us to consider simultaneously both positive and negative weights, and it is particularly sensitive to non-redundancy in path information based on the sign. Finally, for the Betweenness Centrality, we computed the weighted Betweenness Centrality defined as the fraction of all shortest paths in the network passing through a given node (Brandes, 2001). Since these measures are all at the nodal level, to compare simulated vs. experimental centrality, we computed the Pearson correlation coefficient among the obtained profiles of Degree, Clustering Coefficient, and Betweenness Centrality. Although these measures cannot encompass the richness of information contained in the original connectomes, they indeed capture some very important features of the architecture of communication. In fact, large

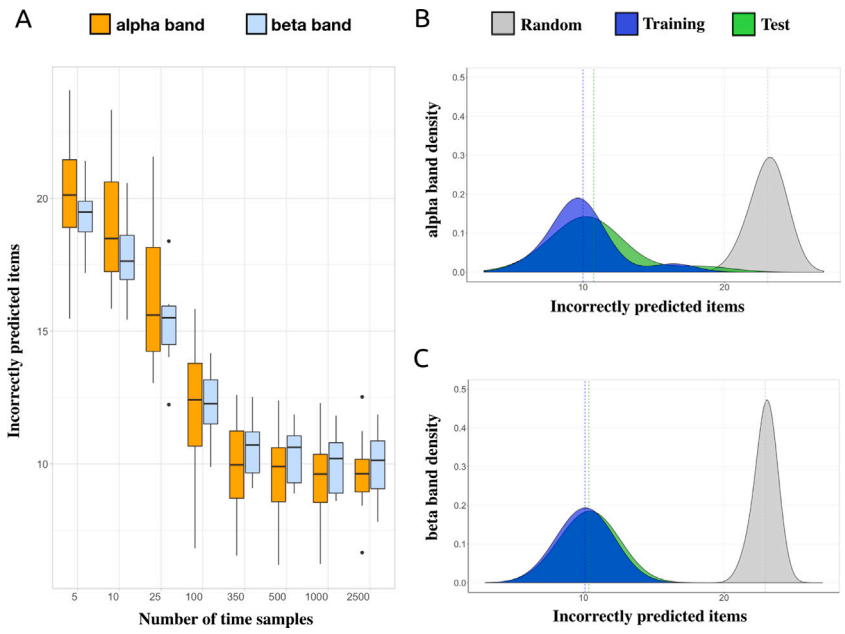

Fig. 3. Algorithm performance. (A) Box plot of incorrectly predicted time samples from the test set for $\bar{J}_{ij}$ matrices estimated for different time intervals as a function of the number of time samples for alpha- and beta-bands. (B-C) Averaged density distributions of incorrectly predicted items for training and test sets in the alpha and beta bands, respectively. The distribution of randomly generated R_{ij} is reported as a reference.

body of literature has shown how efficient models of integration rely on the presence of central regions in the brain, denoted functional hubs. Their spatial location, temporal dynamics, and integration seem to capture fundamental axes of integration across distinct functional domains. In the clinical field, today, many brain pathologies are being reconsidered as disconnection syndromes, *e.g.*, stroke, and the altered behavior of functional hubs seem to correlate with the patients' conditions (see de Pasquale, Chiacchiaretta, et al., 2021). Further, several works showed that cortical hubs captured connectome modifications that could be related to specific genetic variants and disease-related changes. In line with the analyses proposed in this work, previous studies adopted a measure of centrality to obtain an implicit representation of the voxel-level connectivity matrix, see for example Preti and Van De Ville (2017). Basically, connectomes were collapsed into profiles of eigenvector centrality and based on them dynamic states of integration in the brain were extracted. The results show that the obtained dynamic brain parcellations were consistent with those reported in the literature.

3. Results

3.1. Evaluation of algorithm performance

First, we set RHoMM hyperparameters through automatic, data-driven procedures. We already mentioned that we used a learning rate $\alpha = 8{*}10^{-5}$ and a number of training steps $N_{steps} = 1000$, which were obtained from a grid search on all the model hyperparameters taking into account the two bands and the different N_h values (the number of sub-sequences to train the network). To estimate the optimal N_h, the number of intervals to train the network, we minimized the number of incorrectly predicted items as a function of N_h. This value guaranteed a statistically reliable estimate of the average $\bar{J}_{ij}$. As shown in Fig. 3A, the performance reached a stable plateau when we trained the network with a number of time samples larger than 350. When we trained the network with a small number of time samples (*e.g.*, 5), corresponding to a shorter window length, we obtained an average value of incorrectly predicted items only slightly better than from random matrices (shown in Fig. 3B-C). When we increased the number of time samples, the predictive ability of the network grew until it converged to a stable value, which was considerably lower than random matrices (see distributions of incorrectly predicted items for the training and test set using 350 time samples in Fig. 3B-C). In terms of frequency specificity, as it can be noted in Fig. 3A, alpha and beta bands show similar performance and an equivalent optimal number of time samples, corresponding approximately to a window length of 7 s. In Fig. 3B,C, we report the density distribution of the incorrectly predicted items obtained on the $\bar{J}_{ij}$ matrix estimated on 350 time samples both for alpha and beta bands. To assess the algorithm output compared to a random network, for each run and subject, we plot together the density distributions of $\bar{J}_{ij}$ with those randomly generated, R_{ij} (see Materials and Methods). As expected, the random network correctly predicted about 50% of the items, depending on the specific distribution of the on-off states of each run. Interestingly, the networks trained with the BPTT algorithm correctly predicted, on average, more than 70% of items (notwithstanding in most cases, considering the median value of correctly predicted items, the performance reached approximately 80%). Moreover, the resulting estimated networks show a comparable prediction performance both in the training and test sets (Fig. 3B-C), with a slightly better performance in the training set. This assures us that the model is not overparameterized and that it is not overfitting.

3.2. The reliability of the estimated effective connectivity matrices

In Fig. A.2, we show the obtained $\bar{J}_{ij}$ matrices for each subject, run, and frequency band. We first assessed the similarity of the $\bar{J}_{ij}$ matrices for each frequency band by estimating the correlation among the obtained $\bar{J}_{ij}$ within and across subjects (Fig. 4A shows the correlation distributions). The average correlation within-subject was 0.35 ± 0.18 and 0.44 ± 0.14 in the alpha and beta bands, respectively. This similarity value suggests a good consistency across runs in the same subject, and shows that the algorithm produces reliable estimates of the effective connectivity at least over time epochs lasting as the run duration (few minutes), although a non-negligible variability is detected. These correlation values significantly decreased to 0.21 ± 0.12 and 0.29 ± 0.10 respectively when comparing $\bar{J}_{ij}$ matrices across-subjects ($p < 0.004$ and $p < 0.0007$ respectively, Kolmogorov–Smirnov test). In analogy with the previous analysis, we then computed the Pearson correlation among the experimental BLP correlation connectomes, for all subjects and bands (Fig. 4B). The average correlation between runs

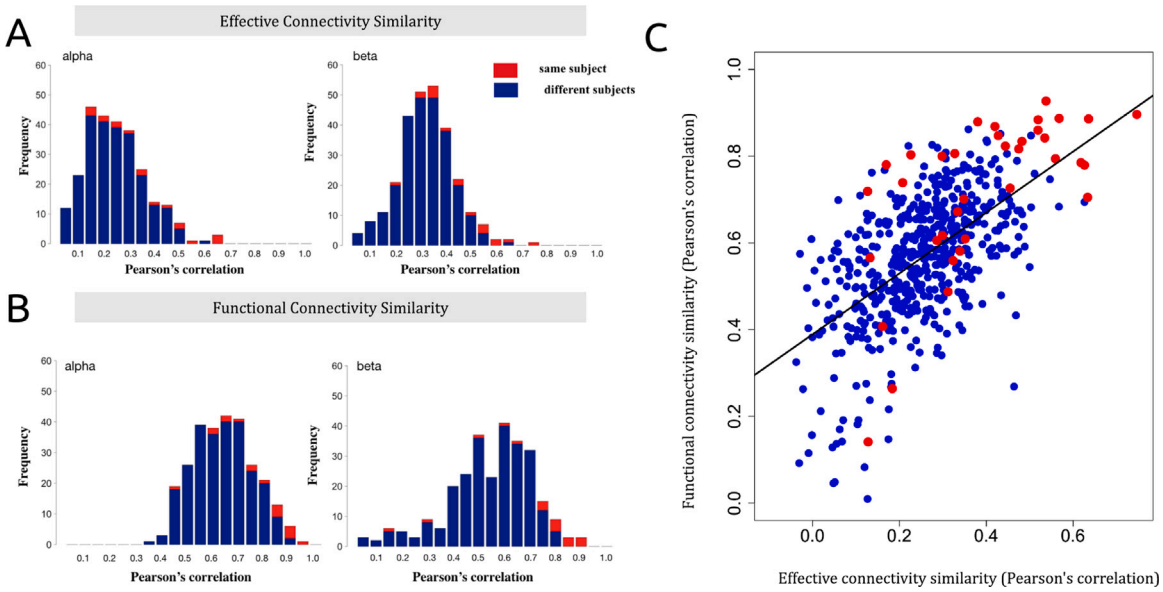

Fig. 4. Comparison between effective connectivity and BLP correlation matrices. Correlation histograms within and between subjects for EC (A) and BLP correlation (B) matrices in alpha (left panel) and beta bands (right panel). The correlations between different runs of the same subjects are reported in red while the correlations across subjects are in blue. (C) Scatter plot of correlation values between EC matrices and correlation values between corresponding BLP correlation matrices. The within-subject comparisons are displayed in red, and the blue dots represent the across-subject comparisons.

from the same subjects was 0.69 ± 0.21 and 0.75 ± 0.15 in the alpha and beta bands, respectively. Also in this case, these values significantly decreased to 0.51 ± 0.15 and 0.61 ± 0.11 respectively, when comparing the BLP correlation matrices across subjects ($p < 0.0001$ and $p < 0.002$ respectively, Kolmogorov–Smirnov test). Notably, the dispersion of the across-run correlation is not small, as in the case of the EC matrices.

Finally, we investigated the relationship between the similarity values obtained over the effective connectivity and the corresponding BLP correlation matrices. In Fig. 4C, we show the scatter plot of the similarity between each pair of EC matrices and the corresponding similarity between BLP correlation matrices. The within-subject comparisons are reported in red. Notably, a clear, significant, linear relationship can be observed ($r = 0.59$, $p < 2\times10^{-16}$). This suggests that lower similarities among the effective connectivity patterns are associated with lower similarities between the corresponding BLP correlation matrices, explaining the across-run variability of the EC matrices even within the same subject. In summary, these control analyses seem to suggest that the effective connectivity estimated through our algorithm describes subject-level dynamics, capturing individual characteristics to be associated with information provided by functional connectivity, which are here measured through BLP correlation.

3.3. Topography of the estimated effective connectivity matrices

Next, we analyzed the topography of the excitatory and inhibitory edges among the network nodes associated with brain parcels. We show in Fig. 5 the output of the developed algorithm. In Fig. 5(A,D), the estimated EC matrices averaged across runs (in the alpha and beta bands) are reported. Qualitatively, in these bands, some common patterns of inhibitory (red tone colors) and excitatory (blue tone colors) connections can be noted. Visual (VIS) and Dorsal Attention (DAN) nodes seem to have a widespread inhibition role in both bands, whereas nodes in the other RSNs seem to play an excitatory role. While in the beta band, we notice few stronger inhibitory links involving the VIS towards the other networks, the alpha inhibition seems more spread than beta, consistently with the inhibition role played by these two bands (Klimesch, 2012). To highlight the most representative effective interactions between pairs of brain regions in a quantitative framework, we thresholded the strongest excitatory and inhibitory edges separately using a common threshold in the two bands, which was selected to show a reasonable number of links (> 50% of the total edges) for the sake of clarity. Then, the EC matrices were visualized as circular graphs for the excitatory (Fig. 5B,E; blue links) and inhibitory (Fig. 5C,F; red links) parts of the network. The main difference is that the excitatory connectivity is more homogeneously distributed across the brain while the inhibitory one is mostly dominated by nodes in the Visual network (and, to a lesser extent by nodes in the DAN and other networks) down-regulating the activity of the other regions.

To quantitatively assess possible differences between the $\bar{J}_{ij}$ matrices obtained in the alpha and beta bands, we compared them through a topological analysis through NBSdirected. As can be seen in Fig. 5G (red directed links), a significant connected component suggested that the $\bar{J}_{ij}$ matrices topologically differ across bands. Specifically, up to a t-value = 3.5, we obtained a significant component from NBSdirected ($p < 0.007$), which at this threshold comprised few edges (10) and nodes (11) representing a stronger inhibition in beta as compared to alpha. Interestingly, this analysis shows that the stronger inhibitory links originating especially from the VIS in the beta EC matrix and circular plots involve edges of connected components, linking VIS/DAN nodes with receiving nodes in the other networks. As far as it concerns the excitatory links, the two bands did not show statistically different results in terms of connected components.

Now, to further characterize the obtained $\bar{J}_{ij}$, we estimated the EC matrices at the RSN-level by averaging the $\bar{J}_{ij}$ columns for each row within and across RSNs, and counting the number of positive and negative contributions (see Materials and Methods). These matrices obtained in the alpha and beta bands are reported in Fig. 6 together with the related bar plots representing the total relative inhibitory and excitatory weight for each RSN. If we focus on the bar plots concerning the excitatory links (blue bars), we notice a considerably reduced contribution of the VIS and DAN compared to the other RSN, for both bands. Moreover, the relative number of excitatory links in beta is larger than in alpha. Conversely, VIS and DAN still play a strong inhibitory role, with a smaller contribution of the DAN in the beta band. Furthermore, the RSN-based inhibition/excitation matrices (Fig. 6B, D) suggest that both bands show a common strong excitatory activity within the VIS, from DAN and DMN onto VIS, from MOT and EXE onto DAN, from VAN, SAL, and EXE onto MOT, from MOT onto VAN,

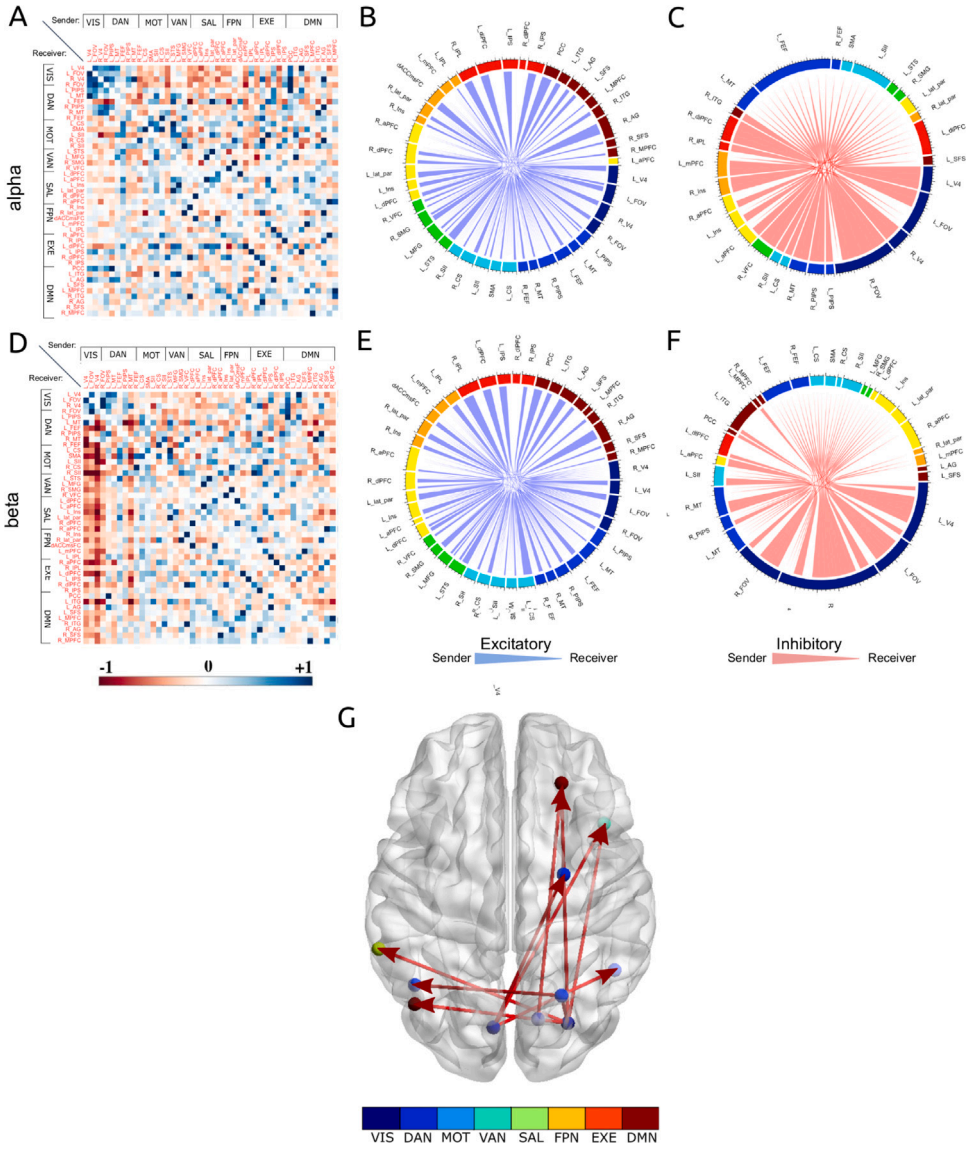

Fig. 5. Resting state effective connectivity. (A-C) Alpha and (D-F) beta bands effective connectivity. (A) shows the averaged EC matrix for the alpha band. The columns represent the sender, and the rows the receivers. (B,C) the circular graphs of the supra-threshold excitatory (blue) and inhibitory (red) networks, respectively. (D) The average effective connectivity for the beta band. (E,F) The circular graphs of respectively the supra-threshold excitatory (blue) and supra-threshold inhibitory (red) networks. (G) Significant graph component representing the few stronger inhibitory connections in the beta compared to the alpha band.

and SAL, from EXE onto DMN. Interestingly, the MOT, SAL, and EXE show widespread excitatory activity towards all the other RSNs in the beta band. This clear excitatory role in the beta compared to the alpha band is obtained only at the RSN-level, and not at the component-level, probably reflecting the variability of excitatory edges across subjects, as expected from the similarity patterns in Fig. 4. Notably, the frequency-specific effective connectivity patterns in the RSN-level matrix agree with the results reported in studies on effective connectivity based on fMRI data (Razi et al., 2017), such as the exclusive within-VIS excitatory role found in alpha, the only mild excitatory role of DMN and the excitatory contribution of SAL and EXE, in beta. As far as it regards the inhibition (red bars) instead, it can be noted a common pattern in the two bands. Apart from itself and DAN, the VIS network seems to inhibit all the other RSNs. A similar inhibition pattern also occurs for the DAN, especially in the alpha band. Further, the EXE seems to show an internal inhibition. Differently from beta, in the alpha band, the VIS is inhibited by the MOT, VAN, SAL, and FPN.

3.4. RHoMM predictive power on synthetic data

In Fig. 7 we show four measures of the ability of RHoMM to recover the "target" network on synthetic data. Two of these are measures of similarity, the Pearson correlation between the target J_{ij} and the estimated $\bar{J}_{ij}$, and the mean squared error between the absolute-value normalized target J_{ij} and the estimated $\bar{J}_{ij}$. The results show that for any of the three target matrices, the similarity is considerably high and asymptotically tending to the optimal value even for 300 time samples. A third parameter is the stability of the estimates over the N_h sequences measured through the mean over all edges i, j of the $\bar{J}^h_{ij}$'s standard deviation. This measure decreases with the number of time samples, suggesting that 350 samples or even some more could provide stable $\bar{J}_{ij}$ estimates. Finally, we show the trend of the number of incorrectly predicted test items, suggesting that 350 time-samples should be adequate to predict the test subsequence.

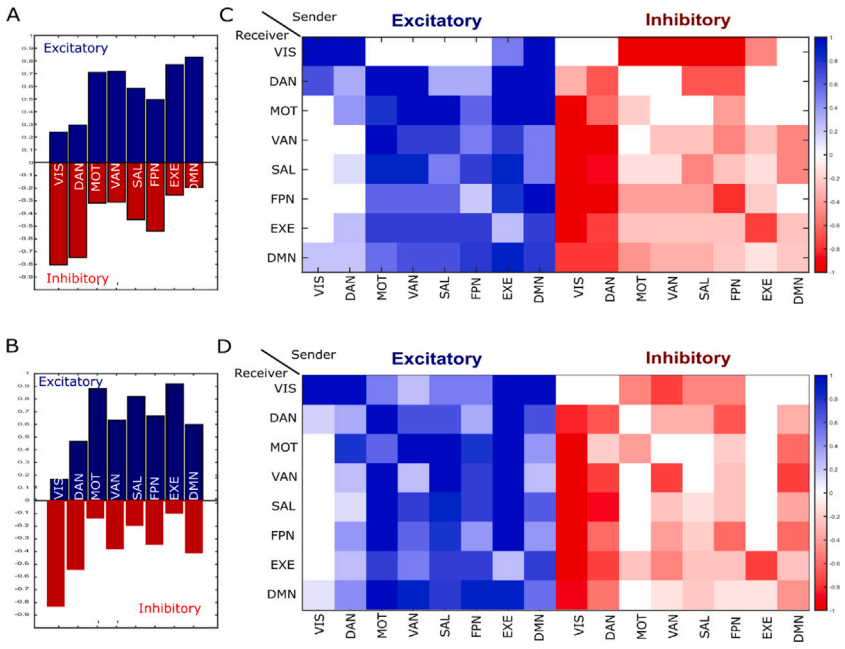

Fig. 6. Excitatory and inhibitory patterns at the RSN-level. (A) Total relative excitatory (blue) and inhibitory (red) weights were obtained as the means over the rows of the RSN-level matrix in the alpha band. (B) same as (A), beta band. A more pronounced inhibitory role is evident in the alpha band, while the beta band plays a stronger excitatory role. (C, D) RSN-level excitation matrices and inhibition matrices representing the normalized number of excitation-inhibition links. Senders are arranged in columns and the receivers are the rows.

3.5. Comparison between simulated and experimental dynamics

To test the ability of the presented method to estimate a dynamical network model reproducing the observed MEG signals, we generated a set of sequences from the RHoMM estimated on each run. Then, we estimated the average binarized-BLP correlation matrix and compared it to the experimental one. To this aim, we investigated their topological properties.

First, for both real and simulated binary data we computed the z-score (zs) of functional connectivity from the correlation matrices. In Fig. 8, we report the average zs matrix in the alpha band from real (Fig. 8A) and simulated (Fig. 8B) data. The same quantities, obtained in the beta band, are reported in Fig. 8C, D. The difference between the real and the simulated matrices are shown in Fig. A.6. To compare the corresponding functional matrices, we applied a Mantel test. For both bands, we obtained a significant similarity through Pearson correlation (corr = 0.76, $p < 0.0001$ in alpha and corr = 0.63, $p < 0.00001$ in beta). We also estimated the difference between the normalized FC patterns, since we expected smaller FC values for the simulated data, due to reduced leakage effects (see Discussion). These differences shown in Fig. A.6, are tiny, consistently with the Mantel test results. Therefore, we conclude that the estimated effective connectivity can reproduce signals with reciprocal statistical dependencies similar to those obtained from real data.

Moreover, to compare the architectures of the two binarized-BLP correlation connectivity matrices, we estimated the weighted Degree Centrality, the weighted Betweenness Centrality, and the weighted Clustering Coefficient at the nodal level. The profiles obtained in the alpha (beta) band are reported in Fig. 9A, B, and C (Fig. 9D, E, and F). We computed their similarity by adopting the Pearson correlation as the distance metric. Notably, these profiles are significantly correlated both in the alpha band (corr = 0.39, $p < 0.04$, corr = 0.31, $p < 0.03$, corr = 0.55, $p < 0.0001$) and in the beta band (corr = 0.48, $p < 0.0001$ for DEG, and corr = 0.67, $p < 0.0001$ for CLU) except for the BC. These findings suggest that the real and simulated data share a similar topological architecture of functional hubs, suggesting that the $\bar{J}_{ij}$ matrix can model the salient characteristics of the nodal dynamics.

Finally, we evaluated the ability to model the experimental dynamics also by comparing the similarity between the FCs obtained from the real and simulated dynamics, for RHoMM and an autoregressive dynamical model using the Langevin form of the DCM evolution equation (SPM 12 manual Ashburner et al., 2021), over a subset of nodes and runs (see details in Appendix B). We chose to compare RHoMM to an autoregressive dynamic model based on the Langevin form of the DCM evolution equation (instead of other DCM models) since this model is more suited for circuits with a size ≥ 10 nodes (as in our case) and for the slowly varying BLP time course (see Appendix B for a more detailed motivation for this selection). The $\bar{J}_{ij}$ matrices obtained with the two methods (see Fig. A.5) show a small overlap (*e.g.*, the inhibitory role of nodes of the VIS and DAN). The FC similarity obtained with RHoMM is in general larger than for the autoregressive dynamical model.

4. Discussion

In this work, we present RHoMM, a generative model which estimates meso-scale effective connectivity from source-level MEG BLP. RHoMM consists of a neural mass model featuring a dynamic evolution determined by a recurrent Hopfield network (Amit et al., 1985; Brunel, 2016; Folli et al., 2018; Gosti et al., 2019; Hillar et al., 2021; Hillar & Tran, 2018; Hopfield, 1982; Leonetti et al., 2020). This model was applied to binary sequences obtained from thresholded BLP time courses which preserve the original functional connectivity topography. We trained RHoMM to predict the evolution in time of the network activity by means of backpropagation through time (BPTT) (Werbos, 1990) and perceptron learning (Carpenter, 1989; Widrow & Hoff, 1960). RHoMM produced EC matrices for each band, showing specific patterns of inhibitory and excitatory links at the nodal- and RSN-level. Finally, RHoMM generated theoretical evolutions of binarized-BLP sequences, and from these, we computed BLP correlation connectomes. Several findings suggest that the developed model is reliable. First, the estimated network outperformed the corresponding random one and captured the individual variability. Second, the effective connectivity revealed patterns consistent with the inhibitory role of the alpha/beta band. Third, the BLP correlation connectomes generated by RHoMM largely reproduced the topological properties of the measured ones.

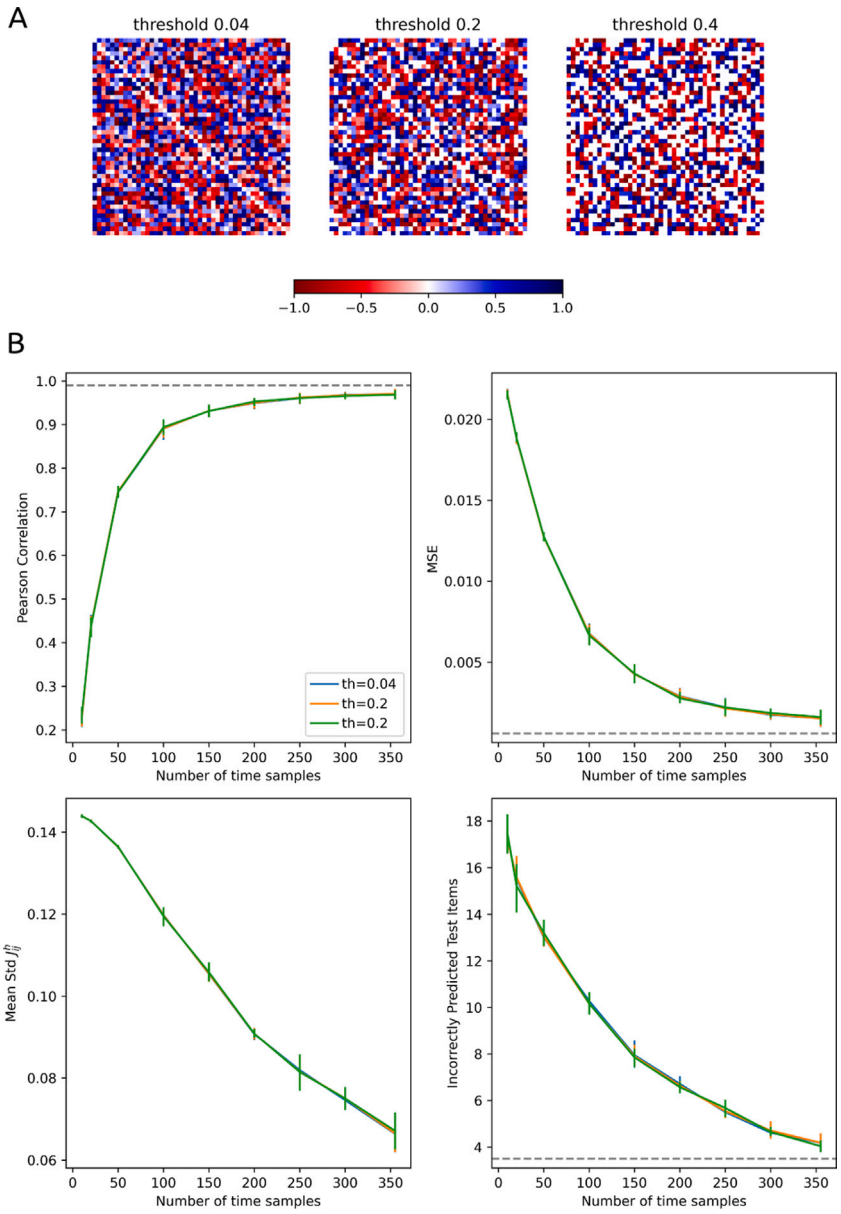

Fig. 7. Comparison between ground-truth and estimated effective connectivity matrices. (A) Examples of randomly generated ground-truth target networks for three different threshold values: 0.04, 0.2, 0.4. (B) Dependence of RHoMM on the number of time samples. For the three "target" networks, obtained with thresholds 0.04 (blue), 0.2 (orange) and 0.4 (cyan), we show the Pearson's correlation between the estimated and target matrix, the mean squared error between the estimated and the absolute-value normalized target matrix, the mean standard deviation over the N_h sequences of the J_{ij} elements, and the incorrectly predicted items. The standard deviation for each measure is estimated from the 5 different generated networks.

Several novel aspects characterize this study. As compared to other models, RHoMM is completely data-driven as it does not require any prior assumptions on the data structure or the underlying generative mechanisms. RHoMM is based on Hopfield networks, but while these have been typically adopted at the microscale level (Amit et al., 1985; Brunel, 2016; Folli et al., 2018; Gosti et al., 2019; Hillar et al., 2021; Hillar & Tran, 2018; Hopfield, 1982; Leonetti et al., 2020), our model was designed for studying excitation/inhibition at the mesoscale level (*e.g.*, in BLP connectomes). It thus allows us to investigate, at different spatial scales, the excitation/inhibition balance which plays a fundamental role in shaping brain dynamics and connectivity. In fact, several computational models show that, in a criticality regime, such balance drives the emergence of neuronal avalanches and joint oscillations that underlie resting state networks (Poil, Hardstone, Mansvelder, & Linkenkaer-Hansen, 2012). These are typically studied through biophysical models that are highly sensitive to parameters' perturbations and need fine-tuning (Abeysuriya et al., 2018). Conversely, our model, being simpler (fewer parameters and hypotheses involved) and scalable to larger sizes, could pave the way for new studies on the mechanisms of interaction in the healthy and affected brain.

4.1. RHoMM validation

4.1.1. Network performance and effective connectivity

Our findings seem to suggest that RHoMM is accurate and can model MEG BLP time sequences. First, we successfully determined, through a data-driven procedure, the minimum number of time samples required to obtain a valid estimate of the effective connectivity. We found that the duration of the time epoch corresponding to this number of samples was approximately 7 s. With this choice, the estimated effective connectivity predicts the activation state systematically better than a random matrix, both for the training and test datasets. Notably, this

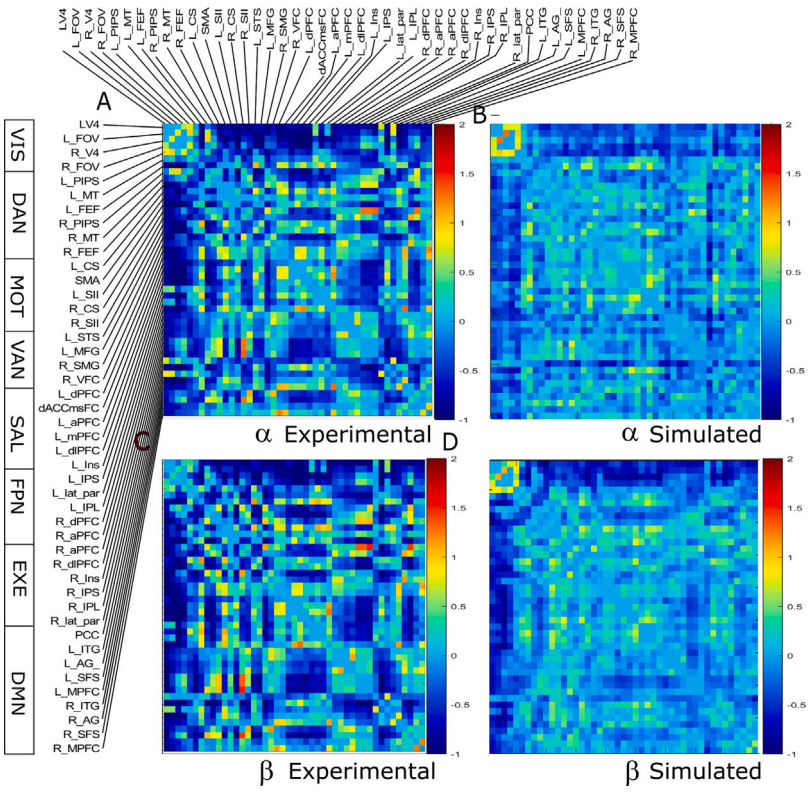

Fig. 8. Comparison between experimental and simulated functional connectivity matrices. The experimental connectivity matrices in alpha (A) and beta (C) bands. The simulated connectivity matrices in the alpha (B) and beta (D) bands.

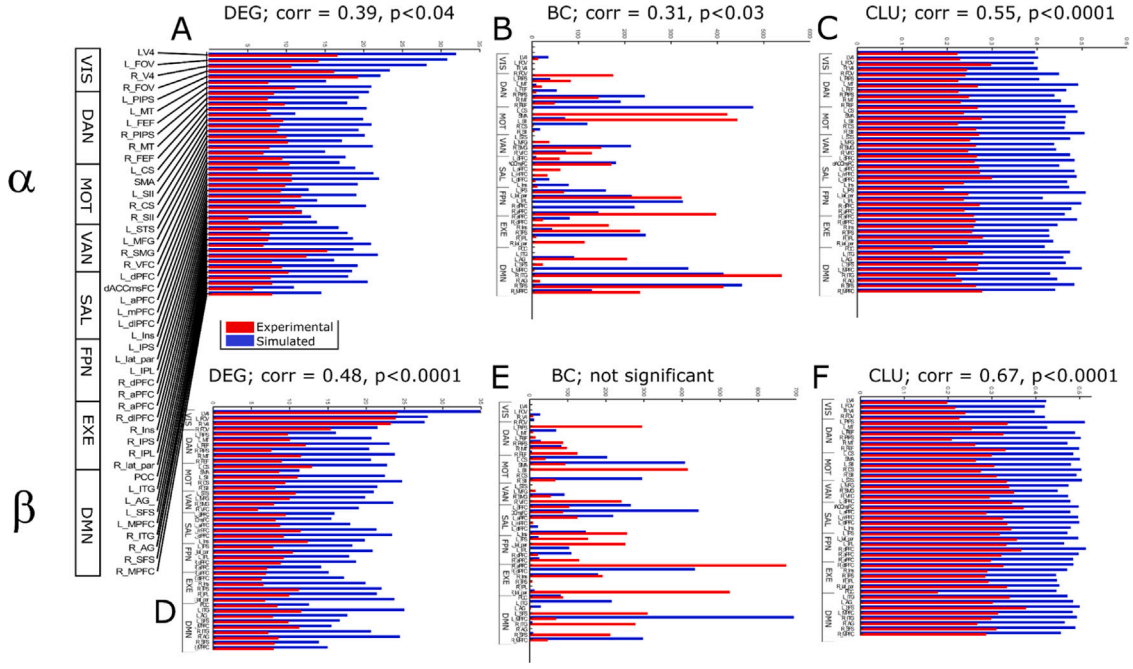

Fig. 9. Comparison between experimental and simulated weighted centrality measures. Experimental (blue) and simulated (red) weighted Degree Centrality in alpha (A) and beta (D) bands. The weighted Betweenness Centrality in the alpha (B) and beta (E) bands. The weighted Clustering Coefficient in the alpha (C) and beta (F) bands. All these centrality measures obtained from the simulated matrices are significantly correlated with the experimental ones.

duration agrees with the reported slow dynamics of RSN functional interactions (Brookes et al., 2011; de Pasquale et al., 2010). Interestingly, we obtained similar time epochs in alpha and beta bands. This is compliant with the experimental findings showing that these RSN slow dynamics are shared in beta and alpha bands, both during rest and task conditions (Betti et al., 2013; de Pasquale et al., 2010; O'Neill et al., 2017).

Second, we obtained promising results on the reliability of the estimated EC matrices. Specifically, we compared the patterns of the effective connectivity obtained from runs recorded from the same

participant and across subjects. We found that the similarity estimated on the same subject was significantly larger than the similarity across subjects. This supports the notion that the pattern of intrinsic connectivity (both effective and functional) is built on individual experience representing priors for task-evoked activity (Betti, Corbetta, de Pasquale, Wens, & Penna, 2018; Raichle, 2011; Spadone et al., 2015) However, over a backbone of stable effective connectivity, we noticed a non-negligible variability within the same individual. Such variability is in line with the variability of the corresponding BLP correlation connectivity matrices (see Fig. 4). Patterns of intrinsic functional connectivity have been reported to be dynamically changing over slow time scales (de Pasquale et al., 2012, 2016; de Pasquale, Spadone, et al., 2021; O'Neill et al., 2018) This might reflect temporal fluctuations and synchronization of different connectivity states (Cocchi, Gollo, Zalesky, & Breakspear, 2017). In this perspective, it is reasonable to expect that the EC matrix obtained by RHoMM, which can replicate the BLP dynamics, may change over time to explain the variability of functional connectivity. The developed RHoMM fulfills this condition. However, since at the current stage, we did not model the temporal dynamics of MEG connectivity, we acknowledge that future studies are needed to investigate time-resolved versions of RHoMM and to evaluate effective connectivity modulations over shorter time scales or induced by task.

Considering this undeniable heterogeneity, we evaluated the significance of the estimated EC. For this reason, we estimated the $\bar{J}_{ij}$ Standard Error (SE) given the null hypothesis that the edge weight is zero. We expected that given a certain degree of normality, the SE would be estimated by the standard deviation of the edges J^k_{ij} obtained for different subsequences. Thus, we confirmed that the SE could be estimated in this way regardless of the underlying heterogeneity (see Appendix C and Fig. A.7). This implies that the edges with weight in absolute value greater than twice the SE are significantly different from zero and have a sign that does not change even if RHoMM is estimated on different subsequences. Figs. A.7B,C show the edge weights that are in absolute value larger than twice the SE.

A final consideration pertains to the higher within- and across-subject similarity of functional connectivity compared to the effective connectivity, together with the different topology of the two types of matrices. This is not surprising. Functional connectivity similarity is also driven by the anatomical distance, while this is not reported for effective connectivity. This dependency on the anatomical distance is in part intrinsic to the data, as already shown also in fMRI studies (Lord, Horn, Breakspear, & Walter, 2012; Salvador et al., 2005) and given the anatomical similarity within and across subjects, this contribution inflates the functional connectivity similarity. Furthermore, functional and effective connectivity provide different quantities and it is plausible that pairwise functional connectivity is detected even if the corresponding effective connectivity is zero. Moreover, the former is symmetrical, while the latter is not. As an example, in Friston (2011) a toy example using DCM and coherence, demonstrated how 2 regions without any causal relationship (EC=0) were functionally connected as measured by coherence (FC was up to 0.78). These connections are called spurious correlations. This profound difference in connectivity patterns also transfers to experimental data, as in our case. Nevertheless, we acknowledge that future studies should deeply investigate possible relationships between FC and EC obtained by RHoMM.

4.1.2. Inhibitory/excitatory patterns in the alpha-beta band

The EC matrices produced by RHoMM showed a clear inhibitory pattern involving the VIS network and the posterior part of the DAN with the other RSNs in both bands with different patterns. While in beta a stronger, localized inhibition was observed from selected nodes in the VIS to the other RSN (see Fig. 5D), a more spread inhibition was observed in alpha mainly in VIS and DAN, and in other networks to a lesser extent (see Fig. 6). These results contribute to corroborating our approach since they are in line with the current literature especially on the inhibitory role of the alpha band in the occipital cortex. Since the seminal work of Berger (1929) showing that alpha power increases/reduces when closing/opening the eyes, a large amount of literature assigned to alpha oscillation an inhibitory role (Haegens, Nácher, Luna, Romo, & Jensen, 2011; Klimesch, 2012; Pfurtscheller, Stancák, & Neuper, 1996). Specifically, the extent and power of alpha oscillatory activity have been associated with cortical disengagement from inputs of the body and the environment (Buzsáki, 2006), supporting the idea of gating by inhibition (Jensen & Mazaheri, 2010). This idea, that alpha oscillations relate to blocking off task-irrelevant pathways, is reflected in the diffuse inhibition exerted by the parieto-occipital cortex on all the other RSNs, as observed in our $\bar{J}_{ij}$ matrices during resting state. In line with our results, the beta band has often been associated with mechanisms typically observed in the alpha band Spitzer and Haegens (2017). It has also been linked to communication feedback (Bastos, Vezoli, et al., 2015), and inhibitory processes (Gajewska-Dendek, Wróbel, Bekisz, & Suffczynski, 2019) involving the VIS.

In general, inhibition is crucial for shaping neural activity and response properties in the mature cortex, such as stimulus discrimination, and network stabilization (for reviews, see Ferguson & Cardin, 2020 and Isaacson & Scanziani, 2011). A large body of psychophysical and physiological evidence has suggested that our visual system encodes the retinal image by means of local mechanisms, responding selectively to bandpass-localized stimuli, such as short bars or Gabor patches. Locally, in the columnar visual cortex, inhibitory neurons actively shape response selectivity (Wilson, Scholl, & Fitzpatrick, 2018). However, recent evidence indicates that visual regions participate in a richer network integrating information over much larger parts of the cortex through excitation and inhibition with a predominance of inhibitory connections, see for example (Adini, Sagi, & Tsodyks, 1997). In network models, this property has been explained by the need of stabilization by inhibition. Even in the absence of any sensory stimulation, in animal models, it has been shown that in some networks such as the visual, somatosensory and motor networks, strong excitatory coupling leads to unstable dynamics unless stabilized by inhibition. These networks are denoted inhibition stabilized networks, see Sanzeni et al. (2020).

Interestingly, we also found a strong within-VIS excitatory role concomitant with the external inhibitory pattern. This excitation/inhibition balance is in line with the evidence that, during vision, the VIS participates in a broader network involving multiple RSNs, where inhibition is predominant (see, *e.g.*, Adini et al., 1997). As an example, using spectral DCM, Razi et al. (2017) found a clear excitation within VIS and a mild inhibition from VIS to the rest of the brain, as in our case. Furthermore, using regression DCM, Frässle et al. (2021) also showed a clear excitatory pattern within VIS, so that this RSN seemed to be extremely segregated due to this balancing between internal excitation and external inhibition. The authors interpreted this pattern in terms of a hierarchical organization ranging from unimodal (*e.g.*, VIS) to transmodal (*e.g.*, FPN, EXE, etc.) networks, as supposed by Mesulam (1998). Notably, functional connectivity studies (*e.g.*, Agcaoglu, Wilson, Wang, Stephen, & Calhoun, 2019) also found a highly internally connected VIS, segregated from the rest of the brain. This pattern, in agreement with our functional and effective connectivity results, was related to the condition of fixation as compared to eyes closed.

Finally, based on our findings, the beta band plays a strong excitatory effect, apart from the VIS network. Within a framework of communication through coherence, it has been suggested that top-down attentional influences, estimated through effective connectivity, are mediated by beta-band synchronization (Fries, 2015). In this perspective, the excitatory role played by the beta band at the RSN-level (see Fig. 6), nicely links with the increased connectivity and the presence of hubs previously reported in this band (Betti, Penna, de Pasquale, & Corbetta, 2021; de Pasquale et al., 2012, 2016; Hipp et al., 2012) .

4.1.3. RHoMM as a generative model

An additional validation of RHoMM tackles the generative capability of the model by (i) comparing the target and estimated EC matrices through a ground-truth simulation; (ii) comparing the BLP correlation connectomes obtained from the acquired data vs. the ones estimated over the synthetic sequences generated by RHoMM. For the first validation, we demonstrated that RHoMM can recover the ground-truth parameters and suggest that our generative model could reliably estimate the EC matrix underlying binarized power time courses. For the second type of validation, we obtained the predicted signals by applying the associated RoHMM multiple times to different starting random states. The results were encouraging: a Mantel test showed a significant similarity between the binary-BLP correlation matrices obtained from measured and generated data. This suggests that the estimated effective connectivity was able to reproduce the functional architecture observed in the real data. This comparison is in analogy with the work of Cabral, Luckhoo, et al. (2014), where the profiles of the functional connectivity generated via the developed model and the real ones were correlated. In this case, depending on the choice of the optimal parameters, the obtained correlation values ranged between 0.3 and 0.4. Notably, these values are even lower than the ones we obtained in the alpha and beta bands.

Furthermore, we analyzed the topological properties of the generated matrices, focusing on integration. In fact, it has been shown that the topology of the architecture of interactions is fundamental in ensuring efficient integration among brain regions (de Pasquale, Spadone, et al., 2021). This relies on the temporal dynamics of a set of central nodes (hubs) ensuring an efficient integration in the brain. For this reason, here we evaluated and compared the patterns of nodal centrality obtained in generated vs. measured connectomes. All the considered measures are significantly correlated, *i.e.*, real and simulated data share a similar architecture of functional hubs. This suggests that RHoMM, although relying on a reductionist model, can model and generate salient characteristics of the nodal dynamics, without applying any oversimplifications. Our results are in line with the work of Betzel et al. (2016), where, to compare simulated vs. real connectomes, the distributions of the same measures of centrality that we adopted here were considered. However, in Betzel et al. (2016) specific network growth mechanisms were considered to constrain either geometric wiring rules based on the node distance or topological features. For this reason, the observed agreement is not unexpected. Differently, in RHoMM no constraint on any network growth mechanisms or topological properties of the expected connections are considered. Thus, the observed similarity between observed vs. simulated centrality seems very promising.

4.2. Methodological considerations

4.2.1. Computational burden

An important aspect of any generative approach that estimates the effective connectivity is the scalability to large network sizes. Notably, RHoMM was applied to source-level MEG BLP time courses, with a temporal resolution of the order of 10 ms, that is a number of points about 100 times larger than fMRI data. This greater temporal resolution implies an increase in computational time since we could use many temporal samples (about 14,700 corresponding to $\sim$ 5 min). This allowed us to divide the recordings into many segments N_h and estimate the mean effective connectivity from them. In our case, we analyzed 45 nodes and the optimal N_h corresponded to sequences of about 350 binary vectors, obtained from a total of about 14,700 temporal samples. To understand how the execution time grows with the number of nodes in RHoMM given a certain MEG BLP time course, we ran a simulation on an 8-core processor using the 30 segments of the training set (comprising 350 transitions each) from a sample run.

For the simulation, we applied RHoMM to the binarized MEG BLP time course of the 45-node set, and then we reduced the number of nodes n to inspect the effect of network size on the execution time. We aimed to obtain a relationship between the number of nodes and execution time, to extrapolate the computational burden for larger network sizes. Specifically, we tested $n = 8, 16, 32, 45$. Since our algorithm solution depends on the selection of the learning rate α and the gradient descent number of steps N_{steps}, we performed an a priori hyperparameter search and found that $N_{steps} = 1000$ worked for all network sizes and the optimal learning rates were $\alpha = 10^{-4}, 10^{-4}, 9 \times 10^{-5}, 8 \times 10^{-5}$ for $n = 8, 16, 32, 45$.

We found that RHoMM takes less than 3 s to converge to a solution for time series of approximately 350 transitions and 45 nodes. This value should be multiplied by 30 because it is repeated on each time series segment, thus it would require 90 s in total for 10,500 temporal samples. From our simulation (see Fig. A.3) we obtained that the execution time scales approximately linearly ($r = 0.9999$) with the network size. In this perspective, for networks of 100 or 150 nodes, which is the size typically used in MEG functional connectivity, RHoMM would take approximately 6 s and 9 s respectively. To provide further predictions on the computational burden for network sizes larger than 45 nodes, we added results from simulated data. We used simulated data since it is easier to scale the network, as we know the ground-truth and we can control RHoMM convergence. We simulated random graph networks with edge weights randomly drawn from a uniform distribution [−1,1] and set to zero the weights below a threshold of 0.04. We selected the suitable hyperparameters ensuring convergence for the different node sizes. Specifically, we used $N_{steps} = 10000$ for all the network sizes and α = [.4, .6, .8, 1.0, 1.5, 2.] for the network sizes n = [8, 16, 32, 45, 100, 150]. The trend of the obtained curve is not linear, but it is close to the experimental trend up to 45 nodes and more conservative computational times are obtained than with the linear assumption.

In contrast, Frässle et al. (2018) reports that, when applied to a 104 region whole-brain BOLD fMRI acquisition of 230 functional images, rDCM with sparsity constraints takes approximately 10 min on a single core and 1 min with 16 processor cores on an HPC cluster. Furthermore, Frässle et al. (2017) reports that on 940 functional images from fMRI acquisitions, the execution time for a 6-node network was around 0.24s with rDCM without sparsity constraints, and around 10 min with the standard DCM variational Bayesian approach under the Laplace approximation, while in Frässle et al. (2021) it took on average 0.6 s on 15-node networks for rDCM without sparsity constraints. It is important to consider that rDCM without sparsity constraints is faster than with sparsity constraints (Frässle et al., 2018, 2017). Notably, these times are approximate indicators as they highly depend on the specific hardware used. Although we acknowledge that a detailed comparison between a gradient descent algorithm (RHoMM) and an algorithm which is based on ordinary least squares (rDCM) is questionable, because the number of iterations has a strong correlation with the convergence to the optimal solution, we can state that the execution rate of RHoMM is competitive, or maybe better than rDCM. In the future, we plan to implement a GPU version of RHoMM that will take advantage of the straightforward parallelization of the BPTT algorithm and will considerably reduce the execution time for networks with more than 100 nodes.

4.2.2. Leakage effects on the effective connectivity matrix

A methodological issue deeply discussed and dealt with when estimating functional connectivity, and in some cases directed connectivity, is the effect of leakage or linear mixing on the interaction patterns. In MEG, first-order leakage is inherent to the inverse approach used to project channel-level signals into the source space, and may produce over-connectivity at 0-phase lag, especially close to the seed. Several methods were proposed to reduce leakage effects on functional interaction measures, either operating on the leakage pattern introduced by the specific inverse approach, *e.g.*, the Geometrical Correction Scheme (Della Penna et al., 2019; Wens et al., 2015) or operating on the signal by reducing 0-lag phase interactions and highlighting lagged

connections, *e.g.*, the orthogonalization method (Brookes, Woolrich, & Barnes, 2012; Hipp et al., 2012) for the slow band-limited power and the imaginary part of coherence or the MIM (Marzetti et al., 2013; Nolte et al., 2004) and PLI (Stam, Nolte, & Daffertshofer, 2007) for the fast signal. In principle, first-order leakage should not affect directed interaction estimates. Since these estimates evaluate the influence of past activity of the seed on the target's future activity, the effects of linear mixing on lagged interactions are inherently leakage-free, as for Phase Transfer Entropy (Lobier, Siebenhühner, Palva, & Palva, 2014) and MPSI (Basti et al., 2018). However, some of the measures estimating causal interactions rely on multivariate autoregressive models of the seed and target signals, which are influenced by their correlation and thus are indirectly affected by leakage. This applies for example to Granger Causality and its spectral version (Bressler & Seth, 2011; Kaminski & Blinowska, 1991; Spadone, Wyczesany, Della Penna, Corbetta, & Capotosto, 2021). In the case of effective connectivity estimation, such as our approach or DCM and any approach fitting to the experimental data to estimate the model parameters, first-order leakage does not affect the results. However, whichever is the connectivity measure, spurious interactions may occur due to second-order leakage effects, caused by induced connectivity among regions close to truly connected ones, occurring also at non-0 lag and thus affecting directed and effective connectivity (He et al., 2019; Palva et al., 2018). Removing such effects, which are inherent to any connectivity estimate, is a hard task, and although a few strategies to remove them were proposed (Wang et al., 2018), these deserve further investigation in terms of stability, reliability, and generalization. As a first step, here we tested how second-order leakage might affect the estimation of the $\bar{J}_{ij}$ matrix, which in principle depends on all the nodes of the connectome. We selected a subset of 16 nodes that were placed at a distance larger than 3.5 cm, using a Monte Carlo method. At this distance, the contribution of first-order leakage was shown to significantly decrease (de Pasquale et al., 2010), and this should reflect in a lower contribution of the second-order leakage. We then estimated the EC matrix again from a subset of 5 subjects and we analyzed the relationship with the corresponding $\bar{J}_{ij}$ values of the full-size matrix, involving the same node pairs. As shown in Fig. A.4, the relationship is linear, as stated by the significant Pearson's correlation coefficient ($r=0.67$, $p< 2.2\times10^{-16}$). Notably, different scaling does not affect the dynamics generated by the $\bar{J}_{ij}$ matrix, thus we can state that the $\bar{J}_{ij}$ subsets are similar, although the range of the $\bar{J}_{ij}$ values is different. Furthermore, if the signals of the nodes excluded from the subset were indirectly influencing the effective connectivity estimation, we would not obtain a linear relationship between the two subsets. Thus, this analysis suggests that these $\bar{J}_{ij}$ values are not biased by second-order leakage and represent reliable effective connectivity.

4.3. Advancement with respect to the current literature

RHoMM is a recurrent neural network, with Hopfield-like dynamics, part of a larger class of generative approaches (Betzel & Bassett, 2017; Lanza, Di Angelantonio, Gosti, Ruocco, & Folli, 2021; Magrans de Abril et al., 2018). Hopfield network models were originally introduced to model memory storage and recall (Hopfield, 1982). Recurrent neural networks were used to describe brain activity during different cognitive tasks, *e.g.*, to model the neuronal population in the prefrontal cortex and elucidate mechanisms of selection and integration of task-relevant inputs (Mante, Sussillo, Shenoy, & Newsome, 2013). Further, in the premotor cortex, they unveiled a mechanism benefitting from temporal expectations in perceptual detection (Carnevale, de Lafuente, Romo, Barak, & Parga, 2015). In general, recurrent neural networks were mostly used to model small-scale neuronal populations. Only recently they were applied to larger scales in a few studies (Hahn et al., 2019; Singh et al., 2020). As an example, a non-predictive model that reproduced functional connectivity using DTI to constrain the network connectivity and gradient descent to minimize the cost function was developed (Hahn et al., 2019). In addition, a recurrent neural network inspired by neural mass models (MINDy) was applied to meso-scale fMRI data, and it was able to estimate the effective connectivity and reproduce brain dynamics at the meso-scale level (Singh et al., 2020). A popular class of generative approaches, widely applied to estimate effective connectivity, is dynamic causal modeling (DCM, Friston, Kahan, Biswal, & Razi, 2014). This is based on linear dynamic equations to model the behavior of neural regions. DCM was typically applied to a small set of brain regions, and successfully modeled brain activity under different conditions (Bastos, Litvak, et al., 2015; Friston et al., 2014; Makuuchi & Friederici, 2013). Recently, a modified version of DCM, named regression DCM, was introduced to analyze a larger number of nodes. This was achieved, especially in fMRI, by introducing a convolution kernel on the BOLD signals to retrieve its functional relationship with the neural activity (Frässle et al., 2021, 2018, 2017). To the best of our knowledge, on MEG data, DCM was used only in the context of small networks and mainly in the comparison of two experimental conditions (David et al., 2006; Jafarian, Litvak, Cagnan, Friston, & Zeidman, 2020; Kiebel et al., 2008; Lu et al., 2012). Here, as a quick comparison with RHoMM over resting state data in the source space, we applied an autoregressive dynamical model using the Langevin form of the DCM evolution equation to retrieve the EC matrix from the source space BLP and generate simulated evolutions. The $\bar{J}_{ij}$s show a small overlap, while it seems that RHoMM outperforms the autoregressive dynamical model in reproducing the FC matrices. However, the comparison is not possible at the level of the predicted sequences, which are different due to binary or not-binary values. We acknowledge that more extensive comparisons between RHoMM and DCM or similar would be of interest and could be the topic of future work.

Analogously to Singh et al. (2020), our recurrent neural network is designed to model and predict meso-scale activity and connectivity through a data-driven approach that intrinsically models the non-linearity of neural dynamics. Moreover, similarly to Singh et al. (2020), RHoMM minimizes the prediction error at the single transition level. Several features make our approach different from the models currently available in the literature. First, RHoMM is substantially less-constrained than other assumption-based models, *e.g.*, DCM, the model in Hahn et al. (2019), and MINDy, since the EC matrices are estimated without any constraint or biases on topology, directionality, and sparseness. Thus, any configuration could potentially be retrieved. Moreover, RHoMM does not assume the gaussianity of the model parameters, thus allowing non-linear dynamics of the state variables, and it could capture features that other linear approaches, such as DCM, might miss. Second, differently from Hahn et al. (2019), RHoMM provides the effective connectivity together with a prediction of the neural dynamics. Third, differently from other approaches, a considerably small number of hyperparameters (only 3, *i.e.*, number of steps in the gradient descent, learning rate and number of epochs to infer the EC matrix) are involved in the model training while no parameters are used for regularization. While these features reduce possible risks of instability in the obtained network, these are tailored for the modeling of MEG BLP, providing a temporal resolution and richness of information considerably higher than fMRI data. Fourth, the network design is optimized for high temporal resolution data (MEG, EEG) through the analysis of binary states (active or inactive) and the dynamics of on/off transitions for each brain node. This allowed the inference of EC matrices (with about 50 nodes), and the prediction of their dynamics, with the possibility of scaling to a larger number of neurons (hundreds). Finally, we designed RHoMM, based on a well-established small-scale model (Hopfield model). This fosters future research on the dynamics of inhibition/excitation at different spatial scales using similar models (*e.g.*, self-similarity between micro and meso-scale). In fact, the application of RHoMM to the meso-scale requires a fundamental change of scale and a different interpretation of the model design. First of all, the activation and rest of McCulloch Pitts neurons in RHoMM are not

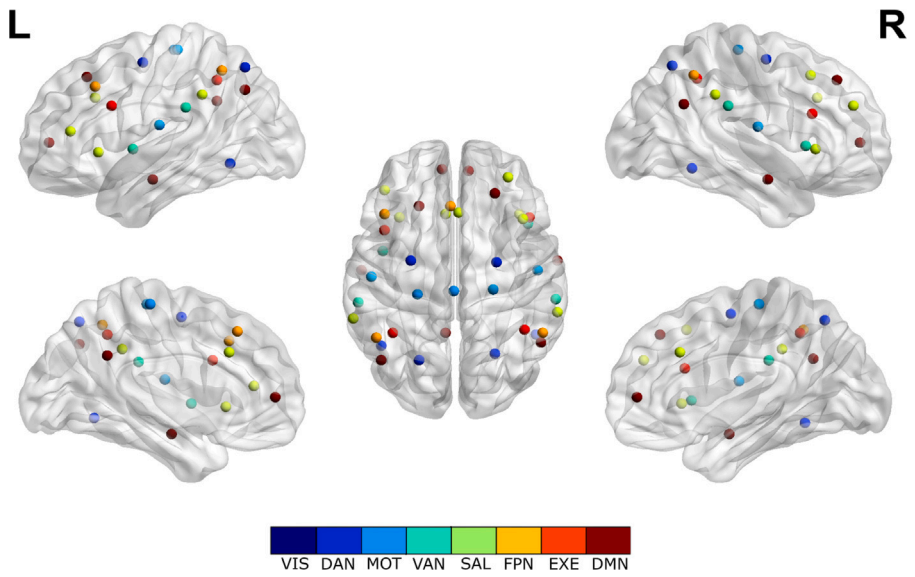

Fig. A.1. Distribution of seed voxels over the MNI template. The seed voxels are color labeled based on the RSN they belong to.

associated anymore with the generation of spike trains from a single cerebral neuron, but to the emergent binary BLP of a brain region (*i.e.*, its above- and below-threshold BLP). For this reason, the binary sequences of the network nodes also consider their inherent nonlinear dynamics. Furthermore, most of the research on Hopfield networks is based on symmetric networks with Hebbian learning. Conversely, to model the meso-scale whole-brain excitation/inhibition pattern, we considered the added complexity of asymmetric networks that, with few exceptions, are often overlooked (Folli et al., 2018; Leonetti et al., 2020). Moreover, we applied BPTT and perceptron learning as opposed to Hebbian learning to generate a model capable of reproducing the brain activity dynamics.

5. Conclusions

In summary, the application of a recurrent Hopfield neural network on source-level MEG power time courses allowed us to estimate the effective connectivity and to generate predictions of the experimental data. RHoMM is a fundamentally data-driven approach that is based on the assumption that BLP time courses can be associated with sequences of binary states with non-linear dynamics. Notably, since recurrent Hopfield neural networks are simple discrete-time recurrent neural networks with binary neurons, RHoMM state space is finite. This property, together with the limited number of model parameters decreases the computational burden of RHoMM, which can be run also on any standard PC. We validated RHoMM on a set of experimental MEG data showing its capability to capture the individual variability and reproduce the salient characteristics of BLP dynamics. Importantly, RHoMM minimalist approach leads to a more constructive framework that could be enriched by additional parameters that might improve the model's predictive ability. This framework is suitable for extensive application in system neuroscience, to derive the inhibition/excitation patterns of larger sets of experiments in order to investigate the complex interplay among different brain regions during different experimental conditions, in the healthy and affected population.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Giorgio Gost reports financial support was provided by Lazio Region. Giancarlo Ruocco reports financial support was provided by European Research Council. Viola Folli reports financial support was provided by D-TAILS srl.

Data availability

Data will be made available on request.

Acknowledgments

The research leading to these results has been also supported by European Research Council Synergy grant ASTRA (n. 855923). ML and GG thank Regione Lazio, LazioInnova, Project LOCALSCENT, Grant PROT. A0375-2020- 36549, Call POR-FESR "Gruppi di Ricerca 2020".

Appendix A. Cost function gradient derivation

Eq. (2) is singular, thus in principle the gradient $\frac{dE}{dJij}$ is not defined. Nevertheless, we can linearize the model and derive an approximate pseudo-derivative for the gradient function. With this aim, we define the activity strength which acts on any node i as the sum of the weighted inputs of the nodes that are linked to i and are active at the time $t-1$:

$$\psi_i(t) = \sum_{j=1}^{n} \bar{J}_{ij} s_j(t-1)$$

This allows us to write the derivative of Eq. (2) with respect to $\bar{J}_{ij}$ as following:

$$\frac{dE}{d\bar{J}_{ij}} = \frac{d}{d\bar{J}_{ij}}[\frac{2}{m}\sum_{t=2}^{m}\sum_{i=1}^{n}(s_i(t)-\theta(\psi_i(t)))^2] \tag{A.1}$$

$$\frac{dE}{d\bar{J}_{ij}} = \frac{2}{m}\sum_{t=2}^{T}(s_i(t)-\bar{s_i}(t))(\frac{d\theta(\psi_i(t))}{d\psi_i(t)})[\frac{d}{d\bar{J}_{ij}}(\bar{J}_{ij}s_j(t-1))]$$

Using the rough linearization $\theta(\psi_i(t_k)) \approx \psi_i(t_k)$,

$$\frac{dE}{d\bar{J}_{ij}} = \frac{2}{m}\sum_{k=2}^{m}(s_i(t)-\bar{s_i}(t))[\frac{d}{d\bar{J}_{ij}}(\bar{J}_{ij}s_j(t-1))]$$

$$\frac{dE}{d\bar{J}_{ij}} = \frac{2}{m}\sum_{k=2}^{m}\left[s_i(t)-\bar{s_i}(t)\right]s_j(t-1)$$

Thus, we obtain the update rule used for the BPTT algorithm, Eq. (5).

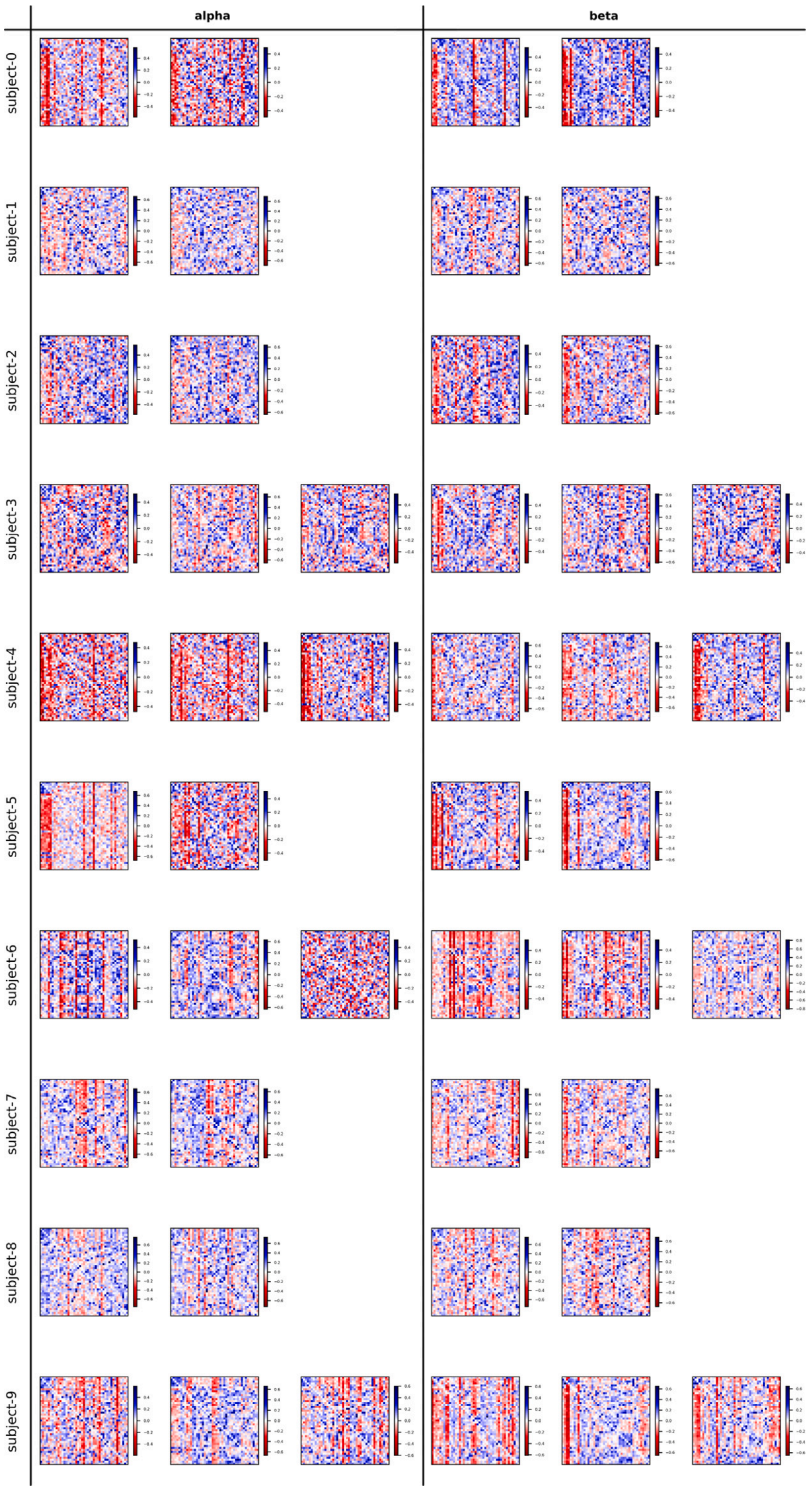

Fig. A.2. Individual effective connectivity matrices. For each subject (row) and run (column) the EC matrices are shown for the alpha (left) and beta (right) bands. Excitatory (positive) values are coded in blue, and inhibitory (negative) in red. The senders are the columns, and the receivers are the rows.

Appendix B. Model comparison

The primary contribution of RHoMM is to propose a nonlinear predictive generative model based on NMMs capable of estimating the EC for BLP source space data. To compare RHoMM with a previously established method we used an autoregressive dynamical model based on the Langevin form of the DCM evolution equation (SPM 12 manual Ashburner et al., 2021). We chose to compare RHoMM to an autoregressive dynamic model based on the Langevin form of the DCM evolution equation (instead of other DCM models) for the following reasons. First, we did not select any of the DCM models for MEG/EEG available in SPM 12 (Ashburner et al., 2021; David et al., 2006; Jafarian et al., 2020; Kiebel et al., 2008; Lu et al., 2012) for two reasons. The first one is that all of them consider small circuits comprising few connections, while we wanted to compare a larger network for the sake of generality. Secondary, these models were designed for the analysis of evoked signals and induced responses, which were not the best choice for our comparison, since we analyzed resting state

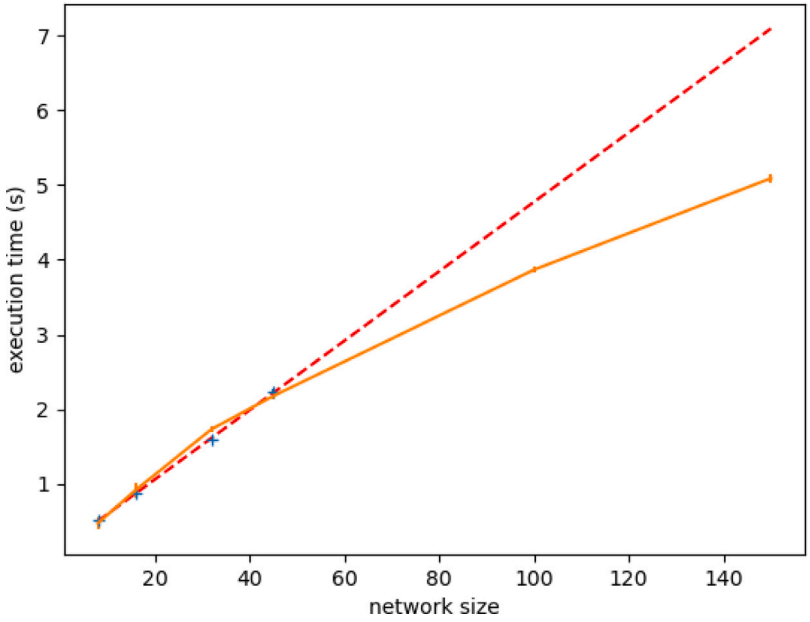

Fig. A.3. Computational burden. This plot shows an estimate of the computational burden when scaling RHoMM. The blue points represent the measured execution time for different sizes of the RHoMM network ($n = 8, 16, 32, 45$). The red lines show the linear fit of the measured values, and it represents an extrapolation of the running times for larger networks. For network sizes larger than 45 nodes, we show the execution time for simulated dynamics obtained from synthetic J_{ij} matrices (orange line).

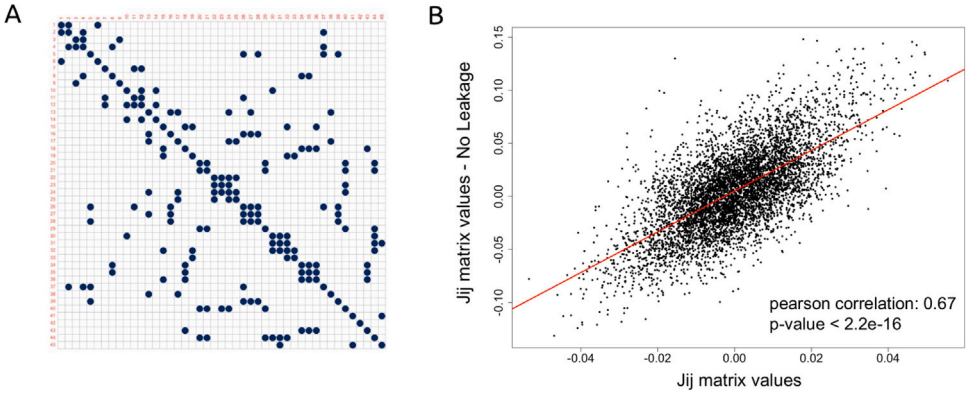

Fig. A.4. Control on second-order leakage effects. (a) Binary matrix based on the spatial euclidean distance calculated between different areas of the brain. The matrix shows how node pairs used in this simulation were selected. Blue elements mark the node pairs closer than 3.5 cm, otherwise, the pairs are white. A subset of 16 nodes at distance > 3.5 cm was selected. (b) Scatterplot between the $\bar{J}_{ij}$ elements of the full-size matrix and the corresponding matrix obtained from the 16 nodes subset. On the vertical axis, we report the $\bar{J}_{ij}$ matrix values obtained using only the MEG signals for which the minimum spatial distance from each other is greater than 3.5 cm. On the horizontal axis we report the values obtained from the original $\bar{J}_{ij}$ connectivity matrix but considering only the rows and columns corresponding to the 16 nodes subset. The linear relationship suggests that these values were similar and that they were not affected by second-order leakage.

data without any external input and/or external condition to explain the spontaneous fluctuation of the oscillatory Band Limited Power, which are only captured as stochastic noise terms. Two other DCM models (cross-spectral DCM (Friston et al., 2012), and phase coupling DCM Penny, Litvak, Fuentemilla, Duzel, & Friston, 2009) were designed for modeling the fast oscillatory activity. While in principle it would be possible to apply these models to our BLP time courses, this is not straightforward and would imply some manipulation of the models, which would not be a benchmark anymore. Inspecting the models for fMRI, the one closer to our requirements resulted in the spectral DCM (spDCM Friston et al., 2014; Razi et al., 2017) which is suitable for modeling intrinsic dynamics during resting state. An important reason is that spDCM can deal with wider circuits than the MEG models. This model consists of a Langevin form of the evolution equation (motion) and a non-linear forward model for the hemodynamic response. Another important reason is that, in our case, the Langevin form of the evolution equation (motion) is a simple equation that seems suitable to be applied to BLP, as the hidden states are assumed to conceptually represent the slow fluctuations of neuronal activity, which are in effect measured through the BLP (see, *e.g.*, Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001 on the link between the BLP and the BOLD signal). We thus used a discretized version of the Langevin evolution equation in DCM, $\Delta x(t) = J x(t) + \upsilon$, and linear regression to estimate the $\bar{J}_{ij}$ matrix.

We used a subset of 16 nodes (the nodes are the same as in Fig. A.4) and a subset of 5 subjects. The first 4 nodes of the subset included 2 nodes from the Visual network and 2 from the DAN. Since we analyzed the BLP time courses associated to nodes in the source space, we did not need forward and inverse equations to invert the ensuing model. Thus, we applied the discretized version of the Langevin evolution equation to the BLP signal in the alpha band. The $\bar{J}_{ij}$ proposed by both approaches for each sample run and their average over the runs are shown in Fig. A.5 together with the similarity between the experimental and simulated FC measured through Pearson's correlation, for RHoMM and the autoregressive dynamical model. In the $\bar{J}_{ij}$ matrices, the first 2 nodes are Visual ones, and the second pair are from the DAN. The

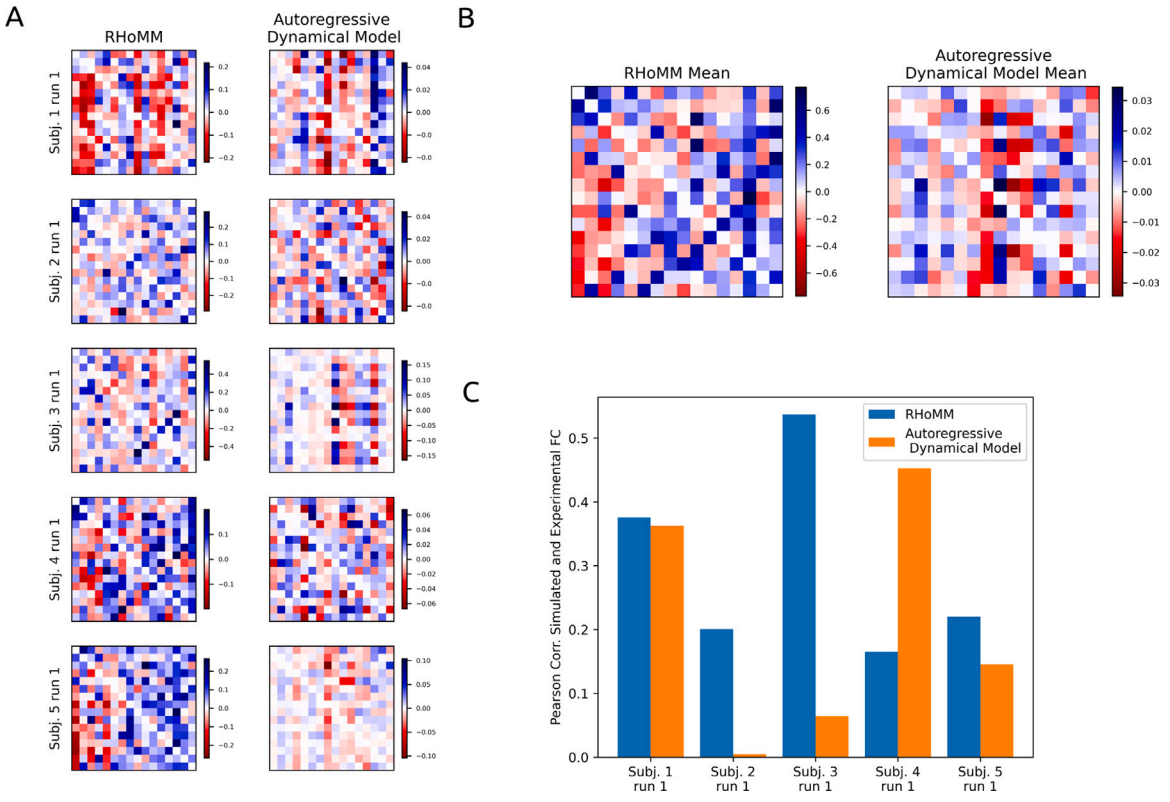

Fig. A.5. Model Comparison. (A) Estimated $\bar{J}_{ij}$ matrices for 5 representative runs, obtained by RHoMM (left) and the Autoregressive Dynamical Model (right). (B) Mean $\bar{J}_{ij}$ matrices over the 5 runs, for RHoMM (left) and Autoregressive Dynamical Model (right). (C) Similarity between the FC matrices obtained with the two generative models, and the experimental FC, estimated by Pearson's correlation.

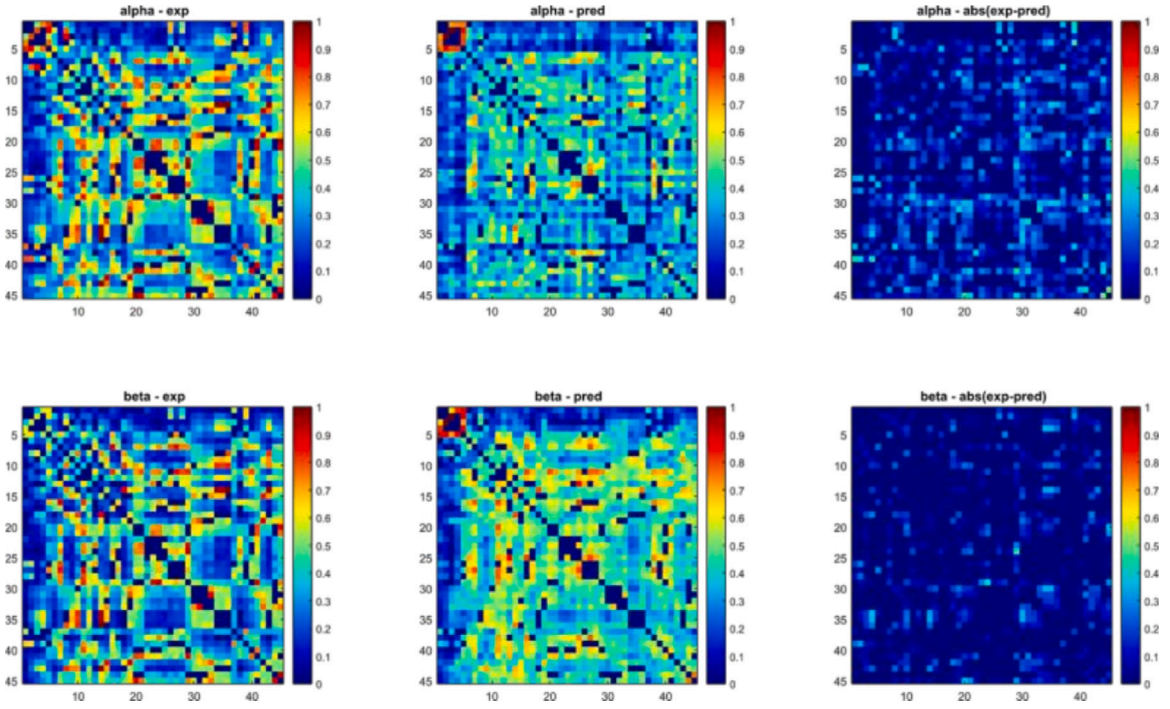

Fig. A.6. Comparison between experimental and simulated FC matrices. Normalized experimental (left) and simulated (middle) matrices, together with their difference (right).

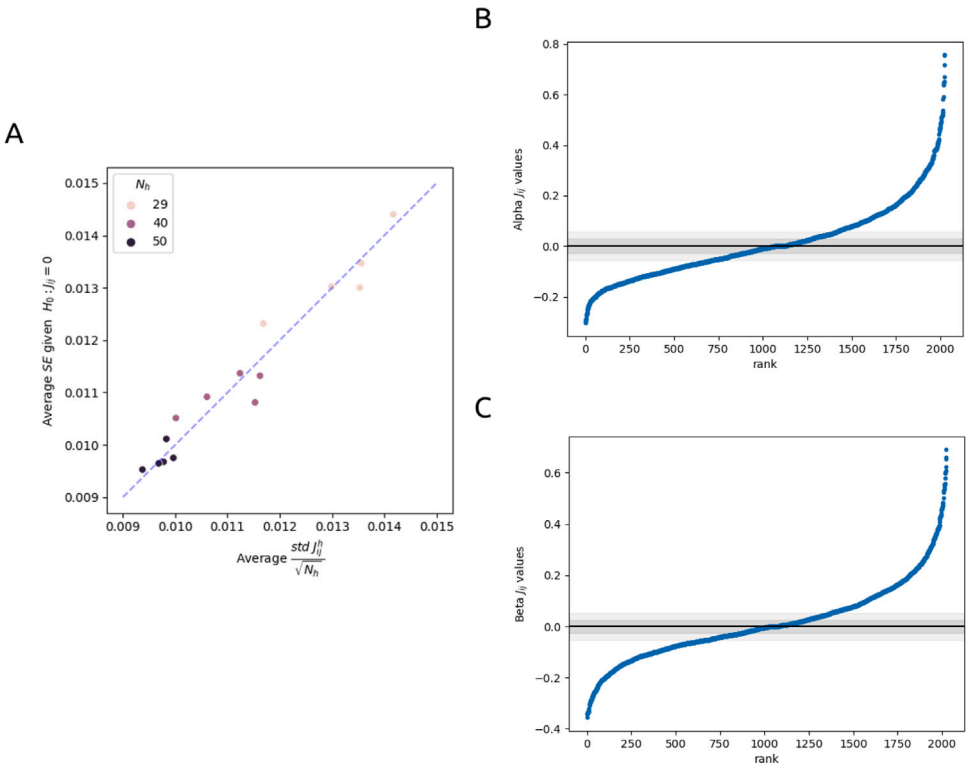

Fig. A.7. Effective Connectivity Significance. (A) Scatter plot of the average SE measured over the target edges that were zero, and the average standard deviation of the J_{ij}^h measured over the N_h subsequences, for different numbers of subsequences N_h = 29 (pink), 39 (purple), 40 (black). The dashed line is the x=y line. (B) and (C) Ranked weights of the experimental average alpha and beta J_{ij}. The gray regions represent the edges within one and two SE.

autoregressive dynamical model partially confirms the inhibitory role on the Visual and DAN networks found in alpha with RHoMM, first and fourth node in the $\bar{J}_{ij}$ matrix, as can be appreciated from the mean $\bar{J}_{ij}$ matrices in Fig. A.5B obtained respectively with RHoMM and the Autoregressive Dynamical Model.

Furthermore, RHoMM generally outperforms the Autoregressive Dynamical Model in retrieving the BLP dynamics, as suggested by the FC similarity bar plot. Notably, there are also some differences among the two matrices, but the comparison between the 2 models is challenging since RHoMM predicts the binary BLP sequences, whereas the Autoregressive Dynamical Model predicts the not-binarized BLP. As a consequence, a comparison at the level of predicted sequences is not possible. Thus, we acknowledge that more extensive comparison between RHoMM and DCM/Autoregressive Dynamical Model or Biophysical Network Model (BNM) would be of interest and could be the topic of multiple papers in the next future, requiring a dedicated research work.

Appendix C. Effective connectivity significance

In this Appendix, we discuss the statistical significance of the estimated edges. For this aim, we demonstrated that the standard deviation of the single estimated connections, std J_{ij}^h, gives us an estimate of the standard error (SE) of the edge population given the null hypothesis $J_{ij} = 0$. We generated synthetic data by constructing an objective network J_{ij} as in Section 2.7. We drew from a uniform distribution $[-1, 1]$ the weights for all J_{ij} connections. We set to zero the edges with weights smaller than the threshold 0.4. Then, we obtained the estimated network training RHoMM on different numbers of subsequences $N_h = 29$, 39, 40. We use $m = 355$ and $N_{steps} = 3{\times}10^5$, and $\alpha = 10^{-6}$. Fig. A.7 shows the average SE of J_{ij}^h conditioned on $J_{ij} = 0$ as a function of the average std $J_{ij}^h/\sqrt{N_h}$. This plot suggests that the obtained values are distributed around the $y = x$ line. This validates our hypothesis that we can use the average std J_{ij}^h to estimate if a single $\bar{J}_{ij}$ is significantly different from zero, that is if its absolute value is larger than the threshold given by 2 SE, with the SE = average (std $J_{ij}^h/\sqrt{N_h}$).

References

Abeysuriya, R. G., Hadida, J., Sotiropoulos, S. N., Jbabdi, S., Becker, R., Hunt, B. A., et al. (2018). A biophysical model of dynamic balancing of excitation and inhibition in fast oscillatory large-scale networks. *PLoS Computational Biology*, *14*(2), Article e1006007.

Adini, Y., Sagi, D., & Tsodyks, M. (1997). Excitatory-inhibitory network in the visual cortex: Psychophysical evidence. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(19).

Agcaoglu, O., Wilson, T. W., Wang, Y. P., Stephen, J., & Calhoun, V. D. (2019). Resting state connectivity differences in eyes open versus eyes closed conditions. *Human Brain Mapping*, *40*.

Amit, D. J., Gutfreund, H., & Sompolinsky, H. (1985). Storing infinite numbers of patterns in a spin-glass model of neural networks. *Physical Review Letters*, *55*(14), 1530–1533.

Ashburner, J., Barnes, G., Chen, C.-C., Daunizeau, J., Flandin, G., Friston, K., et al. (2021). *SPM12 manual the FIL methods group (and honorary members)*: *Tech. rep.*, Functional Imaging Laboratory, Wellcome Centre for Human Neuroimaging, UCL Queen Square Institute of Neurology.

Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, *20*(3), 353–364.

Basti, A., Pizzella, V., Chella, F., Romani, G. L., Nolte, G., & Marzetti, L. (2018). Disclosing large-scale directed functional connections in MEG with the multivariate phase slope index. *NeuroImage*, *175*, 161–175.

Bastos, A., Litvak, V., Moran, R., Bosman, C., Fries, P., & Friston, K. (2015). A DCM study of spectral asymmetries in feedforward and feedback connections between visual areas V1 and V4 in the monkey. *NeuroImage*, *108*, 460–475.

Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J. M., Oostenveld, R., Dowdall, J. R., et al. (2015). Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron*, *85*.

Berger, H. (1929). Über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, *87*, 527–570.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, *13*(10), 281–305.

Betti, V., Corbetta, M., de Pasquale, F., Wens, V., & Penna, S. D. (2018). Topology of functional connectivity and hub dynamics in the beta band as temporal prior for natural vision in the human brain. *Journal of Neuroscience*, *38*(15), 3858–3871.

Betti, V., DellaPenna, S., de Pasquale, F., Mantini, D., Marzetti, L., Romani, G. L., et al. (2013). Natural scenes viewing alters the dynamics of functional connectivity in the human brain. *Neuron*, *79*, 782–797.

Betti, V., Penna, S. D., de Pasquale, F., & Corbetta, M. (2021). Spontaneous beta band rhythms in the predictive coding of natural stimuli. *Neuroscientist*, *27*, 184–201.

Betzel, R. F., Avena-Koenigsberger, A., Goñi, J., He, Y., de Reus, M. A., Griffa, A., et al. (2016). Generative models of the human connectome. *NeuroImage*, *124*, 1054–1064.

Betzel, R. F., & Bassett, D. S. (2017). Generative models for network neuroscience: prospects and promise. *Journal of the Royal Society Interface*, *14*, Article 20170623.

Bottou, L. (2012). Stochastic gradient descent tricks. *Neural Networks, Tricks of the Trade, Reloaded*, *7700*, 430–445.

Brandes, U. (2001). A faster algorithm for betweenness centrality*. *Journal of Mathematical Sociology*, *25*(2), 163–177.

Breakspear, M. (2017). Dynamic models of large-scale brain activity. *Nature Neuroscience*, *20*, 340–352.

Bressler, S. L., & Seth, A. K. (2011). Wiener–Granger Causality: A well established methodology. *NeuroImage*, *58*(2), 323–329.

Brookes, M., Woolrich, M., & Barnes, G. (2012). Measuring functional connectivity in MEG: A multivariate approach insensitive to linear source leakage. *NeuroImage*, *63*(2), 910–920.

Brookes, M. J., Woolrich, M., Luckhoo, H., Price, D., Hale, J. R., Stephenson, M. C., et al. (2011). Investigating the electrophysiological basis of resting state networks using magnetoencephalography. *Proceedings of the National Academy of Sciences*, *108*(40), 16783–16788.

Brunel, N. (2016). Is cortical connectivity optimized for storing information? *Nature Neuroscience*, *19*(5), 749–755.

Bullmore, E., & Sporns, O. (2012). The economy of brain network organization. *Nature Reviews Neuroscience*, *13*, 336–349.

Buzsáki, G. (2006). *Rhythms of the brain*. New York: Oxford University Press.

Cabral, J., Hugues, E., Sporns, O., & Deco, G. (2011). Role of local network oscillations in resting-state functional connectivity. *NeuroImage*, *57*, 130–139.

Cabral, J., Kringelbach, M. L., & Deco, G. (2014). Exploring the network dynamics underlying brain activity during rest. *Progress in Neurobiology*, *114*, 102–131.

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.

Cabral, J., Luckhoo, H., Woolrich, M., Joensson, M., Mohseni, H., Baker, A., et al. (2014). Exploring mechanisms of spontaneous functional connectivity in MEG: How delayed network interactions lead to structured amplitude envelopes of band-pass filtered oscillations. *NeuroImage*, *90*, 423–435.

Carbo, E. W. S., Hillebrand, A., van Dellen, E., Tewarie, P., de Witt Hamer, P. C., Baayen, J. C., et al. (2017). Dynamic hub load predicts cognitive decline after resective neurosurgery. *Scientific Reports*, *7*, 42117.

Carnevale, F., de Lafuente, V., Romo, R., Barak, O., & Parga, N. (2015). Dynamic control of response criterion in premotor cortex during perceptual detection under temporal uncertainty. *Neuron*, *86*(4), 1067–1077.

Carpenter, G. A. (1989). Neural network models for pattern recognition and associative memory. *Neural Networks*, *2*(4), 243–257.

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.

Costantini, G., & Perugini, M. (2014). Generalization of clustering coefficients to signed correlation networks. In A. de la Fuente (Ed.), *PLoS One*, *9*(2), Article e88669.

David, O., Kiebel, S. J., Harrison, L. M., Mattout, J., Kilner, J. M., & Friston, K. J. (2006). Dynamic causal modeling of evoked responses in EEG and MEG. *NeuroImage*, *30*(4), 1255–1272.

de Pasquale, F., Chiacchiaretta, P., Pavone, L., Sparano, A., Capotosto, P., Grillea, G., et al. (2021). Brain topological reorganization associated with visual neglect after stroke. *Brain Connectivity*.

de Pasquale, F., Corbetta, M., Betti, V., & Della Penna, S. (2018). Cortical cores in network dynamics. *NeuroImage*, *180*, 370–382.

de Pasquale, F., Della Penna, S., Snyder, A. Z., Lewis, C., Mantini, D., Marzetti, L., et al. (2010). Temporal dynamics of spontaneous MEG activity in brain networks. *Proceedings of the National Academy of Sciences*, *107*(13), 6040–6045.

de Pasquale, F., Della Penna, S., Snyder, A. Z., Marzetti, L., Pizzella, V., Romani, G. L., et al. (2012). A cortical core for dynamic integration of functional networks in the resting human brain. *Neuron*, *74*(4), 753–764.

de Pasquale, F., Della Penna, S., Sporns, O., Romani, G. L., & Corbetta, M. (2016). A dynamic core network and global efficiency in the resting human brain. *Cerebral Cortex*, *26*(10), 4015–4033.

de Pasquale, F., Spadone, S., Betti, V., Corbetta, M., & Della Penna, S. (2021). Temporal modes of hub synchronization at rest. *NeuroImage*, *235*, Article 118005.

Deco, G., Jirsa, V. K., & McIntosh, A. R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature Reviews Neuroscience*, *12*, 43–56.

Della Penna, S., Corbetta, M., Wens, V., & de Pasquale, F. (2019). The impact of the geometric correction scheme on MEG functional topology at rest. *Frontiers in Neuroscience*, *13*, 1114.

Della Penna, S., Delgratta, C., Granata, C., Pasquarelli, A., Pizzella, V., Rossi, R., et al. (2000). Biomagnetic systems for clinical use. *Philosophical Magazine B*, *80*(5), 937–948.

Favaretto, C., Spadone, S., Sestieri, C., Betti, V., Cenedese, A., Penna, S. D., et al. (2021). Multi-band MEG signatures of BOLD connectivity reorganization during visuospatial attention. *NeuroImage*, *230*, Article 117781.

Ferguson, K. A., & Cardin, J. A. (2020). Mechanisms underlying gain modulation in the cortex. *Nature Reviews Neuroscience*, (2).

Folli, V., Gosti, G., Leonetti, M., & Ruocco, G. (2018). Effect of dilution in asymmetric recurrent neural networks. *Neural Networks*, *104*, 50–59.

Folli, V., Leonetti, M., & Ruocco, G. (2017). On the maximum storage capacity of the hopfield model. *Frontiers in Computational Neuroscience*, *10*, 144.

Frässle, S., Harrison, S. J., Heinzle, J., Clementz, B. A., Tamminga, C. A., Sweeney, J. A., et al. (2021). Regression dynamic causal modeling for resting-state fMRI. *Human Brain Mapping*, *42*(7), 2159–2180.

Frässle, S., Lomakina, E. I., Kasper, L., Manjaly, Z. M., Leff, A., Pruessmann, K. P., et al. (2018). A generative model of whole-brain effective connectivity. *NeuroImage*, *179*, 505–529.

Frässle, S., Lomakina, E. I., Razi, A., Friston, K. J., Buhmann, J. M., & Stephan, K. E. (2017). Regression DCM for fMRI. *NeuroImage*, *155*, 406–421.

Fries, P. (2015). Rhythms for cognition: Communication through coherence. *Neuron*, *88*, 220–235.

Friston, K. J. (2011). Functional and Effective Connectivity: A Review. *Brain Connectivity*, *1*(1), 13–36.

Friston, K., Bastos, A., Litvak, V., Stephan, K., Fries, P., & Moran, R. (2012). DCM for complex-valued data: Cross-spectra, coherence and phase-delays. *NeuroImage*, *59*, 439–455.

Friston, K. J., Kahan, J., Biswal, B., & Razi, A. (2014). A DCM for resting state fMRI. *NeuroImage*, *94*, 396–407.

Friston, K. J., Li, B., Daunizeau, J., & Stephan, K. E. (2011). Network discovery with DCM. *NeuroImage*, *56*(3), 1202–1221.

Friston, K., Preller, K. H., Mathys, C., Cagnan, H., Heinzle, J., Razi, A., et al. (2019). Dynamic causal modelling revisited. *NeuroImage*, *199*, 730–744.

Gajewska-Dendek, E., Wróbel, A., Bekisz, M., & Suffczynski, P. (2019). Lateral inhibition organizes beta attentional modulation in the primary visual cortex. *International Journal of Neural Systems*, *29*.

Gosti, G., Folli, V., Leonetti, M., & Ruocco, G. (2019). Beyond the maximum storage capacity limit in Hopfield recurrent neural networks. *Entropy*, *21*(8), 726.

Gu, S., Pasqualetti, F., Cieslak, M., Telesford, Q. K., Yu, A. B., Kahn, A. E., et al. (2015). Controllability of structural brain networks. *Nature Communications*, *6*(1), 8414.

Haegens, S., Nácher, V., Luna, R., Romo, R., & Jensen, O. (2011). α-Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. *Proceedings of the National Academy of Sciences*, *108*(48), 19377–19382.

Hahn, G., Skeide, M. A., Mantini, D., Ganzetti, M., Destexhe, A., Friederici, A. D., et al. (2019). A new computational approach to estimate whole-brain effective connectivity from functional and structural MRI, applied to language development. *Scientific Reports*, *9*(1), 8479.

He, B., Astolfi, L., Valdes-Sosa, P. A., Marinazzo, D., Palva, S. O., Benar, C.-G., et al. (2019). Electrophysiological brain connectivity: Theory and implementation. *IEEE Transactions on Biomedical Engineering*, *66*(7), 2115–2137.

Hebb, D. O. (1949). *the Organization of Behavior; A Neuropsychological Theory*. New York: Wiley.

Hillar, C., Chan, T., Taubman, R., & Rolnick, D. (2021). Hidden hypergraphs, error-correcting codes, and critical learning in hopfield networks. *Entropy 2021, Vol. 23, Page 1494*, *23*, 1494.

Hillar, C. J., & Tran, N. M. (2018). Robust exponential memory in hopfield networks. *Journal of Mathematical Neuroscience*, *8*.

Hipp, J. F., Hawellek, D. J., Corbetta, M., Siegel, M., & Engel, A. K. (2012). Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nature Neuroscience*, *15*(6), 884–890.

Honey, C. J., Kötter, R., Breakspear, M., & Sporns, O. (2007). Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proceedings of the National Academy of Sciences*, *104*(24), 10240–10245.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, *79*(8), 2554–2558.

Isaacson, J. S., & Scanziani, M. (2011). How inhibition shapes cortical activity. *Neuron*, *72*(2).

Jafarian, A., Litvak, V., Cagnan, H., Friston, K. J., & Zeidman, P. (2020). Comparing dynamic causal models of neurovascular coupling with fMRI and EEG/MEG. *NeuroImage*, *216*, Article 116734.

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Frontiers in Human Neuroscience*, *4*, 186.

Kaminski, M. J., & Blinowska, K. J. (1991). A new method of the description of the information flow in the brain structures. *Biological Cybernetics*, *65*, 203–210.

Kiebel, S. J., Garrido, M. I., Moran, R. J., & Friston, K. J. (2008). Dynamic causal modelling for EEG and MEG. *Cognitive Neurodynamics*, *2*, 121–136.

Kim, D.-H., Park, J., & Kahng, B. (2017). Enhanced storage capacity with errors in scale-free hopfield neural networks: An analytical study. In C. Dovrolis (Ed.), *PLoS One*, *12*, Article e0184683.

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617.

Lansner, A. (2009). Associative memory models: from the cell-assembly theory to biophysically detailed cortex simulations. *Trends in Neurosciences*, *32*(3), 178–186.

Lanza, E., Di Angelantonio, S., Gosti, G., Ruocco, G., & Folli, V. (2021). A recurrent neural network model of C. Elegans responses to aversive stimuli. *Neurocomputing*, *430*, 1–13.

Larson-Prior, L., Oostenveld, R., Della Penna, S., Michalareas, G., Prior, F., Babajani-Feremi, A., et al. (2013). Adding dynamics to the human connectome project with MEG. *NeuroImage*, *80*, 190–201.

Leonetti, M., Folli, V., Milanetti, E., Ruocco, G., & Gosti, G. (2020). Network dilution and asymmetry in an efficient brain. *Philosophical Magazine*, *100*(20), 2544–2555.

Lobier, M., Siebenhühner, F., Palva, S., & Palva, J. M. (2014). Phase transfer entropy: A novel phase-based measure for directed connectivity in networks coupled by oscillatory interactions. *NeuroImage*, *85*, 853–872.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*.

Lord, A., Horn, D., Breakspear, M., & Walter, M. (2012). Changes in community structure of resting state functional connectivity in unipolar depression. *PLoS One*, *7*(8).

Lu, Q., Li, H., Luo, G., Wang, Y., Tang, H., Han, L., et al. (2012). Impaired prefrontal–amygdala effective connectivity is responsible for the dysfunction of emotion process in major depressive disorder: A dynamic causal modeling study on MEG. *Neuroscience Letters*, *523*(2), 125–130.

Magrans de Abril, I., Yoshimoto, J., & Doya, K. (2018). Connectivity inference from neural recording data: Challenges, mathematical bases and research directions. *Neural Networks*, *102*, 120–137.

Makuuchi, M., & Friederici, A. D. (2013). Hierarchical functional connectivity between the core language system and the working memory system. *Cortex*, *49*(9), 2416–2423.

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, *503*(7474), 78–84.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*(2), 209–220.

Mantini, D., Penna, S. D., Marzetti, L., de Pasquale, F., Pizzella, V., Corbetta, M., et al. (2011). A signal-processing pipeline for magnetoencephalography resting-state networks. *Brain Connectivity*, *1*(1), 49–59.

Marzetti, L., Della Penna, S., Snyder, A. Z., Pizzella, V., Nolte, G., de Pasquale, F., et al. (2013). Frequency specific interactions of MEG resting state activity within and across brain networks as revealed by the multivariate interaction measure. *NeuroImage*, *79*, 172–183.

Mesulam, M. M. (1998). From sensation to cognition. *Brain*, *121*.

Monti, M., Fiorentino, J., Milanetti, E., Gosti, G., & Tartaglia, G. G. (2022). Prediction of time series gene expression and structural analysis of gene regulatory networks using recurrent neural networks. *Entropy 2022, Vol. 24, Page 141*, *24*(2), 141.

Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., & Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clinical Neurophysiology*, *115*(10), 2292–2307.

O'Neill, G. C., Tewarie, P. K., Colclough, G. L., Gascoyne, L. E., Hunt, B. A., Morris, P. G., et al. (2017). Measurement of dynamic task related functional networks using MEG. *NeuroImage*, *146*, 667–678.

O'Neill, G. C., Tewarie, P., Vidaurre, D., Liuzzi, L., Woolrich, M. W., & Brookes, M. J. (2018). Dynamics of large-scale electrophysiological networks: A technical review. *NeuroImage*, *180*, 559–576.

Onnela, J.-P., Saramäki, J., Kertész, J., & Kaski, K. (2005). Intensity and coherence of motifs in weighted complex networks. *Physical Review E*, *71*(6), Article 065103.

Palva, J. M., Wang, S. H., Palva, S., Zhigalov, A., Monto, S., Brookes, M. J., et al. (2018). Ghost interactions in MEG/EEG source space: A note of caution on inter-areal coupling measures. *NeuroImage*, *173*, 632–643.

Penny, W. D., Litvak, V., Fuentemilla, L., Duzel, E., & Friston, K. (2009). Dynamic causal models for phase coupling. *Journal of Neuroscience Methods*, *183*.

Pereira, U., & Brunel, N. (2018). Attractor dynamics in networks with learning rules inferred from in vivo data. *Neuron*, *99*(1), 227–238.e4.

Pfurtscheller, G., Stancák, A., & Neuper, C. (1996). Event-related synchronization (ERS) in the alpha band — An electrophysiological correlate of cortical idling: A review. *International Journal of Psychophysiology*, *24*(1–2), 39–46.

Poil, S. S., Hardstone, R., Mansvelder, H. D., & Linkenkaer-Hansen, K. (2012). Critical-state dynamics of avalanches and oscillations jointly emerge from balanced excitation/inhibition in neuronal networks. *Journal of Neuroscience*, *32*(29), 9817–9823.

Preti, M. G., & Van De Ville, D. (2017). Dynamics of functional connectivity at high spatial resolution reveal long-range interactions and fine-scale organization. *Scientific Reports*, *7*(1).

Raichle, M. E. (2011). The restless brain. *Brain Connectivity*, *1*(1), 3–12.

Razi, A., Seghier, M. L., Zhou, Y., McColgan, P., Zeidman, P., Park, H. J., et al. (2017). Large-scale DCMs for resting-state fMRI. *Network Neuroscience*, *1*(3), 241.

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, *52*(3), 1059–1069.

Salvador, R., Suckling, J., Coleman, M. R., Pickard, J. D., Menon, D., & Bullmore, E. (2005). Neurophysiological architecture of functional magnetic resonance images of human brain. *Cerebral Cortex*, *15*(9).

Sanzeni, A., Akitake, B., Goldbach, H. C., Leedy, C. E., Brunel, N., & Histed, M. H. (2020). Inhibition stabilization is a widespread property of cortical networks. *eLife*, *9*.

Sebastiani, V., de Pasquale, F., Costantini, M., Mantini, D., Pizzella, V., Romani, G. L., et al. (2014). Being an agent or an observer: Different spectral dynamics revealed by MEG. *NeuroImage*, *102*(P2), 717–728.

Singh, M. F., Braver, T. S., Cole, M. W., & Ching, S. N. (2020). Estimation and validation of individualized dynamic brain models with resting state fMRI. *NeuroImage*, *221*, Article 117046.

Spadone, S., Penna, S. D., Sestieri, C., Betti, V., Tosoni, A., Perrucci, M. G., et al. (2015). Dynamic reorganization of human resting-state networks during visuospatial attention. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 8112–8117.

Spadone, S., Wyczesany, M., Della Penna, S., Corbetta, M., & Capotosto, P. (2021). Directed flow of beta band communication during reorienting of attention within the dorsal attention network. *Brain Connectivity*, *11*(9), 717–724.

Spitzer, B., & Haegens, S. (2017). Beyond the status quo: A role for beta oscillations in endogenous content (RE)activation. *eNeuro*, *4*.

Sporns, O. (2013). Network attributes for segregation and integration in the human brain. *Current Opinion in Neurobiology*, *23*(2), 162–171.

Sporns, O., & Kötter, R. (2004). Motifs in brain networks. In K. J. Friston (Ed.), *PLoS Biology*, *2*(11), Article e369.

Stam, C. J., Nolte, G., & Daffertshofer, A. (2007). Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Human brain mapping*, *28*(11), 1178–1193.

Tagliazucchi, E., Balenzuela, P., Fraiman, D., & Chialvo, D. R. (2012). Criticality in large-scale brain fMRI dynamics unveiled by a novel point process analysis. *Frontiers in Physiology*, *3*, 15.

Tang, W., Liu, H., Douw, L., Kramer, M. A., Eden, U. T., Hämäläinen, M. S., et al. (2017). Dynamic connectivity modulates local activity in the core regions of the default-mode network. *Proceedings of the National Academy of Sciences*, *114*, 9713–9718.

Thomas Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165.

Tu, C., Rocha, R. P., Corbetta, M., Zampieri, S., Zorzi, M., & Suweis, S. (2018). Warnings and caveats in brain controllability. *NeuroImage*, *176*.

Vidaurre, D., Smith, S. M., & Woolrich, M. W. (2017). Brain network dynamics are hierarchically organized in time. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(48), 12827–12832.

Wang, S. H., Lobier, M., Siebenhühner, F., Puoliväli, T., Palva, S., & Palva, J. M. (2018). Hyperedge bundling: A practical solution to spurious interactions in MEG/EEG source connectivity analyses. *NeuroImage*, *173*, 610–622.

Wens, V., Marty, B., Mary, A., Bourguignon, M., Op de Beeck, M., Goldman, S., et al. (2015). A geometric correction scheme for spatial leakage effects in MEG/EEG seed-based functional connectivity mapping. *Human Brain Mapping*, *36*(11), 4604–4621.

Werbos, P. J. (1990). Backpropagation through time: What it does and how to do it. *Proceedings of the IEEE*, *78*(10), 1550–1560.

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. In *1960 IRE WESCON convention record* (pp. 96–104).

Wilson, D. E., Scholl, B., & Fitzpatrick, D. (2018). Differential tuning of excitation and inhibition shapes direction selectivity in ferret visual cortex. *Nature*, *560*(7716).

Zalesky, A., Fornito, A., & Bullmore, E. T. (2010). Network-based statistic: Identifying differences in brain networks. *NeuroImage*, *53*, 1197–1207.

Zang, C., & Wang, F. (2020). Neural dynamics on complex networks. In *Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining, vol. 11, no. 20* (pp. 892–902). Association for Computing Machinery.

Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, *4*(1).

Zhang, S., Jiang, H., Wei, S., & Dai, L.-R. (2015). Rectified linear neural networks with tied-scalar regularization for LVCSR. In *Proceedings of the annual conference of the international speech communication association, vol. 2015-Janua* (pp. 2635–2639). International Speech and Communication Association.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., et al. (2020). Graph neural networks: A review of methods and applications. *AI Open*, *1*, 57–81.