Parameter inference on brain network models with unknown node dynamics and spatial heterogeneity

Viktor Sip*, Spase Petkoski, Meysam Hashemi, Viktor Jirsa*

Aix Marseille Univ, INSERM, INS, Inst Neurosci Syst, Marseille, France

Abstract

Model-based data analysis of whole-brain dynamics links the observed data to model parameters in a network of neural masses. In recent years a special focus was placed on the role of regional variance of model parameters for the emergent activity. Such analyses however depend on the properties of the employed neural mass model, which is often obtained through a series of major simplifications or analogies. Here we propose a data-driven approach where the neural mass model needs not to be specified. Building on the recent progresses in identification of dynamical systems with neural networks, we propose a method to infer from the functional data both the neural mass model representing the regional dynamics as well as the region- and subject-specific parameters, while respecting the known network structure. We demonstrate on two synthetic data sets that our method is able to recover the original model parameters, and that the trained generative model produces dynamics resembling the training data both on the regional level and on the whole-brain level. The present approach opens a novel way to the analysis of resting-state fMRI with possible applications in understanding the changes of whole-brain dynamics during aging or in neurodegenerative diseases.

Keywords: large-scale brain network modeling, model discovery, parameter inference, resting-state fMRI

1. Introduction

One avenue for analysis of resting-state functional magnetic resonance imaging (fMRI) is the use of computational models of large-scale brain network dynamics (Breakspear, 2017; Suárez et al., 2020). A general goal of this approach is to relate the observed brain activity to the dynamical repertoire of the computational model, possibly via identification of optimal model parameters, leading to a better mechanistic interpretation of the observations. Such

*Corresponding author

Email addresses: viktor.sip@univ-amu.fr (Viktor Sip), viktor.jirsa@univ-amu.fr (Viktor Jirsa)

models are network-based, where the nodes represent brain regions and the edges the structural connections between them. Models constrained by individual brain imaging data are referred to as virtual brains and typically use diffusion-weighted imaging data for the edges and neural mass models for the local dynamics of a brain region. Neural masses are low-dimensonal models of neuronal population activity.

When linking the models with the observations, until recently studies focused only on a small number of parameters - such as the global coupling strength - due to the computational costs associated with the exploration of a high-dimensional parameter space. In recent years, however, several works utilized the whole-brain modeling framework in order to explore the role of spatial heterogeneity of model parameters. Specifically, the studies found that the whole-brain models can better reproduce the features of resting-state fMRI when the regional variability is constrained by the MRI-derived estimates of intracortical myelin content (Demirtaş et al., 2019), functional gradient (Kong et al., 2021), or gene expression profiles (Deco et al., 2021), and similar regional variability was found even without prior restrictions (Wang et al., 2019).

Neural mass models employed in these studies (such as the dynamic mean field model of conductance-based spiking neural network (Deco et al., 2013) or Hopf bifurcation model of neural excitability (Deco et al., 2017)) are derived through a series of major simplifications or built upon loose mathematical analogies. It can thus be questioned to what degree the dynamical structure embodied in these models is sufficient to capture the essential elements of the neural dynamics manifesting in the observed data. Would two different neural mass models lead to the same conclusions, or do the results strongly depend on the exact model form? Such questions are not yet sufficiently answered.

Meanwhile, novel techniques to learn the models of nonlinear dynamical systems from the data itself are being developed and applied in various fields of physical and life sciences (Linderman et al., 2017; Duncker et al., 2019; Roeder et al., 2019; Nassar et al., 2019; Schmidt et al., 2021), including in neuroscience on all scales (Pandarinath et al., 2018; Koppe et al., 2019; Singh et al., 2020). The common assumption in these approaches is that the observed data are generated by an unknown dynamical system of reasonably low dimensionality, which can be represented with a flexible artificial neural network. The parameters of this network are learned during training, so that it best reproduces the data.

These developments raise the question whether a similar approach can be applied in the context of whole-brain modeling: Can we learn a dynamical system representing a neural mass at each node of a large-scale brain network? Such approach would allow to side-step the issue of reliance on a specific neural mass models which lie at the heart of the large-scale modeling, and instead extract this model directly from the functional data. That is what we aim to investigate in this work. Using the known network structure, derived from diffusion-weighted imaging, and the observed resting-state fMRI, we infer the dynamical system representing the neural masses in the nodes of the network. To account for the regional and subject heterogeneity, we allow this (initially unknown) neural mass model to depend on a region-specific and subject-specific parameters. These parameters we also infer from the observations together with the model, obtaining the map of dynamically-relevant parameters.

To do so, we utilize the framework of amortized variational inference, or variational autoencoders (Kingma and Welling, 2019), inspired in particular by its application for inferring neural population dynamics (Pandarinath et al., 2018) and for dynamical systems with hierarchical structure (Roeder et al., 2019). In brief, our system is composed of an encoding network, mapping the observed time series to the subject- and region-specific parameters and to the trajectory in the source space, a neural network representing the dynamical system, and the observation model acting as the decoder from the source to the observation space. These are jointly trained to maximize the evidence lower bound, so that the predictions of the trained model closely resemble the original data.

In this work we test our method on two synthetic data sets, generated with the two models commonly used in large-scale brain modeling: the mean field model of conductance-based spiking neural network, or mean field model for short (Deco et al., 2013), and the Hopf bifurcation model (Deco et al., 2017). For both test cases we use a cohort of eight subjects with realistic structural connectomes, and with model parameters varying across subjects and brain regions. We show that the trained generative model can reproduce many features of the original data set, and we demonstrate that the method can extract regional and subject-specific parameters strongly related to the original parameters used for the simulation.

2. Methods

2.1. Structural connectomes

The synthetic data sets were generated using the structural connectomes of eight subjects from Human Connectome Project (Van Essen et al., 2012). Specifically, eight subjects from HCP 1200 Subjects cohort were used (ID numbers 100307, 100408, 101107, 101309, 101915, 103111, 103414, and 103818). For those, *Structural Preprocessed* and *Diffusion Preprocessed* packages were downloaded (Glasser et al., 2013). Next, the structural connectomes were built for the cortical regions of Desikan-Killiany parcellation (Desikan et al., 2006) using MRtrix 3.0 (Tournier et al., 2012). To do so, first the response function for spherical deconvolution was estimated using the *dhollander* algorithm (Dhollander et al., 2016). Next, fibre orientation distribution was estimated using multi-shell multi-tissue constrained spherical deconvolution (Jeurissen et al., 2014). Then 10 million tracks were generated using the probabilistic iFOD2 (second-order integration over fiber orientation distributions) algorithm (Tournier et al., 2010). These were then filtered using the SIFT algorithm (Smith et al., 2013). Finally, the connectome were built by counting the tracks connecting all pairs of brain regions in the parcellation. The connectome matrices were normalized so that the largest element in each was equal to one.

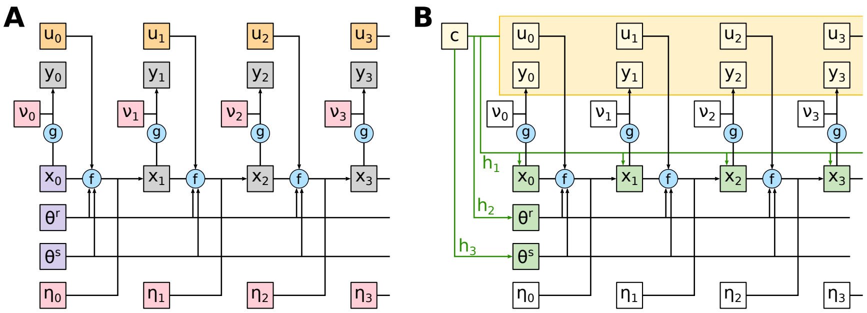

Figure 1: Overview of the method architecture visualized for one brain region. In the sketches we drop the region indices for simplicity, and keep only the time indices. (A) Generative model. With known functions f and g, and given initial conditions $\boldsymbol{x}_0$ and parameters $\boldsymbol{\theta}^r$ and $\boldsymbol{\theta}^s$, the model can be simulated in forward fashion, influenced by the system noise $\boldsymbol{\eta}$ and observation noise ν. The network input for region j at time k is calculated on the fly from the current states of other regions, $u_{j,k} = \sum_{i=1}^{n} w_{ji} y_{i,k}$. (B) Inference model. The data (observation time series $\boldsymbol{y}$, precomputed network input time-series $\boldsymbol{u}$ and one-hot vector $\boldsymbol{c}$ identifying the subject) are mapped through the encoder functions h_1, h_2, and h_3 onto the system states $\boldsymbol{x}$, region-specific parameters $\boldsymbol{\theta}^r$ and subject-specific parameters $\boldsymbol{\theta}^s$. The observation function g appears in the likelihood function, while the system evolution function f enters the prior on the states. The noise $\boldsymbol{\eta}$ and ν is present only implicitly via the likelihood and the prior functions. The inference problem amounts to the maximization of the resulting ELBO over the parameters of the generative model f, g, encoder functions h_1, h_2, h_3, and variance of the system and observation noise.

The perturbed connectomes were constructed by taking the original connectome W and adding a matrix with elements from random normal distribution, scaled by the perturbation magnitude ϵ, i.e. $W_\epsilon = W + \epsilon A$. For each value of perturbation magnitude, four different perturbed connectomes were built. The log-scaled connectome was calculated as $W_{\log} = \log_{10}(W + 10^q)$ with $q = -3$. In all cases the matrices were also normalized so that the maximal element was equal to one.

2.2. Amortized variational inference for networks of nonlinear dynamical systems

Overview. Before we delve into the details of the method, we introduce the general ideas behind our approach. We follow the general framework of large-scale brain network modeling, and we assume that for a specific subject the observations $y_j(t)$ of a brain region j are generated by a dynamical system

$$\dot{\boldsymbol{x}}_j(t) = f\left(\boldsymbol{x}_j(t), \boldsymbol{\theta}_j^r, \boldsymbol{\theta}^s, u_j(t)\right) + \boldsymbol{\eta}_j(t), \tag{1}$$

$$y_j(t) = g(\boldsymbol{x}_j(t)) + \nu_j(t) \tag{2}$$

where $\boldsymbol{x}_j(t) \in \mathbb{R}^{n_s}$ is the state at time t, $\boldsymbol{\theta}_j^r \in \mathbb{R}^{m_r}$ and $\boldsymbol{\theta}^s \in \mathbb{R}^{m_s}$ are the region-specific and subject-specific parameters, and

$$u_j(t) = \sum_{i=1}^{n} w_{ji} g_c(\boldsymbol{x}_j(t)) \tag{3}$$

is the network input with $\{w_{ij}\}_{i,j=1}^{n}$ being the structural connectome matrix of the network with n nodes. The functions f, g, and g_c are initially unknown, and $\boldsymbol{\eta}_j(t)$ and $\nu_j(t)$ is the system and observation noise, respectively (Fig. 1A).

From the observed time series of multiple subjects we wish to infer both the model functions f and g (and g_c, which for simplicity we assume is identical to g), shared across the subjects, as well as region- and subject-specific parameters $\boldsymbol{\theta}_j^r$ and $\boldsymbol{\theta}^s$. To do so, we adopt the general framework of amortized variational inference (Kingma and Welling, 2019) with hierarchical structure in parameters (Roeder et al., 2019) (Fig. 1B). We consider the states $\boldsymbol{x}_j$, and the parameters $\boldsymbol{\theta}_j^r$ and $\boldsymbol{\theta}^s$ the latent variables, and we seek their approximate posterior distribution represented by multivariate Gaussian distributions. In the spirit of amortized variational inference, we do not optimize their parameters directly, but through encoder functions h_1, h_2, and h_3, which transform the data in the latent variables (system states, regional, and subject parameters respectively). The assumption that the observation and coupling functions are identical, $g \equiv g_c$, allows us to effectively decouple the network problem to uncoupled regions with known network input, and so we can consider time-series of one region of one subject as a single data point. We represent the nonlinear function f with a generic artificial neural network, and function g as a linear transformation. The inference problem is ultimately transformed into optimization of the cost function, evidence lower bound (ELBO), which is to be maximized over the weights of f, g, h_1, h_2, and h_3, and over the variances of the system and observation noise. After the optimization, we obtain the description of the dynamical system in terms of functions f and g, probabilistic representation of the regional and subject parameters $\boldsymbol{\theta}_j^r$ and $\boldsymbol{\theta}^s$, as well as projections of the observations in the state space $\boldsymbol{x}_j$. The inferred parameters $\boldsymbol{\theta}_j^r$ and $\boldsymbol{\theta}^s$ will not have a mechanistic meaning; however, they can provide a measure of (dis)similarity of the regions and subject, and they can be interpreted via the inferred dynamical system f.

Generative dynamical system. As outlined above, to make the inference problem more tractable, we simplify the problem and assume that the nodes are coupled through the observed variable y_j. More precisely, we assume that in Eqs. (1-3) $g \equiv g_s$, and that the observation noise term ν_j is small enough that it can be included in the coupling. Then the network input has the form

$$u_j(t) = \sum_{i=1}^{n} w_{ji} y_i(t). \tag{4}$$

This form has the advantage that the network input is independent of any hidden variables and can be computed directly from the known observations y_j. This effectively decouples the time series in different nodes so that they can be processed separately, as described below.

For the purpose of the inference, we use the time-discretized form of Eqs. (1-

2) utilizing the Euler method,

$$\begin{aligned} \boldsymbol{x}_{j,k+1} &= \boldsymbol{x}_{j,k} + \Delta_t f\left(\boldsymbol{x}_{j,k}, \boldsymbol{\theta}_j^r, \boldsymbol{\theta}^s, u_{j,k}\right) + \boldsymbol{\eta}_{j,k}, \\ y_{j,k} &= g(\boldsymbol{x}_{j,k}) + \nu_{j,k}, \end{aligned}$$

where we denote the time step with the index k.

Evidence lower bound. As usual in variational inference, we aim to maximize the evidence lower bound (ELBO), and by doing so at the same time minimize the Kullback-Leibler divergence between the true posterior and the approximate posterior q. In the following text, we consider only a single data point from one subject and one region, and we omit the region indexing for brevity.

A single data point $\{\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{c}\}$ representing the data from a one region is composed of the observed time series $\boldsymbol{y} \in \mathbb{R}^{n_t}$, network input time series $\boldsymbol{u} \in \mathbb{R}^{n_t}$, and one-hot vector $\boldsymbol{c} \in \mathbb{R}^{n_{sub}}$, that is, a vector with zeros everywhere except i-th position with value one, encoding the identity of subject i. For this data point the ELBO can be expressed as follows. (For details see Supplementary Information.)

$$\begin{aligned} L = {} & \mathbb{E}_q[\log p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}^r, \boldsymbol{\theta}^s, \boldsymbol{u})] && (5) \\ & + \mathbb{E}_q[\log p(\boldsymbol{x}|\boldsymbol{\theta}^r, \boldsymbol{\theta}^s, \boldsymbol{u})] + \mathbb{E}_q[\log p(\boldsymbol{\theta}^r)] + \frac{1}{n}\mathbb{E}_q[\log p(\boldsymbol{\theta}^s)] && (6) \\ & - \mathbb{E}_q[\log q(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{c})] - \mathbb{E}_q[\log q(\boldsymbol{\theta}^r|\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{c})] - \frac{1}{n}\mathbb{E}_q[\log q(\boldsymbol{\theta}^s|\boldsymbol{c})] && (7) \end{aligned}$$

Here the first line represents the decoder loss, second line the priors for states $\boldsymbol{x}$ and region- and subject-specific parameters $\boldsymbol{\theta}^r$ and $\boldsymbol{\theta}^s$, and the third line the approximate posteriors again for states, region-, and subject-specific parameters.

Decoder, or the observation model. We assume that the observation model can be modeled as a linear transformation of the system state with Gaussian noise, $y = g(\boldsymbol{x}) + \nu = \boldsymbol{a} \cdot \boldsymbol{x} + b + \nu$. This forward projection essentially represents the decoder part of the encoder-decoder system, and so the likelihood in Eq. (5) can be expanded over time as

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}^r, \boldsymbol{\theta}^s, \boldsymbol{u}) = \prod_{k=1}^{n_t} p(y_k|\boldsymbol{x}_k) = \prod_{k=1}^{n_t} N(y_k|\boldsymbol{a} \cdot \boldsymbol{x}_k + b, \sigma_o^2), \tag{8}$$

where $N(y|\mu, \sigma^2)$ stands for normal distribution with mean μ and variance σ^2. The parameters of the observation model, which are to be optimized, are the coefficients of the linear projection $\boldsymbol{a}$ and b, together with the observation noise variance σ_o^2.

Prior on the system states. The first term in Eq. (6) represents the prior function on the system states $\boldsymbol{x}$ given the input $\boldsymbol{u}$ and parameters $\boldsymbol{\theta}^r$ and $\boldsymbol{\theta}^s$, and

Parameter	**Value**
State space dimension n_s	2
Number of region-specific parameters m_{reg}	2
Number of subject-specific parameters m_{sub}	1
Number of hidden units in f	32
Number of LSTM units in h_1 and h_2	32
Batch size	16
Number of samples to evaluate expectation over the approximate posterior	8

Table 1: Method parameters used in the test cases on synthetic data.

it is here where the dynamical system f appears in the ELBO. This term can be expanded over time as

$$\begin{aligned} p(\boldsymbol{x}|\boldsymbol{\theta}^r, \boldsymbol{\theta}^s, \boldsymbol{u}) &= p(\boldsymbol{x}_0) \prod_{k=1}^{n_t} p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k, \boldsymbol{\theta}^r, \boldsymbol{\theta}^s, u_k) \\ &= N(\boldsymbol{x}_0|\boldsymbol{0}, \boldsymbol{I}) \prod_{k=1}^{n_t} N(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k + \Delta_t f(\boldsymbol{x}_k, \boldsymbol{\theta}^r, \boldsymbol{\theta}^s, u_k), \operatorname{diag}(\boldsymbol{\sigma}_s^2)). \end{aligned} \tag{9}$$

Here we use the standard normal distribution as a prior for the initial state $\boldsymbol{x}_0$, and then evolve the system over time according to the function f. We represent the function f as a two-layer neural network, with a Rectified Linear Unit (ReLU) activation function in the hidden layer. The weights of the network are to be optimized, together with the system noise standard deviation $\boldsymbol{\sigma}_s$. The number of hidden units is given in the Tab. 1.

Prior on the parameters. For the region- and subject-specific parameters we utilize the standard normal distribution as a prior, as is often used in variational autoencoders. The priors in the second and the third term in Eq. (6) can thus be written as $p(\boldsymbol{\theta}^r) = N(\boldsymbol{\theta}^r|\boldsymbol{0}, \boldsymbol{I})$ and $p(\boldsymbol{\theta}^s) = N(\boldsymbol{\theta}^s|\boldsymbol{0}, \boldsymbol{I})$.

Approximate posteriors. We follow the standard approach and utilize multivariate normal distributions for the approximate posteriors in Eq. (7). For the states $\boldsymbol{x}$ and region-specific parameters $\boldsymbol{\theta}^r$ we use the idea of amortized variational inference and instead of representing the parameters directly, we train a recurrent neural network to extract the means and the variances from the time series of the observations $\boldsymbol{y}$, time series of the network input $\boldsymbol{u}$, and the one-hot vector $\boldsymbol{c}$ encoding the subject identity:

$$(\boldsymbol{\mu}_x, \log \boldsymbol{\sigma}_x^2) = h_1(\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{c}), \tag{10}$$

$$q(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{c}) = N(\boldsymbol{x}|\boldsymbol{\mu}_x, \operatorname{diag}(\boldsymbol{\sigma}_x^2)), \tag{11}$$

and

$$(\boldsymbol{\mu}_r, \log \boldsymbol{\sigma}_r^2) = h_2(\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{c}), \tag{12}$$

$$q(\boldsymbol{\theta}^r | \boldsymbol{y}, \boldsymbol{u}, \boldsymbol{c}) = N(\boldsymbol{\theta}^r | \boldsymbol{\mu}_r, \operatorname{diag}(\boldsymbol{\sigma}_r^2)). \tag{13}$$

Specifically, we use Long Short-Term Memory (LSTM) networks for both functions h_1 and h_2. The input to the networks at step k is the concatenated observation y_k and the network input u_k, to which is also appended the time-independent one-hot vector $\boldsymbol{c}$.

The subject-specific parameters $\boldsymbol{\theta}^s$ depend only on the subject identity encoded in the one-hot vector $\boldsymbol{c}$. They are represented directly in the matrices of means $\boldsymbol{M}_s$ and log-variances $\boldsymbol{V}_s$. For a specific subject the relevant values are extracted through the product with the one-hot vector $\boldsymbol{c}$,

$$(\boldsymbol{\mu}_s, \log \boldsymbol{\sigma}_s^2) = h_3(\boldsymbol{c}) = (\boldsymbol{M}_s \cdot \boldsymbol{c}, \boldsymbol{V}_s \cdot \boldsymbol{c}) \tag{14}$$

$$q(\boldsymbol{\theta}^s | \boldsymbol{c}) = N(\boldsymbol{\theta}^s | \boldsymbol{\mu}_s, \operatorname{diag}(\boldsymbol{\sigma}_s^2)) \tag{15}$$

Optimization. The optimization target is the negative dataset ELBO,

$$L_{\text{dataset}} = \sum_{i=1}^{n_{sub}} \sum_{j=1}^{n} L_{ij}, \tag{16}$$

where L_{ij} is the ELBO associated with a subject i and region j, defined by Eqs. (5-7). We minimize the cost function over the weights of the LSTM networks h_1, h_2, weights of the neural network f, means and variances of the subject-specific parameters $\boldsymbol{M}_s$, $\boldsymbol{V}_s$, system and observation noise variances $\boldsymbol{\sigma}_s^2$ and σ_o^2 (in log-scale), and forward projection parameters $\boldsymbol{A}$ and $\boldsymbol{b}$.

The method is implemented in Keras 2.4 (Chollet et al., 2015). The parameters of the method and of optimization procedure are given in Tab. 1. For optimization we use the Adam algorithm (Kingma and Ba, 2017). The expectations in Eqs. (5-7) are approximated using samples drawn from the approximate posterior distribution. The optimization is run for 2000 epochs with learning rate 0.003 and then for additional 1000 epochs with learning rate 0.001. To make the optimization more stable we use gradient clipping with limits (-1000, 1000). To better guide the optimization procedure, we follow the previous works (Pandarinath et al., 2018) with initial ELBO relaxation: The terms corresponding to the priors and approximate posteriors for states $\boldsymbol{x}$ and parameters $\boldsymbol{\theta}^r$ and $\boldsymbol{\theta}^s$ (Eqs. (6-7)) are scaled by a coefficient β, which linearly increases from 0 to 1 between the first and 500th epoch.

Two regularization terms are added to the cost function. First is a L2 regularization on the kernel weights biases and of the neural network representing function f, $\alpha_f \left(\sum_{i=1}^{n_w} (w_i^f)^2 + \sum_{i=1}^{n_b} (b_i^f)^2 \right)$, where n_w and n_b is the number of kernel weights w_i^f and bias coefficients b_i^f, respectively. Second is on the states $\boldsymbol{x}$, $\alpha_x \sum_{i=1}^{n_s} \sum_{k=1}^{n_s} x_{k,i}^2$. We set $\alpha_f = 0.01$ and $\alpha_x = 0.01$.

The initial conditions for the optimization are set as follows. The log variances of the system noise are set to -2, and the log variances of the observation noise to 0. The projection vector $\boldsymbol{a}$ is initialized randomly drawing from normal distribution (mean 0, std 0.3), and the projection bias b is set to zero. Matrices for subject-specific parameters $\boldsymbol{M}_s$ and $\boldsymbol{V}_s$ are initialized randomly drawing from a normal distribution (mean 0, std 0.01). All layers of employed neural networks use the default initialization provided by Keras.

2.3. Whole-brain network models for simulated data sets

Hopf bifurcation model. The Hopf model of large-scale brain dynamics (Deco et al., 2017) is built by placing a neural mass near supercritical Hopf bifurcation at each node of a brain network. Each neural mass i is described by two parameters: bifurcation parameter a_i and intrinsic frequency f_i. For $a_i < 0$ the uncoupled neural mass has one stable fixed point, and for $a_i > 0$ the neural mass has a stable limit cycle indicating sustained oscillations with frequency f_i. The bifurcation exists at the critical value $a_i = 0$. The dynamics of each node in the network are given by a set of two coupled nonlinear stochastic differential equations,

$$\dot{x}_i =(a_i - x_i^2 - y_i^2)x_i - \omega_i y_i + G \sum_{j=1}^{n} w_{ij}(x_j - x_i) + \beta \eta_i^x(t), \tag{17}$$

$$\dot{y}_i =(a_i - x_i^2 - y_i^2)y_i + \omega_i x_i + G \sum_{j=1}^{n} w_{ij}(y_j - y_i) + \beta \eta_i^y(t), \tag{18}$$

where $\omega_i = 2\pi f_i$, $G > 0$ is the scaling of the coupling, w_{ij} is the weight of connection from node j to node i. Additive Gaussian noise η is included in the equations, with standard deviation β.

To generate the synthetic dataset, we use the structural connectome matrices of human subjects as described above. We simulate eight subjects, with increasing coupling coefficient G spaced linearly between 0 and 0.7. The intrinsic frequency f_i of all nodes is sampled randomly from uniform distribution on $[0.03, 0.07]$ Hz. The bifurcation parameter a is sampled randomly from uniform distribution $[-1, 1]$. The initial conditions of the system for all subjects and both variables are chosen randomly from normal distribution $N(0, 0.3)$, the system is then simulated for 205 s. First 25 s are then discarded to avoid the influence of the initial conditions, leaving 180 s of data. The system is simulated with Euler method with time step $\Delta_t = 0.02$ s. As the observed variable we take the first of the two variables in each node (i.e., x_i), downsampled to 1 Hz, therefore every timeseries contain 180 time points. The data are normalized to zero mean and variance equal to one (calculated across the whole data set).

Parametric mean field model. The parametric mean field model (pMFM) was derived as a reduction from a spiking neural model (Deco et al., 2013). The resulting model is described by one nonlinear stochastic differential equation in each node of the brain network,

$$\dot{S}_i = -\frac{S_i}{\tau_s} + (1 - S_i)\gamma H(x_i) + \sigma\eta_i(t), \tag{19}$$

$$H(x_i) = \frac{ax_i - b}{1 - \exp\left(-d(ax_i - b)\right)}, \tag{20}$$

$$x_i = r_i J S_i + I_o + I_n, \tag{21}$$

where x_i is the total input current, $H(x_i)$ is the population firing rate, and S_i the average synaptic gating variable. The total input current depends on the recurrent connection strength r_i, synaptic coupling strength $J = 0.2609\,\mathrm{nA}$, excitatory subcortical input $I_o = 0.295\,\mathrm{nA}$, and the regional coupling $I_n = G\sum_{j=1}^{n} w_{ij}S_j$, scaled by the global scaling coefficient G. The strength of the coupling between region j and i is proportional to the structural connection strength w_{ij}. The kinetic parameters of the models are the decay time constant $\tau_s = 100\,\mathrm{ms}$ and $\gamma = 0.641/1000$. Values for the input-output function $H(x_i)$ are $a = 270\,\mathrm{nC}^{-1}$, $b = 108\,\mathrm{Hz}$, $d = 0.154\,\mathrm{s}$. Depending on the parameter values and the strength of the network coupling, the system can be either in monostable downstate regime at low firing-rate values, bistable regime with two stable fixed points, or monostable upstate regime at high firing-rate values. The stochastic transitions between states are driven by the additive Gaussian noise η_i with standard deviation σ.

The initial conditions for S_i were chosen randomly from uniform distribution on $[0.2, 0.8]$. The system was simulated for eight subjects with connectome matrices described above. For each subject, a specific value of coupling coefficient G producing the strongest functional connectivity was used. This was determined by performing 4 minute long simulations with subject-specific connectome and fixed regionally heterogeneous parameters, repeated for 31 values of G between 0.17 and 0.22 (where optimal value was expected to lie), and picking the value where the mean of functional connectivity from the last 2 minutes was the highest. With this value of G, the activity of each subject was simulated for 16.4 minutes, first two of which were discarded to avoid the influence of the initial conditions. The Euler method with time step $\Delta_t = 10\,\mathrm{ms}$ was used for the simulation. The resulting time series of S_i were temporally averaged over windows of size 0.72 seconds, leaving 1200 time points in every time series. The data are normalized to zero mean and variance equal to one (calculated across the whole data set).

3. Results

Evaluation workflow. We test the proposed method on two synthetic data sets, where the data are generated by models commonly used in whole-brain modeling. First is the Hopf bifurcation model (Deco et al., 2017), shown on Fig. 2. That is a two-equation neural mass model, where depending on the value of the bifurcation parameter a_i the dynamics is either noise-driven around a stable fixed point (for $a_i < 0$) or oscillatory with frequency f_i (for $a_i > 0$). In the

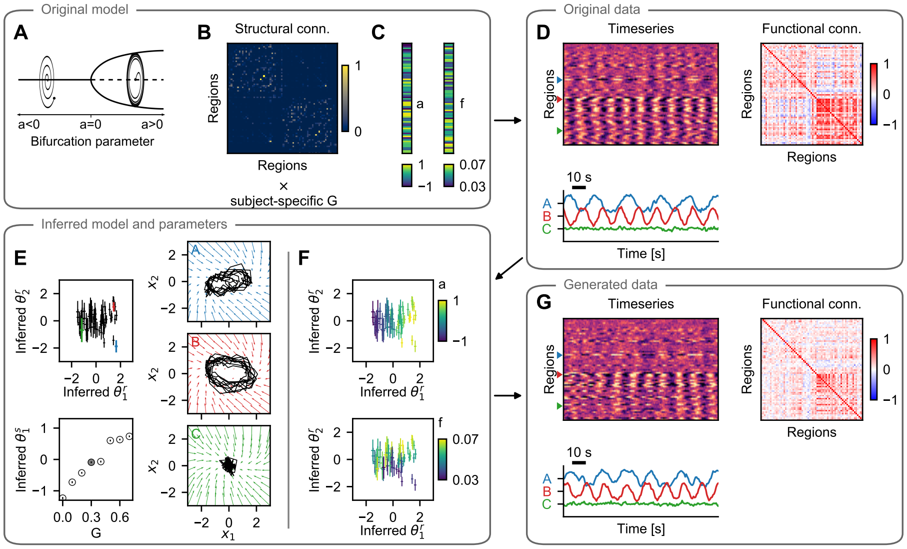

Figure 2: Hopf model test case: example subject. (A-C) The training data are simulated using a network model of brain dynamics, where in each node a Hopf neural mass model is placed (A). The nodes are coupled through a connectome derived from diffusion-weighted imaging (B) scaled by a subject-specific coupling parameter G. The values of bifurcation parameter a_i and intrinsic frequency f_i vary across brain regions (C). (D) Timeseries generated with the original model with three examples (bottom) and the calculated functional connectivity (right). (E) Inferred regional parameters for all regions (top left, example nodes highlighted in color) and inferred subject-specific parameter (bottom left, in gray among parameters for all subjects in the dataset). The span of the crosses/lines corresponds to two standard deviations of the inferred Gaussian distribution. In the bottom panel circles are added for visual aid due to the small standard deviations. The inferred dynamics in state space of the three example nodes are on the right. The vector field is evaluated assuming zero network input and using the inferred region- and subject-specific parameters. (F) Inferred regional parameters colored by the ground truth values of the bifurcation parameter a_i (top) and frequency f_i (bottom). The bifurcation parameter correlates with inferred θ_2^r, while frequency correlates with θ_1^r, but only for regions in the oscillatory regime, i.e. where $a_i > 0$. (G) Timeseries generated with the trained model and using the inferred parameters. Important features of the data are preserved both the level of single regions (amplitude, frequency) as well as on the network level (functional connectivity).

synthetic data set, these two parameters are randomly varied across regions. Second model is the parametric mean field model (pMFM Deco et al., 2013), shown on Fig. 3. That is an one-equation model, and depending on the network input, it can be pushed into monostable down- or up-state, or a bistable regime. The switching between the states is noise driven, and we vary the noise strength across brain regions.

Both models are used to generate synthetic data for eight subjects, each with individual structural connectome containing 68 cortical regions of the Desikan-Killiany parcellation (Desikan et al., 2006). The connectome is scaled by the global coupling strength G which we set to increase linearly across subjects for the Hopf model, or which we set to the optimal value (in terms of highest

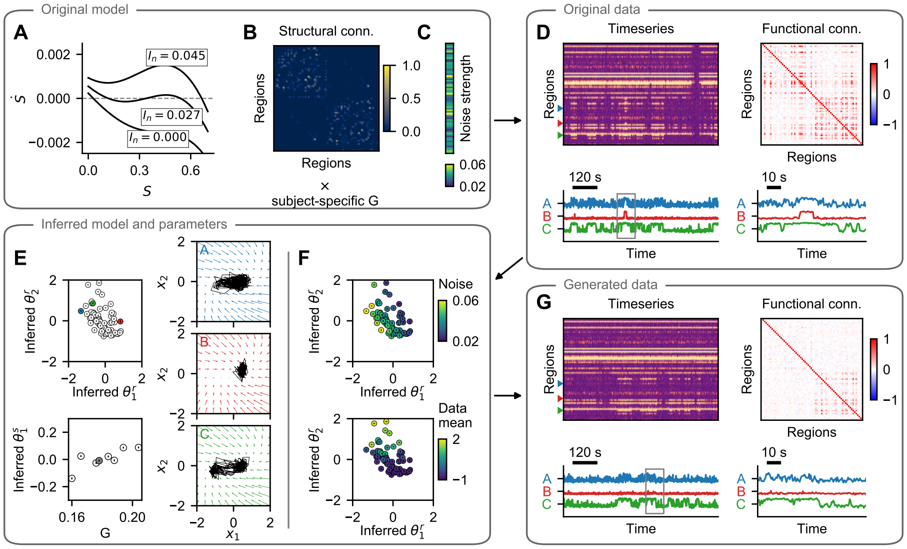

Figure 3: Parametric mean field model test case: example subject. Layout same as in Fig. 2. (A-C) The training data are simulated using a network of pMFM neural masses. Depending on the network input, these can be forced into the monostable regime (down or up state) or into the bistable regime. The dynamics is noise driven, with noise strength varying across regions. (D) Timeseries generated with the original model and the functional connectivity. Three examples shown in the bottom panel, with a window of hundred seconds on the right. (E) Inferred regional parameters (top left) and subject-specific parameter. Circles are added for visual aid due to the small standard deviations. The inferred dynamics in state space of the three examples are on the right. (F) Inferred regional parameters colored by the ground truth values of the noise strength parameter (top); the original parameter is encoded along the diagonal of the inferred parameters. Bottom panel shows coloring according to the mean of the original timeseries, which does not represent an original model parameter, rather a data feature. (G) Timeseries generated with the trained model and using the inferred parameters. Region-specific features (switching between states, noisiness) are well preserved. Structure of the regional correlations is also reproduced, but the correlations are weaker compared to the original.

produced functional connectivity), different for every subject, with pMFM.

To establish the performance of the described method, we proceed as follows. First, we simulate the data with the original model and random values of regional parameters (Fig. 2D and Fig. 3D). Next, using the whole data set of eight subjects, we train the model, obtaining at the same time the trained generative model described by the function f of the dynamical system, and also the probabilistic representation of subject- and region-specific parameters (Fig. 2E and Fig. 3E). Taking random samples from the posterior distributions of the parameters, and using random system and observation noise, we repeatedly generate new timeseries using the trained model (Fig. 2G and Fig. 3G).

We evaluate the quality of the trained model based on the following criteria. First, we establish whether the inferred parameters are related to the original parameters of the model (Figs. 2F, 3F, 4A,E). Second, we wish to evaluate

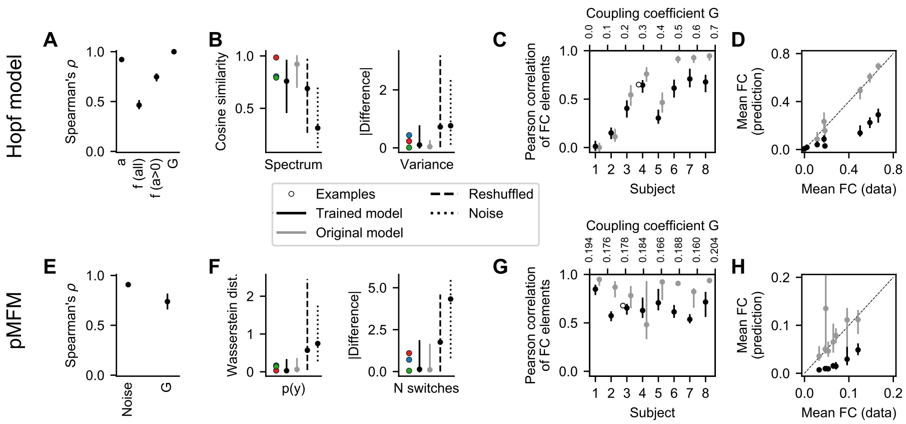

Figure 4: Quantitative evaluation of the synthetic test cases. Top row - Hopf model, bottom row - parametric mean field model. (A, E) Nonlinear correlation between the original parameters and the optimal projection of the inferred region-specific parameters (bifurcation parameter a and frequency f for Hopf model, noise strength for pMFM) and subject-specific parameter (coupling strength G). (B, F) Fit between the regional features of the original timeseries and those generated by the trained model. We show the cosine similarity of the timeseries spectra and the difference in variance for the Hopf model, and Wasserstein distance of the distributions in the observation space and the difference in logarithm of number of switches for pMFM. These are evaluated for the examples from Fig. 2 and Fig. 3, all timeseries generated by the trained model, and the surrogates described in the main text. (C, G) Fit between the functional connectivity of the original and generated timeseries. (D, H) Mean value of non-diagonal elements of functional connectivity matrices. For both models, the correlation strength is underestimated, even if the structure is preserved. In all panels, the bars show the (5, 95) percentile interval with the dot representing the median value. The statistics are computed from 50 samples of the posterior distribution for 8 subjects (grouped together in A, B, E, F) and 68 regions (for region-specific parameters and features). The statistics of the surrogate distributions using the original model are also calculated from 50 samples.

whether the features of the generated timeseries resemble those of the original timeseries, both on the regional level (Fig. 4B,F) and on the network level (Fig. 4C,D,G,H). We explore these aspects in the following paragraphs.

Inferred parameters encode the original model parameters. The example on Fig. 2 shows how are the original regional parameters encoded in the inferred parameters $\boldsymbol{\theta}^r$ for the Hopf model. The bifurcation parameter a is encoded in the inferred parameter θ_1^r (upper panel), while the frequency f is encoded in θ_r^2 (lower panel). The latter is however true only for the regions in the oscillatory regime, i.e. with $a > 0$. That is not a deficiency of the proposed method: in the fixed point regime the activity is mainly noise-driven, and the value of the frequency parameter has small to negligible influence (see the example C on Fig. 2D). In other words, the parameter is not identifiable from the data. That is reflected in the inferred parameters. For the regions with $a > 0$ (or equivalently with $\theta_1^r > 0$) the inferred parameters θ_2^r have low variance, and their mean encodes the original frequency parameter. For the regions with $a < 0$, however, inferred θ_r^2 have high variance, close to the prior value 1, and overlapping distributions, indicating that not much information is encoded in θ_r^2 in this regime.

Also for the pMFM test case the noise strength parameter is well identified (Fig. 3F), however the second dimension of the region-specific parameter θ_2^r is used to encode the mean of the regional timeseries. Presumably, this is so that the parameter θ_2^r can compensate for the weaker network coupling, which we discuss later. For both examples the subject-specific coupling strength is encoded in the subject parameter $\boldsymbol{\theta}^s$ (Figs. 2E, 3E, lower panels).

The quantitative analysis of the goodness-of-fit is shown on Fig. 4A,E. To evaluate it, for each of the original parameters we first identified the direction in the parameter space along which the parameter is encoded by taking a linear regression of the posterior distribution means. Then, we repeatedly took samples from the posterior distributions of the parameters, projected them on the identified subspace, and calculated the nonlinear Spearman's correlation coefficient ρ. For most parameters the values are close to the optimal value of 1, indicating that the original parameters are indeed accurately recovered in the inferred parameters. The exception is the frequency f due to the above discussed non-identifiability. If, however, we restrict the regions only to those where the bifurcation parameter is positive, the correlation markedly increases, as expected based on the discussed example.

On Fig. S1 we further evaluate how the goodness-of-fit changes with the increased coupling in the Hopf model. Presumably, as the coupling increases, the regional timeseries are more affected by the activity of the connected regions and less by its internal parameters, and it is thus more difficult to recover the original parameters from the data. Indeed, that is the trend that we observe both for the bifurcation parameter a and frequency f of the nodes in oscillatory regime.

Trained model reproduces the features of regional timeseries. A crucial test of the trained model is an evaluation whether the generated data resemble those used for the initial training. This resemblance should not be understood as reproducing the timeseries exactly, since they depend on a specific noise instantiation, rather that the features we consider meaningful should be preserved. For both test cases, we evaluate the similarity of two features. For the Hopf model with its oscillatory dynamics we evaluate the cosine similarity of the spectra of the original and generated timeseries, and the difference between the variance of the timeseries, since the variance differs greatly between the nodes in oscillatory and fixed-point regimes (Fig. 4B). For the pMFM, we compare the timeseries based on the distribution in the 1D observation space (that is, taking the samples collapsed across time) using the Wasserstein distance (called also Earth mover's distance) of two distributions. Second feature of pMFM timeseries is the log-scaled number of switches between the up- and down-state, capturing the temporal aspect of the switching dynamics (Fig. 4F).

We evaluate the measures for 50 different noise instantiations, leading to 50 different time series for each region, obtaining a distribution of goodness-of-fit metrics. The same metrics are evaluated also for three surrogates: First is the original computational model, run with different noise instantiations. That provides an optimistic estimate of what can be achieved in terms of goodness-of-fit, considering that the features will necessarily depend on the specific noise instantiation used in the initial simulations. Second surrogate is obtained by randomly reshuffling the original data between regions and subjects. Third surrogate is simply white noise with zero mean and variance equal to one (which, due to the data normalization, is equal to the mean and variance of the original data set taken across all subjects and regions).

In most measures, the trained model performs comparably or slightly worse than the original model and markedly better than the surrogates. The exception is cosine similarity of the spectra with the Hopf model. That is due to the very strong coupling in some subjects (for coupling coefficients $G \geq 0.5$) leading to close to homogeneous activity across brain regions, and thus comparable performance of the reshuffling surrogate.

Functional network structure is reproduced, but with lowered strength. Just as the well trained model should be able to reproduce the features of the original data on the level of single regions, it should also be able to reproduce the relevant features on the network level. Specifically, we evaluate how well is the functional connectivity reproduced. In general, functional connectivity (FC) quantifies the statistical dependencies between the time series from brain regions. While there are multiple ways to measure it, the most ubiquitous is the linear (Pearson's) correlation of the time series, which we use here as well. This static FC captures the spatial structure of statistical similarities, however, it has its limitations, notably it ignores the temporal changes in FC structure (Preti et al., 2017; Lurie et al., 2020).

The examples for both investigated models indicate that the FC structure is indeed well reproduced, but with lower strength, particular in the case of pMFM example (Figs. 2G and 3G).

This is further analyzed for all subjects on Fig. 4, and visualized on Fig. S2 and Fig. S3. For the Hopf model, the coupling coefficient was increased between subjects. For low coupling values, the FC structure is not reproduced (as measured by Pearson correlation between the non-diagonal elements of original FC and trained model FC; Fig. 4C). That is however true also for the original model due to the FC elements being close to zero and noise-dependent. For stronger coupling, the structure is preserved better, although the trained model plateaus around values of 0.7 for the correlation between the FC matrices, even when the correlations between the original model increases further. The comparison of the mean value of non-diagonal FC elements furthermore reveals that the strength of the correlations is considerably underestimated with the trained model (Fig. 4D).

For the pMFM, the coupling coefficient was set to optimal value (in the sense of maximal FC), specific to each subject. Also there we can see well reproduced structure of the correlations (Fig. 4G), although too with reduced strength (Fig. 4H).

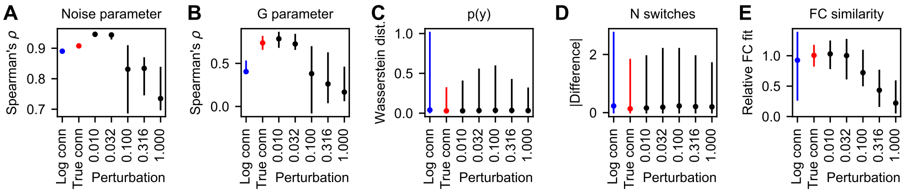

Figure 5: Effect of the connectome perturbation. (A) Spearman's correlation coefficient for the recovery of the noise parameter for the log-scaled connectome (blue), original connectome (red), and five perturbed connectomes (black). (B) Spearman's correlation coefficient for the recovery of the coupling parameter G. (C) Wasserstein distance of the distributions in the observation space of the original data and the data generated by the trained model. (D) Difference of the logarithm of number of switches between down- and up-state of the regional timeseries. (E) Relative FC fit, that is, normalized Pearson's correlation coefficient between the non-diagonal elements of the original FC and the FC generated by the trained model. The normalization is performed by dividing the coefficients by the mean of values obtained for the true connectome for every subject separately. The normalization is done in order to make the values comparable across subjects. For all panels, data were generated using four different connectome perturbations for each magnitude value, and one connectome for the original and log-scaled connectome. In panels A and B, 100 samples were drawn from the parameter distributions for each trained model. In panels C-E 50 simulations were performed to calculate the measures of goodness-of-fit for each model. These were then aggregated across all subjects (and across regions apart from panels B and E). Each line represents the 5 to 95 percentile range, with the dot representing the median.

These results indicate that while the trained model can discover the existence of the network coupling, it systematically underestimates its strength. Given that in the pMFM the strength of the network input can shift a single neural mass from the monostable down-state to bistable regime and to monostable up-state, the underestimated coupling leads to the necessity of utilizing the regional parameter to compensate for the missing coupling (Fig. 3F).

Large perturbations of the connectome lead to reduced performance. To assess the influence of the inexact structural connectome on the goodness-of-fit, we have trained the model on the pMFM data set with perturbed connectomes. That is, instead of the original connectivity matrix W we have trained the model with $W_\epsilon = W + \epsilon A$, where A is matrix with elements drawn from standard normal distribution, and $\epsilon > 0$ is the perturbation magnitude. In addition we have also used a log-scaled connectivity matrix.

Fig. 5 shows how are the indicators of goodness-of-fit from Fig. 4 modified by these perturbed connectomes. High perturbation magnitudes reduces the recovery of regional and subject parameters (Fig. 5A,B) as well as the similarity of the generated functional connectivity (Fig. 5E). The regional features, on the other hand, are reproduced similarly well even for large perturbations (Fig. 5C,D). Using the log-scaled connectome has a similar negative effect, although less pronounced.

4. Discussion

Summary. In this work we have introduced a method for analysis of whole-brain dynamics based on a model of networked dynamical systems. Using the structure of the network and the functional data, the method allows to infer both the unknown generative dynamical system, and the parameters varying across regions and subjects.

We have tested the method on two synthetic data sets, one generated by a virtual brain model composed of nodes with a Hopf model (Fig. 2) and one generated by virtual brains with a parametric mean field model (Fig. 3). Detailed analysis of the results has shown that the proposed method can recover the original parameters as well as reproduce the important features of the original data both the single region level and on the network level (Fig. 4). With these results in hand, the planned next step is evaluation of the method on human resting-state fMRI data remains, which is yet to be performed.

Importance of dynamically-relevant parameters. Large-scale brain dynamics during resting-state is altered in neurodegenerative diseases (Hohenfeld et al., 2018) and in normal aging (Ferreira and Busatto, 2013). Myriads of regionally varying parameters that can plausibly influence the large-scale dynamics can be measured either in vivo or post mortem, such as cell density, cell type composition, local connectivity structure, connectivity to subcortical structures, or receptor densities, to name just a few. But which ones are in fact relevant for large-scale brain dynamics, and how do they influence it? Construction of bottom-up mechanistic models that would include all possible parameters and allow to investigate their role is unfeasible due to the complexity of human brain with its dynamics spanning multiple temporal and spatial scales, even if the parameters were in fact accurately measured (Frégnac, 2017).

Our approach instead pursues this understanding from the opposite direction. We use the amortized inference framework to learn the dynamical system driving the dynamics, and with it also the parameters varying across regions and subjects. Since these parameters are inferred from the functional data in unsupervised fashion, they are by construction the parameters relevant for the large-scale dynamics. Given the abstract nature of the inferred model, the mechanistic meaning of these dynamically-relevant parameters is not self-evident, yet they still provide a measure of similarity of brain regions and different subjects and their effect on the dynamics can be investigated through the trained model. Furthermore, given large enough data set, the dynamically-relevant parameters may be linked to the measured quantities (or their combinations). Such link may provide insights into the origin of neurodegenerative diseases if the dynamically-relevant parameters differ between the disease stages.

Importantly, the link between dynamically-relevant parameters and the measurable quantities can be estimated from a preexisting patient cohort, and then only applied to single subject. That is advantageous if the measurement is difficult, costly, or impossible to perform in clinical setting (such as for cell type composition estimated from post mortem studies); in such cases, the dynamically-relevant parameters may instead be estimated from easy-to-obtain resting-state fMRI and then mapped using the known link.

Learning complex dynamics. We have tested the method on synthetic data generated using the Hopf model and parametric mean field model as neural masses embedded in a whole-brain network. Admittedly, these two models, while often used in whole-brain modeling, are dynamically quite simple - after all, they are represented by one or two differential equations per node. And even for these models some shortcomings of the method are noticeable, in particular the insufficiently captured network interactions leading to weakened functional connectivity in the generated data.

One can ask whether the method would be able to handle more complex dynamics, generated by higher dimensional models, with many coexisting fixed points and limit cycles, and possibly acting on multiple time scales. In principle, the present method can be applied to more complex data, and, if the state space is set to be sufficiently large and the hidden layer in function f sufficiently wide, arbitrarily complex dynamics can be represented by the architecture. Whether such system can be successfully discovered through the optimization process is however a different question, one that we are not able to answer here, since designing neural network architectures for novel tasks is notoriously difficult problem without robust theoretical guidelines.

Considerable amount of other architectures were explored in related works, and although they were not applied in a networked setting, their elements could be incorporated in our framework to improve its performance for more complex dynamics. For instance, Duncker et al. (2019) relied on Gaussian processes conditioned on set of fixed points to learn the system dynamics, and demonstrated its efficacy on multistable dynamical systems. Nassar et al. (2019) used a tree structure to partition the state space and approximate the system in each partition with linear dynamics. Koppe et al. (2019) used piecewise linear recurrent neural network to analyze fMRI data. Schmidt et al. (2021) later expanded on this work introducing an approach for better approximation of systems with multiple time scales through creation of slow manifolds in the state space using a regularization scheme of the dynamical system.

Imperfect connectome. Our method assumes that the structural connectome through which the local dynamics is coupled is known. What we can obtain, however, is only an estimate from diffusion tractography, suffering from a range of biases (Rheault et al., 2020; Grisot et al., 2021). Our results indicate that while the method can handle small perturbations of the connectome, larger perturbations or different scaling can considerably degrade its performance (Fig. 5). To some extent this might be overcome by running the method with several connectomes using different scalings (linear, logarithmic) or different corrections for known biases and choosing the optimal connectome via model comparison methods.

If that would not produce results of sufficient quality, alternate approach can be pursued, one that would use the estimated structural connectome not as hard data but only as a soft prior for the effective connectivity of the model. Such approach was described for whole-brain dynamics generated by the multivariate Ornstein-Uhlenbeck process, using the thresholded structural connectivity as a topological mask for the inferred effective connectivity (Gilson et al., 2019, 2020). The model connectivity may be inferred even without any prior anatomical constraints, as demonstrated by the MINDy method that relies on a simple one-equation neural mass model (Singh et al., 2020).

Acknowledgements

This work was supported by a public grant overseen by the French National Research Agency (ANR) as part of the second “Investissements d’Avenir” program (reference: ANR-17-RHUS-0004) and by the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 945539 (SGA3) Human Brain Project. Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. The authors wish to thank Jayant Jha and Anirudh N. Vattikonda for helpful discussions.

Competing interests

The authors declare no competing interests.

References

Breakspear, M., Mar. 2017. Dynamic models of large-scale brain activity. Nature Neuroscience 20 (3), 340–352.

Chollet, F., et al., 2015. Keras. https://github.com/fchollet/keras.

Deco, G., Kringelbach, M. L., Arnatkeviciute, A., Oldham, S., Sabaroedin, K., Rogasch, N. C., Aquino, K. M., Fornito, A., Jul. 2021. Dynamical consequences of regional heterogeneity in the brain’s transcriptional landscape. Science Advances 7 (29), eabf4752.

Deco, G., Kringelbach, M. L., Jirsa, V. K., Ritter, P., Dec. 2017. The dynamics of resting fluctuations in the brain: Metastability and its dynamical cortical core. Scientific Reports 7 (1), 3095.

Deco, G., Ponce-Alvarez, A., Mantini, D., Romani, G. L., Hagmann, P., Corbetta, M., Jul. 2013. Resting-State Functional Connectivity Emerges from Structurally and Dynamically Shaped Slow Linear Fluctuations. Journal of Neuroscience 33 (27), 11239–11252.

Demirtaş, M., Burt, J. B., Helmer, M., Ji, J. L., Adkinson, B. D., Glasser, M. F., Van Essen, D. C., Sotiropoulos, S. N., Anticevic, A., Murray, J. D., Mar. 2019. Hierarchical Heterogeneity across Human Cortex Shapes Large-Scale Neural Dynamics. Neuron 101 (6), 1181–1194.e13.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., Killiany, R. J., Jul. 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31 (3), 968–980.

Dhollander, T., Raffelt, D., Connelly, A., 2016. Unsupervised 3-tissue response function estimation from single-shell or multi-shell diffusion MR data without a co-registered T1 image. In: ISMRM Workshop on Breaking the Barriers of Diffusion MRI.

Duncker, L., Bohner, G., Boussard, J., Sahani, M., May 2019. Learning interpretable continuous-time models of latent stochastic dynamical systems. In: International Conference on Machine Learning. PMLR, pp. 1726–1734.

Ferreira, L. K., Busatto, G. F., Mar. 2013. Resting-state functional connectivity in normal brain aging. Neuroscience & Biobehavioral Reviews 37 (3), 384–400.

Frégnac, Y., Oct. 2017. Big data and the industrialization of neuroscience: A safe roadmap for understanding the brain? Science 358 (6362), 470–477.

Gilson, M., Kouvaris, N. E., Deco, G., Mangin, J.-F., Poupon, C., Lefranc, S., Rivière, D., Zamora-López, G., Nov. 2019. Network analysis of whole-brain fMRI dynamics: A new framework based on dynamic communicability. NeuroImage 201, 116007.

Gilson, M., Zamora-López, G., Pallarés, V., Adhikari, M. H., Senden, M., Campo, A. T., Mantini, D., Corbetta, M., Deco, G., Insabato, A., Apr. 2020. Model-based whole-brain effective connectivity to study distributed cognition in health and disease. Network Neuroscience 4 (2), 338–373.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., Jenkinson, M., Oct. 2013. The minimal preprocessing pipelines for the Human Connectome Project. NeuroImage 80, 105–124.

Grisot, G., Haber, S. N., Yendiki, A., Oct. 2021. Diffusion MRI and anatomic tracing in the same brain reveal common failure modes of tractography. NeuroImage 239, 118300.

Hohenfeld, C., Werner, C. J., Reetz, K., Jan. 2018. Resting-state connectivity in neurodegenerative disorders: Is there potential for an imaging biomarker? NeuroImage: Clinical 18, 849–870.

Jeurissen, B., Tournier, J.-D., Dhollander, T., Connelly, A., Sijbers, J., Dec. 2014. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. NeuroImage 103, 411–426.

Kingma, D. P., Ba, J., Jan. 2017. Adam: A Method for Stochastic Optimization. arXiv:1412.6980 [cs].

Kingma, D. P., Welling, M., 2019. An Introduction to Variational Autoencoders. Foundations and Trends® in Machine Learning 12 (4), 307–392.

Kong, X., Kong, R., Orban, C., Peng, W., Zhang, S., Anderson, K., Holmes, A., Murray, J. D., Deco, G., van den Heuvel, M., Yeo, B. T., Mar. 2021. Anatomical and Functional Gradients Shape Dynamic Functional Connectivity in the Human Brain. Preprint, Neuroscience.

Koppe, G., Toutounji, H., Kirsch, P., Lis, S., Durstewitz, D., Aug. 2019. Identifying nonlinear dynamical systems via generative recurrent neural networks with applications to fMRI. PLOS Computational Biology 15 (8), e1007263.

Linderman, S., Johnson, M., Miller, A., Adams, R., Blei, D., Paninski, L., Apr. 2017. Bayesian Learning and Inference in Recurrent Switching Linear Dynamical Systems. In: Artificial Intelligence and Statistics. PMLR, pp. 914–922.

Lurie, D. J., Kessler, D., Bassett, D. S., Betzel, R. F., Breakspear, M., Kheilholz, S., Kucyi, A., Liégeois, R., Lindquist, M. A., McIntosh, A. R., Poldrack, R. A., Shine, J. M., Thompson, W. H., Bielczyk, N. Z., Douw, L., Kraft, D., Miller, R. L., Muthuraman, M., Pasquini, L., Razi, A., Vidaurre, D., Xie, H., Calhoun, V. D., Feb. 2020. Questions and controversies in the study of time-varying functional connectivity in resting fMRI. Network Neuroscience 4 (1), 30–69.

Nassar, J., Linderman, S. W., Bugallo, M., Park, I. M., Jun. 2019. Tree-Structured Recurrent Switching Linear Dynamical Systems for Multi-Scale Modeling. arXiv:1811.12386 [cs, stat].

Pandarinath, C., O'Shea, D. J., Collins, J., Jozefowicz, R., Stavisky, S. D., Kao, J. C., Trautmann, E. M., Kaufman, M. T., Ryu, S. I., Hochberg, L. R., Henderson, J. M., Shenoy, K. V., Abbott, L. F., Sussillo, D., Oct. 2018. Inferring single-trial neural population dynamics using sequential auto-encoders. Nature Methods 15 (10), 805–815.

Preti, M. G., Bolton, T. A., Van De Ville, D., Oct. 2017. The dynamic functional connectome: State-of-the-art and perspectives. NeuroImage 160, 41–54.

Rheault, F., Poulin, P., Caron, A. V., St-Onge, E., Descoteaux, M., Feb. 2020. Common misconceptions, hidden biases and modern challenges of dMRI tractography. Journal of Neural Engineering 17 (1), 011001.

Roeder, G., Grant, P. K., Phillips, A., Dalchau, N., Meeds, E., Oct. 2019. Efficient Amortised Bayesian Inference for Hierarchical and Nonlinear Dynamical Systems. arXiv:1905.12090 [cs, stat].

Schmidt, D., Koppe, G., Monfared, Z., Beutelspacher, M., Durstewitz, D., Mar. 2021. Identifying nonlinear dynamical systems with multiple time scales and long-range dependencies. arXiv:1910.03471 [cs, q-bio, stat].

Singh, M. F., Braver, T. S., Cole, M. W., Ching, S., Nov. 2020. Estimation and validation of individualized dynamic brain models with resting state fMRI. NeuroImage 221, 117046.

Smith, R. E., Tournier, J.-D., Calamante, F., Connelly, A., Feb. 2013. SIFT: Spherical-deconvolution informed filtering of tractograms. NeuroImage 67, 298–312.

Suárez, L. E., Markello, R. D., Betzel, R. F., Misic, B., Apr. 2020. Linking Structure and Function in Macroscale Brain Networks. Trends in Cognitive Sciences 24 (4), 302–315.

Tournier, J.-D., Calamante, F., Connelly, A., 2010. Improved probabilistic streamlines tractography by 2nd order integration over fibre orientation distributions. In: Proceedings of the International Society for Magnetic Resonance in Medicine. Vol. 18. p. 1670.

Tournier, J.-D., Calamante, F., Connelly, A., Feb. 2012. MRtrix: Diffusion tractography in crossing fiber regions. International Journal of Imaging Systems and Technology 22 (1), 53–66.

Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T. E. J., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S. W., Della Penna, S., Feinberg, D., Glasser, M. F., Harel, N., Heath, A. C., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S. E., Prior, F., Schlaggar, B. L., Smith, S. M., Snyder, A. Z., Xu, J., Yacoub, E., Oct. 2012. The Human Connectome Project: A data acquisition perspective. NeuroImage 62 (4), 2222–2231.

Wang, P., Kong, R., Kong, X., Liégeois, R., Orban, C., Deco, G., van den Heuvel, M. P., Thomas Yeo, B., Jan. 2019. Inversion of a large-scale circuit model reveals a cortical hierarchy in the dynamic resting human brain. Science Advances 5 (1), eaat7854.

Appendix A. Supplementary information

Evidence lower bound (ELBO)

For a single subject, the observations contain the time series from all n regions, $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$. They are complemented by the region time series for the network input, $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)$, and the one-hot vector $\boldsymbol{c}$ encoding the subject identity. The latent variables $\boldsymbol{Z}$ contain the state time series $\boldsymbol{x}_j$ for all regions j, region-specific parameters $\boldsymbol{\theta}^r_j$, and subject specific parameters $\boldsymbol{\theta}^s$, that is, $\boldsymbol{Z} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \boldsymbol{\theta}^r_1, \ldots, \boldsymbol{\theta}^r_n, \boldsymbol{\theta}^s)$. Our goal is to minimize the Kullback-Leibler divergence between the approximate and true posterior, which can be rewritten as a sum of subject ELBO and evidence itself,

$$\begin{aligned}&\mathrm{KL}(q(\boldsymbol{Z}|\boldsymbol{Y},\boldsymbol{U},\boldsymbol{c}) \mid\mid p(\boldsymbol{Z}|\boldsymbol{Y},\boldsymbol{U},\boldsymbol{c})) = \mathbb{E}_q[\log q(\boldsymbol{Z}|\boldsymbol{Y},\boldsymbol{U},\boldsymbol{c})] - \mathbb{E}_q[\log p(\boldsymbol{Z}|\boldsymbol{Y},\boldsymbol{U},\boldsymbol{c})] \\ &= \underbrace{\mathbb{E}_q[\log q(\boldsymbol{Z}|\boldsymbol{Y},\boldsymbol{U},\boldsymbol{c})] - \mathbb{E}_q[\log p(\boldsymbol{Y}|\boldsymbol{Z},\boldsymbol{U},\boldsymbol{c})] - \mathbb{E}_q[\log p(\boldsymbol{Z}|\boldsymbol{U},\boldsymbol{c})]}_{-L_{\text{subject}}} + \mathbb{E}_q[\log p(\boldsymbol{Y}|\boldsymbol{U},\boldsymbol{c})].\end{aligned}$$

Maximizing the ELBO then minimizes the KL divergence. We can factorize all terms of the ELBO across n brain regions: the approximate posterior,

$$q(\boldsymbol{Z}|\boldsymbol{Y},\boldsymbol{U},\boldsymbol{c}) = \prod_{j=1}^{n} q(\boldsymbol{x}_j|\boldsymbol{y}_j,\boldsymbol{u}_j,\boldsymbol{c}) \prod_{j=1}^{n} q(\boldsymbol{\theta}^r_j|\boldsymbol{y}_j,\boldsymbol{u}_j,\boldsymbol{c}) q(\boldsymbol{\theta}^s|\boldsymbol{c}),$$

the data likelihood,

$$p(\boldsymbol{Y}|\boldsymbol{Z},\boldsymbol{U},\boldsymbol{c}) = \prod_{j=1}^{n} p(\boldsymbol{y}_j|\boldsymbol{x}_j,\boldsymbol{\theta}^r_j,\boldsymbol{\theta}^s,\boldsymbol{u}_j,\boldsymbol{c}),$$

and the prior,

$$\begin{aligned}p(\boldsymbol{Z}|\boldsymbol{U},\boldsymbol{c}) &= p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n|\boldsymbol{\theta}^r_1,\ldots,\boldsymbol{\theta}^r_n,\boldsymbol{\theta}^s,\boldsymbol{U},\boldsymbol{c})\ p(\boldsymbol{\theta}^r_1,\ldots,\boldsymbol{\theta}^r_n,\boldsymbol{\theta}^s|\boldsymbol{U},\boldsymbol{c}) \\ &= \prod_{j=1}^{n} p(\boldsymbol{x}_j|\boldsymbol{\theta}^r_j,\boldsymbol{\theta}^s,\boldsymbol{u}_j,\boldsymbol{c}) \prod_{j=1}^{n} p(\boldsymbol{\theta}^r_j|\boldsymbol{u}_j,\boldsymbol{c}) p(\boldsymbol{\theta}^s|\boldsymbol{c}).\end{aligned}$$

We require that the data likelihood and priors depend on the subject identity only through the latent variables, so we remove the dependence on $\boldsymbol{c}$. We also require that the priors of $\boldsymbol{\theta}^r_j$ do not depend on the external input $\boldsymbol{u}_j$. Then we define the region ELBO as

$$\begin{aligned}L_j = \;&\mathbb{E}_q[\log p(\boldsymbol{y}_j|\boldsymbol{x}_j,\boldsymbol{\theta}^r_j,\boldsymbol{\theta}^s,\boldsymbol{u}_j)] \\ &+ \mathbb{E}_q[\log p(\boldsymbol{x}_j|\boldsymbol{\theta}^r_j,\boldsymbol{\theta}^s,\boldsymbol{u}_j)] + \mathbb{E}_q[\log p(\boldsymbol{\theta}^r_j)] + \frac{1}{n}\mathbb{E}_q[\log p(\boldsymbol{\theta}^s)] \\ &- \mathbb{E}_q[\log q(\boldsymbol{x}_j|\boldsymbol{y}_j,\boldsymbol{u}_j,\boldsymbol{c})] - \mathbb{E}_q[\log q(\boldsymbol{\theta}^r_j|\boldsymbol{y}_j,\boldsymbol{u}_j,\boldsymbol{c})] - \frac{1}{n}\mathbb{E}_q[\log q(\boldsymbol{\theta}^s|\boldsymbol{c})]\end{aligned}$$

so that the subject ELBO is the sum of region ELBOs, $L_{\text{subject}} = \sum_{j=1}^{n} L_j$.

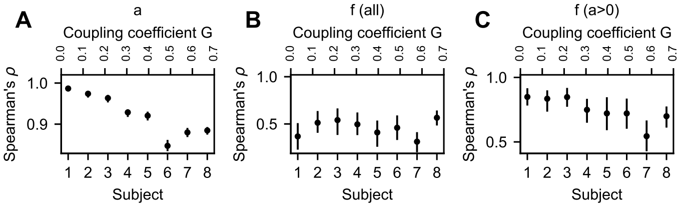

Figure S1: Recovery of the region-specific parameters in the Hopf model for different subjects with different coupling coefficient G. The figure contains the data from Fig. 2A, separated for the individual subjects.

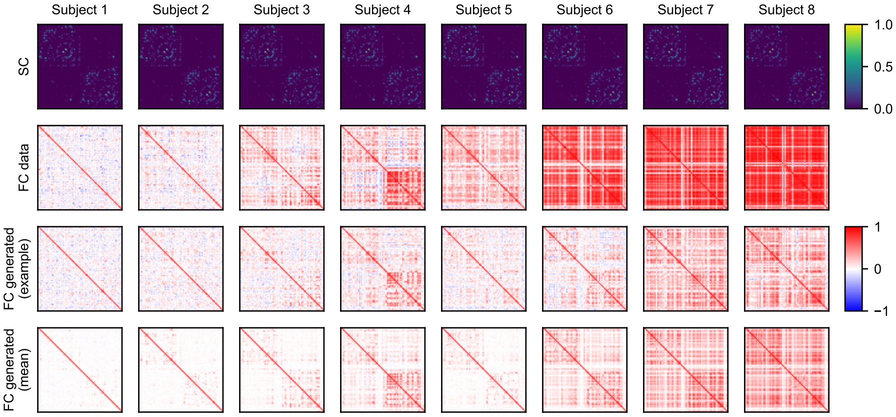

Figure S2: Structural and functional connectivity matrices for all subjects in the Hopf model test case. First row: structural connectivity. Second row: Functional connectivity of the original data used for the training. Third row: Functional connectivity of the example data generated with the trained model. Fourth row: Functional connectivity of the data generated with the trained model, averaged over 50 samples.

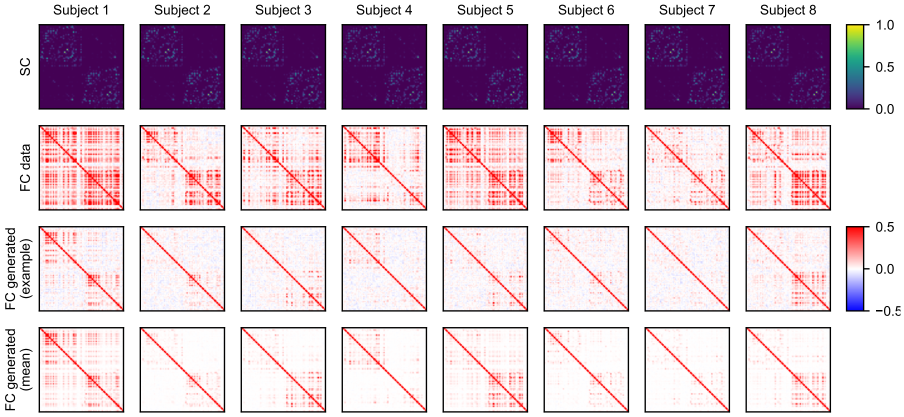

Figure S3: Structural and functional connectivity matrices for all subjects in the pMFM test case. Layout the same as in Fig. S2