Identification of modulated whole-brain dynamical models from nonstationary electrophysiological data

Journal of Neural Engineering

PAPER

OPEN ACCESS

RECEIVED
30 April 2025

REVISED
24 August 2025

ACCEPTED FOR PUBLICATION
29 September 2025

PUBLISHED
10 October 2025

Identification of modulated whole-brain dynamical models from nonstationary electrophysiological data

Addison Schwamb*, Zongxi Yu and ShiNung Ching

Washington University in St. Louis, St. Louis, MO, United States of America
* Author to whom any correspondence should be addressed.

E-mail: a.l.schwamb@wustl.edu, y.zongxi@wustl.edu and shinung@wustl.edu

Keywords: neural modeling, data-driven modeling, dynamical systems, modulation

Abstract

Objective. Understanding the mechanisms underlying brain dynamics is a long-held goal in neuroscience. However, these dynamics are both individualized and nonstationary, making modeling challenging. Here, we present a data-driven approach to modeling nonstationary dynamics based on principles of neuromodulation, at the level of individual subjects. *Approach.* Previously, we developed the mesoscale individualized neural dynamics (MINDy) modeling approach to capture individualized brain dynamics which do not change over time. Here, we extend the MINDy approach by adding a modulatory component which is multiplied by a set of baseline, stationary connectivity weights. We validate this model on both synthetic data and publicly available electroencephalography data in the context of anesthesia, a known modulator of neural dynamics. *Main results.* We find that our modulated MINDy approach is accurate, individualized, and reliable. Additionally, we find that our models yield biologically interpretable inferences regarding the effects of propofol anesthesia on mesoscale cortical networks, consistent with previous literature on the neuromodulatory effects of propofol. *Significance.* Ultimately, our data-driven modeling approach is reliable and scalable, and provides insight into mechanisms underlying observed brain dynamics. Our modeling methodology can be used to infer insights about modulation dynamics in the brain in a number of different contexts.

1. Introduction

An important and persistent challenge in the analysis of recorded mesoscale neural activity (i.e. commensurate with externally recorded fields and potentials) is the inference of latent neurophysiological mechanisms that underlie overt observations. Indeed, while there are myriad tools and methods to analyze brain electrophysiological recordings (e.g. power spectral density estimation [1, 2]), there are fewer that provide direct inference of circuit-level mechanisms (i.e. the interaction of excitatory and inhibitory neural subpopulations, from which said recordings originate). Identifying and understanding these mechanisms is a difficult task, however, as they are not directly observable via typical mesoscale recording modalities. For instance, electroencephalography (EEG) measures electrical potential non-invasively at the scalp, and hence does not provide direct access to neuronal subpopulation-level activity [3]. Parametric dynamical systems modeling offers a methodological path to obviating this issue. Such models, via their mathematical formulation, embed mechanistic hypotheses or inductive biases regarding how neural activity and secondary observables such as EEG are generated. While dynamical systems modeling is potentially powerful in this context, there are several extant challenges that remain unsolved regarding the construction of such models from neural recordings.

First, neural activity patterns vary between individuals, indicating that the underlying mechanisms also vary on an individual level. For example, the posterior dominant rhythm is a classic example of neural dynamics, which manifests as a strong alpha-band (i.e. 10–14 Hz) oscillation localized to the posterior of the scalp in EEG recordings (i.e. overlaying occipital cortex). While the general characteristics of the posterior dominant rhythm (its frequency and spatial location) are consistent across individuals [4], the exact frequency and power of the posterior

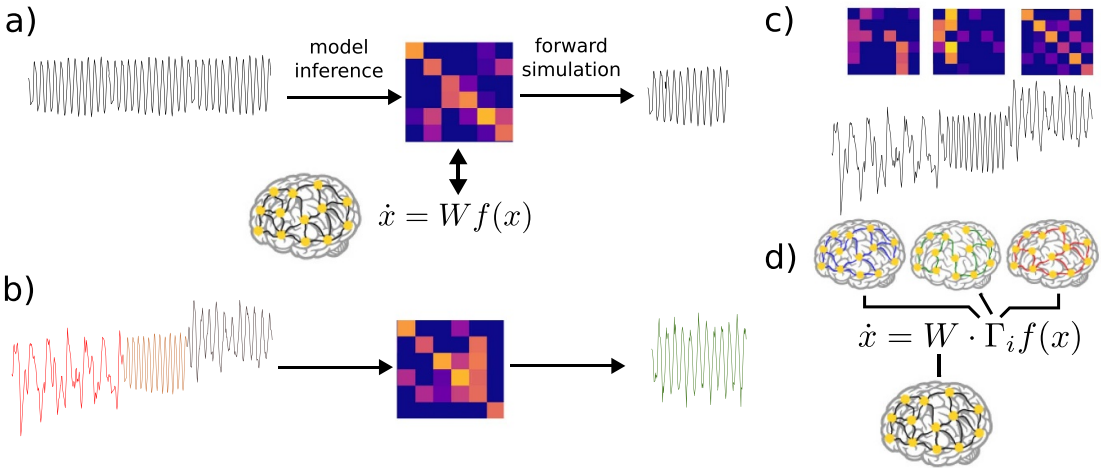

Figure 1. Approaches to fitting individualized models from data. (a) In a stationary setting, a single set of (time-invariant) model parameters (here, a connectivity/weight matrix) is fit to data, yielding a corresponding stationary prediction/forward simulation. (b) When data is non-stationary, fitting a single stationary model will in general lead to erroneous or averaged modeled dynamics. (c) A typical way to address non-stationary data is to fit separate models (e.g. distinct weight matrices) to the distinct non-stationary regimes. (d) In the proposed approach, we seek to fit non-stationary models that decompose parameters into a baseline connectivity and regime-specific modulatory matrices.

dominant rhythm is specific to individuals [5]. To address this challenge of individuality, work has been done to create data-driven models based on single-subject data [6, 7], providing individualized neural dynamics models which can be used to infer mechanisms underlying a person's brain activity [8–10]. This approach is schematized in figure 1(a).

In addition to varying between individuals, neural dynamics are also highly non-stationary, i.e. they vary temporally based on many factors such as modulation of neurophysiologic states (e.g. sleep vs. wake [11], rest vs. task [12, 13], and healthy activity vs. pathology [14]). Therefore, an individual dynamical systems model of the kind mentioned above can typically only offer insights for a specific physiological regime, and/or for a relatively narrow epoch of time during which dynamics are stationary [15]. There is an unmet need for data-driven modeling methods that can capture non-stationarity in dynamics associated with multiple neurophysiologic states. If an approach is taken that does not account for this non-stationarity, the model which is returned will be a poor fit for any of the regimes present in the data (figure 1(b)).

To overcome this challenge, several approaches to time- or state-dependent dynamical systems models for neural activity have been suggested. In essence, these models embed a mechanism to modify the model parameters in a manner that captures categorical changes in neural dynamics (e.g. sleep vs. wake). For instance, the switching linear dynamical system framework [16, 17] embeds multiple linear dynamical systems, of which a single model is 'active' at any given time [18–20], as shown in figure 1(c). Switches between models are often enacted through a latent model, such as a hidden Markov model (HMM) [21] that provides the dynamics behind switches [22, 23]. Through such a mechanism, a model can embed a number of distinct dynamical regimes. A similar approach is taken in [24, 25], but here the researchers use a recurrent neural network (RNN) model, rather than a linear dynamical system. They construct a discrete number of RNNs (each with a different recurrent weight matrix), thus implementing a different set of dynamics for each RNN.

It is important to note that modeling non-stationarity can be understood as a problem with two phases: (i) modeling when dynamical regimes change and (ii) modeling what in the latent dynamics has changed. Our proposed approach tackles the latter phase. Specifically, we infer the changes in latent dynamics as changes in mesoscale *neuromodulation* (figure 1(d)), rather than comparing distinct dynamic regimes (figure 1(c)). In this paper, we take a similar approach to [24, 25], beginning with a biologically interpretable RNN model describing population-level neural activity. However, in contrast to the approach of having a discrete set of recurrent weight matrices, we implement a modulation architecture (figure 1(d)). To our knowledge, data-driven inference of neuromodulatory effects on mesoscale dynamics has not been previously pursued in this way.

Our model consists of a single weight matrix common to all dynamical regimes, which is then multiplied element-wise by one of a discrete number of modulatory matrices corresponding to different dynamical regimes. Such an assumption is based, schematically, on the actions of neuromodulators on neural circuits that modulate synaptic efficacy. We assume that such modulation occurs on a slower

timescale than the timescale of neural activity. By using this modulation structure, we are able to separate the aspects of an individual's dynamics which are common across time and neurophysiologic state from those which are variable. Additionally, we impose specific excitatory and inhibitory sub-structures on our parameters, in order to preserve the biological interpretability of our fitted models. All model parameters are estimated directly from data in an individualized manner.

Below, we develop and present the proposed methodology. We first introduce our modulated RNN architecture, building from our prior data-driven modeling approach of mesoscale individualized neural dynamics (MINDy) [9, 10] to enact the aforementioned modulation architecture. We then validate this on simulated data, to test the accuracy and reliability of our model in recovering known ground truth parameters. We then test our model on the ability to infer distinct neural mechanisms associated with levels of general anesthesia, a pharmacological modulation of neural circuits. These data allow us to further validate the reliability and individuality of our fitted models. We then reconcile the inferred models and mechanisms with prior observations regarding the modulatory effects of anesthesia on cortical networks.

2. Methods

2.1. Model formation

Our model is adapted from our prior whole-brain dynamical systems framework in [9, 10]. The base (i.e. unmodulated) dynamics are governed by:

$$\boldsymbol{x}_{t+1} - \boldsymbol{x}_t = \boldsymbol{W}\tanh(\boldsymbol{S}\boldsymbol{x} + \boldsymbol{V}) + \boldsymbol{D}\boldsymbol{x} + \boldsymbol{C} + \epsilon_t \quad (1)$$

$$\boldsymbol{y}_t = \boldsymbol{H}\boldsymbol{x}_t + \nu_t, \quad (2)$$

where $\boldsymbol{x}_t \in \mathbb{R}^n$ represents the neural activity of n neural populations at time t, and $\{\boldsymbol{W}, \boldsymbol{S}, \boldsymbol{V}, \boldsymbol{D}, \boldsymbol{C}\}$ are the tunable model parameters. Here, $\boldsymbol{W}$ is the connectivity matrix, $\boldsymbol{D}$ is a diagonal matrix representing the decay (leak) in neural activity for each population, $\boldsymbol{S}$ and $\boldsymbol{V}$ parameterize the slope and offset of the sigmoidal activation function, and $\boldsymbol{C}$ represents a baseline bias of each neural population. $\boldsymbol{H}$ is the lead field model which translates the population-level neural activity into the measured data, and ϵ_t and ν_t represent the process noise and measurement noise, respectively.

Note that this model is mathematically comparable to a vanilla RNN, but arranged and constrained to reflect biologically interpretable relationships between excitatory and inhibitory neural populations. By assigning neural populations to spatial locations in the brain and constraining the parameters to specific valence (i.e. positive/excitatory, negative/inhibitory), we can examine changes in these parameters associated with physiological changes. Specifically, we enforce excitatory and inhibitory substructures onto the neural population activity vector $\boldsymbol{x}_t$ and the connectivity matrix $\boldsymbol{W}$ like so:

$$\boldsymbol{x}_t = \begin{bmatrix} \boldsymbol{x}_{t,\mathrm{exc}} \\ \boldsymbol{x}_{t,\mathrm{inh}} \end{bmatrix}, \quad \boldsymbol{W} = \begin{bmatrix} \boldsymbol{W}_{\mathrm{ee}} & \boldsymbol{W}_{\mathrm{ie}} \\ \boldsymbol{W}_{\mathrm{ei}} & \boldsymbol{W}_{\mathrm{ii}} \end{bmatrix}. \quad (3)$$

Entries in $\boldsymbol{W}_{\mathrm{ee}}$ and $\boldsymbol{W}_{\mathrm{ei}}$ are constrained to be positive or 0, as they represent the connections from excitatory neural populations. Additionally, $\boldsymbol{W}_{\mathrm{ee}}$ and $\boldsymbol{W}_{\mathrm{ei}}$ are full submatrices. Conversely, $\boldsymbol{W}_{\mathrm{ie}}$ and $\boldsymbol{W}_{\mathrm{ii}}$ have entries which are negative or 0, as they represent the connections from inhibitory neural populations. Since inhibitory neurons only have local connections [26], these submatrices are constrained to be diagonal.

We adapted and generalized this model by adding neuromodulation via matrices $\boldsymbol{\Gamma}_i$, $i = 1, \ldots, m$ multiplied element-wise by the connectivity matrix $\boldsymbol{W}$:

$$\boldsymbol{x}_{t+1} - \boldsymbol{x}_t = (\boldsymbol{W} \odot \boldsymbol{\Gamma}_i)\tanh(\boldsymbol{S}\boldsymbol{x} + \boldsymbol{V}) + \boldsymbol{D}\boldsymbol{x} + \boldsymbol{C} + \epsilon. \quad (4)$$

Here, m represents the number of modulatory states to be modeled. This modulation structure is schematized in figure 1(d). To preserve the signed connectivity structure of $\boldsymbol{W}$, entries in each $\boldsymbol{\Gamma}_i$ matrix are constrained to be positive or 0. This enables a modulation which scales entries in $\boldsymbol{W}$ by varying amounts, without affecting the base structure of the connectivity.

The specification of $\boldsymbol{H}$ is made to reflect the specifics of the data modality being used for model construction. In our case, because we are using cortical data (EEG), we account for prevailing hypotheses regarding the contributions of different cell types to surface-level potentials. In this context, it is generally believed that inhibitory neurons are not close enough to the surface of the cortex, nor do they possess the spatial organization, to generate fields detectable via EEG [27]. Therefore, we construct our lead field matrix $\boldsymbol{H}$ to have zeros in the submatrix multiplied by the inhibitory component of $\boldsymbol{x}$:

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{H}_{\mathrm{exc}} & \boldsymbol{0} \end{bmatrix}. \quad (5)$$

2.2. Model fitting procedure

Because $\boldsymbol{H}$ is not invertible, we now face what is sometimes termed a dual inference problem: i) estimate $\boldsymbol{x}_t$ and ii) estimate the parameters $\{\boldsymbol{W}, \boldsymbol{\Gamma}, \boldsymbol{D}, \boldsymbol{S}, \boldsymbol{V}, \boldsymbol{C}\}$, from the observable (i.e. EEG) data.

To address this, we adopt the iterative estimation approach detailed in [10, 28]: first we apply a Kalman filter [29, 30] to a small window of data in order to estimate the state. We then evolve the model forward from the (estimated) state to generate a forward prediction. We then backpropagate the error through

Table 1. Parameter values or random distributions used to create synthetic data models. **I** denotes the identity matrix, and **1** denotes the vector of all ones.

Parameter	Description	Initialization value
$\boldsymbol{W}_s$	Sparse part of connectivity matrix	$(16/20)\mathcal{U}(0,1)^3$
$\boldsymbol{W}_{l1}, \boldsymbol{W}_{l2}$	Low rank part of connectivity matrix	$\mathcal{U}(0,1)^3 + 0.2\mathcal{U}(0,1)$
$\boldsymbol{S}_{exc}$	Excitatory nonlinearity slope	2.5
$\boldsymbol{S}_{inh}$	Inhibitory nonlinearity slope	1
$\boldsymbol{V}$	Nonlinearity offset	0
$\boldsymbol{D}_{exc}$	Excitatory decay	$0.65 + 0.02\mathcal{U}(0,1)$
$\boldsymbol{D}_{inh}$	Inhibitory decay	$0.8 + 0.02\mathcal{U}(0,1)$
$\boldsymbol{C}$	Baseline neural activity	0
$\boldsymbol{H}_{exc}$	Lead field for excitatory populations	$\mathcal{N}(0,1)$
$\boldsymbol{Q}$	Measurement noise covariance	$0.25\mathbf{I}$
$\boldsymbol{R}$	Process noise covariance	$(0.2 + 0.1\mathcal{U}(0,1))\mathbf{I}$
nS	Number of modulation states	3
μ_i	Mean of $\boldsymbol{\Gamma}_{i,k}$	$\mathcal{N}(1, 0.1)$
σ_i (uniform)	Variance of $\boldsymbol{\Gamma}_{i,k}$	$\mathcal{N}(0.4, 0.1)$
σ_i (normal)	Variance of $\boldsymbol{\Gamma}_{i,k}$	$\mathcal{N}(0.05, 0.01)$
$\boldsymbol{\Gamma}_{i,k}$ (uniform)	Vector constructing $\boldsymbol{\Gamma}_i$	$\mathcal{U}(\mu_i - 0.5\sigma_i, \mu_i + 0.5\sigma_i)$
$\boldsymbol{\Gamma}_{i,k}$ (normal)	Vector constructing $\boldsymbol{\Gamma}_i$	$\mathcal{N}(\mu_i, \sigma_i)$
$\boldsymbol{A}$	HMM transition probability	$0.999\mathbf{I} + (0.001/\text{nS})(\mathbf{1}\mathbf{1}^\top - \mathbf{I})$
$\boldsymbol{\pi}$	Initial HMM state probability	$(1/\text{nS})\mathbf{1}$

both the free simulation steps and the Kalman filtering steps, calculating the error gradient for each parameter at each step. In addition to calculating the parameter error gradients (i.e. error gradients of $\{\boldsymbol{W}, \boldsymbol{\Gamma}, \boldsymbol{D}, \boldsymbol{C}, \boldsymbol{S}, \boldsymbol{V}\}$), we also calculate the error gradients of the estimated covariances of the noise terms ϵ and ν. These noise covariances are necessary for the Kalman filter to estimate the current state $\boldsymbol{x}_t$, and are optimized alongside the model parameters. By backpropagating the error in this manner and fitting both the parameters and the noise covariances, we fit a model which produces the most accurate state estimator which can best predict future measurements.

After backpropagation through the free simulation and Kalman filtering steps, we then update the model parameters and the estimates of the noise covariances. Then, the process is repeated with a new small window in the epoch of fitting data. This process of Kalman filtering, free simulation and backpropagation is repeated with new data windows until the Kalman and free simulation errors converge. Multiple windows of the data epoch are used so that the model captures the general dynamics and statistical properties of the entire epoch, to avoid overfitting to a few timesteps of data. These windows are selected at random to avoid biasing the fit toward any one period within the epoch (e.g. the beginning or end of the epoch).

In the unmodulated MINDy model, the parameters do not vary temporally, so all parameters are updated in all steps of the model. In the expanded modulated MINDy model proposed here, only one of the $\boldsymbol{\Gamma}_i$ matrices is used at each timepoint. To account for this, we select the appropriate $\boldsymbol{\Gamma}_i$ at each time t using the modulation labels and use that $\boldsymbol{\Gamma}_i$ for model evolution and gradient calculation.

To aid in the tractability of the problem, we implement several constraints on $\boldsymbol{W}$ and $\boldsymbol{\Gamma}$. We constrain $\boldsymbol{W}_{ee}$ and $\boldsymbol{W}_{ei}$ to have 75% of their non-diagonal connections be zero, enforcing a prior level of sparsity in connections. We also constrain each $\boldsymbol{\Gamma}_i$ to be rank 1, i.e. the outer product of two n-dimensional vectors. This assumption is motivated by the premise that neuromodulators act in a spatially diffuse manner [31]. Because each $\boldsymbol{\Gamma}_i$ is multiplied element-wise by $\boldsymbol{W}$, the excitatory submatrices of each $\boldsymbol{\Gamma}_i$ also effectively have 75% of their elements set equal to 0.

To fit the model parameters, we use NADAM [32], implemented with PyTorch's Autograd engines [33]. This improves fitting efficiency and scalability relative to our prior work, by allowing GPU acceleration, which can be especially beneficial for processing large-scale EEG recordings. Notably, this significantly reduces the number of iterations needed for the convergence of the backpropagated Kalman filter approach. With Autograd, both the covariance update and local linearization process are included in the backward operation.

2.3. Simulation and actual data

2.3.1. Synthetic data

To enable the validation of our fitting procedure, we created synthetic data with known parameters. To create this synthetic data, we established models with a combination of fixed and random parameters. The fixed values and distributions for the model parameters are listed in table 1. It should be noted that our $\boldsymbol{W}$

submatrices were constructed as a linear combination of a sparse and low-rank matrix via:

$$\boldsymbol{W}_{\text{ee}} = \boldsymbol{W}_{\text{s}} + \boldsymbol{W}_{\text{l1}}\boldsymbol{W}_{\text{l2}}^{\top} + \boldsymbol{W}_{\text{diag}}, \tag{6}$$

where $\boldsymbol{W}_{\text{s}} \in \mathbb{R}^{n\times n}$ is a sparse matrix, $\boldsymbol{W}_{\text{l1}}, \boldsymbol{W}_{\text{l2}} \in \mathbb{R}^{n\times n/4}$ are low rank matrices, and $\boldsymbol{W}_{\text{diag}} \in \mathbb{R}^{n}$ is a vector specifying the diagonal self-connection weights. Note that equation (6) specifies $\boldsymbol{W}_{\text{ee}}$, but both $\boldsymbol{W}_{\text{ee}}$ and $\boldsymbol{W}_{\text{ei}}$ were constructed in this way. $\boldsymbol{W}_{\text{ie}}$ and $\boldsymbol{W}_{\text{ii}}$ were constructed as diagonal matrices with diagonal values directly sampled from their distributions.

We initialized each $\boldsymbol{\Gamma}_i$ matrix as the outer product of a vector drawn from a unique distribution. To construct each unique distribution, we first randomly generate its mean, μ_i, from $\mathcal{N}(1, 0.1)$. We then generate a random binary digit indicating whether we should use a uniform distribution, or a normal distribution. Then, we generate a variance σ_i for the distribution. If we are using a uniform distribution, $\sigma_i \sim \mathcal{N}(0.4, 0.1)$, and if we are using a normal distribution, $\sigma_i \sim \mathcal{N}(0.05, 0.01)$. Then, we generate a vector, $\Gamma_{i,k}$, from the constructed distribution: if uniform, $\Gamma_{i,k} \sim \mathcal{U}(\mu_i - (\sigma_i/2), \mu_i + (\sigma_i/2))$; if normal, $\Gamma_{i,k} \sim \mathcal{N}(\mu_i, \sigma_i)$. Then, $\Gamma_i = \Gamma_{i,k}\Gamma_{i,k}^{\top}$.

We chose the distributions for μ_i and σ_i such that there could be variation in the distributions of $\boldsymbol{\Gamma}_i$, while also maintaining values close to 1. If the values of Γ are very large or very small, Γ will overwhelm $\mathbf{W}$ in the effective neural connectivity $\mathbf{W} \odot \Gamma_i$. In other words, our assumed modulation does not rescale synaptic weights by large amounts.

To generate state changes in our synthetic data, we initialized a HMM with transition probability matrix $\boldsymbol{A}$ and initial state probability $\boldsymbol{\pi}$. At each timestep, we calculated the probability of the next state based on $\boldsymbol{A}$ and then randomly selected the state index weighted by the calculated probability.

Once we had generated our random models, we forward simulated them for 20 000 timesteps (equivalent to 80 s of 250 Hz EEG) to create synthetic data. When fitting on this synthetic data, the true lead field matrix and noise covariances ($\boldsymbol{H}$, $\boldsymbol{Q}$, and $\boldsymbol{R}$) were used to initialize the models, but the other model parameters were initialized randomly and compared to the true values after fitting, to test parameter recovery when ground truth is known. Since the true value of $\boldsymbol{W}$ is known, however, we create a mask zeroing out the same entries in the fit $\boldsymbol{W}$, to avoid zeroing out connections which are actually present in the synthetic models.

2.3.2. EEG data

For our second experiment, we used EEG data of 20 subjects dosed with propofol published in [34] and available online [35]. In this data, subjects were recorded at four levels of sedation: (i) resting baseline with no anesthesia, (ii) mild sedation (defined as 0.6 μg ml^{-1} target blood plasma concentration), (iii) moderate sedation (1.2 μg ml^{-1} target blood plasma concentration), and iv) return to mild sedation. Each sedation level had approximately 7 min of data, recorded at 250 Hz after 10 min of allowing the blood plasma level to reach a steady state and was saved as a separate EEG file. This data is highly compatible for our purposes because the four pharmacological regimes above can be used as labels for constructing and validating our modulated model.

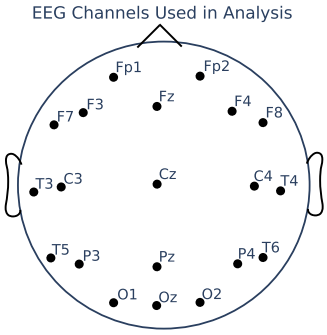

Figure 2. EEG channels used in analysis of propofol dosage EEG data.

We filtered the data between 0.5 and 15 Hz, subtracted the median of each channel, and divided by the mean absolute deviation of each channel. We combined all four sedation state recordings into a single array for each subject, and added a sedation state index $i \in \{1, 2, 3, 4\}$ for each timepoint of the array. We also used a subset of the recorded channels in [34], using only 20 channels, shown in figure 2. These channels make up the 10–20 system, a standard montage commonly used in EEG analysis—the remaining 108 channels are at less frequently used locations on the scalp.

The data can be modeled with all 128 channels if desired, however the parameter search space will be significantly larger than with 20 channels and as a result the computational time will be longer and the quality of the fit may degrade. Because clinical EEG uses lower density montages (typically in the range of 18–25 channels), we conducted our analysis with a montage of channels in this range.

Any channel montage will suffice for the model, provided the montage yields enough spatial information for the analysis in question. If the activity under investigation is frontal in nature, for example, then a simple 4-channel frontal montage may suffice. For whole-brain analyses, the montage should have a good distribution across the surface of the scalp. Analyses that require very granular spatial information will need a higher number of channels to be modeled than those requiring only approximate spatial information. In clinical settings, it is likely that the

montage used will be limited by the channel setup for EEG used in the clinic. For basic scientific investigations, there may be more freedom to choose a montage based on the specific needs of the analysis.

Our modeling methodology does not depend on the sampling frequency of the data used; however it can only encode frequency information that is present in the data. Therefore it is required that any frequencies under analysis be no greater than half the sampling rate of the data, according to the Nyquist sampling criterion [36].

Our $\boldsymbol{H}$ matrix was constructed as in equation (5), with

$$\boldsymbol{H}_{\text{exc}} = \mathbf{I} - 0.05\mathbf{1}\mathbf{1}^{\top}, \tag{7}$$

where $\mathbf{I}$ denotes the 20×20 identity matrix and $\mathbf{1}$ denotes the 20-dimensional vector of all ones. We constructed $\boldsymbol{R}$ and $\boldsymbol{Q}$ as diagonal matrices, with $\boldsymbol{R} = 1.2\mathbf{I}$ and $\boldsymbol{Q} = 0.25\mathbf{I}$. The other model parameters were initialized randomly, as in experiment 1. In this experiment, we do not have a ground truth of the 75% of non-diagonal connections in the $\boldsymbol{W}_{\text{ee}}$ and $\boldsymbol{W}_{\text{ei}}$ submatrices which are zero as we did in the synthetic data case. To continue enforcing this constraint, we selected a random 75% of non-diagonal connections in $\boldsymbol{W}_{\text{ee}}$ and $\boldsymbol{W}_{\text{ei}}$, and set these to 0 for all subjects. Finally, because we wanted to explore modulations relative to baseline (prior to propofol dosing), we enforce $\boldsymbol{\Gamma}_1 = \mathbf{1}\mathbf{1}^{\top}$, i.e. $\boldsymbol{W} \odot \boldsymbol{\Gamma}_1 = \boldsymbol{W}$. The other $\boldsymbol{\Gamma}$ matrices ($i \in \{2,3,4\}$) are fit as all $\boldsymbol{\Gamma}_i$ in the synthetic data experiment.

3. Results

3.1. Modulated MINDy recovers ground truth connections and modulations

Our first experiment tested whether modulated MINDy could accurately recover the connectivity and modulation matrices in models with known ground-truth parameters. To benchmark this, we compared the accuracy of parameter inferences for the modulated MINDy architecture in the presence of multiple non-stationary regimes to the performance of the unmodulated MINDy architecture (previously validated in [10]) on data in which there is only a single, stationary regime. That is, for the unmodulated MINDy results, we fit unmodulated MINDy models to single-regime synthetic data generated by using the simulated data setup specified in [10].

As shown in figure 3(a), we achieve ground truth correlations of nearly the same level as in the unmodulated MINDy problem. The unmodulated MINDy models achieve a median correlation coefficient of $r = 0.9771$ (IQR: 0.9720–0.9866) for the full $\boldsymbol{W}$ correlation, whereas the modulated MINDy models yield a correlation coefficient of $r = 0.9280$ (IQR: 0.8539–0.9362). The EE and EI submatrices had similar correlations: MINDy EE, $r = 0.9873$ (IQR: 0.9519–0.9957); Modulated MINDy EE, $r = 0.9202$ (IQR: 0.8238–0.9408); MINDy EI, $r = 0.9715$ (IQR: 0.9506–0.9895), Modulated MINDy EI, $r = 0.9227$ (IQR: 0.8615–0.9495). Since we are fitting not only one connectivity matrix but also several $\boldsymbol{\Gamma}$ matrices, a slight decrease in fit quality is expected relative to the unmodulated problem, due to the increase in the connectivity parameter search space from $n \times n$ (MINDy) to $n \times n \times m$ (modulated MINDy).

Additionally, we achieved similarly high ground truth correlations for the EE and EI components of the modulation matrices $\boldsymbol{\Gamma}$: $\boldsymbol{\Gamma}_1$ $r = 0.9269$ (IQR: 0.8902–0.9434) EE, $r = 0.9160$ (IQR: 0.8618–0.9385) EI; $\boldsymbol{\Gamma}_2$ $r = 0.9228$ (IQR: 0.8864–0.9421) EE, $r = 0.9152$ (IQR: 0.7665–0.9343) EI; $\boldsymbol{\Gamma}_3$ $r = 0.9215$ (IQR: 0.9011–0.9411) EE, $r = 0.9071$ (IQR: 0.8216–0.9435) EI, as shown in figure 3(b). These results indicate that we can accurately recover both the true base connectivity and the true modulations of that connectivity, and are summarized in table 2.

Figure 4 provides examples of individual connection and modulation matrices and their recovered estimates via modulated MINDy (i.e. the proposed method). As shown, the spatial patterns of high- and low-amplitude connections in the true matrices are replicated in the estimated matrices. The estimated $\boldsymbol{W}$ tends to have a slightly higher amplitude than the true $\boldsymbol{W}$, and the estimated $\boldsymbol{\Gamma}$ tends to have a slightly lower amplitude than the true $\boldsymbol{\Gamma}$. This is to be expected, since we only regularized the structure of $\boldsymbol{W}$ and $\boldsymbol{\Gamma}$, and did not regularize their amplitude. In summary, we can conclude that the modulated MINDy model can estimate the connectivity and modulation parameters to within a potential scaling factor of the true value.

This scaling factor arises from the decomposition of the connectivity into two matrices, $\boldsymbol{W} \odot \boldsymbol{\Gamma}_i$. Since there are an infinite number of ways to factor a number into the product of two other numbers, we cannot assure that the elements of the estimated $\boldsymbol{W}$ and $\boldsymbol{\Gamma}_i$ are not off by a constant from the 'true' values, i.e. $\hat{\boldsymbol{W}} = \gamma \boldsymbol{W}$, $\hat{\boldsymbol{\Gamma}}_i = (1/\gamma)\boldsymbol{\Gamma}_i$, where γ represents the scaling factor. To reduce the impact of this, we constrain the parameter search space by enforcing a low-rank constraint onto $\boldsymbol{\Gamma}_i$. However, this does not entirely resolve the issue, and it is an important limitation to consider when analyzing the models returned by modulated MINDy. For many use cases, this ambiguity is irrelevant. For example in this analysis, we analyze parameter values in comparison to one another, i.e. by comparing $\boldsymbol{\Gamma}_i$ with $\boldsymbol{\Gamma}_j$, in which case the scaling factor cancels out.

3.2. Modulated MINDy is robust to noise

To test modulated MINDy's robustness to noise, we repeated our ground truth recover experiment

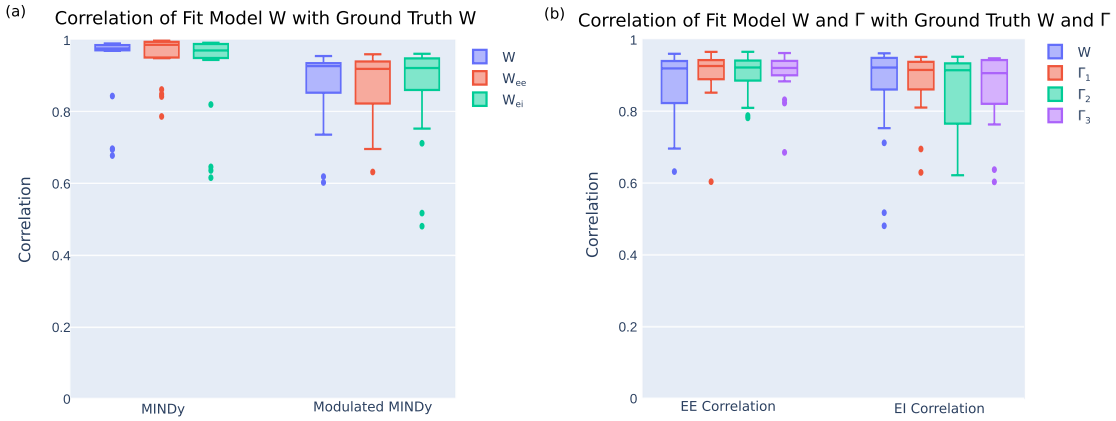

Figure 3. Modulated MINDy returns models with high ground truth correlations. (a) Ground truth correlations for the full connectivity, excitatory–excitatory connections, and excitatory–inhibitory connections for both unmodulated and modulated MINDy problems. (b) Ground truth correlations for the connectivity matrix ($\boldsymbol{W}$) and the connectivity modulation matrices ($\boldsymbol{\Gamma}$) in the modulated MINDy problem.

Table 2. Correlation of fit parameters with ground truth parameter values for MINDy and modulated MINDy.

Parameter, model	Correlation (r)	Inter-quartile range (IQR)
$\boldsymbol{W}$, MINDy	0.9771	0.9720–0.9866
$\boldsymbol{W}$, modulated MINDy	0.9280	0.8359–0.9362
$\boldsymbol{W}_{ee}$, MINDy	0.9873	0.9519–0.9957
$\boldsymbol{W}_{ee}$, modulated MINDy	0.9202	0.8238–0.9408
$\boldsymbol{W}_{ei}$, MINDy	0.9715	0.9506–0.9895
$\boldsymbol{W}_{ei}$, modulated MINDy	0.9227	0.8615–0.9495
$\boldsymbol{\Gamma}_{1,ee}$, modulated MINDy	0.9160	0.8618–0.9385
$\boldsymbol{\Gamma}_{2,ee}$, modulated MINDy	0.9228	0.8864–0.9421
$\boldsymbol{\Gamma}_{3,ee}$, modulated MINDy	0.9215	0.9011–0.9411
$\boldsymbol{\Gamma}_{1,ei}$, modulated MINDy	0.9160	0.8618–0.9385
$\boldsymbol{\Gamma}_{2,ei}$, modulated MINDy	0.9152	0.7665–0.9343
$\boldsymbol{\Gamma}_{3,ei}$, modulated MINDy	0.9071	0.8216–0.9435

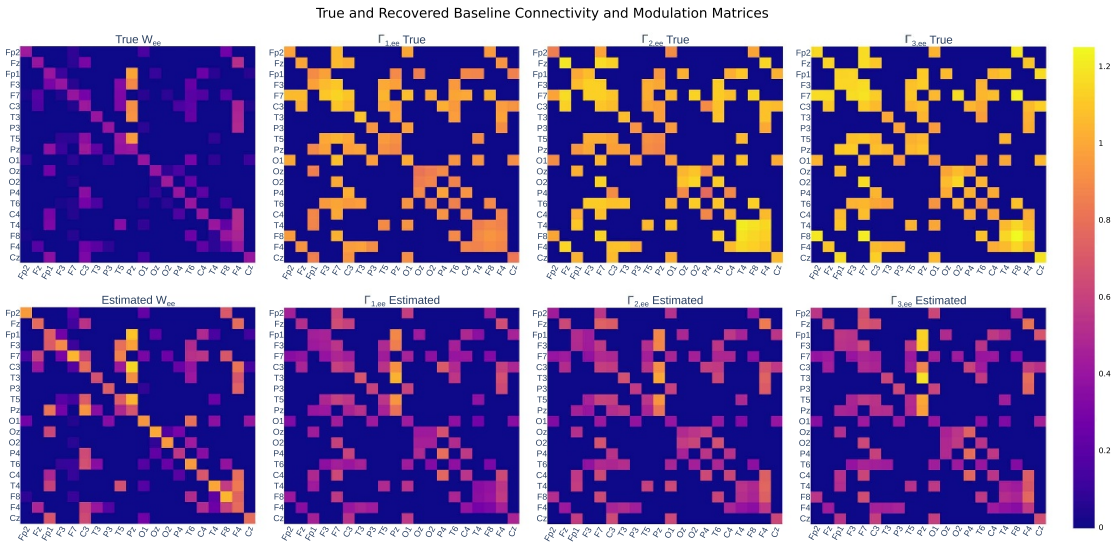

Figure 4. Modulated MINDy returns connectivity and modulation matrices accurate to a potential scaling factor. True (top row) and estimated (bottom row) $\boldsymbol{W}$ and $\boldsymbol{\Gamma}$ matrices from a single subject.

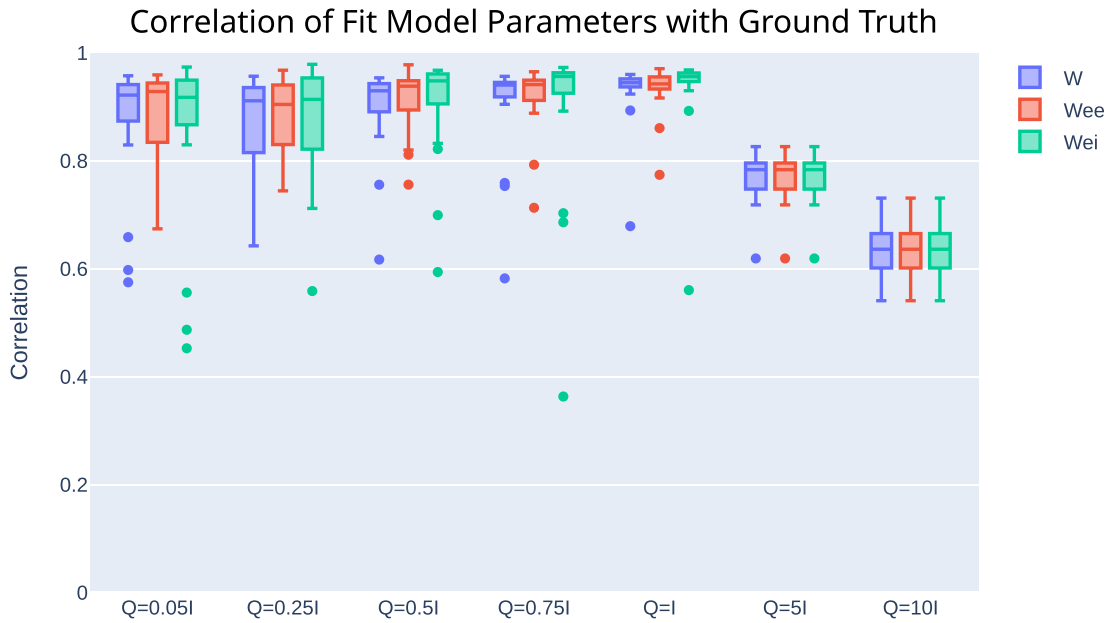

Figure 5. Modulated MINDy is robust to noise. Correlation of inferred $\boldsymbol{W}$ (blue), $\boldsymbol{W}_{ee}$ (red), and $\boldsymbol{W}_{ei}$ (green) with ground truth values, for differing noise covariances.

Table 3. Modulated MINDy's correlation with ground truth in the presence of noise.

Parameter	Corr. (IQR), $Q = 0.05\mathbf{I}$	Corr. (IQR), $Q = 0.25\mathbf{I}$	Corr. (IQR), $Q = 0.5\mathbf{I}$	Corr. (IQR), $Q = 0.75\mathbf{I}$	Corr. (IQR), $Q = \mathbf{I}$	Corr. (IQR), $Q = 5\mathbf{I}$	Corr. (IQR), $Q = 10\mathbf{I}$
$\boldsymbol{W}$	0.9217 (0.8736–0.9412)	0.9112 (0.8152–0.9567)	0.9297 (0.8907–0.9430)	0.9403 (0.9184–0.9453)	0.9447 (0.9368–0.9518)	0.7836 (0.7476–0.7959)	0.6361 (0.6013–0.6652)
$\boldsymbol{W}_{ee}$	0.9283 (0.8342–0.9441)	0.9044 (0.8301–0.9405)	0.9382 (0.8942–0.9483)	0.9412 (0.9119–0.9494)	0.9425 (0.9329–0.9554)	0.7836 (0.7476–0.7959)	0.6361 (0.6013–0.6652)
$\boldsymbol{W}_{ei}$	0.9175 (0.8668–0.9494)	0.9137 (0.8215–0.9535)	0.9481 (0.9054–0.9609)	0.9562 (0.9252–0.9631)	0.9559 (0.9468–0.9627)	0.7836 (0.7476–0.7959)	0.6361 (0.6013–0.6652)

with varying levels of noise covariance in our generated data ($Q = [0.05\mathbf{I}, 0.25\mathbf{I}, 0.5\mathbf{I}0.75\mathbf{I}, \mathbf{I}, 5\mathbf{I}, 10\mathbf{I}]$. We found that modulated MINDy accurately inferred the ground truth parameters for levels of noise up to $\boldsymbol{Q} = \mathbf{I}$, and then the performance declined (figure 5, table 3).

3.3. Modulated MINDy is reliable and individualized

In our second experiment, we tested modulated MINDy on actual EEG data to understand how it functioned on a population of individuals. We split each subject's data in half temporally within each sedation level, in order to perform a test–retest reliability analysis.

We observed that models were reliable, showing high correlations within a single subject ($\boldsymbol{W}$: $r = 0.7670$ (IQR: 0.7187–0.8118); $\boldsymbol{W}_{ee}$: $r = 0.7930$ (IQR: 0.7276–0.8746); $\boldsymbol{W}_{ei}$: $r = 0.6855$ (IQR: 0.6054–0.7202; figure 6). Additionally, correlations across subjects were significantly lower ($\boldsymbol{W}$: $r = 0.6230$ (IQR: 0.5722–0.6802), $p = 1.97\text{e}^{-11}$; $\boldsymbol{W}_{ee}$: $r = 0.6510$ (IQR: 0.5760–0.7252), $p = 1.04\text{e}^{-8}$; $\boldsymbol{W}_{ei}$: $r = 0.5144$ (IQR: 0.4275–0.5821), $p = 1.61\text{e}^{-8}$), indicating that models are individualized, a requisite condition for capturing dynamics and mechanisms which vary across individuals. These results are summarized in table 4.

3.4. Models identify consistent modulation associated with pharmacology

Having established validity in ground-truth settings, we proceeded to apply the method to labeled EEG data from stages of propofol anesthesia. Specifically, for each of 20 individuals, we inferred a base connectivity matrix for the pre-sedation regime, as well as modulation matrices for the regimes of mild anesthesia, moderate anesthesia and return to mild anesthesia. To test our method's ability to provide insight into the observed changes into dynamics, we analyzed the modulation ($\boldsymbol{\Gamma}$) matrices at each sedation level

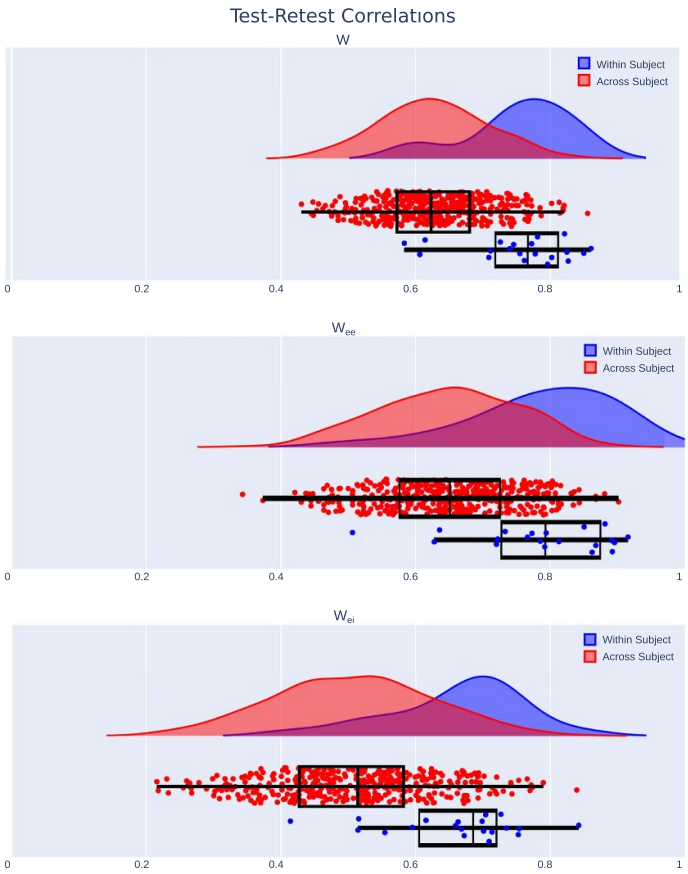

Figure 6. Modulated MINDy is reliable and individualized. Test–retest correlations on split-half data are significantly higher within subject than across subject in the full $\boldsymbol{W}$ connectivity matrix, as well as the EE and EI submatrices.

Table 4. Test–retest results for modulated MINDy.

Parameter	Within-subject correlation	Within-subject IQR	Across-subject correlation	Across-subject IQR	p-value
$\boldsymbol{W}$	0.7670	0.7187–0.8118	0.6230	0.5722–0.6802	$1.97e^{-11}$
$\boldsymbol{W}_{ee}$	0.7930	0.7276–0.8746	0.6510	0.5760–0.7252	$1.04e^{-8}$
$\boldsymbol{W}_{ei}$	0.6855	0.6054–0.7202	0.5144	0.4275–0.5821	$1.61e^{-8}$

compared to baseline. An exemplar subject is shown in figure 7.

Of note, modulated MINDy is detecting changes that are not clear to the naked eye in either the raw EEG or the power spectrogram (figure 7). This indicates that modulated MINDy is picking up on true *latent* dynamics which cannot be observed directly from the data. By capturing these latent dynamics, modulated MINDy is capable of providing greater insight into the mechanisms that generate changes in neural state, in this case, changes in consciousness.

We performed a two-way ANOVA on the post-synaptic modulation of excitatory connections to analyze the effects of channel location and sedation state. We excluded the pre-sedation state from this analysis since $\boldsymbol{\Gamma}_1 = \mathbf{1}\mathbf{1}^\top$ and so there will not be any variance across subjects or spatial location in that sedation state. The ANOVA indicated that channel location and sedation state both significantly affected post-synaptic modulation of excitatory connections, as did their interaction (ANOVA values given in table 5).

In analyzing the models at the population level, we observed first that the vast majority of the values within the matrices fall within the range $[0, 1)$, though a small percentage ($\boldsymbol{\Gamma}_2$: 0.169%, $\boldsymbol{\Gamma}_3$: 0.150%, $\boldsymbol{\Gamma}_4$: 0.128%) are greater than 1 (figure 8). Since these values are multiplying the base connection weights $\boldsymbol{W}$, a fractional value in $\boldsymbol{\Gamma}$ indicates a relative decrease of connection strength, or a relative increase in inhibition. In the population average, all values are fractional. This pattern of the modulations increasing inhibition is consistent with the widespread inhibitory effects of propofol [37, 38].

Importantly, we went beyond these bulk inferences and also analyzed the spatiotemporal changes

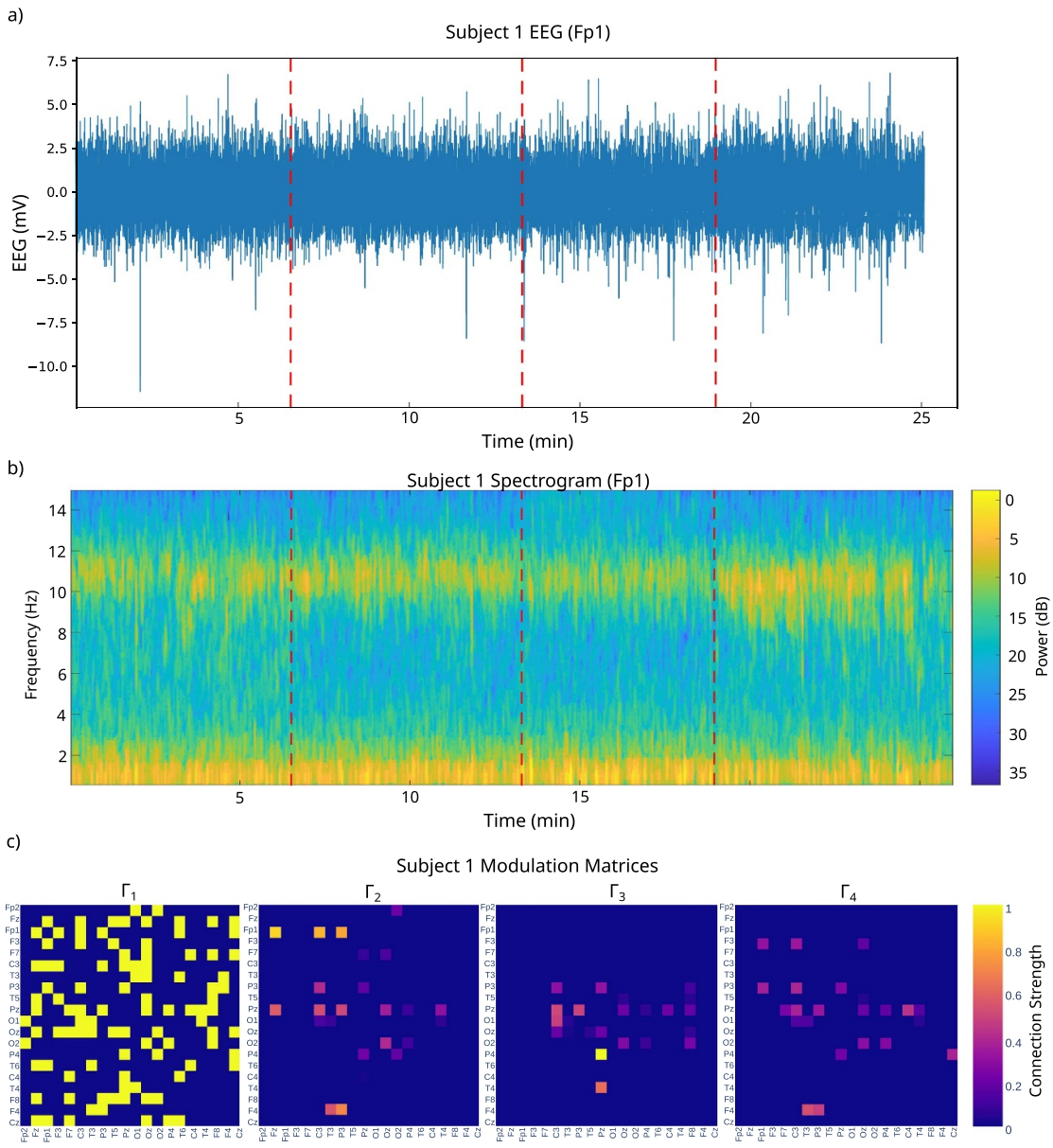

Figure 7. Modulated MINDy detects changes not obvious by eye. EEG timeseries data (a), power spectrogram (b) and $\mathbf{\Gamma}$ matrices for an exemplar subject. Dashed red lines indicate times when the anesthesia state changes.

Table 5. ANOVA of post-synaptic modulation values inferred by modulated MINDy.

	Sum of squares	Degrees of freedom	*F*-value	*p*-value
Channel	62.75	19	7.24	b.m.p.
Sedation state	89.70	2	98.33	b.m.p.
Channel:Sedation state	113.54	38	6.55	b.m.p.
Residual	519.97	1140		

b.m.p.: below machine precision.

associated with each modulation regime. We found that in the population mean, the modulation values which were highest (i.e. smallest relative inhibition) were typically connections to populations underlying posterior channels, particularly in the moderate sedation state (figure 9(a)). This is compatible with prior characterizations of the effect of propofol on attenuating the posterior regions of the scalp [39, 40]. Also note the substantial decrease in post-synaptic modulation value from $\mathbf{\Gamma}_1$ to $\mathbf{\Gamma}_2$, $\mathbf{\Gamma}_3$, and $\mathbf{\Gamma}_4$, further indicating propofol's general inhibitory effects. We also noted that each modulation state introduced specific spatial changes to the effective connectivity. Note that in this paper we use the term 'effective connectivity' differently from how it is used in connectomics papers, e.g. [41, 42]. That field uses 'effective

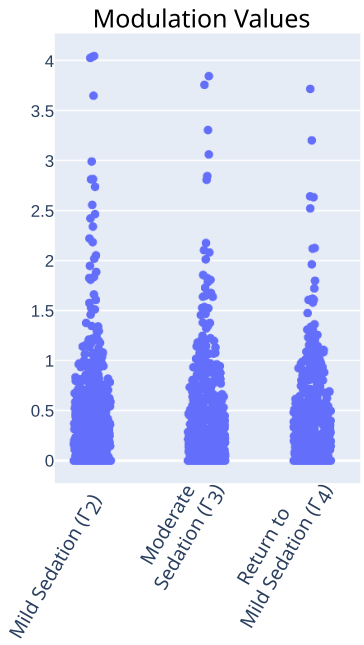

Figure 8. Modulated MINDy identifies inhibitory modulation during sedation. Distributions of all modulation values in each sedation state.

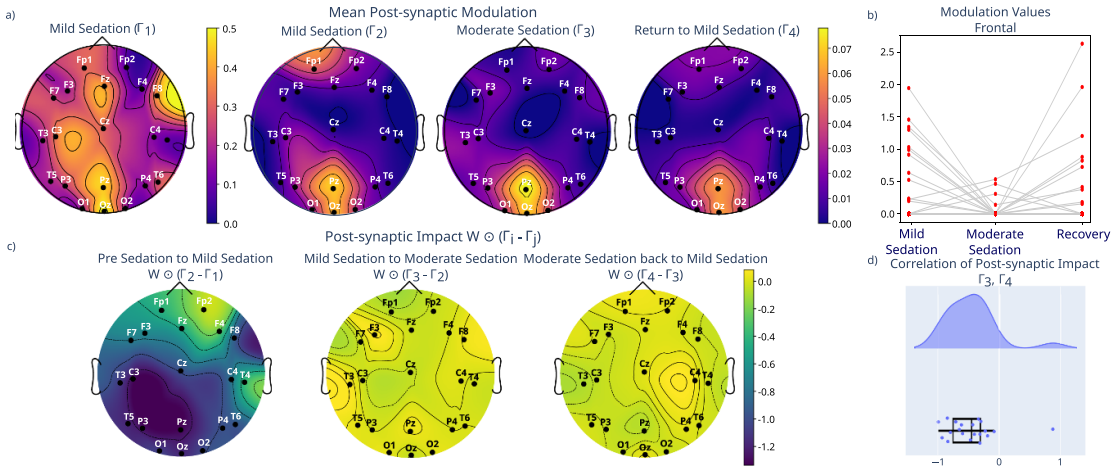

Figure 9. Modulated MINDy identifies inhibitory, spatially focused modulation during sedation. (a) Mean post-synaptic modulation in population mean of modulation matrices in sedation states. (b) Post-synaptic modulation impact (difference between two temporally adjacent modulation matrices, scaled by $\boldsymbol{W}$) across transitions between sedation states (population mean). (c) Distributions of EE modulation values in frontal channels. (d) Correlation between moderate sedation post-synaptic impact and return to mild sedation post-synaptic impact.

connectivity' as a technical term, analogous to functional connectivity or structural connectivity. We use 'effective connectivity' to refer to the constant connectivity $\boldsymbol{W}$ multiplied by the current modulation $\boldsymbol{\Gamma}_i$. In other words, 'effective connectivity' in this paper refers to the *modulated* connectivity $\boldsymbol{W} \odot \Gamma_i$. To quantify the spatial changes of the effective connectivity in each modulatory state, we define a measure of the modulation's impact, defined as the difference between two temporally adjacent modulations matrices scaled by $\boldsymbol{W}$, e.g. $\boldsymbol{W} \odot (\boldsymbol{\Gamma}_2 - \boldsymbol{\Gamma}_1)$. Thus, we account for both the magnitude differences in $\boldsymbol{\Gamma}$ as well as the base magnitude of the connection weights which are being modulated. A positive impact indicates a strengthened connection, and a negative impact indicates a weakened connection.

As the subjects move from pre-sedation to mild sedation, there is a widespread weakening of connections across the whole brain, but again particularly in the posterior areas (figure 9(c), panel 1). As the subjects progress from mild to moderate sedation and from moderate sedation to the return to mild, the changes are of smaller, less significant magnitude (figure 9(c)). The change from mild to moderate sedation is characterized by a further weakening of the frontal connections (figure 9(b)). This

general observation is consistent with the observation of anteriorization of neural activity in the time before loss of consciousness [39, 40, 43], in which the loss of posterior EEG power is accompanied by a potentiation of activity in frontal regions (note the posterior–anterior dichotomy in inferrered modulatory impact, especially from baseline to mild sedation, panel 1 of figure 9(c). Anteriorization has been modeled as a differential effect of inhibition along the posterior–anterior axis [44, 45]. Interestingly, as the subjects transition from moderate sedation back to mild sedation, there is a reversal of the effects seen in the transition from mild sedation to moderate sedation (figure 9(c)), highlighted by the negative correlations between the impact of moderate sedation and the impact of the return to mild sedation (figure 9(d)). This indicates a gradual decrease in EE connection strength as the subjects are dosed with anesthesia, followed by a return to a state similar to the mild sedation in the return to mild phase.

Overall, these results indicate that the methodology is identifying modulatory effects that are consistent with what we would expect from prior detailed studies on the mesoscale effects of propofol anesthesia, thus supporting the validity of the proposed approach.

Importantly, this modeling methodology identifies changes in electrophysiological dynamics that may or may not be observable by eye. Currently, identifying changes in electrophysiological dynamics, even when the times state changes occur are labeled, requires electrophysiologists examining long EEG recordings by hand. This process is time-consuming, and algorithms to categorize changes in neural activity can reduce reliance on it. Additionally, modeling methodologies such as this one can produce *mechanistic* insights into the underlying circuit dynamics that cannot be identified by simple descriptive analyses or observed by eye.

4. Discussion

4.1. Data-driven inference of modulation for understanding non-stationary brain dynamics

Brain dynamics are fundamentally nonstationary, changing as individuals switch between varying tasks, cycle between sleep and wake states, or as pathologies within the brain improve or worsen. Such nonstationarity is mediated, at least in part, by processes of neuromodulation that potentiate or attenuate synaptic connections within and between brain areas. Our goal in this paper was to develop a data-driven, parametric modeling framework for identifying such modulation at whole-brain scales.

Our specific approach was to formulate modulation within a physiologically interpretable RNN construct, where a baseline set of synaptic weights—common to all non-stationary regimes—is scaled by regime-specific modulation. Within this framework are several nontrivial technical challenges. Specifically, because our model is formulated at the level of latent (unobserved) neural populations, we faced the dual-estimation problem of fitting model parameters and estimating state variables at the same time. Compounding this is the need to fit not one connection matrix, but rather a family of such matrices. Since the base connection weights $\boldsymbol{W}$ are multiplied by the modulations $\boldsymbol{\Gamma}_i$ $(i = 1, 2, \ldots, m)$, this is a fundamentally ill-posed problem. However, we showed that by imposing appropriate priors on the construction of both $\boldsymbol{W}$ and $\boldsymbol{\Gamma}$, the problem can be made tractable, leading to interpretable results.

At a mathematical level, our modeling setup involves, in essence, a matrix factorization problem: there are a number of distinct effective connectivity matrices $(\boldsymbol{W} \odot \boldsymbol{\Gamma}_i)$, which are then factorized into a component common across all regimes $(\boldsymbol{W})$, and a modulatory component which varies discretely based on regime $(\boldsymbol{\Gamma}_i)$. While there are algorithms for the decomposition of a matrix into the Hadamard product of two matrices [46, 47], these algorithms solve a problem which is constructed differently from ours. The algorithms in [46, 47] decompose a given matrix into two or more low-rank matrices. In our problem, we are principally concerned with finding finding a decomposition with a component matrix *common* to all given matrices, as motivated by our specific domain application context. Importantly, we allow our common component $(\boldsymbol{W})$ to be full-rank, rather than decomposing into two low-rank matrices.

As noted, data-driven decomposition of neuromodulatory effects on mesoscale dynamical models has not been widely studied in the computational neural modeling community. Several authors have, however, done work with a similar motivation for taking a neuromodulation approach to modeling. Li *et al* [23] developed a statistical generalized linear model (GLM) construct that embodies a modulatory nonstationary architecture for neuronal-level modeling. They specifically imposed a level of similarity between switched models by incorporating a Gaussian prior onto the weight matrices of their GLMs. They further incorporated a decomposition of connection strength from connection direction—decomposing their weight matrices into a direction component in $\{-1, 0, 1\}$, and a strength component in $\mathbb{R}_{>0}$. Having done this decomposition, they also impose a Gumbel–Softmax prior on the direction component, minimizing the number of connections that switch direction with different regimes. In contrast, working at the meso-scale, we enforce a direction mask onto our effective connectivity matrices which does not change over time, and decompose the strength of each matrix into a common component $\boldsymbol{W}$ and a switched component $\boldsymbol{\Gamma}$.

There are certainly many biophysical models that have engaged neuromodulation from a bottom-up

perspective, especially at neuronal and small-circuit scales, e.g. [48–51], which can generate inferences and predictions at the meso- and macro-scale [52, 53]. We view our contribution here as a methodological enabling of such approaches in the sense that our framework takes mesoscale data and performs an inference problem to arrive at a mesoscale model of neural modulation.

4.2. Modulated MINDy provides individualized inference of modulated whole-brain dynamics

Our developed approach represents a generalization of our mesoscopic individualized neural dynamics (MINDy) framework [9, 10]. Specifically, rather than identifying a single dynamical regime, the generalized framework proposed here allows for spatially relevant representations of both the base connectivity common to all regimes as well as modulation that may vary with time and brain state. In this sense, this modulated model architecture is also multi-timescale, as the population activity state $\boldsymbol{x}$ changes on a much faster timescale than the selection index i. Importantly, our model architecture is biologically interpretable, providing insight into excitatory–inhibitory dynamics not provided by vanilla RNN architectures. Modulated MINDy accurately estimates known latent dynamics from synthetic data, and infers reliable and individualized models when tested on human data. Modulated MINDy is also scalable, requiring only ∼10 min to fit 30 min of 20-channel EEG data with 3 modulation matrices.

4.3. Modulated MINDy as a tool to infer interpretable neuromodulation

As a proof of concept, we tested modulated MINDy on open-source, labeled EEG recordings of subjects receiving the general anesthetic drug propofol. We found that our models had modulation structures consistent with prior literature on propofol sedation, including promoting inhibition along the posterior–anterior spatial axis. Thus, modulated MINDy provides spatiotemporal models which are not only accurate and reliable, but can be interpreted to gain mechanistic insight into an individual's brain dynamics and how they are modulated over time or as a function of exogenous factors or inputs. We emphasize that our applicative example in anesthesia was not intended to make a specific scientific point about propofol *per se*, but rather as a methodological validity test.

In this validity test, we did not explicitly model $GABA_A$ and $GABA_B$ receptors separately, but modeled inhibitory populations as a single type type compared to excitatory populations, regardless of their subtype. To model $GABA_A$ and $GABA_B$ separately, this model can be extended to include two inhibitory populations, one corresponding to $GABA_A$ and one corresponding to $GABA_B$. (See for example, [28], which used MINDy to model two excitatory and two inhibitory populations per EEG channel, though the inhbitory populations did not explicitly correspond to different GABA types.)

We envision a number of applicative contexts for the proposed method, for both basic scientific and clinical questions. Modulated MINDy can be used in a task context, to infer modulations unique to specific cognitive functions. It could be used, as in this proof of concept, to characterize the spatiotemporal effects of an exogenous intervention, either pharmacological or otherwise. Additionally, it could be used in clinical contexts, to associate different modulation patterns with various states of pathology, e.g. seizures, coma, or ischemia. A key aspect of modulated MINDy is that it infers *individual* changes in brain dynamics from EEG recordings. This enables comparison of the patients based on characteristics of their brain dynamics, rather than by external measures such as propofol concentration. Since pharmacodynamics are highly individualized, comparison at this level could yield interesting results related to the mechanisms of propofol onto individual brains. Ultimately, modulated MINDy is a modeling tool to infer changes that underlie non-stationary brain recordings, where the non-stationary regimes are labeled in the data.

4.4. Limitations

We note a few important limitations of the proposed approach. Most notably, as outlined in our introduction, our approach tackles the problem of inferring what is modulated within network dynamics, and not the companion problem of inferring when such modulation has occurred. In other words, we require here known demarcation or labeling of nonstationary regimes. Future work will engage the challenge of inferring both modulation and regime engagement, simultaneously. Our model itself is formulated at the mesoscale, with clear abstraction of cellular and sub-cellular dynamics. These assumptions could in principle be generalized, though that would lead to increasingly computational complexity regarding the ensuing inference problem. Additionally, the nature of the Hadamard decomposition results in the inferred $\boldsymbol{W}$ and $\boldsymbol{\Gamma}$ parameters being subject to a potential scaling factor, as discussed in 3.1. This means that any direct analysis of parameter values should be applied with caution. This scaling factor is irrelevant, however, for analyses comparing parameters (as in this paper), analyses of the product $\boldsymbol{W} \odot \boldsymbol{\Gamma}_i$, or analyses of the overall dynamics of the model.

4.5. Conclusion

In conclusion, we have presented modulated MINDy, a framework for fitting multiscale, modulated, mesoscale models of brain dynamics to individual data, and have validated it on both ground truth synthetic and actual human EEG data. In the future, we plan to extend this model by enabling modulated inference on data where the modulatory states are unlabeled,

i.e. where there is a need to infer also the points at which the non-stationary regimes within the data change. We anticipate that modulated MINDy's ability to give mechanistic inference will make it a powerful tool for analysis in many neuroscientific and clinical contexts.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://www.repository.cam.ac.uk/handle/1810/252736.

Acknowledgment

Portions of this work were supported by Grants R01NS130693 and 5T32NS126157-02 from the US National Institutes of Health.

ORCID iDs

Addison Schwamb 0009-0003-3017-030X
Zongxi Yu 0009-0001-7016-0013
ShiNung Ching 0000-0003-4063-7068

References

[1] Babadi B and Emery N B 2014 A review of multitaper spectral analysis *IEEE Trans. Biomed. Eng.* **61** 1555–64
[2] Liu S, Wang J, Shanshan Li and Cai L 2023 Epileptic seizure detection and prediction in EEGs using power spectra density parameterization *IEEE Trans. Neural Syst. Rehabil. Eng.* **31** 3884–94
[3] Blinowska K and Durka P 2006 Electroencephalography (EEG) *Wiley Encyclopedia of Biomedical Engineering* vol 10 p 9780471740360
[4] Louis E K S, Frey L C, Britton J W, Hopp J L, Korb P, Koubeissi M Z, Lievens W E and Pestana-Knight E M 2016 The normal EEG *Electroencephalography (EEG): an introductory text and atlas of normal and abnormal findings in adults, children and infants [internet]* (American Epilepsy Society)
[5] Marcuse L V, Schneider M, Mortati K A, Donnelly K M, Arnedo V and Grant A C 2008 Quantitative analysis of the EEG posterior-dominant rhythm in healthy adolescents *Clin. Neurophysiol.* **119** 1778–81
[6] Sanzleon P, Stuart A K, Woodman M M, Domide L, Mersmann J, Anthony R M and Jirsa V 2013 The virtual brain: a simulator of primate brain network dynamics *Front. Neuroinform.* **7** 10
[7] Schirner M, Rothmeier S, Jirsa V K, Mcintosh A R and Ritter P 2015 An automated pipeline for constructing personalized virtual brains from multimodal neuroimaging data *NeuroImage* **117** 343–57
[8] Breakspear M 2017 Dynamic models of large-scale brain activity *Nat. Neurosci.* **20** 340–52
[9] Singh M F, Braver T S, Cole M W and Ching S 2020 Estimation and validation of individualized dynamic brain models with resting state fMRI *NeuroImage* **221** 117046
[10] Singh M F, Braver T S, Cole M W and Ching S 2025 Precision data-driven modeling of cortical dynamics reveals person-specific mechanisms underpinning brain electrophysiology *PNAS* **122** 1–12
[11] Nir Y and de Lecea L 2023 Sleep and vigilance states: embracing spatiotemporal dynamics *Neuron* **111** 1998–2011
[12] Grigg O and Grady C L 2010 Task-related effects on the temporal and spatial dynamics of resting-state functional connectivity in the default network *PLoS One* **5** e13311
[13] Kringelbach M L, Perl Y S, Tagliazucchi E and Deco G 2023 Toward naturalistic neuroscience: mechanisms underlying the flattening of brain hierarchy in movie-watching compared to rest and task *Sci. Adv.* **9** eade6049
[14] Manuca R, Casdagli M C and Savit R S 1998 Nonstationarity in epileptic EEG and implications for neural dynamics *Math. Biosci.* **147** 1–22
[15] Tyrcha J, Roudi Y, Marsili M and Hertz J 2013 The effect of nonstationarity on models inferred from neural data *J. Stat. Mech.* **2013** 03005
[16] Glaser J I, Whiteway M, Cunningham J P, Paninski L and Linderman S W 2020 Recurrent switching dynamical systems models for multiple interacting neural populations *Advances in Neural Information Processing Systems* vol 2020-Decem pp 1–12
[17] Amber H, Zoltowski D, Nair A, Anderson D, Duncker L and Linderman S 2024 Modeling latent neural dynamics with Gaussian process switching linear dynamical systems 1–31 (arXiv: 2408.03330)
[18] Song C Y, Hsieh H L, Pesaran B and Shanechi M M 2022 Modeling and inference methods for switching regime-dependent dynamical systems with multiscale neural observations *J. Neural Eng.* **19** 066019
[19] Song C Y and Shanechi M M 2023 Unsupervised learning of stationary and switching dynamical system models from Poisson observations *J. Neural Eng.* **20** 066029
[20] Weng G, Clark K, Akbarian A, Noudoost B and Nategh N 2024 Time-varying generalized linear models: characterizing and decoding neuronal dynamics in higher visual areas *Front. Comput. Neurosci.* **18** 1–18
[21] Lawrence R R 1989 A tutorial on hidden Markov models and selected applications in speech recognition *Proc. IEEE* **77** 257–86
[22] Mingjian H, Das P, Hotan G and Purdon P L 2023 Switching state-space modeling of neural signal dynamics *PLoS Comput. Biol.* **19** e1011395
[23] Li C, Kim S H, Rodgers C, Choi H and Anqi W 2024 One-hot generalized linear model for switching brain state discovery *12th Int. Conf. on Learning Representations, ICLR 2024* pp 1–16
[24] Karniol-Tambour O, Zoltowski D M, Diamanti E M, Pinto L, Brody C D, Tank D W and Pillow J W 2024 Modeling state-dependent communication between brain regions with switching nonlinear dynamical systems *12th Int. Conf. on Learning Representations, ICLR 2024* pp 1–21
[25] Zhang Y and Saxena S 2024 Inference of neural dynamics using switching recurrent neural networks *Advances in Neural Information Processing Systems* vol 37
[26] Kätzel D, Zemelman B V, Buetfering C, Wölfel M and Miesenböck G 2011 The columnar and laminar organization of inhibitory connections to neocortical excitatory cells *Nat. Neurosci.* **14** 100–9
[27] Buzsáki G, Anastassiou C A and Koch C 2012 The origin of extracellular fields and currents-EEG, ECoG, LFP and spikes *Nat. Rev. Neurosci.* **13** 407–20
[28] Schwamb A, Singh M, Guerriero R and Ching S N 2024 Data-driven modeling of neural dynamics from EEG to track physiological changes *Proc. Annual Int. Conf. IEEE Engineering in Medicine and Biology Society, EMBS* pp 1–4
[29] Kalman R E 1960 A new approach to linear filtering and prediction problems *J. Basic Eng.* **82** 35–45
[30] Julier S J and Uhlmann J K 1997 New extension of the Kalman filter to nonlinear systems *Signal Processing, Sensor Fusion and Target Recognition VI* **3068** 182
[31] Marder E and Thirumalai V 2002 Cellular, synaptic and network effects of neuromodulation *Neural Netw.* **15** 479–93
[32] Dozat T 2016 Incorporating Nesterov Momentum into Adam *ICLR Workshop* pp 2013–6
[33] Ansel J *et al* 2024 PyTorch 2: faster machine learning through dynamic Python bytecode transformation and graph

compilation *Proc. of the 29th ACM Int. Conf. on Architectural Support for Programming Languages and Operating Systems* vol 2 (Association for Computing Machinery) pp 929–47
[34] Chennu S, O'Connor S, Adapa R, Menon D K and Bekinschtein T A 2016 Brain connectivity dissociates responsiveness from drug exposure during propofol-induced transitions of consciousness *PLoS Comput. Biol.* **12** 1–17
[35] Chennu S, O'Connor S, Adapa R, Menon D K and Bekinschtein T A 2015 Research data supporting 'brain connectivity during propofol sedation' (available at: www.repository.cam.ac.uk/handle/1810/252736)
[36] Nyquist H 1928 Certain topics in telegraph transmission theory *Trans. Am. Inst. Electr. Eng.* **47** 617–44
[37] Trapani G, Altomare C, Sanna E, Biggio G and Liso G 2000 Propofol in anesthesia. Mechanism of action, structure-activity relationships and drug delivery *Curr. Med. Chem.* **7** 249–71
[38] Vanlersberghe C and Camu F 2008 *Propofol* (Springer) pp 227–52 (available at: https://doi.org/10.1007/978-3-540-74806-9_11)
[39] Murphy M *et al* 2011 Propofol anesthesia and sleep: a high-density EEG study *Sleep* **34** 283–91
[40] Weiner V S, Zhou D W, Kahali P, Stephen E P, Peterfreund R A, Aglio L S, Szabo M D, Eskandar E N, Salazar-Gomez A F and Sampson A L *et al* 2023 Propofol disrupts alpha dynamics in functionally distinct thalamocortical networks during loss of consciousness *Proc. Natl Acad. Sci.* **120** e2207831120
[41] David O, Guillemain I, Saillet S, Reyt S, Deransart C, Segebarth C and Depaulis A 2008 Identifying neural drivers with functional MRI: an electrophysiological validation *PLoS Biol.* **6** e315
[42] Pedro A V-S, Roebroeck A, Daunizeau J and Friston K 2011 Effective connectivity: influence, causality and biophysical modeling *NeuroImage* **58** 339–61
[43] Purdon P L *et al* 2013 Electroencephalogram signatures of loss and recovery of consciousness from propofol *Proc. Natl Acad. Sci. USA* **110** E1142–51
[44] Vijayan S, Ching S, Purdon P L, Brown E N and Kopell N J 2013 Thalamocortical mechanisms for the anteriorization of alpha rhythms during propofol-induced unconsciousness *J. Neurosci.* **33** 11070–5
[45] Ching S, Cimenser A, Purdon P L, Brown E N and Kopell N J 2010 Thalamocortical model for a propofol-induced α-rhythm associated with loss of consciousness *Proc. Natl Acad. Sci.* **107** 22665–70
[46] Wertz S, Vandaele A and Gillis N 2025 Efficient algorithms for the hadamard decomposition (arXiv:2504.13633)
[47] Ciaperoni M, Gionis A and Mannila H 2024 The Hadamard decomposition problem *Data Min. Knowl. Discov* **38** 2306–47
[48] Fellous J-M and Linster C 1998 Computational models of neuromodulation *Neural Comput.* **10** 771–805
[49] Goldman M S, Golowasch J, Marder E and Abbott L F 2001 Global structure, robustness and modulation of neuronal models *J. Neurosci.* **21** 5229–38
[50] Marder E 2011 Variability, compensation and modulation in neurons and circuits *Proc. Natl Acad. Sci.* **108** 15542–8
[51] Jacquerie K, Minne C and Drion G 2022 Neuromodulation of excitability and synaptic plasticity: an underestimated challenge for computational models *Annual Meeting of Belgian Society for Neuroscience (BSN 2022)*
[52] Jones S R, Pritchett D L, Sikora M A, Stufflebeam S M, Hämäläinen M and Moore C I 2009 Quantitative analysis and biophysically realistic neural modeling of the MEG mu rhythm: rhythmogenesis and modulation of sensory-evoked responses *J. Neurophysiol.* **102** 3554–72
[53] Shine J M, Müller E J, Munn B, Cabral J, Moran R J and Breakspear M 2021 Computational models link cellular mechanisms of neuromodulation to large-scale neural dynamics *Nat. Neurosci.* **24** 765–76