2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) | 979-8-3503-7149-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/EMBC53108.2024.10781777

Data-driven modeling of neural dynamics from EEG to track physiological changes

Addison Schwamb[1], Matthew Singh[1], Réjean Guerriero[2], and ShiNung Ching[1]

***Abstract*—A key challenge in studying brain function is gaining insight into the mechanisms which drive neural activity. In this paper, we seek to address this challenge by developing a framework for generative, individualized models based on EEG data which can give insight into the neural functions which drive observed electrophysiological activity. The models created using this framework are accurate, reliable, individualized, and capable of tracking changes in neural activity during the physiological changes occurring during cardiac arrest. Due to the biophysical significance of the model structure, we can gain insight into the mechanisms driving these changes in neural activity, e.g., lowered excitatory inputs across the brain.**

***Index Terms*—neural dynamics, modeling, EEG**

I. Introduction

A key aspect of understanding brain function involves clarifying the generative mechanisms that give rise to neuronal firing and brain electrical activity more generally. A typical paradigm in this regard will involve the analysis of recorded brain activity (e.g., EEG) in order to illuminate embedded structure within these signals using tools from statistical signal processing, e.g., power spectral density estimation [1]–[3]. However, these descriptive approaches fall short of providing a direct generative mechanism that gives rise to observed brain activity. On the other hand, dynamical systems modeling approaches can be used to posit the biophysical mechanisms within brain circuits that can generate overt measurements such as EEG [4], [5]. Such models provide mechanistic insight by providing the biological interactions (e.g., interplay of excitatory and inhibitory neurons) and dynamics underlying the signals which are being generated. While powerful, this paradigm brings with it several challenges. First, such models are typically constructed in a bottom-up fashion, relying on low-level characterizations of neuroanatomy and biophysics which may or may not be known *a priori*. Furthermore, the large number of parameters in these models makes them difficult to deploy at large (whole-brain) scales and, especially, to account for individual variation in brain electrophysiology.

Recently, effort has been directed at combining these approaches toward the creation of data-driven generative models that can be constructed directly from measurements of brain activity such as EEG, MEG or fMRI [6]–[9]. This work can enable the creation of generative models on an individual basis, where each model's parameters reflect subject-specific brain architecture and dynamics. The overall goals of this paper are to: (i) adapt and test a framework for building subject-specific dynamical models of whole-brain activity from EEG recordings, and (ii) demonstrate the methodology by tracking underlying changes in physiology in a clinical context, specifically the occurance of cardiac arrest.

A. Data-driven model identification

The current paper builds on our recent development of the Mesoscopic Individualized Neural Dynamics (MINDy) modeling framework, originally developed to model functional neuroimaging data [9] and later for MEG data [10]. The major technical bottleneck addressed in this framework pertains to the so-called 'dual estimation' problem of system identification. In the current context, EEG provides an indirect 'output' measurement emanating from the underlying neural populations we seek to model. Hence, the modeling methodology must simultaneously estimate the state of these populations from the measured signal, and detect the parameters of the model which dictates the dynamics of said state. This is the aforementioned 'dual estimation problem' (identifying states and parameters at the same time), and in the current work we will adapt the techniques of [10] into the EEG domain.

B. Tracking latent physiological changes

Dynamic models of latent neural populations can provide insights into how brain function adapts to different physiologic states, including pathology. By capturing neural data during multiple physiologic states and creating models from each of these states, we can interrogate changes that occur. Such changes may occur in the model parameters (e.g., in the relative strength of interactions between brain regions), or in the corresponding model dynamics (i.e., how those parameters impact the generation of brain activity). A premise of the current work is that data-driven modeling can identify changes associated with patient physiology in clinical settings. We will test this premise by applying the methodology to data from patients who experience cardiac arrest – a significant physiological event – while on EEG recording.

The rest of the paper is organized as follows: In section II, we present the methodology used to create individualized generative models from EEG data. In section III we analyze the performance of this methodology. In section IV, we discuss the EEG data taken from children experiencing a cardiac arrest event, the models obtained from these patients over the course

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

[1]Addison Schwamb, Matthew Singh, and ShiNung Ching are with the Department of Electrical & Systems Engineering, 1 Brookings Dr., Washington University in St. Louis, Saint Louis, MO 63130, USA a.l.schwamb@wustl.edu, f.singh@wustl.edu shinung@wustl.edu

[2]Rejean Guerriero is with the Department of Neurology, 660 S. Euclid Ave., Washington University School of Medicine, Saint Louis, MO 63110, USA r.guerriero@wustl.edu

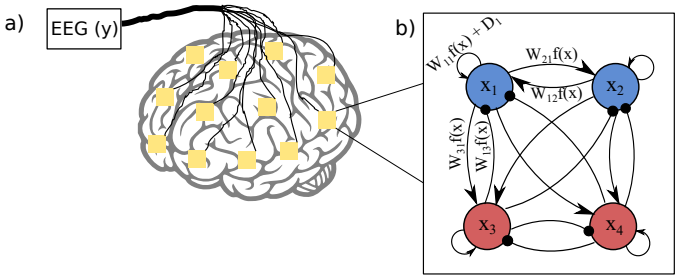

Fig. 1. Modeling approach. EEG signal is measured at several spatial locations in the brain (a). We model each of these spatial locations as having two excitatory and two inhibitory underlying neural populations (b) which interact according to eq.1.

of their arrests, and the possibilities for gaining mechanistic insight into this process through these models. Finally, in section V we give conclusions and discuss future work.

II. Methods

A. Model structure

We seek a general dynamical brain network model with biologically interpretable parameters that can be fit to each individual's EEG data (Fig. 1). In this regard, we consider a general formulation of a mean-field neural mass type model:

$$\dot{\mathbf{x}} = \mathbf{W}f(\mathbf{x}) + \mathbf{D}\mathbf{x} + \mathbf{C}, \tag{1}$$

$$f(\mathbf{x}) = \tanh(\mathbf{S}\mathbf{x} + \mathbf{V}), \tag{2}$$

Here, $\mathbf{x}$ is the firing rate of each neural population. We choose the number of populations as two excitatory populations and two inhibitory populations for each spatial location (as defined by the location of each EEG lead – see Fig. 1). This firing rate is linearly mapped into EEG signal space via a lead field matrix $\mathbf{H}$:

$$\mathbf{y} = \mathbf{H}\mathbf{x}, \tag{3}$$

where $\mathbf{y}$ is the modeled EEG signal and $\mathbf{H}$ is specified *a priori* based on the geometry of the head and a restriction that excitatory populations primarily contribute to the measure EEG signal. The matrices $\mathbf{W}$, $\mathbf{S}$, $\mathbf{V}$, $\mathbf{D}$, and $\mathbf{C}$ parameterize the dynamics of the latent neural populations' activation. Here, $\mathbf{W}$ is the neural connectivity matrix, and is constrained such that excitatory neurons may have long and short range connections, but inhibitory neurons may have only local, short range connections. $\mathbf{D}$ is a diagonal matrix representing the decay (leak) in neural activity for each population. $\mathbf{S}$ and $\mathbf{V}$ parameterize the slope and offset of the sigmoidal activation function, and $\mathbf{C}$ represents the baseline bias of each neural population.

B. Model intepretation

Note that this model is mathematically comparable to a vanilla type recurrent neural network, but arranged and contrained to reflect biologically interpretable relationships between excitatory and inhibitory neural populations. Indeed, a key feature of adopting this model structure is the biological significance of each of the parameters. By assigning neural populations to spatial locations in the brain and constraining the parameters, we can examine changes in these parameters associated with physiological changes. For example, if the long-range connection weights in $\mathbf{W}$ increase relative to the short-range connection weights, we can infer that there is an increase in inter-region activity across the brain.

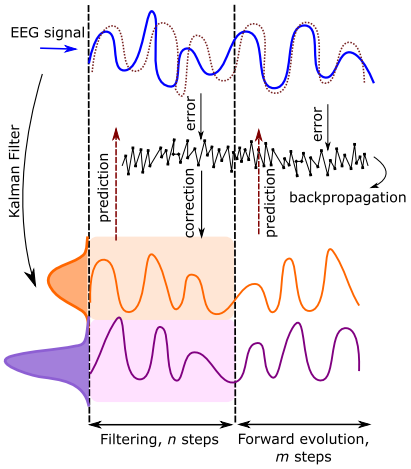

Fig. 2. Backpropagated Kalman filtering approach. The state is approximated using an n-step Kalman filter, and then the model is free-simulated for m steps. Error collected at each step of the process is then backpropagated in order to update both the model parameters and the noise covariances, and the entire process is repeated.

C. Fitting procedure

To fit these parameters, we adopt the backpropagated Kalman filtering (BKF) approach that we previously developed to fit MINDy models to MEG data [10], [11] and adapt it to EEG signals. The Kalman filter is a classical solution for extracting latent states from noisy data under a known model and characteristics of noise [12]. In our case, however, we are solving a dual estimation problem of learning both the state (i.e., $\mathbf{x}$) and the dynamics (i.e., $\mathbf{W}$, etc.). Because of this, we adopt an iterative approach: first apply a Kalman filter to a small window of data in order to approximate the state. We then evolve the model forward from this approximated state. By backpropagating the error through the free simulation steps and the Kalman filtering steps, the error in the model parameters can be separated from the error in the noise covariances. The model and the noise covariance are then updated, and the steps repeated with a new small window in the epoch of fitting data (see Fig. 2). 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).

This fitting method is agnostic with regard to data modality, but some adjustments were needed in order to adapt it from working with MEG to EEG data. Many of these adjustments were simply to hyperparameters in the fitting method, e.g., the number of Kalman filtering and free simulation steps in each fit, or the parameters of the NADAM optimizer used to adjust the parameters based on the backpropagated error. A more significant challenge, however, is in ensuring the model can capture the low frequency signals we observed in the raw EEG data. To solve this, we first lowpass filtered our data to 30Hz and below so that the model was only fitting on frequencies that are physiologically meaningful in the context of EEG. This prevented the model from fitting to high frequency signals that may be arising from artifacts. Additionally, we fit our models on a much longer (multiple seconds long) epoch of data, in order to capture low frequency dynamics which would not be evident in an epoch of less than one second.

D. EEG data and patient cohort

To examine the changing dynamics due to cardiac arrest, we examined EEG recordings from 8 patients who experienced a cardiac arrest event while inpatients at St. Louis Children's Hospital, for whom we had both EEG and ECG recordings. These patients were 50% (n = 4) male, and 2 were under two years old. The remaining 6 were 7 years old or older. 37.5% (n = 3) of patients survived their cardiac arrest.

III. Model Performance

A. Modeled timeseries are predictive of frequency content and spatial distribution

Once the parameters have been fit to the data, we have a generative model that captures the dynamics of the brain activity during the epoch it is trained on. The model can then be excited by noise and evolved forward to produce a time signal with the same statistical properties as the original measured signal. By transforming the measured and predicted signals into the frequency domain, we observe that their frequency content is quite similar (Fig. 3). This is encouraging since the model has not been directly trained to match the data in the frequency domain, but rather, these similarities emerge from fitting the model parameters using one-step-ahead prediction in the time domain. This verifies that the model can recapitulate the frequency content of the EEG recordings from which it is fit.

Additionally, these models replicate spatial distribution of EEG activity. In a comparison of the global coherence (a measure of spatial frequency correlations [13]), the global coherence at each spatial location calculated from the model's predictions closely matched the global coherence from the measured data (e.g., Fig. 4).

B. Models are reliable and individualized

To test reliability and resistance to overfitting in our modeling paradigm, we conducted a test-retest analysis. Models were fit on two adjacent 3-min epochs of data at the beginning of recordings, and the correlation of the connection weight matrices W was calculated. Additionally, we calculated the correlation of the measured power spectral densities from each epoch to provide a baseline of the reliability of the data itself. We also calculated the cross-subject test-retest correlation, in order to compare the individualization of each model (Fig. 5). While not as reliable as the PSD itself, models achieve reliability around 0.65, and are highly individualized since across subject similarity is around 0.5.

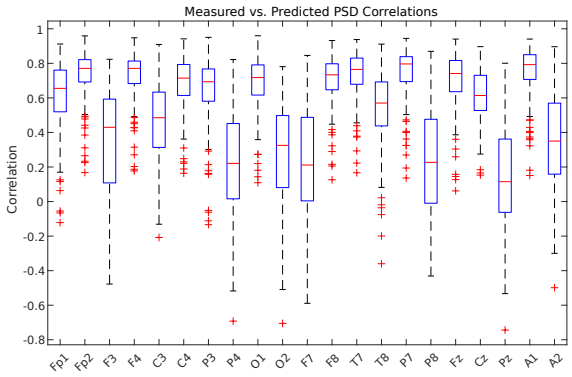

Fig. 3. Correlations of measured and predicted power spectral densities, organized by EEG channel location.

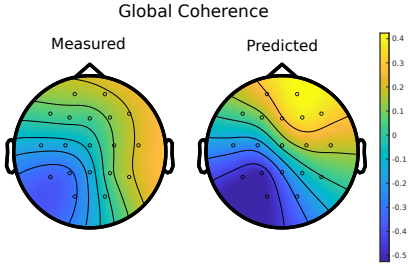

Fig. 4. Global coherence analysis of the measured and predicted EEG signals. The predicted EEG signals preserve the spatial structures of the measured signals.

IV. Potential for Insights into Cardiac Arrest

To understand the capability of this approach to capture changes in the latent neural populations when subjects ex-

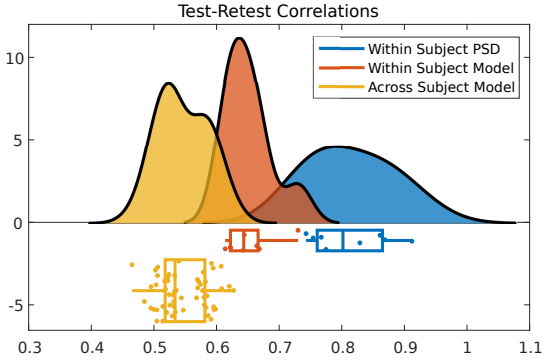

Fig. 5. Test-retest analysis correlation distributions.

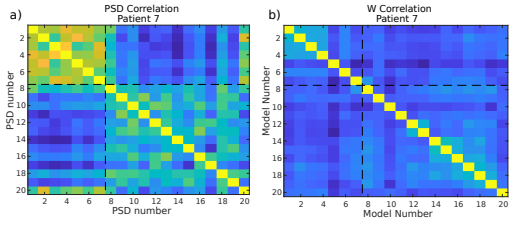

Fig. 6. a) Correlation of power spectral densities in 3-min epochs surrounding cardiac arrest. Black dashed lines represent time of arrest b) Correlation of connection weight matrices in the same epochs. Black dashed lines again represent time of arrest.

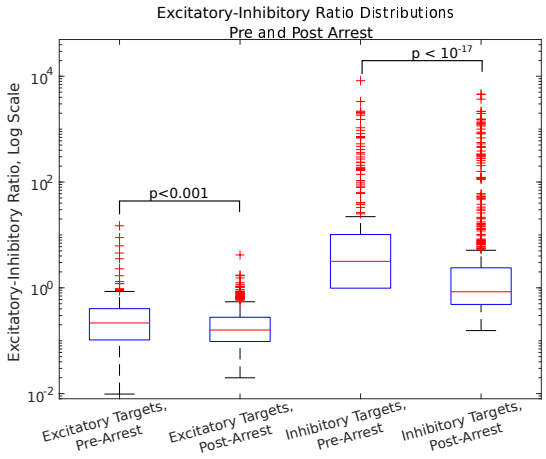

Fig. 7. Distributions of excitatory-inhibitory ratios for excitatory and inhibitory targets, before and after arrest.

perienced cardiac arrest, we analyzed and fit models on 3-min nonoverlapping epochs of data beginning 20min prior to cardiac arrest (defined as time of CPR) and ending either 40min after the arrest or at the end of the recording. This resulted in 20 models per subject, with arrest occurring during model 7. There is a clear change in the EEG data at the time our subjects experienced cardiac arrest. This is born out in the correlations of both the power spectral densities and the connectivity weights of the model from epoch to epoch (e.g., exemplar patient in Fig. 6), where there is a clear block structure dividing pre-arrest epochs and post-arrest epochs.

The ability of these models to capture changes which occur in the brain during cardiac arrest gives us significant leverage to understand how brain physiology may react during this severe event. While we used the models' ability to generate timeseries to validate our method, their key advantages lie in the biophysical interpretability of their parameters and in their ability to capture the dynamics of latent neural populations. In this patient, there was a significant decrease in the excitatory-inhibitory ratio of connection weights (Fig. 7), indicating a general decrease of excitatory connection strength relative to inhibitory connection strength, particularly for inhibitory targets.

V. Conclusion

We have extended the MINDy modeling approach [9], [10] to EEG data, and validated it using EEG recordings from cardiac arrest patients. The models we fit capture frequency characteristics of the patients' EEG activity, and are both reliable and individualized. Most promising, however, is their ability to capture physiological changes. In the future, we will conduct a more thorough analysis of how both the model parameters and the latent neural dynamics change when the arrest occurs and as patients recover. To further extend this work, we plan to integrate multimodal data with which EEG can be temporally aligned, such as pulse, respiratory rate, and cerebral oxygenation. By gaining access to the dynamics of neural populations underlying electrophysiology, these modeling methods may provide important leverage on tracking neurophysiological changes associated with injury and hence patient care.

References

[1] C. Bentes, H. Martins, A. R. Peralta, C. Morgado, C. Casimiro, A. C. Franco, A. C. Fonseca, R. Geraldes, P. Canhão, T. Pinho e Melo, T. Paiva, and J. M. Ferro, "Early EEG predicts poststroke epilepsy," *Epilepsia Open*, vol. 3, no. 2, pp. 203–212, 2018.

[2] A. D. Engell, N. Y. Kim, and G. Mccarthy, "Sensitivity to Faces with Typical and Atypical Part Configurations within Regions of the Face-processing Network : An fMRI Study," *J Cogn Neurosci*, no. 2012, pp. 963–972, 2018.

[3] R. L. Buckner, F. M. Krienen, and B. T. Yeo, "Opportunities and limitations of intrinsic functional connectivity MRI," *Nat. Neurosci.*, vol. 16, no. 7, pp. 832–837, jul 2013.

[4] H. Okano, E. Sasaki, T. Yamamori, A. Iriki, T. Shimogori, Y. Yamaguchi, K. Kasai, and A. Miyawaki, "Brain/MINDS: A Japanese National Brain Project for Marmoset Neuroscience," *Neuron*, vol. 92, pp. 582–590, 2016.

[5] H. Markram, "The Blue Brain Project," *Nat. Rev. Neurosci.*, vol. 7, pp. 153–160, 2006. [Online]. Available: www.nature.com/reviews/neuro

[6] M. Breakspear, "Dynamic models of large-scale brain activity," *Nat. Neurosci.*, vol. 20, no. 3, pp. 340–352, 2017.

[7] P. Sanzleon, S. A. Knock, M. M. Woodman, L. Domide, J. Mersmann, A. R. Mcintosh, and V. Jirsa, "The virtual brain: A simulator of primate brain network dynamics," *Frontiers in Neuroinformatics*, vol. 7, no. MAY, jun 2013.

[8] M. Schirner, S. Rothmeier, V. K. Jirsa, A. R. Mcintosh, and P. Ritter, "An automated pipeline for constructing personalized virtual brains from multimodal neuroimaging data," *NeuroImage*, vol. 117, pp. 343–357, 2015.

[9] M. F. Singh, T. S. Braver, M. W. Cole, and S. Ching, "Estimation and validation of individualized dynamic brain models with resting state fMRI," *NeuroImage*, vol. 221, 2020.

[10] M. F. Singh, M. Wang, M. W. Cole, and S. N. Ching, "Efficient identification for modeling high-dimensional brain dynamics," *2022 American Control Conference (ACC)*, vol. 2022-June, pp. 1353–1358, 2022.

[11] M. F. Singh, T. S. Braver, M. W. Cole, and S. Ching, "Precision data-driven modeling of cortical dynamics reveals idiosyncratic mechanisms underlying canonical oscillations," *bioRxiv*, pp. 2023–11, 2023.

[12] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. Series D, pp. 35–45, 1960.

[13] A. Cimenser, P. L. Purdon, E. T. Pierce, J. L. Walsh, A. F. Salazar-Gomez, P. G. Harrell, C. Tavares-Stoeckel, K. Habeeb, and E. N. Brown, "Tracking brain states under general anesthesia by using global coherence analysis," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 108, no. 21, pp. 8832–8837, may 2011.