METHODS

Explicit Modeling of Brain State Duration Using Hidden Semi Markov Models in EEG Data

NELSON J. TRUJILLO-BARRETO[1], **DAVID ARAYA GALVEZ**[2,3],
ALAND ASTUDILLO[4], **AND WAEL EL-DEREDY**[3,5,6]

[1]Faculty of Biology, Medicine and Health, School of Health Sciences, The University of Manchester, M13 9PL Manchester, U.K.
[2]Facultad de Ingeniería, Instituto de Tecnología para la Innovación en Salud y Bienestar (ITISB), Universidad Andrés Bello, Viña del Mar 2531015, Chile
[3]Centro de Investigación y Desarrollo en Ingeniería en Salud, Universidad de Valparaíso, Valparaíso 2362905, Chile
[4]NICM Health Research Institute, Western Sydney University, Penrith, NSW 2751, Australia
[5]Department of Electronic Engineering, School of Engineering, Universitat de València, 46010 València, Spain
[6]ValgrAI: Valencian Graduate School and Research Network of Artificial Intelligence, 46022 València, Spain

Corresponding author: Nelson J. Trujillo-Barreto (nelson.trujillo-barreto@manchester.ac.uk)

ABSTRACT We consider the detection and characterization of brain state transitions based on ongoing electroencephalography (EEG). Here, a brain state represents a specific brain dynamical regime or mode of operation that produces a characteristic quasi-stable pattern of activity at the topography, sources, or network levels. These states and their transitions over time can reflect fundamental computational properties of the brain, shaping human behavior and brain function. The hidden Markov model (HMM) has emerged as a useful tool for uncovering the hidden dynamics of brain state transitions based on observed data. However, the limitations of the Geometric distribution of states' durations (dwell times) implicit in the standard HMM, make it sub-optimal for modeling brain states in EEG. We propose using hidden semi Markov models (HSMM), a generalization of HMM that allows modeling the brain states duration distributions explicitly. We present a Bayesian formulation of HSMM and use the variational Bayes framework to efficiently estimate the HSMM parameters, the number of brain states, and select among candidate brain state duration distributions. We assess HSMM performance against HMM on simulated data and demonstrate that the accurate modeling of state durations is paramount for making reliable inference when the task at hand requires accurate model predictions. Finally, we use actual resting-state EEG data to illustrate the benefits of the approach in practice. We demonstrate that the possibility of modeling brain state durations explicitly provides a new way for investigating the nature of the neural dynamics that generated the EEG data.

INDEX TERMS Brain state, hidden semi Markov model, state duration, EEG.

I. INTRODUCTION

There is great current interest in the identification and characterization of transiently stable and recurrent patterns of activity in resting-state EEG at topography [1], [2]

The associate editor coordinating the review of this manuscript and approving it for publication was Humaira Nisar.

sources [3], [4], or functional connectivity levels [5], [6]. These switching patterns are often interpreted as traces of hidden *brain states* or modes of operation with functional and clinical relevance [2], [7], [8]. Although the relationship between such "phenomenological" description and the causal mechanisms of brain dynamics is not clear, the described brain state switching might still reflect fundamental

computational properties of the brain, which shape human behavior and brain function [9], [10]. Of particular interest is the duration of the brain states (dwell-time of each brain state). A heavy-tailed duration distribution of transient interhemispheric synchronization patterns can be a sign of metastable brain dynamics [10], and longer brain state durations are found in patients of Lewy body dementia compared to Alzheimer's disease and healthy controls [11]. However, current methods of inferring brain state transitions (i.e. dynamic brain state allocation [4]) in EEG are not ideally suited to model the brain state duration. This study advances the state of the art by proposing a new method that overcomes this limitation.

Current brain state allocation methods are dominated by descriptive rather than explanatory methods, including temporal sliding windows [12], [13] and adaptive segmentations using clustering [14], [15], [16], among others [8], [17]. The shortcomings of these methods are well established [17], [18] and stem from the fact that the properties of brain state dynamics are not considered during brain state allocation, but are assessed post-hoc. On the contrary, explanatory methods go about brain state allocation by inverting an assumed model of the generative process by which the hidden brain state dynamics gives rise to data [5]. The switching and recurrent behavior of brain state dynamics suggests using a Hidden Markov Model (HMM) to represent the generative process. Although HMM has proved to be an effective generative model for EEG brain state allocation [3], [4], [6], [10], [19], it still imposes restrictions on the type of brain state dynamics that can be modeled. In brief, the HMM generates data sequentially by using a Markov chain to model the hidden brain state transitions, while one observation is emitted by the active brain state at each time point. The Markov chain however, satisfies the first-order Markov property, which requires that, given the current brain state, the probability of a future brain state is independent of the past history of the chain. This induces a memoryless Geometric distribution (discretized exponential) of brain state durations, so that the probability of staying in the same brain state for a fix duration does not depend on the time already spent in the brain state. Moreover, the Geometric distribution places higher probability on shorter brain state durations [20], [21]. Therefore, the dynamics prescribed by a HMM are characterized by short-term correlations (short memory) and fast brain state switching, which is not consistent with observed features of resting-state EEG data.

Recent findings suggest that resting-state EEG is a lot more complex typically showing multiscale dynamics. On one hand, using HMM and microstates analysis [22] to study power amplitude fluctuations and time-dependent functional connectivity, has revealed brain state switching in the 50-100 ms time-scale [2], [3], [4], [5], [15], [19], [23]. On the other hand, haemodynamic correlates of this fast dynamics show spatial patterns that resemble the resting-state networks observed in functional Magnetic Resonance Imaging (fMRI),

which vary in the seconds scale [1], [8], [24]. The link between such disparate time-scales is hypothesized to be mediated by underlying mechanisms of dynamic instability (see [25] for a review), such as multistability [26], metastability [10] and criticality [27], which endow brain state dynamics with long-term dependencies (memory) and scale-invariance across several timescales [27], [28], [29], [30]. For example, [30] found signs of monofractality (scale-invariance) over 6 dyadic time-scales (256ms-10s range) in EEG microstate sequences. Notably, the degree of long-range dependency was destroyed when microstate durations were equalized, suggesting that important information about the nonlinear mechanisms underlying resting-state EEG is carried in the brain state switching times. The complex dynamics described above can induce brain state duration distributions which are far from the memoryless Geometric distribution predicted by HMM.

Realistic computational simulations of large-scale spontaneous brain activity have demonstrated the emergence of metastable traveling patterns of activity with heavy-tailed duration distributions [10]. This result was in broad agreement with the duration distribution of phenomenological brain state obtained from resting-state MEG data [6]. More generally, heavy-tailed and power-law distributions also occur in human behavioral data, such as the duration of percepts in perceptual multi-stability [31], [32], [33] and the duration of active periods during human activity [34]. These distributions are consistent with persistent (long-lived) brain states ("trapping" effect) and the existence of long-term dependencies in the data. Therefore, while standard HMM seems like a natural fit to the dynamic brain state allocation, it inadequately models the temporal properties of the resting-state EEG and can potentially bias the inference towards unrealistically fast brain state switching.

The present paper proposes to extend HMM into a more flexible model called Hidden Semi Markov Model (HSMM) [35]. HSMM is a generalization of HMM where the Markov assumption is relaxed to allow for explicit modeling of the brain state duration distribution. By choosing an appropriate duration distribution HSMM can naturally encapsulate prior beliefs that brain states are persistent, while at the same time allowing brain state switching and recurrence. This flexibility comes at the cost of an increased model complexity. We will show however that in many situations, the added complexity is compensated for by the explanatory power of the HSMM model. Using simulations, we investigate the sort of problems the HMM assumptions can have a negative impact on, and demonstrate how HSMM resolves these issues. We use a variational Bayes formulation of HSMM [36], which allows for coherent and efficient estimation of the parameters of the model, the hidden brain state sequence and its dimensionality (number of brain state modes). The proposed Bayesian framework also allows for selection of the appropriate brain state duration

distribution based on data by means of formal Bayesian model comparison. This is demonstrated in practice using actual resting-state EEG data from 5 subjects to arbitrate between different brain state duration distributions to model the data. A MATLAB (Mathworks Ⓡ) implementation of our Brain State Dynamics (BSD) toolbox is also available (https://github.com/daraya78/BSD)

The paper is organized as follows. Sections II-A and II-B describe the technical aspects of HSMM highlighting its main differences with respect to the standard HMM. Sections II-C and II-D summarize the variational Bayes framework used to infer the parameters of the models. Sections II-E and II-F present the computational simulations and the performance evaluation metrics used to compare HSMM vs HMM. Sections II-G and II-H describe the EEG data used in this paper, its preprocessing and the methodology used for its analysis with HSMM. Finally, sections III and IV describe respectively the results obtained, and a discussion of the main findings in relation to the existing literature, as well as the limitations of the work and future research directions

II. METHODS

The proposed method is motivated by hybrid dynamical systems' (HDS) theory for multivariate time series analysis [36] applied to EEG. The principles of HSMM are summarized and interested readers are referred to Appendices and previous descriptions of these models for additional mathematical and algorithmic details.

A. HIDDEN SEMI MARKOV MODEL AS A GENERALIZATION OF HIDDEN MARKOV MODELS

HSMM and HMM are successful models of HDS. A HDS can be thought of as a continuous system with some hidden discrete logic so that it produces continuous observations while its hidden dynamics can rapidly transition or switch among a discrete set of states or modes of operation. HSMM and HMM both use two stochastic processes to capture such dual behavior, which is shown in their Dynamic Bayesian Network (DBN) representations in figure 1 (a, b). The underlying discrete process generates the hidden states of the system. This influences the continuous process that generates the observed data or emissions [21], [37].

However, HSMM and HMM differ in the way the state duration is considered. In HMM, the hidden process is a Markov chain where the state emits one observation at a time (figure 1a). In this case the state duration is implicitly captured as consecutive transitions to the same state, which induce a Geometric distribution over the state duration [21]. In HSMM, the hidden process is a semi-Markov chain where the state is a "super-state" that emits a segment of observations of random length (figure 1b). The segment's length determines the state's duration explicitly and is assumed to be sampled from a duration distribution that can be of arbitrary form. In order to uniquely define the state duration, self-transitions between super-states are not permitted. The super-state representation does not lead to
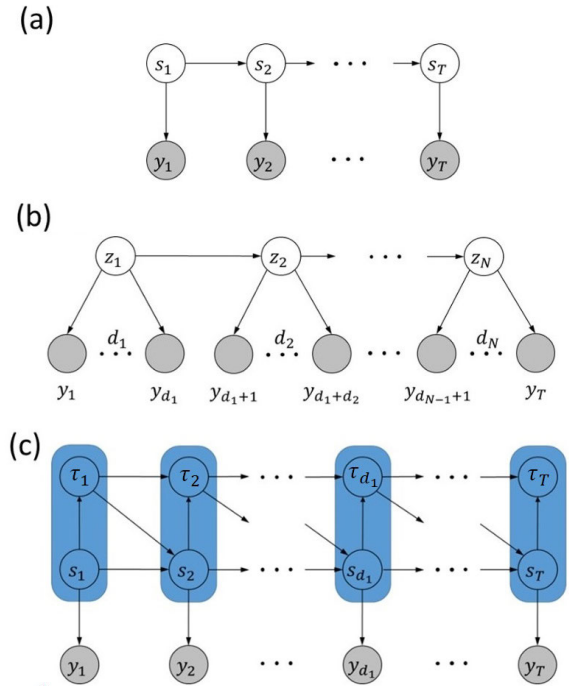

FIGURE 1. Dynamic Bayesian Network (DBN) representations of the hybrid system models used in this paper. Shaded grey circles represent observations, clear circles represent hidden state variables. Parameters have been omitted for simplicity. (a) HMM: A hidden Markov chain (s_t) emitting one observation (y_t) at each time point. (b) Super-state representation of HSMM: A hidden Markov chain on a set "super-states" (z_l) where each super-state visited emits a segment of observations of random duration d_l. (c) HMM embedding of HSMM: A hidden Markov chain on the augmented state-space of states and residual times (s_t, τ_t) emitting one observation at each time point. The augmented joint-state is highlighted in blue. When the chain first enters the state s_t at time t, the residual time τ_t is set to a duration sampled from the state's duration distribution; it then deterministically counts down to 1 at which point the state is free to transition and τ_t is reset to a duration sampled from the duration distribution of the new state.

a valid DBN, because the number of observation nodes of each segment is random and then so is the structure of the DBN (figure 1b). Therefore, standard inference methods developed for DBN are not applicable [38]. Instead, we use the equivalent representation in figure 1c, where the HSMM is embedded in a standard HMM [36], [38]. The state-space is augmented to consider the joint process of states and their residual times (time left in the state) as a single Markov chain. The HMM embedding allows the efficient inference algorithms developed for standard HMM to be used for HSMM. In this case, the data are generated in the following way. When the chain first enters a state, the residual time is set to a duration sampled from the state's duration distribution and it then deterministically counts down to 1. At this point, the chain is free to switch to a different state, and the residual time is reset to a duration sampled from the duration distribution of the new state. The state's value is copied across every time slice spent in the state to ensure a regular structure. There are several modeling approaches to semi-Markovianity [35], [39]. Our focus is on the setting where

each state is given an explicit parametric duration distribution (so-called Explicit-Duration HMM).

B. HSMM AS GENERATIVE MODEL FOR BRAIN STATE ALLOCATION

We consider the brain as a HDS that can switch between a discrete set of hidden brain states while emitting continuous EEG signals or related features. This is modeled using a HSMM (figure 1c), where $\mathbf{y}_{1:T}$ ($\mathbf{y}_t \in R^n$) denotes the n-variate time-series of EEG signals of length T at n emission channels; $\mathbf{s} = s_{1:T}$ is the sequence of hidden brain states taking values in the set $S = \{1, 2, \ldots, K\}$ of possible brain states at each time point ($s_t \in S$); and $\boldsymbol{\tau} = \tau_{1:T}$ with ($\tau_t \in D$) is the sequence of residual times taking values in the set $\Delta = \{1, 2, \ldots, D\}$ of possible brain state durations. The HSMM is then defined by four distributions: the initial, the transition, the duration and the emission distributions of the brain states; which will be specified as follows.

1) BRAIN STATE TRANSITION DISTRIBUTION

In the present setting it is assumed that: (i) the transition to a brain state is independent of the duration of the previous brain state, (ii) the duration is only conditioned on the brain state; (iii) the observation at a given time point only depends on the active brain state; and (vi) the transition probability of each segment is constant. The joint process $(s, \tau)_t$ taking values in $S \times D$, then evolves according to a stationary Markov chain with a transition kernel such that the probability of transition from brain state i with residual time d' at time t to brain state j with residual time d at time $t+1$, is

$$\Phi_{(i,d')(j,d)} \triangleq p((s, \tau)_{t+1} = (j, d)|(s, \tau)_t = (i, d'))$$

$$= \begin{cases} I(i = j)I(d = d' - 1) & \text{if } d' > 1 \overset{remain}{\underset{decrement}{}} \\ a_{ij}p_j(d) & \text{if } d' = 1 \overset{transition}{\underset{reset}{}} \end{cases}$$
(1)

where a_{ij} are the elements of the $K \times K$ brain state transition matrix $\mathbf{A}$ with diagonal elements $a_{ii} = 0$; $p_j(d) = p(d|\boldsymbol{\lambda}^{(j)})$ is the duration distribution of brain state j with parameters $\boldsymbol{\lambda}^{(j)}$; and the indicator function $I(x) = 1$ if x is true and zero otherwise. In words, equation (1) means that during a transition, the brain dynamics evolves in two stages. If the brain is in state i, and the end of the segment has not been reached ($d' > 1$), then it remains in that state and decrements by 1 its residual time ($d = d' - 1$). Instead, if the end of the segment has been reached ($d' = 1$), then the brain probabilistically transitions to a different state $j(a_{ii} = 0)$ with transition probability a_{ij} and the residual time is reset to a duration d sampled from the duration distribution $p_j(d)$ of state j. Note that the transition kernel is factorized over states and durations.

2) BRAIN STATE DURATION DISTRIBUTIONS

We model duration distributions based on Normal and Lognormal densities. Since EEG measurements are always

in discrete time (encoded as positive integers) and states' durations have a pre-defined maximum duration, the duration probabilities were then obtained by discretisation and truncation of the corresponding probability density. That is, the duration probabilities of the duration distribution with probability density function g_k of state k, is obtained as:

$$p_k(d) = \frac{\int_{u-1}^u g_k(x)dx}{\int_0^D g_k(x)dx}, \quad k = 1, \ldots, K, \quad u = 1, \ldots, D$$
(2)

Almost all probability densities can be used in this way to model the states' duration probabilities. In this paper we use a Normal density to model states' durations in the simulated data

$$g_k(x) = N(x; \nu^{(k)}, \rho^{(k)^{-1}})$$
(3)

and compare its performance to using a Lognormal density in the empirical data section

$$g_k(x) = LogN(x; \nu^{(k)}, \rho^{(k)^{-1}})$$
(4)

where $\nu^{(k)}$ and $\rho^{(k)^{-1}}$ are respectively location and dispersion parameters of the two densities for the $k - th$ brain state. Other options of duration distributions are available in our MATLAB (Mathworks ®) toolbox.

3) EMISSION/OBSERVATION DISTRIBUTIONS

Since our focus is on the modeling of brain state durations, without loss of generality we assume that the brain states are characterized by simple features where the observations emitted during the k-th brain state are independently sampled from a multivariate Normal distribution of dimension n with state-specific mean $\boldsymbol{\mu}^{(k)}$ and precision $\Sigma^{(k)}$ [1], [3], [4], [40]. The conditional probability of observing y_t given the brain was in state k at time t is then

$$p(\mathbf{y}_t|s_t = k, \boldsymbol{\theta}^{(k)}) = N_n(y_t; \boldsymbol{\mu}^{(k)}, \Sigma^{(k)^{-1}})$$
(5)

where $\boldsymbol{\theta}^{(k)} = [\boldsymbol{\mu}^{(k)}, \Sigma^{(k)}]$

4) INITIAL DISTRIBUTION AND BOUNDARY CONDITIONS

The initial brain state distribution $\pi = \{\pi_{kd}\}$ is defined so that

$$p((s, \tau)_1 = (k, d)|\boldsymbol{\pi}) = \pi_{kd}$$
(6)

and represents the probability of the brain being in state k with residual time d at the beginning of the sequence. To completely specify a HSMM, boundary conditions at the beginning and end of the observation sequence must also be defined. Commonly used simplifying conditions assume that observations start and/or end at a state segment boundary [35], [41]. Here we assume the more realistic case that observations may start and end at any time during a state segment [35], [37].

C. BAYESIAN INFERENCE OF HSMM

We followed the variational Bayes (VB) approach to HSMM in [36], and adapted it to the model specification used here. We present the main principles of VB for HSMM, and refer the reader to Appendix for mathematical details. The specification of the HSMM defines a complete likelihood function of the form

$$
\begin{aligned}
&p(\mathbf{y}_{1:T}, \mathbf{s}, \boldsymbol{\tau} | \boldsymbol{\Theta}, M) \\
&= p(s_1, \tau_1 | \boldsymbol{\pi}) p(\mathbf{y}_1 | s_1, \tau_1, \boldsymbol{\theta}^{(s_1)}) \\
&\quad \cdot \prod_{t=2}^{T} p(\mathbf{y}_t | s_t, \boldsymbol{\theta}^{(s_t)}) p(s_t | s_{t-1}, \tau_{t-1}, \mathbf{A}) p(\tau_t | s_t, \tau_{t-1}, \boldsymbol{\lambda}^{(s_t)})
\end{aligned}
\tag{7}
$$

where $\boldsymbol{\theta} = \boldsymbol{\theta}^{(1:K)}$ and $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(1:K)}$ denote the collection of parameters of the emission and duration distributions of all the states, respectively; M is the model class defined by the choice of the model type (HSMM or HMM) or by the choice of the number of brain states given a model type, and $\boldsymbol{\Theta} = \{\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\theta}, \boldsymbol{\lambda}\}$ denotes all the parameters that specify a model within the class. HMM is completely defined by the reduced set of parameters $\boldsymbol{\Theta} = \{\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\theta}\}$, because the duration distribution is not explicitly defined for that model. For notational simplicity, dependency on M in the right-hand-side of equation (7) and in the rest of the paper is implicit and only used when strictly necessary.

The Bayesian approach requires defining prior distributions on all the parameters. After combination with the likelihood via the Bayes rule, the joint posterior distribution of parameters and hidden states given the data is obtained, based on which statistical inference can proceed. We assume that all parameters are independent a priori so that the posterior is:

$$
\begin{aligned}
p(\mathbf{s}, \boldsymbol{\tau}, \boldsymbol{\Theta} | \mathbf{y}_{1:T}, M) &\propto p(\mathbf{y}_{1:T}, \mathbf{s}, \boldsymbol{\tau} | \boldsymbol{\Theta}, M) \\
&\quad \cdot p(\boldsymbol{\pi}) p(\mathbf{A}) p(\boldsymbol{\lambda}) p(\boldsymbol{\theta})
\end{aligned}
\tag{8}
$$

where the prior distributions $p(\boldsymbol{\pi})$, $p(\mathbf{A})$ and $p(\boldsymbol{\theta})$ are assumed to be conjugate (see Appendix B).

Exact Bayesian inference based on (8) is not possible due to the dependence between the parameters and the hidden states. However, the augmented state representation of HSMM allows using the efficient VB algorithm developed for HMM [36], [42]. VB allows making inference on parameters, hidden states, and models by approximating the joint posterior with a simpler variational density. The usual Mean Field Approximation [43] requires the approximate posterior to factorize over subsets of parameters and states:

$$
\begin{aligned}
q(s_{1:T}, \tau_{1:T}, \boldsymbol{\Theta}) &= q(s_{1:T}, \tau_{1:T}) q(\boldsymbol{\Theta}) \\
&= q(s_{1:T}, \tau_{1:T}) q(\pi) q(\mathbf{A}) q(\boldsymbol{\lambda}) q(\boldsymbol{\theta})
\end{aligned}
\tag{9}
$$

Note that the structure of the HSMM implies an exact posterior that is already factorized over $\boldsymbol{\pi}$, $\mathbf{A}$ and $\boldsymbol{\theta}$ (equations (7) and (8)), therefore only the factorization over parameters and states is an actual approximation in (9). In the

present paper additional factorizations were required over the parameters of the duration, and the emission distributions (see Appendix A).

The VB algorithm minimizes the Kullback-Leibler (KL) divergence [44] between the true and the approximate posterior. This is equivalent to maximizing the Negative Free Energy (NFE), a lower bound on the logarithm of the model's evidence $\log(p(\mathbf{y}_{(1:T)}|M)))$. By approximating the model's evidence, the NFE becomes a measure of model quality that trades model accuracy for model complexity and can be used for both model class selection and for monitoring the convergence of the VB algorithm.

D. 2.4 ESTIMATION OF THE NUMBER OF BRAIN STATES

The number of states is assumed to be known (i.e. a structural parameter), therefore inferring its optimal value given the data amounts to a model class selection task. In the VB framework proposed here, we used a model selection strategy based on the NFE, similar to [5]. First, a model space is created by inferring multiple models with varying numbers of states. The optimal number of states is then inferred by selecting the model with maximal NFE. In practice we infer models with an increasing number of states until the NFE peaks. Additionally, during each model's inference, an automatic state "resetting" (ASR) was implemented within the VB algorithm, whereby at each VB iteration, all parameters of states not receiving support from the data, were reset to their prior values. The interim irrelevant states were defined as those getting negligible posterior probability at all time points after each iteration:

$$
s_0 = \{k \in S : \sum_{t=1}^{T} \gamma_t^{(k)} \le \epsilon\}
\tag{10}
$$

where $\gamma_t^{(k)} = q(s_t = k) = \sum_{d=1}^{D} \gamma_t^{(i,d)}$ is the marginal posterior probability of being at state k at time t, marginalized over all possible durations; and ϵ is a small threshold (0.001 in our case). After convergence of VB, the remaining irrelevant states were eliminated, which effectively amounts to an automatic relevance determination of state. Note also that the ASR makes the VB algorithm more robust to falling in local minima. This is because, in subsequent VB iterations, those states reset to their priors, can account for data points which have not been well explained by the other states.

E. SIMULATION FRAMEWORK

We used computational simulations as a test ground for evaluating the performance of HSMM vs HMM for brain state allocation. The simulations pipeline is shown in figure 2. The data in all simulations were generated from a HSMM with uniform distributions over state durations. However, the HSMMs that are then fitted to this data are not provided with knowledge of this distribution, they rather assumed a Normal distribution. All simulations were based on 3 brain states, each characterized by a specific EEG topography

map and a specific data covariance matrix. The brain state topographies were obtained using a generative process, whereby a simulated source activity was projected to the sensor space using the linear EEG Forward Model [45]. A realistic lead field matrix relating the source and sensor spaces was computed based on the digital brain phantom developed at the Montreal Neurological Institute (MNI) [46]. The source space consisted of a mesh of $G=20000$ current dipoles located at the vertices of the tessellated grey/white matter interface. The sensor space consisted of the standard 128 electrodes BioSemi system, which was digitally placed on the scalp of the phantom. In each simulation experiment, the source activity image associated with each brain state was sampled from a G-variate Normal distribution with zero mean and covariance matrix $\Sigma_J = (\mathbf{L}^T\mathbf{L})^{-1}$, where $\mathbf{L}$ is the discrete surface Laplacian operator [47] defined on the nodes of the vertices in source space. This choice of Σ_J is consistent with the assumptions of the popular LORETA source reconstruction method [48] and ensured producing spatially smooth distributed source activity. After projection to the sensor space, the EEG map associated with each brain state were normalized to have unit norm.

The temporal dynamics of the EEG signals where generated considering the brain as a HDS using a HSMM with uniform brain state duration distributions, where at each brain state, the brain emitted a segment of observations of length sampled from a uniform duration distribution centred on a brain state-specific mean duration $\langle d \rangle$ and with a brain state-specific width w. During the lifetime of a brain state, state emissions were generated by sampling from a multivariate normal distribution with constant mean and diagonal covariance matrix specific to that brain state. The total length of the observed data in each realization of a simulation experiment was $T=400$ in all cases. The mean of the brain state emissions was the scalp-projected source activity associated with the brain state, and the covariance matrix modeled the (possibly brain state-specific) zero mean Gaussian iid observation noise. The variance of the noise was manipulated to achieve the required Signal-to-Noise Ratio (SNR) in each simulation experiment (see Appendix C for details).

Brain state transitions were designed to emulate the quasi-periodicity of empirically observed EEG microstate sequences [49]. At the end of a brain state segment, a new brain state s was randomly generated based on the asymmetric transition matrix:

$$\begin{bmatrix} 0 & 0.8 & 0.2 \\ 0.2 & 0 & 0.8 \\ 0.8 & 0.2 & 0 \end{bmatrix} \tag{11}$$

This choice of matrix promotes cyclic brain state transitions ($s1 \rightarrow s2 \rightarrow s3$). Without loss of generality, in all simulations the system was assumed to be in the brain state 1 at the beginning of the sequence.

Since our focus is on the modeling of brain state durations, brain state maps were generated ensuring a fair degree of dissimilarity between the maps of different brain states. That way, differences in performance between the evaluated methods could be attributed to the differences in how brain state duration is modeled, rather than to inability to discriminate between similar brain state maps.

F. EVALUATION OF MODEL PERFORMANCE

The mathematical details of the evaluation scores used to assess the performance of the models are summarized in Appendix D. Following [49], the Auto Information Function (AIF) (Appendix D) was used to demonstrate that the HSMM was capable of producing features of state dynamics similar to those observed in EEG microstate sequences. The AIF of a brain state sequence measures the amount of information about the future of the brain state sequence that is contained in its own past, therefore providing an indicator of the long-term dependencies in the sequence. The Sequence Accuracy Score (SAS) (Appendix D.2) was used to compare the accuracy of HSMM vs HMM in solving the brain state allocation problem. SAS measured the similarity between the inferred and the true brain state sequences expressed in percentage, so that a value of 100% indicated identical sequences, while 0% indicated no matching between any of the elements of the two sequences. The explanatory power of a model was assessed in terms of the extent to which new state sequences predicted from the trained models, showed a brain state duration distribution similar to the one used to generate the original training dataset. The Jensen-Shannon Divergence (JSD) [50] (Appendix D) was used to measure the dissimilarity between the histograms of the brain state durations distributions of the training and the predicted sequences. The JSD is bounded between 0 (perfect similarity) and 1 (perfect dissimilarity).

Model quality measures were used to evaluate the relative performance of two given models (or model classes) based on information theoretic principles, which do not require comparison to a gold standard. Two model classes were compared based on the Log-Bayes Factor $\log BF(M_1, M_0)$ [51] (Appendix D.4) that measures the evidence provided by the data in favor of the model class M_1 (e.g. HSMM) against M_0 (e.g. HMM). Models within the same model class were also evaluated based on their Discriminative Power Score $DSP(H_0, H_1)$ [52] (Appendix D.5). DSP is a directed divergence that measures the ease of discriminating between two models (H_0 vs H_1) of the same model class (e.g. two HSMMs with different set of parameters $\mathbf{\Theta}_0$ and $\mathbf{\Theta}_1$) given an observed data sequence. This divergence is always positive and is zero if the two models are the same ($\mathbf{\Theta}_0 = \mathbf{\Theta}_1$).

G. EMPIRICAL DATA ACQUISITION AND ANALYSIS

EEG data collection was approved by the ethics committee of the University of Manchester, UK (ref. 14374) and was conducted according to the declaration of Helsinki. Ten minutes of spontaneous task-free EEG were recorded from
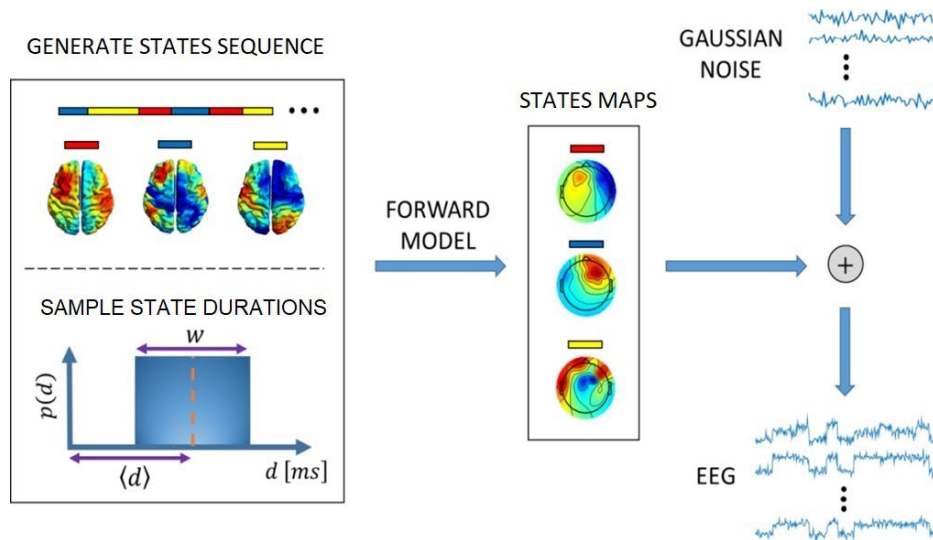
FIGURE 2. Simulation pipeline. EEG activity is simulated considering the brain as a hybrid dynamical system. The hidden discrete state process represents fast transitions between source activity configurations, which remain stable over a time segment. The length (duration) of the time segments are randomly sampled from uniform distribution with state-specific mean $\langle d \rangle$ and width w. The neural source activity is instantaneously projected to the sensors' space using the linear EEG Forward model and subsequently corrupted with Normal noise to produce the observed EEG signals.

five healthy subjects who were asked to sit comfortably with eyes open during the experimental session. The EEG data were recorded from 64 scalp electrodes using the ActiveTwo system (BioSemi, Amsterdam, Netherlands) and Actiview® acquisition software (Biosemi, Amsterdam, Netherlands) at a sampling rate of 2048 Hz.

1) DATA PRE-PROCESSING
For consistent comparisons, we followed a similar pre-processing pipeline as in [3], but slightly modified to the case of EEG data. The pipeline was implemented in MATLAB (Mathworks ®) as follows. Channels and periods of data containing obvious artefacts were visually identified and discarded (5%). The signals were band-pass filtered between 0.1 Hz and 120 Hz using a high-pass and low-pass FIR filter in sequence; and were subsequently de-trended and down-sampled to 512 Hz. Artefact detection and correction was carried out by decomposing the data into 30 temporally independent components using independent component analysis. Stereotyped components related to ocular, cardiac, motion, and mains interference artefacts were classified and rejected based on their topographic, temporal and spectral features. The remaining components were then used to reconstruct the clean data. Bad channels were interpolated using spline interpolation and the data was subsequently re-referenced to average reference for further analysis.

2) DATA PREPARATION FOR HSMM ANALYSIS
Following previous studies, the HSMM was applied to the power envelopes of the EEG signals [1], [3], [40]. The

pre-processed data was first band pass (FIR) filtered between 4 and 30 Hz. The amplitude envelope of the oscillatory activity at each channel was then derived by computing the magnitude of the analytic signal, obtained from the Hilbert transformed data. For computational efficiency, the envelope amplitudes were down-sampled to 40 Hz using a polyphase anti-aliasing filter as implemented in the MATLAB signal processing toolbox. Principal Component Analysis (PCA) was then used on the normalized envelopes (zero mean and unit standard deviation) for whitening and dimension reduction to keep the first 40 principal components (99.21% explained variance).

3) HSMM SETUP
Two HSMMs were inferred using either a Normal or a Lognormal duration distribution, and maximum brain state duration of 5 s. In all models, the brain states were assumed to emit observations using each a 40-variate Normal distribution (the first 40 principal components of the envelope data), specified by their corresponding mean vector and covariance matrix.

Group level parameters were inferred by sharing parameters across subjects [21], [42] by assuming that the observed EEG sequence of different subjects are independent realizations of the same hidden semi Markov process. This should be contrasted with other works where subjects' data sequences are concatenated in time and a single HMM model is fit to the full concatenated data sequence [1]. The parameter sharing approach avoids introducing potential artificial state switches at the boundaries of concatenated data, as well as allows accommodating boundary conditions for each subjects' data sequence.

4) BRAIN STATE TOPOGRAPHIC MAPS

Due to the PCA dimensionality reduction, the means and covariance matrices characterising the brain states are not straightforwardly interpretable since they are estimated in the low dimensional space spanned by the principal components. To obtain interpretable scalp activity maps associated with each brain state, we followed a procedure similar to [3]. The partial correlation of the estimated brain states time courses with the amplitude envelope data at all the electrodes, was computed. brain state time course correspond a binary vector indicating whether the brain states is on or off at each point of the most probable brain state sequence, which was estimated using the Viterbi algorithm for HSMM [37]. The obtained correlation maps represent the unique contribution of each brain state to the oscillatory activity power, while components of the envelope signals that are common across brain states are minimized.

The partial correlation was computed using a General Linear Model (GLM) analysis [53] with the brain state time course as predictors and the full-rank envelope data (prior to dimension reduction and whitening) as response variables. Given the brain state time course $\hat{s}_{1:T}$, each element of the $[T \times K]$ GLM's matrix of predictors $\mathbf{X}$ is

$$X_{tk} = \begin{cases} 1, & \hat{s}_t = k \\ 0, & \hat{s}_t \neq k \end{cases}$$

The columns of $\mathbf{X}$ and the envelope data were both z-scored prior to GLM analysis. The K vectors of estimated regression coefficients defined scalp spatial maps representing estimates of the partial correlation coefficient between each brain state and the envelope data. To facilitate interpretation, brain states from different models or model classes were assigned the same label if they had similar topographic maps.

H. INITIALISATION OF THE VB ALGORITHM

To account for dependence of the inference on the initial conditions (e.g. convergence to local minima), the HSMM and the HMM VB algorithms were ran 10 times each using different starting points. Each starting point was determined using k-means to cluster all data points and defining an initial brain state sequence by assigning each data point to the closest cluster. The k-means algorithm was itself initialized 3 times using random data points as initial centroid positions and the solution producing the most compact clusters was chosen as the optimal one. Finally, the sample estimates of the models' parameters obtained based on the obtained k-means brain state sequence were used to initialize the VB iterations. The optimal VB initialisation was determined as the one with highest NFE after 10 iterations of the algorithm. In practice, this strategy produced identical results as compared to running the VB fully until convergence for every starting point and then choosing the solution with highest NFE; but it was cheaper computationally.

III. RESULTS

A. NON-GEOMETRIC BRAIN STATE DURATIONS CAN GENERATE "SENSIBLE" BRAIN STATE DYNAMICS

We first demonstrated that the brain state sequences underlying the simulated data showed dynamical properties of empirically observed EEG microstates sequences. We focused primarily on the relatively long but finite memory (temporal dependencies) demonstrated via distinct periodicities of the AIF of microstate sequences, with a tail that converges asymptotically to Markovian memoryless in the long run [54]. Figure 3 shows the AIF of brain state sequences generated by HSMM with Uniform durations and the corresponding HMM predictions; as functions of the mean ($\langle d \rangle$ in figure 3a,b) and width (w in figure 3c,d) of the generating duration distributions. The HSMM used 3 brain states with identical duration distributions and the transition matrix defined in section II-E. The HMM predictions were generated using the empirical transition matrix computed based on the matching HSMM sequence. The AIF of HSMM showed oscillations with amplitude decaying slowly with the time lag. This behavior corresponds to a process with memory (time dependencies), but Markovian "memoryless" in the long run; which is consistent with empirical results of EEG microstates analysis [49]. Both the frequency and amplitude of the AIF oscillations depended on the brain states' mean duration, with higher mean durations corresponding to slower but stronger oscillations (figure 3a). Changes in the width of the duration distribution however, only affected the amplitude of the AIF oscillations, so that increased widths were associated with reduced amplitudes (figure 3c). The AIF of the predicted HMM sequences showed monotonic fast decays with the time lag in all cases. The decay rate was higher for smaller mean duration values (figure 3b), while it was insensitive to changes in the width of the duration distribution (figure 3d).

B. MODELING BRAIN STATE DURATIONS IMPROVES GENERALIZATION POWER

The generalization power of any model is underpinned by the accuracy of its predictions. A simulation study was then carried out to investigate whether the correct modeling of the brain state duration distribution had an important effect on the ability of HSMM vs HMM to make accurate out-of-sample predictions. For this analysis, synthetic EEG data sequences of length 4000 a.u. were first generated based on HSMMs with different uniform brain state durations but identical transition matrix as in previous section. The duration distribution used to generate each dataset was varied by changing the mean duration of a target brain state (20-140 a.u. with 20 a.u. intervals), while the mean of the other two brain states was fixed (50 a.u.). All brain state duration distributions had identical width (6 a.u.). Two SNR levels (30% and 60%) were also used, for a total of 8 simulated EEG datasets (5 mean durations x 2 SNR levels). Inference on each dataset was carried out by training models with two different
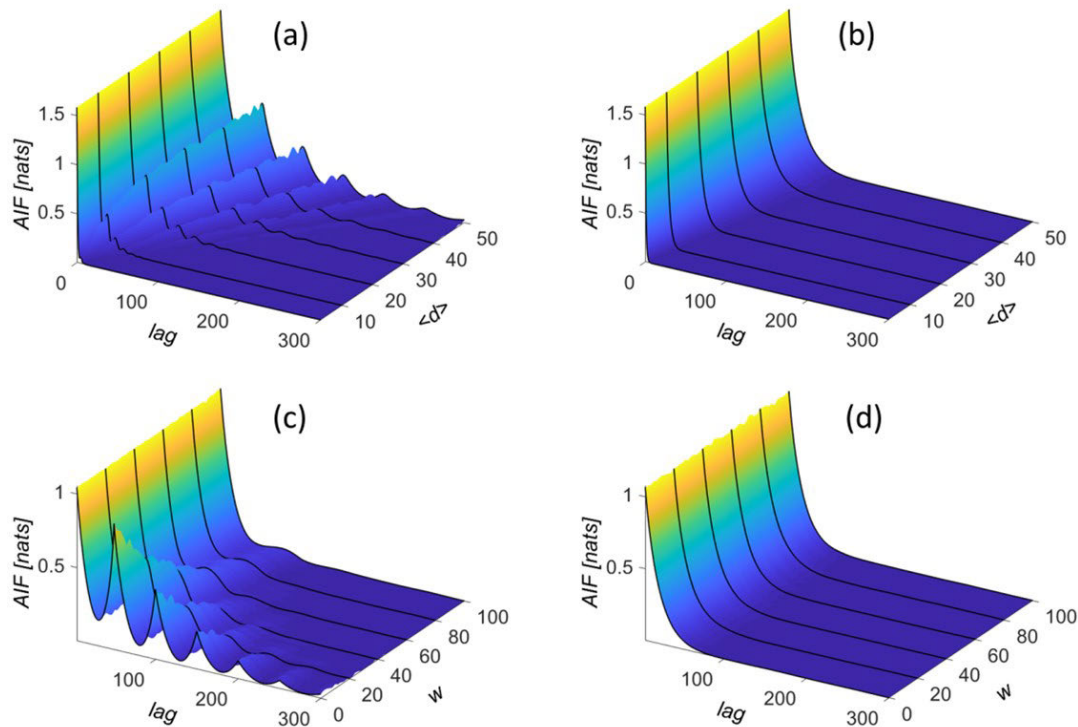
FIGURE 3. Reproducing microstates features in simulated brain state sequences. Auto Information Function (AIF) of the brain state sequences generated using HSMM (a & c) and HMM (b & d) without emissions, as a function of the mean and width of the duration distributions of the brain states. In (a) and (b), the mean duration of the three brain states was varied while the widths were fixed at 6 a.u. In (c) and (d), the widths were varied while the duration means were fixed at 50 a.u. The periodicities and the asymptotic decay in the case of the HSMM resemble those of the empirically observed EEG microstates, as illustrated by [49]. The AIF is measured in natural units of information (nats).

brain state duration distributions, a HSMM with Normal distribution for brain state durations and a HMM with an implicit Geometric distribution for durations. After training, new (predicted) brain state sequences were generated from each of the trained models and the resultant histograms of durations of the target brain state were compared to the true duration distribution that generated the training dataset.

Figure 4 shows the true and predicted histograms for one exemplary dataset. As expected, the Geometric distribution implicit in HMM allocated nonzero probability mass across the whole range of possible duration values with higher probability at shorter durations. On the contrary, the explicit Normal duration distribution in HSMM concentrated its probability mass in the correct range of duration values.

Plots of the JSD between the true and predicted histograms of each model are depicted in figure 5 as a function of the brain state mean duration and SNR of the training dataset. HSMM was consistently more accurate and noise-robust at regenerating the true duration distributions, as evidenced by the lower JSD values across all simulations. Notably, HMM reproducibility deteriorated as mean duration increased. This primarily reflects the effect of the unrealistic Geometric duration model implicit in HMM. That is, underlying duration distributions with longer mean durations have to be captured
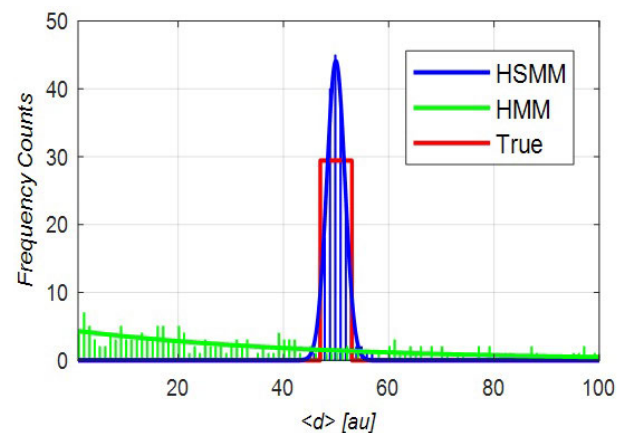

FIGURE 4. Histograms of durations of a test brain state, generated from trained HSMM and HMM. HSMM and HMM were trained on the same dataset simulated using the uniform duration distribution in red (mean 50 a.u. and width 6 a.u.). Subsequently, new sequences of brain states were generated from each trained model and the respective duration histograms corresponding to one brain state were computed and shown. The geometric distribution implicit in HMM, produced brain state durations across the whole range of possible values (green bars). The explicit normal duration distribution in HSMM concentrated the durations in the correct range of values (blue bars). The best fitted geometric (green line) and normal (blue line) brain state duration distributions (respectively for the HMM and the HSMM), are also shown.

by a HMM with a "flatter" Geometric duration distribution (i.e. higher self-transition probability), which will produce

more widespread duration histograms when used to generate new sequences after training. A Geometric distribution that encourages short brain state durations may also lead to less accurate estimation of HMM parameters in the training phase on data produced by long brain state durations. That is, given a fixed length of data, a longer brain state duration translates into a lower number of state transitions and therefore less data is available for a reliable estimation of transition probabilities. In this situation, the ability of HSMM to model brain state durations helps to regularize and stabilize the inference. No apparent effect of the SNR was found for any of the two models.
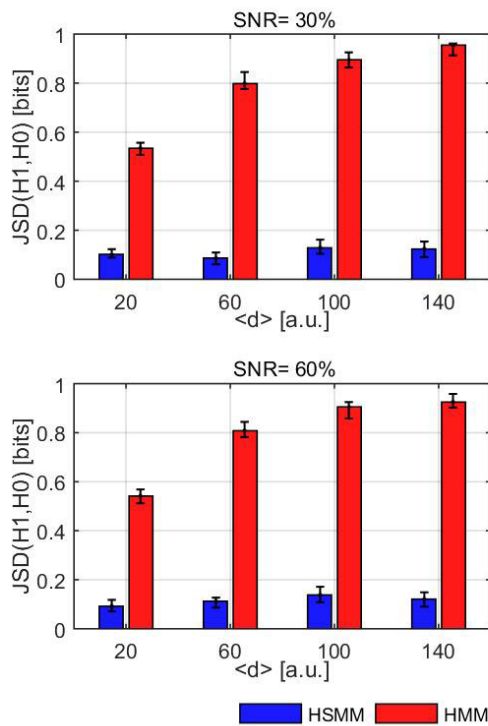

FIGURE 5. Jensen-Shannon Divergence (JSD) between the true (H0) and predicted (H1) brain state duration histograms of HSMM and HMM. A lower JSD value indicates higher similarity between the histograms of the predicted (out-of-sample) durations by a model after training, and of the true durations underlying the training dataset. HSMM out-of-sample predictions consistently and robustly reproduce the brain state durations of the underlying system. The simulations were the same as in figure 4, except the mean duration of one of the brain states in the training sequences was now varied from 20 a.u. to 140 a.u. (step 40 a.u.) at two SNR values. The error bars over 100 realizations of the experiment are indicated.

These results demonstrate that, beyond good data fitting, an accurate modeling of the brain state durations is important when prediction and interpretation is the goal. In this case, using HMM to model a system with non-Geometric durations (as seems to be the case with the brain) could lead to wrong conclusions. In the next sections we investigated how this problem can affect the performance of HSMM vs HMM, in some typical EEG applications.

C. BRAIN STATE DECODING UNDER DATA UNCERTAINTY

Common applications of EEG brain state allocation involve brain state decoding. This is relevant e.g. for developing systems for Brain Computer Interfaces, online EEG monitoring of patients in intensive care units, identification of epileptic seizures, sleep monitoring and study of memory retrieval in movie viewing, among others. Typically, a model is first trained on a dataset representative of the dynamics of interest; and then the trained model is used for decoding the brain states in a new dataset in an unsupervised fashion. In such a situation, accurate predictions are important for the model-based monitoring system to perform robustly in a noisy environment. We tested the impact of incorrectly modeling brain state durations, on the brain state decoding task. For this, the accuracy of HSMM (with Normal durations) and HMM (Geometric durations) was evaluated when used to decode data generated from a system with uniformly distributed brain state durations. Given a previously trained HSMM or HMM, with estimated parameters Θ, and a new observation sequence $\mathbf{y}_{(1:T)}$ (test dataset), we were interested in the most likely brain state sequence that produced the new observations. This is solved using the well-known Viterbi algorithm [37]. In this simulation experiment, the same model was used to generate the training and test datasets, but the test dataset was noisier. This mimics a situation in which brain state monitoring is performed under uncontrolled observation noise fluctuations. Two sets of training and test SNR values and 100 pairs of training and test datasets in each case were simulated. All parameters of the simulation setup were the same as in the previous section, except for the mean duration of the two fixed brain states, which was 10 a.u. in this case.

As expected, using the more suitable Normal duration model resulted in a better brain state decoding performance. Figure 6 shows the SAS for HSMM with Normal durations and HMM as a function of the mean duration of the target brain state, and for different SNR values of the training and the test datasets. In all cases, HSMM showed higher accuracy than HMM. For increased difference between the SNR of the training and the test datasets, the accuracy of the HMM reduced significantly; while HSMM remained stable with a SAS above 90% in almost all cases. Interestingly, in the case of greatest reduction in data quality (figure 6 lower panel) the accuracy of the HSMM increased for increasing values of the mean brain state duration, while the HMM deteriorated.

D. CLASSIFICATION OF BRAIN STATE TEMPORAL STRUCTURES

It has been shown that the presence (or absence) of specific brain states during rest or task conditions can have cognitive or clinical relevance, because a brain state, as defined here, could represent short lasting but meaningful modes of brain operation [55]. Similarly, the temporal structure of brain state sequences, might contain additional information, representing "phenotypes" of longer lasting experimental,
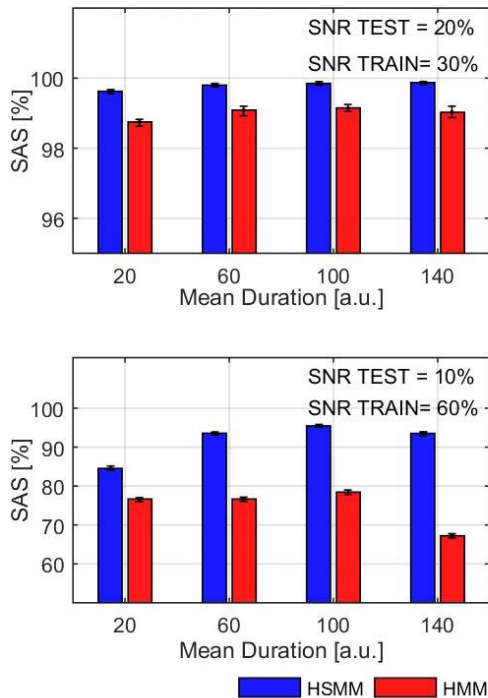
FIGURE 6. Accuracy of brain state allocation based on unreliable test data. HSMM and HMM were first trained on data produced by a HDS with non-geometric states' duration distributions, and were then used to decode the brain state sequence underlying a new test dataset from the same system but with reduced SNR. The mean duration of a test brain state was varied at two levels of SNR of the training dataset (SNR_0) and the test dataset (SNR). The accuracy of the decoding was evaluated using the Sequence Accuracy Score (SAS), between the decoded and the true brain state sequence. Higher SAS indicates a more accurate decoding. HSMM achieved higher SAS for all duration values explored, suggesting robustness to the noise fluctuations. The error bars over 100 realizations of the experiment are indicated.

clinical or environmental conditions, for example anesthesia, sleep or effect of drugs. Therefore, HSMM and HMM can enable "condition phenotyping", via model-based classification of data streams (e.g. task conditions or clinical groups) corresponding to brain state sequences with distinctive temporal structures.

The temporal structure of the brain state sequence refers to the brain states temporal order or brain state "syntax" (given by the transition probabilities), and the dwell time in each brain state (given by the duration distributions). We focussed on evaluating the impact of modeling brain state durations on condition phenotyping. That is, given the generative assumptions (a specific duration model), which condition is most likely to be associated with previously unseen data. That is, given previously trained HSMMs (or HMMs) each representing one condition, with estimated parameters $\hat{\boldsymbol{\Theta}}_{1:C}$ (one for each condition $c = 1, \ldots, C$), and given a new sequence of observations $\mathbf{y}_{1:T}$, we are interested in the probabilities $p(\mathbf{y}_{1:T}|\hat{\boldsymbol{\Theta}}_c, M)$ that the new data were generated by each of the models. These probabilities can be used for model-based Naïve Bayes classification [56] to allocate the new data to the most probable model (condition).

HSMM and HMM were then compared in terms of their ability to classify a test data sequence as belonging to the correct condition among a set of candidate conditions. The conditions differed in the mean duration of the underlying brain states. This experiment mimics a situation in which the models are used as e.g. diagnostic tools by allocating a new patient to one among a set of clinical conditions. Each clinical condition is characterized by a phenotypical alteration of the mean time that the brain spends in each brain state. We used a semi-supervised setting for classification, where labeled data were first used to train different models, with each label denoting a specific condition. The new unlabeled test data sequence was then allocated by assigning the label corresponding to the model with highest probability of having generated the test data. The training set consisted of N=11 simulated data sequences (length T=4000 a.u.), one for each condition, and their corresponding labels. All conditions were simulated based on HSMMs with the same transition matrix as in previous sections and Uniform duration distributions for the 3 brain states. The mean duration of all brain states within each condition was the same, but varied across conditions from 2 a.u. to 22 a.u. (step-size 2), which determined each condition's label. The unlabeled test sequence was generated from the model corresponding to condition 10. The widths of all brain state duration distributions in the training and the test data were fixed to 6 a.u.
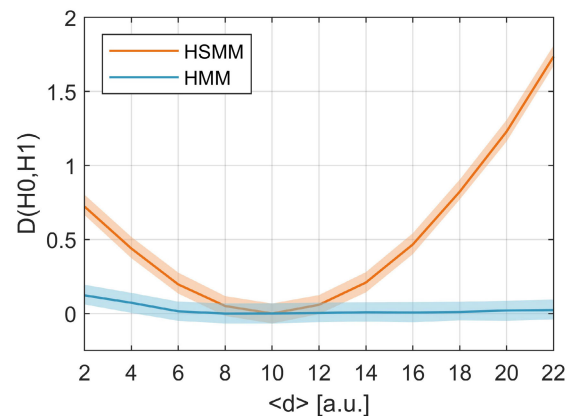

FIGURE 7. Discriminative Power Score (DPS) of HSMM vs HMM. A previously unseen data sequence (with mean brain states duration of 10 a.u) is classified as being generated by one of 11 HSMM and HMM models which were trained on data that differed in the mean duration of the underlying brain states, from 2-22 a.u. in steps of 2. Each point on the curves is a measure of the discrimination of the hypotheses H0 that the test data was generated by its true model (⟨d⟩=10) against the hypothesis H1 that it was generated by the candidate model with duration on the x-axis. A value of DPS=0 indicates no discrimination between the two hypotheses, that is the candidate model is very close to the truth. The shadowed areas represent the 95% confidence interval for discrimination. HSMM with Normal durations is able to allocate the test sequence to the correct model, whereas HMM has no discrimination power.

HSMM (with Normal durations) and HMM (Geometric durations) were used for training and classification in 100 repetitions of the experiment as described above. Figure 7 shows the ability of each model to discriminate between the test sequence and each condition, measured in terms of

the DPS. Each point of the curve represents the average information per test data sample for discrimination of the hypothesis H_0 that the test sequence was generated by the correct model (mean brain state duration 10 a.u.) against the hypothesis H_1 that it was generated by the candidate condition model (mean duration in the x-axis). The zero value represents no ability to discriminate between the two hypotheses and therefore indicates the condition the test sequence is allocated to. As expected, using the more suitable Normal duration model resulted in better classification performance. Discrimination based on HSMM allowed for allocating the test sequence to the condition associated with the correct mean brain state duration. In the case of HMM, there was not enough information in the test data for discrimination between the two hypotheses in any of the conditions, except for the one with the shortest mean brain state duration (2 a.u.). That is, classification based on HMM would allocate the test data sequence to any of the conditions with equal probability except for the case of shortest mean brain state duration (high sensitivity but low specificity). This is consistent with the inability of the implicit Geometric duration distribution to model long durations accurately.

E. RESTING-STATE EEG: LIGHT-TAILED VS HEAVY-TAILED DURATION DISTRIBUTIONS

We have so far shown that accurate modeling of brain state durations is paramount for a meaningful brain state allocation. A key question then is which duration distribution should we use? The choice of the duration Probability Density Function (PDF) will ultimately define the kind of brain state dynamics that the model can explain and therefore it effectively imposes constraints on the temporal properties of the brain state process. For example, a light-tailed "bell-shaped" PDF is typical of a process with a characteristic time-scale, while a heavy-tailed right skewed PDF corresponds to a process with long-term dependencies and possibly temporal scaling law [28], [29], [57]. A Geometric (or Exponential) distribution in turn describes a memoryless process, and separates heavy-tailed from light-tailed distributions. That is, heavy-tailed (light-tailed) distributions are those whose tail decays slower (faster) than the Geometric distribution. We refer the reader to the last section for a more detailed discussion on the relationship between the shape of the distribution and the types of dynamics of the underlying system.

We use real resting state EEG data to select among two alternative duration models. Specifically, we compare using HSMM with Lognormal (heavy-tailed and skewed) vs Normal (light-tailed and symmetric) duration distributions to model the fluctuations of the EEG amplitude's envelopes.

First, we investigated whether there is evidence in the data supporting one duration model vs the other. Two HSMMs, each using one of the duration distributions with maximum duration $D=5000$ ms, were trained on the actual data as described in section II-G. In the two cases, 5 brain states were
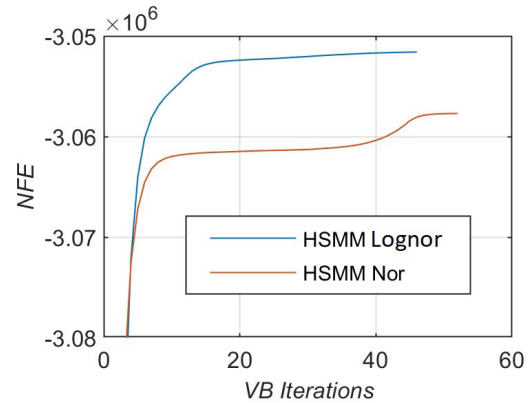

FIGURE 8. Convergence of the VB algorithm. Negative Free Energy (NFE) of the HSMM for the two duration models used in the EEG resting-state data. The HSMM with Lognormal durations converged faster and reached a higher NFE value.

automatically inferred from data using model comparison. Figure 8 shows the convergence curve of the NFE for the two optimal models with 5 brain states. The HSMM with Lognormal durations converged faster and showed higher NFE with a log-Bayes Factor of 400 in its favor after convergence.
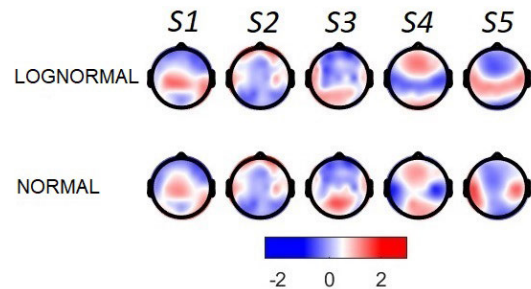

FIGURE 9. Brain states allocation in resting-state EEG. Partial correlation maps of the estimated brain states in each duration model. Based on the pair-wise similarity between the maps of the two duration models, the brain states were relabeled and matched based on these maps by visual inspection. The analyzed EEG topographic maps display a distribution of brain electrical activity in which clearly localized and differentiated areas of positive and negative polarities are observed. The regions with positive polarity, indicated by the red color, and the regions with negative polarity, represented by the blue color.

The partial correlation maps associated with each brain state are shown in figure 9. Based on the pair-wise similarity between the maps of the two duration models, the brain states were relabeled and matched based on these maps by simple visual inspection. When analyzing the EEG topographic maps brain state 4 and brain state 5, it is notable that both maps exhibit identical activation patterns, but with inverse polarities in all corresponding areas. This symmetry and inversion of polarities suggest a possible functional operation with a complementary relationship. When comparing the EEG topographic maps associated with normal duration and lognormal duration models, notable differences in brain activation patterns can be observed. The most prominent

discrepancies are found in maps brain state 4 and brain state 5. In the case of the normal duration model, the band that extends horizontally from the left temporal area to the right temporal area, crossing the central region of the brain, is interrupted at the center of the head.
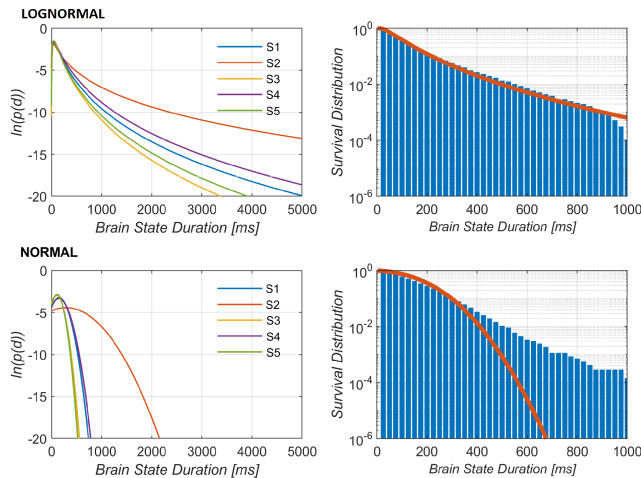

FIGURE 10. Estimated brain state duration distributions. The left panels show the logarithm of the approximate posterior distribution of duration for all brain states. The right panels show (in log-scale) the empirical survival distribution of duration ($P(X > x)$) for each duration model (Normal and Lognormal) for the exemplary state S2, as well as their corresponding estimated parametric distributions (red curve).

Figure 10 summarizes the temporal characteristics of the inferred brain states in terms of their duration distributions induced by the posterior expected values of the parameters of the Lognormal and the Normal duration models. It can be observed that the Lognormal distribution allows for assigning probability mass to longer durations, while the Normal distribution allocates the mass only to shorter durations. In addition, it can be seen that the empirical durations of brain state 2 show asymmetric heavy-tailed histograms in the two cases, which do not follow a Normal distribution, whereas the Lognormal distribution shows an accurate fit.

IV. DISCUSSION
We have proposed Hidden Semi-Markov Model (HSMM) as a generative model for brain state allocation in EEG, which allows for the explicit modeling of the brain state duration distribution. The added flexibility circumvents a well-known limitation of the standard HMM where the implicit distribution of brain state duration is Geometric. The Geometric distribution places more probability mass at shorter than at longer brain state duration, making HMM intrinsically biased towards fast brain state switching. A more fundamental problem is that the Geometric distribution is memoryless. This property implies that the probability of waiting time until the next brain state transition is independent of the time already spent in the current brain state. For a model to describe such behavior, it must continuously forget the state the system is currently in (hence memoryless). A Geometric duration is then not consistent with the long-range temporal

dependency (''long memory'') and near scale-invariance (self-similarity) ubiquitous in real EEG data [29], [57], [58], [59], [60]. This poses limitations on using HMM for modeling EEG data accurately [49], as demonstrated in the model generalization study (figure 3). More importantly, we showed that modeling brain state duration distributions explicitly with HSMM allows capturing EEG dynamical features that are inconsistent with the Geometric duration implicit in HMM. This qualifies HSMM as a more suitable generative model for brain state allocation when solving problems where prediction accuracy is important.

Another usually overlooked limitation of HMM is that it places a small penalty on adding a new state, and no penalty on the new state having similar features to another state [20]. Consequently, using model comparison to infer the number of brain states based on EEG data, where persistent brain states are expected, can lead to the creation of unnecessary extra states and unrealistically fast switching, in order to reproduce the long brain state durations. Indeed, prior EEG studies using HMM for brain state allocation noted a problematic increase in NFE with an increasing number of states, rendering estimation of the number of states through NFE maximization impractical [40]. This issue is crucial as extra states may produce unreliable parameter estimates and pose a risk of overfitting due to insufficient data per state. To enhance inference stability, HMM-based brain state allocation has been commonly applied to extensive datasets with hundreds of subjects [61]. Nevertheless, this approach fails to address the challenge of optimal state number selection. The results in this study demonstrate the successful application of the proposed HSMM framework for analyzing EEG data from a limited number of subjects, while still allowing estimation of the optimal number of brain states through NFE maximization.

We note that the potentially redundant brain states and fast switching in HMM is not necessarily a problem if model averaging is the goal [4]. However, if interpretation and prediction is important, one would like to incorporate prior knowledge that a ''slow'' switching dynamics is more likely. HMM does not incorporate this kind of prior, while in HSMM this is done explicitly by choosing a suitable duration distribution to model the data. We therefore compared the performance of these model in two tasks which rely on accurate model predictions. The first task tested whether an accurate modeling of brain state durations helped brain state allocation based on data with increased external noise. The (Bayesian) rationale is that if the data is unreliable, the accuracy and robustness of the brain state decoding is determined by the ability of the model to represent the actual dynamics of the system producing the observations. We found that HSMM with Normal brain state durations allowed robust and accurate brain state allocation for different data uncertainty levels and different mean brain state durations of the underlying system. This is because the Normal duration used in HSMM was a good approximation of the true brain state duration distribution. In contrast, the Geometric

brain state durations in HMM led to non-robust brain state inference dominated by the noise fluctuations. The second task involved identifying data segments whose underlying state sequence differed in their temporal structure (states' syntax and durations), rather than the individual states' emission parameters. We argue that brain state sequences with a characteristic temporal structure could serve as phenotypes of complex mental processes, which live in a time-scale beyond that of an individual brain state. Similar concepts have been used to interpret both task-related [62], [63] and resting-state [2], [61] activity. The target phenotype to be identified differed from the rest of the candidate phenotypes only in the mean duration of their component brain states. This task tested whether the implicit Geometric distribution in HMM was "good enough" to discriminate between phenotypes, or if the more accurate modeling of the state duration in HSMM was required. Our results showed that, HSMM allowed the correct identification of the target phenotype, whereas HMM could not discriminate between the target and any of the candidate phenotypes. It can be argued that in the identification of these long-lived complex mental processes, the temporal structure of the underlying brain state sequence might play an even more important role than in the identification of individual short-lived brain states, and therefore HSMM is recommended over HMM.

We argue that the choice of duration PDF determines the temporal properties of the brain state process, which will ultimately impose constraints on the kind of dynamics that the model can explain. In other words, the choice of duration PDF can be used to represent our beliefs about specific nonlinear dynamical mechanisms (instabilities) generating the data, such as multi-stability, meta-stability or criticality (see [27]). In a multi-stable system driven by noise, the system jumps erratically between different attractors where it gets briefly trapped in each basin of attraction. Consequently, the time series produced by a multi-stable system shows a relatively heavy-tailed (stretched exponential) dwell distribution. In a meta-stable system, there are no attractors, but rather a sequence of linked weakly unstable fixed points, such that the system dwells in the neighborhood of each, but does not show trapping. In this case, the distribution of sequential dwell times show a unimodal but highly skewed heavy-tailed distribution characteristic of long term dependencies. In a critical system, a single fixed point is very weakly attracting or neutral, so that the system noise leads to long and unstructured excursions corresponding to scale-free fluctuations and corresponding power-law statistics (e.g. power-law duration distributions). We can therefore use the explicit brain state duration PDF in HSMM to express alternative hypotheses of temporal dependency and scaling-law of the brain state sequence, and select the best distribution given the EEG data in a principled way via Bayesian Model Selection. This approach can help bridge between the observed brain state data features and the associated underlying neural mechanisms, which would allow interrogating the data

about such mechanisms. To illustrate the approach we used actual resting-state EEG data to select between a heavy-tailed (Lognormal) and a light-tailed (Normal) brain state duration PDF for brain state allocation using HSMM. Model comparison demonstrated evidence in favor of the Lognormal PDF.

The results in figure 10 also showed that the histogram of brain state durations obtained by fitting a HSMM with Normal brain state durations did not actually conform to a Normal distribution. This indicates a "mismodeling" of brain state durations by the Normal model. Additionally, the results in figure 9 showed differences in the brain maps inferred based on the two models, which indicates differences in the identification of the active brain state at each time point. As both models have the same emission model, then these differences arise from the type of duration model assumed. These findings suggest that, in certain time periods, the information derived from the emission model may be ambiguous and not allow for accurate identification of brain states. In these cases, having an appropriate duration model becomes relevant.

Finally, our results suggest that brain states' fractional occupancy (FO), a widely used post-hoc summary statistic [3], [6], [61], is difficult to interpret because it conflates brain state duration and transition probabilities. A clear advantage of the proposed HSMM is that it allows disambiguating the individual contribution of state duration transitions probabilities to FO because the two parameters are modelled explicitly and independent of each other. This implies that experimental or clinical conditions linked to changes in brain state durations, can be distinguished from those associated with changes in transition probabilities between brain states. In contrast, in HMM, this is unfeasible due to the confounding of brain state transitions and durations.

The framework proposed here enables several lines of work in the study of brain function and related translational applications. Future research may delve into refining the subtle aspects of brain state modeling introduced by HSMM. Of particular interest, investigating different forms of the explicit brain state duration distribution within HSMM could be used to express hypotheses concerning non-linear dynamical mechanisms, such as multistability, meta-stability, or criticality inherent in neural processes. In this context, the proposed application of Bayesian model comparison for selecting optimal duration distributions presents an avenue for principled exploration, linking observed EEG data features with underlying neural mechanisms. Coupling this approach with integrative research exploring the extension of the HSMM to other functional imaging modalities (e.g. MEG, fMRI, fNIRS) or behavioral data, could deepen our understanding of brain function. Additionally, evaluation of the proposed approach for multi-subject inference and addressing challenges associated with extensive data collection could also enhance its utility in large-scale studies. Future translational research may focus on assessing the

clinical relevance of the proposed framework, particularly in identifying aberrant brain state dynamics associated with neurological disorders. In this context, novel therapeutic interventions may be developed by adapting HSMM to real-time EEG-based applications such as brain-computer interfaces and neurofeedback systems.

V. CONCLUSION

The present study introduced and extensively evaluated the Hidden Semi-Markov Model as a robust and flexible generative model for brain state allocation in EEG data. By explicitly modeling the distribution of brain state durations, HSMM overcomes limitations of the widely used (conventional) Hidden Markov Model, particularly its inherent reliance on a Geometric distribution for representing brain state durations. The findings presented herein highlight the advantages of HSMM in capturing the nuanced dynamics of EEG data, especially in scenarios demanding accurate predictions and meaningful interpretations, indicating its potential significance in neuroscientific and clinical research. Additionally, the demonstrated superiority of HSMM in tasks involving increased external noise and the identification of complex mental processes suggests its potential utility in real-world applications. In summary, the introduced HSMM framework is a promising avenue for advancing EEG-based brain state modeling, and future research endeavours can build upon this foundation to address emerging challenges and extend its applicability to diverse domains.

(Nelson J. Trujillo-Barreto and David Araya Galvez are co-first authors.)

APPENDIX A
MEAN FIELD APPROXIMATION AND NEGATIVE FREE ENERGY

The mathematics of the VB framework for HSMM is described in detail in [36]. The mean-field approximation used here has the following form:

$$q(\mathbf{s}, \boldsymbol{\tau}, \boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\theta}, \boldsymbol{\lambda}) = q(\mathbf{s}, d, \boldsymbol{\tau})q(\boldsymbol{\pi})\left(\prod_{k=1}^{K} q(\mathbf{a}_{k:})\right)$$
$$\cdot q(\boldsymbol{\mu})q(\boldsymbol{\Sigma})q(\boldsymbol{\delta})q(\boldsymbol{\rho}) \quad (12)$$

where, in addition to the factorization between hidden states and parameters (see main text) we have assumed factorization over the parameters of both the duration and the emission distributions; as well as over the rows of the transition matrix [42]. This factorization induces the following form of the NFE:

$$\text{NFE}(q(\mathbf{s}, \boldsymbol{\tau}, \boldsymbol{\Theta})) = -\text{KL}(q(\boldsymbol{\pi})\|p(\boldsymbol{\pi}))$$
$$- \sum_{k=1}^{K} \text{KL}(q(\mathbf{a}_{k:})\|p(\mathbf{a}_{k:}))$$
$$- \sum_{k=1}^{K} \text{KL}(q(\boldsymbol{\mu}^{(k)})\|p(\boldsymbol{\mu}^{(k)})))$$

$$- \sum_{k=1}^{K} \text{KL}(q(\boldsymbol{\Sigma}^{(k)})\|p(\boldsymbol{\Sigma}^{(k)}))$$
$$- \sum_{k=1}^{K} \text{KL}(q(\nu^{(k)})\|p(\nu^{(k)})))$$
$$- \sum_{k=1}^{K} \text{KL}(q(\rho^{(k)})\|p(\rho^{(k)}) + \log(C(s_{1:T}))) \quad (13)$$

where $KL(q(x)\|p(x))$ denotes the Kullback-Leibler (KL) divergence between the approximate variational posterior $q(x)$ and the prior $p(x)$; and $C(s_{1:T})$ is a normalization constant that can be computed after each VB-E step [36]. Note that due to conjugacy of the priors, all KL divergences are between known standard distributions and can be computed analytically. We refrain from including the KL expressions here, since these are standard results available somewhere else.

APPENDIX B
PRIORS AND POSTERIOR UPDATES FOR ALL PARAMETERS

This section presents the update equations of the VB-M-Step. It uses the following marginal statistics obtained from the Forward-Backward algorithm implemented in the VB-E step:

- The joint posterior probability of the joint process (s_t, τ_t) being at (i,d) at time $t-1$ and at (j, d') at time t:

$$\xi_t(i, d, j, d') = q((s_{t-1}, \tau_{t-1})$$
$$= (i, d), (s_t, \tau_t) = (j, d')) \quad (14)$$

- The posterior probability of the joint process (s_t, τ_t) being at (i,d) at time t:

$$\gamma_t(i, d) = q((s_t, \tau_t) = (i, d))$$
$$= \sum_{d'=1}^{D} \sum_{j=1}^{K} \xi_t(j, d', i, d) \quad (15)$$

Given the structure of the HSMM presented here, these statistics can usually be reduced respectively to $\xi_t(i, 1, j, d)$, the posterior probability of a transition occurring at time t; and $\gamma_t(i) = q(s_t = i) = \sum_{d=1}^{D} \gamma_t(i, d)$, the posterior probability of state i being active at time t.

1) MULTIVARIATE NORMAL EMISSION MODELS

The signal emitted from the k-th brain state is modeled as a n-variate Normal distribution with mean $\mu^{(k)}$ and precision matrix $\Sigma^{(k)}$.

$$p(\mathbf{y}_t|s_t = k, \boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)}) = \text{N}_n\left(\mathbf{y}_t; \boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)^{-1}}\right) \quad (16)$$

For the simulation experiments, the data emitted from each state was assumed to be independent over channels with precision matrix $\Sigma^{(k)} = \text{diag}(\boldsymbol{\sigma}^{(k)})$, where $\boldsymbol{\sigma}^{(k)}$ is the $n \times 1$ vector of precisions (inverse variances) of each channel's

data. For the empirical data analysis a full precision matrix was assumed.

- **Prior distribution for means of emissions**
  For both the simulations and the actual data we assumed that all brain state means are independent across states and i.i.d a priori with identical conjugate weakly informative priors

$$p(\boldsymbol{\mu}|\boldsymbol{\mu_0}, \boldsymbol{\Sigma}_0) = \prod_{k=1}^{K} N_n(\boldsymbol{\mu}^k; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0^{-1}),$$

$$\boldsymbol{\mu}_0 = \mathbf{0}_{n\times 1}, \quad \boldsymbol{\Sigma}_0 = \sigma_0 \mathbf{I}_n \quad (17)$$

where $\mathbf{0}_{n\times 1}$ is a $n\times 1$ vector of zeros, $\mathbf{I}_n$ is the identity matrix of size $n \times n$. We used $\sigma_0 = 0.1$ for the simulations and $\sigma_0 = 0.001$ for the empirical data.

- **Prior distribution for precision of emissions**
  For the emission model used in simulations, precisions were assumed to be independent over states and channels a priori, with identical conjugate Gamma priors and weakly informative shape and scale parameters a and b, respectively

$$p(\boldsymbol{\sigma}|a_0, b_0) = \prod_{k=1}^{K} \prod_{i=1}^{n} \text{Gam}(\sigma_i^{(k)}; a_0, b_0) \quad (18)$$

where $a_0 = 0.001$ and $b_0 = 1000$. For the full precision model used in the actual data, precision matrices are independent over states a priori with identical conjugate n-dimensional Wishart distributions with degrees of freedom a and scale matrix B

$$p(\boldsymbol{\Sigma}|a_0, \mathbf{B_0}) = \prod_{k=1}^{K} Wishart(\boldsymbol{\Sigma}^{(k)}; a_0, \mathbf{B_0}) \quad (19)$$

where $a_0 = n$ and $\mathbf{B}_0 = n\mathbf{I}_n$

- **Posterior updates for means of emissions**
  Due to conjugacy of the priors, the variational densities of the states' means of the two emission models have both Gaussian forms. For the independent emission model the posterior is also independent over channels:

$$q(\boldsymbol{\mu}) = \prod_{k=1}^{K} N_n\left(\boldsymbol{\mu}^{(k)}; \hat{\boldsymbol{\mu}}^{(k)}, (diag(\hat{\boldsymbol{\sigma}}^{(k)})^{-1}\right) \quad (20)$$

with posterior mean and precisions given by

$$\hat{\mu}_i^{(k)} = (\sigma_0\mu_{0i} + \hat{a}_i^{(k)}\hat{b}_i^{(k)} \sum_{t=1}^{T} \gamma_t(k)y_{it})$$

$$\cdot \left(\sigma_0 + N^{(k)}\hat{a}_i^{(k)}\hat{b}_i^{(k)}\right)^{-1} \quad (21)$$

$$\hat{\sigma}_i^{(k)} = \sigma_0 + N^{(k)}\hat{a}_i^{(k)}\hat{b}_i^{(k)} \quad (22)$$

where $\hat{a}_i^{(k)}$ and $\hat{b}_i^{(k)}$ are the shape and scale parameters of the approximate posterior of the precisions, associated with emissions from state k (see below) at channel i;

and $N^{(k)}$ is the expected state's counts (i.e. the expected number of state's visits)

$$N^{(k)} := \sum_{t=1}^{T} \gamma_t^{(k)}, \quad k = 1, \dots, K \quad (23)$$

In the case of the emission model with full precisions, the variational posterior for the means are independent over states so that

$$q(\boldsymbol{\mu}) = \prod_{k=1}^{K} N_n\left(\boldsymbol{\mu}^{(k)}; \hat{\boldsymbol{\mu}}^{(k)}, \hat{\boldsymbol{\Sigma}}^{(k)^{-1}}\right) \quad (24)$$

with posterior parameters

$$\hat{\boldsymbol{\mu}}^{(k)} = \left(\Sigma_0 + N^{(k)}\hat{a}^{(k)}\hat{\mathbf{B}}^{(k)}\right)^{-1}$$

$$\cdot (\boldsymbol{\Sigma}_0\boldsymbol{\mu}_0 + \hat{a}^{(k)}\hat{\mathbf{B}}^{(k)} \sum_{t=1}^{T} \gamma_t^{(k)}\mathbf{y}_t) \quad (25)$$

$$\hat{\boldsymbol{\Sigma}}^{(k)} = \boldsymbol{\Sigma_0} + N^{(k)}\hat{a}^{(k)}\hat{\mathbf{B}}^{(k)} \quad (26)$$

- **Posterior updates for precisions of emissions**
  Again, conjugacy of the priors ensures that the variational posteriors have the same functional form as the prior. In the case of the independent observation model, the variational posterior of the precisions factorizes over states and channels so that

$$q(\boldsymbol{\sigma}) = \prod_{k=1}^{K} \prod_{i=1}^{n} Gam\left(\sigma_i^{(k)}; \hat{a}_i^{(k)}, \hat{b}_i^{(k)}\right) \quad (27)$$

with posterior shape and scale parameters

$$\hat{a_i^{(k)}} = a_0 + \frac{1}{2}N^{(k)} \quad (28)$$

$$\hat{b_i^{(k)}} = b_0 + \sum_{t=1}^{T} \left(\gamma_t^{(k)}(y_{ti} - \hat{\mu}_i^{(k)})^2 + \hat{\sigma}_i^{(k)^{-1}}\right) \quad (29)$$

In the case of the emission model with full precision matrices, the variational posterior of the precisions factorizes over states with state-specific Wishart distributions so that

$$q(\Sigma) = \prod_{k=1}^{K} \text{Wishart}_n\left(\Sigma^{(k)}, \hat{a}^{(k)}, \hat{\mathbf{B}}^{(k)}\right) \quad (30)$$

with posterior degrees of freedom and scale parameters

$$\hat{a}^{(k)} = a_0 + N^{(k)} \quad (31)$$

$$\hat{\mathbf{B}}^{(k)} = (\mathbf{B}_0^{-1} + \sum_{t=1}^{T} \gamma_t^{(k)}(\mathbf{y}_t - \hat{\mu}^{(k)})(\mathbf{y}_t - \hat{\mu}^{(k)})^T)^{-1}$$

$$(32)$$

2) DURATION MODELS
- **Normal duration model**
  We impose independent conjugate and weakly informative prior distributions on the mean and the precision of the normal duration distribution of each state.
  - Prior distribution for the parameters
    Independent Normal prior distributions with identical mean v_0 and precision η_0 were assumed on the mean duration of each state so that the full prior factorizes over states:

$$p(\mathbf{v}) = \prod_{k=1}^{K} N\left(v^{(k)}; v_0, \eta_0^{-1}\right) \quad (33)$$

where $v_0 = 1$ ($v_0 = 100$ for the empirical data) and $\eta_0 = 10^{-5}$. In the same way, independent weakly informative Gamma prior distributions with identical shape and scale parameters ((u_0 and v_0)) were assumed for the precision of the duration of each state

$$p(\boldsymbol{\rho}) = \prod_{k=1}^{K} Gam\left(\rho^{(k)}; u_0, v_0\right) \quad (34)$$

where $u_0 = 0.001$ and $v_0 = 1000$, in order to achieve non-informative priors.
  - Posterior updates for the parameters
    Conjugacy and independence of the priors induce variational posterior densities of the same functional form as the priors, which are also factorized over states, so that

$$q(\mathbf{v}) = \prod_{k=1}^{K} N\left(v^{(k)}; \hat{v}^{(k)}, \hat{\eta}^{(k)^{-1}}\right) \quad (35)$$

where the posterior mean and precision of the duration of state k are given by

$$\hat{v}^{(k)} = \left(\eta_0\varphi_0 + \hat{u}^{(k)}\hat{v}^{(k)} \sum_{d=1}^{D} \gamma_d^{(k)} d\right) \quad (36)$$
$$\cdot \left(\eta_0 + \hat{u}^{(k)}\hat{v}^{(k)} C^{(k)}\right)^{-1}$$
$$\hat{\eta}^{(k)} = \eta_0 + \hat{u}^{(k)}\hat{v}^{(k)} C^{(k)} \quad (37)$$

Similarly, the variational densities of the precision of the durations have the same functional form as their priors and also factorize over states, so that

$$q(\rho) = \prod_{k=1}^{K} Gam\left(\rho^{(k)}; \hat{u}^{(k)}, \hat{v}^{(k)}\right) \quad (38)$$

where the posterior shape and scale parameters of the posterior Gamma distribution associated with state k are given by

$$\hat{u^{(k)}} = u_0 + \frac{1}{2} C^{(k)} \quad (39)$$

$$v^{\hat{(k)}} = \left(v_0^{-1} + \frac{1}{2}\sum_{d=1}^{D} \gamma_d^{(k)}((d - \hat{\varphi}^{(k)})^2 + \hat{\eta}^{(k)^{-1}})\right)^{-1} \quad (40)$$

In the above expressions, $\gamma_d^{(k)}$ and $C^{(k)}$ are derived from the sufficient statistics obtained in the VB-E step of the VB algorithm

$$\gamma_d^{(k)} = \sum_{t=1}^{T} \sum_{\substack{i=1 \\ i \neq k}}^{K} \xi_t(i, 1, k, d) \quad (41)$$

$$C^{(k)} = \sum_{d=1}^{D} \gamma_d^{(k)} \quad (42)$$

- **Lognormal duration model**
  Similar to the Normal case, we impose independent conjugate and weakly informative prior distributions on the parameters of the Lognormal distribution associated with each state. Given that the conjugate priors for the parameters of the Lognormal distribution are the same as for the Normal distribution, the expressions for the priors and their variational posteriors are the same as for the Normal case, but replacing the mean and precision of the Normal model with the first and second parameters, respectively of the Lognormal model. In this case, the values of the parameters of the priors (hyperparameters) were $v_0 = 4.5$, $\eta_0 = 0.2$, $u_0 = 0.001$ and $v_0 = 1000$

3) TRANSITION MODEL
From the definition of HSMM, the transitions between state segments are governed by a Markov process without self-transitions. Therefore, the transition distribution between states is Multinomial. We then use the usual conjugate priors and assumptions as in the standard HMM [42].

- **Prior distribution for the transition probabilities**
  Using standard results from the Bayesian inference of HMM, we assume that the rows $\mathbf{a}_{i:}$ of the transition matrix are independent and identically distributed a priori. We then choose the same conjugate non-informative prior distribution for all $\mathbf{a}_{i:}$, which is a symmetric non-informative Dirichlet distribution with concentration parameter α_0 and zero mass on self-transitions,

$$p(\mathbf{A}) = \prod_{i=1}^{K} Dir(\mathbf{a}_{i:}; \boldsymbol{\alpha}_0) \quad (43)$$

where $\mathbf{a}_{i:} = (a_{i1}, \ldots, a_{ii-1}, a_{ii+1}, \ldots, a_{iK})$ is the i-th row of the transition matrix $\mathbf{A}$ without its i-th element; and $\boldsymbol{\alpha}_0 = \alpha_0 \mathbf{1}_{1 \times (K-1)}$ with $\alpha_0 > 0$. A non-informative prior is achieved by using $\alpha_0 = 1$.
- **Posterior updates for the transition probabilities**
  Due to the conjugacy of the prior, the variational posterior of all the rows are Dirichlet, so that

$$q(\mathbf{A}) = \prod_{i=1}^{K} Dir(\mathbf{a}_{i:}; \hat{\boldsymbol{\alpha}}) \quad (44)$$

with posterior concentration parameters

$$\alpha_{ij} = \alpha_a + \sum_{t=1}^{T} \sum_{d=1}^{D} \hat{\xi}_t(i, 1, j, d) \qquad \forall i \neq j \qquad (45)$$

4) INITIAL PROBABILITIES
Similar to the transition distribution, the initial probability distribution π is also Multinomial and therefore can be treated in a similar way as one row of the transition matrix.

- Prior distribution for the initial probabilities
  In the general case (no simplifying initial condition) a conjugate symmetrical Dirichlet prior can be imposed on the initial joint state (s_1, τ_1) as

$$p(\boldsymbol{\pi}) = \text{Dir}(\pi_{11}, \ldots, \pi_{1D}, \ldots, \pi_{KD}; \boldsymbol{\zeta}_0) \qquad (46)$$

  where $\boldsymbol{\zeta}_0 = \zeta_0 \mathbf{1}_{1 \times KD}$ and $\zeta_0 > 0$. A non-informative prior is implemented using $\zeta_0 = 1$. Under the simplifying initial condition $\tau_1 = 1$ (so that $\pi_{kd} = 0$ for $d \neq 1$) then the prior can be reduced to

$$p(\pi) = Dir(\pi_{11}, \ldots, \pi_{K1}; \boldsymbol{\zeta}_0) \qquad (47)$$

  where $\boldsymbol{\zeta}_0 = \zeta_0 \mathbf{1}_{1 \times K}$ and $\zeta_0 > 0$.
- Posterior updates for the initial probabilities
  Due to the conjugacy of the prior, the variational posterior of initial probabilities is also Dirichlet

$$q(\boldsymbol{\pi}) = \text{Dir}(\pi_{11}, \ldots, \pi_{1D}, \ldots, \pi_{KD}; \hat{\boldsymbol{\zeta}}) \qquad (48)$$

  with concentration parameters

$$\hat{\zeta}_{kd} = \zeta_0 + \zeta_1(k, d) \qquad (49)$$

  or under the simplified initial condition

$$\zeta_k = \zeta_0 \hat{+} \zeta_1(k, 1) \qquad (50)$$

APPENDIX C
DEFINITION OF SIGNAL-TO-NOISE RATIO
For a given data sequence, the SNR was defined as the percentage of noise-free signal energy present in the data:

$$\text{SNR}[\%] = \frac{\|signal\|^2}{\|signal\|^2 + \|noise\|^2} 100\% \qquad (51)$$

For iid Gaussian noise in each channel and normalized brain state maps, it can be demonstrated that

$$\bar{s}^2 = \frac{T}{nT - 1} \left(\frac{100\%}{SNR[\%]} - 1 \right) \qquad (52)$$

where $\bar{s}^2$ is the unbiased sample variance estimate and n and T are the number of channels and the length of the data sequence respectively. This expression was then used to compute the simulated noise variance required to achieve a selected SNR value.

APPENDIX D
EVALUATION SCORES
1) AUTO INFORMATION FUNCTION
The Auto Information Function (AIF) is analogous to the autocorrelation function, but for symbolic sequences such as the brain states sequences. It measures the dependence between different time points with a given time lag l. This is done by measuring the Kullback-Leibler divergence between the symbol distributions at time t and $t+l$. That is, the AIF for time lag l is defined as:

$$\text{AIF}(l) = H(s_{t+l}) - H(s_{t+l}|s_t) \qquad (53)$$

This is the difference between the entropies H of the distributions $p(s_{t+l})$ and $p(s_{t+l}|s_t)$. For a stationary Markov chain the AIF can be obtained analytically [54].

2) SEQUENCE ACCURACY SCORE
Denoting the brain state decoded at time t as $\hat{s}_t$ and the true one as $\tilde{s}_t$, the Sequence Accuracy Score (SAS) between the sequences $s_{1:T}$ and $\tilde{s}_{1:T}$ of equal length T is defined as

$$SAS(\hat{s}, \tilde{s}) = \frac{1}{T} \left(T - HD(\hat{s}_{1:T}, \tilde{s}_{1:T}) \right) \times 100\%$$
$$= \left(\frac{1}{T} \sum_{t=1}^{T} \delta(\hat{s}, \tilde{s}) \right) \times 100\% \qquad (54)$$

where $\delta(\cdot, \cdot)$ is the Kronecker's delta function and $HD(x, y)$ denotes the Hamming distance between sequences x and y [64]. Given two sequences of equal length, the Hamming Distance is an edit distance defined as the minimum number of substitutions required to change one sequence into the other. A SAS of 100% means the two sequences are identical (no substitutions needed), while 0% indicates no matching between any of the elements of the two sequences (all elements need to be substituted).

3) JENSEN-SHANNON DIVERGENCE
In probability theory, the Jansen-Shannon Divergence (JSD) [50] is used to measure the similarity or dissimilarity between two probability distributions or histograms. JSD is a symmetrized version of the Kullback-Leibler divergence $KL(P\|Q)$ between two distributions or histograms P and Q. It is defined as

$$\text{JSD}(D, Q) = \frac{1}{2}\text{KL}(P\|M) + \frac{1}{2}\text{KL}(Q\|M) \qquad (55)$$

where $H = \frac{1}{2}(P + Q)$. The JSD is bounded between 0 (perfect similarity) and 1 (perfect dissimilarity) when the KL is expressed in base 2 logarithm (bits); and its square root is a metric known as the Jensen-Shannon distance.

4) LOG-BAYES FACTOR
Given a data sequence $\mathbf{y}_{1:T}$ and two given model classes M_1 and M_0, the log-Bayes factor [51] is defined as

$$\log\text{BF}(M_1, M_0) = \log \left(\frac{p(\mathbf{y}_{1:T}|M_1)}{p(\mathbf{y}_{1:T}|M_0)} \right)$$
$$\approx \text{NFE}(M_1) - \text{NFE}(M_0) \qquad (56)$$

where $p(\mathbf{y}_{1:T}|M_k)$ is the evidence of model class M_k. Since the evidence cannot be calculated in closed form, we used its lower bound approximation in terms of the (variational) Negative Free Energy, $p(\mathbf{y}_{1:T}|M_k) \approx \text{NFE}(M_k)$.

5) DISCRIMINATIVE POWER SCORE

Given two models (of the same model class) each defined by the set of parameters $\mathbf{\Theta}_0$ (model 0) or $\mathbf{\Theta}_1$, (model 1), and given a data sequence $\mathbf{y}_{(1:T)}$ that was generated by the model 0, the Discriminative Power Score is defined in terms of the directed divergence [52]

$$\text{DPS}(H_0, H_1) = \lim_{T \to \infty} \frac{1}{T}\left(\log \frac{p(\mathbf{y}_{1:T}|\mathbf{\Theta}_1, M)}{p(\mathbf{y}_{1:T}|\mathbf{\Theta}_0, M)}\right) \quad (57)$$

where $p(\mathbf{y}_{1:T}|\mathbf{\Theta}_k, M)$ is the probability that the data sequence was generated by model k (i.e. the model with parameters $\mathbf{\Theta}_k$), after marginalization over all possible brain state paths. For a sufficiently long sequence, if H_0 and H_1 are the hypotheses that the data sequence was generated by model 0 and model 1, respectively, $\text{DPS}(H_0, H_1)$ is the average information per data sample for discrimination of H_0 against H_1. This divergence is always positive and is zero if the two models are the same ($\mathbf{\Theta}_0 = \mathbf{\Theta}_1$).

ACKNOWLEDGMENT

REFERENCES

[1] B. Hunyadi, M. W. Woolrich, A. J. Quinn, D. Vidaurre, and M. De Vos, "A dynamic system of brain networks revealed by fast transient EEG fluctuations and their fMRI correlates," *NeuroImage*, vol. 185, pp. 72–82, Jan. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S105381191831958X

[2] C. M. Michel and T. Koenig, "EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: A review," *NeuroImage*, vol. 180, pp. 577–593, Oct. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S10538 1191731008X

[3] A. P. Baker, M. J. Brookes, I. A. Rezek, S. M. Smith, T. Behrens, P. J. P. Smith, and M. Woolrich, "Fast transient networks in spontaneous human brain activity," *eLife*, vol. 3, pp. 1–18, Mar. 2014.

[4] M. W. Woolrich, A. Baker, H. Luckhoo, H. Mohseni, G. Barnes, M. Brookes, and I. Rezek, "Dynamic state allocation for MEG source reconstruction," *NeuroImage*, vol. 77, pp. 77–92, Aug. 2013. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S10 53811913002851

[5] I. Olier, N. J. Trujillo-Barreto, and W. El-Deredy, "A switching multi-scale dynamical network model of EEG/MEG," *NeuroImage*, vol. 83, pp. 262–287, Dec. 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811913003935

[6] D. Vidaurre, R. Abeysuriya, R. Becker, A. J. Quinn, F. Alfaro-Almagro, S. M. Smith, and M. W. Woolrich, "Discovering dynamic brain networks from big data in rest and task," *NeuroImage*, vol. 180, pp. 646–656, Oct. 2018, doi: 10.1016/j.neuroimage.2017.06.077.

[7] R. Becker, D. Vidaurre, A. J. Quinn, R. G. Abeysuriya, O. P. Jones, S. Jbabdi, and M. W. Woolrich, "Transient spectral events in resting state MEG predict individual task responses," *NeuroImage*, vol. 215, Jul. 2020, Art. no. 116818.

[8] M. G. Preti, T. A. Bolton, and D. Van De Ville, "The dynamic functional connectome: State-of-the-art and perspectives," *NeuroImage*, vol. 160, pp. 41–54, Oct. 2017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1053811916307881

[9] P. Ritter, V. K. Jirsa, A. R. McIntosh, and M. Breakspear, "Editorial: State-dependent brain computation," *Frontiers Comput. Neurosci.*, vol. 9, Jun. 2015, Art. no. 77. [Online]. Available: http://journal.frontiersin.org/Article/10.3389/fncom.2015.00077/abstract

[10] J. A. Roberts, L. L. Gollo, R. G. Abeysuriya, G. Roberts, P. B. Mitchell, M. W. Woolrich, and M. Breakspear, "Metastable brain waves," *Nature Commun.*, vol. 10, no. 1, p. 1056, Mar. 2019. [Online]. Available: http://www.nature.com/articles/s41467-019-08999-0

[11] J. Schumacher, L. R. Peraza, M. Firbank, A. J. Thomas, M. Kaiser, P. Gallagher, J. T. O'Brien, A. M. Blamire, and J.-P. Taylor, "Dysfunctional brain dynamics and their origin in Lewy body dementia," *Brain*, vol. 142, no. 6, pp. 1767–1782, Jun. 2019.

[12] E. C. A. Hansen, D. Battaglia, A. Spiegler, G. Deco, and V. K. Jirsa, "Functional connectivity dynamics: Modeling the switching behavior of the resting state," *NeuroImage*, vol. 105, pp. 525–535, Jan. 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S10538 11914009033

[13] G. C. O'Neill, P. K. Tewarie, G. L. Colclough, L. E. Gascoyne, B. A. E. Hunt, P. G. Morris, M. W. Woolrich, and M. J. Brookes, "Measurement of dynamic task related functional networks using MEG," *NeuroImage*, vol. 146, pp. 667–678, Feb. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811916304530

[14] M. Hassan, P. Benquet, A. Biraben, C. Berrou, O. Dufor, and F. Wendling, "Dynamic reorganization of functional brain networks during picture naming," *Cortex*, vol. 73, pp. 276–288, Dec. 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0010945215003251

[15] A. Khanna, A. Pascual-Leone, C. M. Michel, and F. Farzan, "Microstates in resting-state EEG: Current status and future directions," *Neurosci. Biobehavioral Rev.*, vol. 49, pp. 105–113, Feb. 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0149763414003492

[16] A. Mheich, M. Hassan, M. Khalil, C. Berrou, and F. Wendling, "A new algorithm for spatiotemporal analysis of brain functional connectivity," *J. Neurosci. Methods*, vol. 242, pp. 77–81, Mar. 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0165027015000035

[17] G. C. O'Neill, P. Tewarie, D. Vidaurre, L. Liuzzi, M. W. Woolrich, and M. J. Brookes, "Dynamics of large-scale electrophysiological networks: A technical review," *NeuroImage*, vol. 180, pp. 559–576, Oct. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811917308169

[18] R. Hindriks, M. H. Adhikari, Y. Murayama, M. Ganzetti, D. Mantini, N. K. Logothetis, and G. Deco, "Can sliding-window correlations reveal dynamic functional connectivity in resting-state fMRI?" *NeuroImage*, vol. 127, pp. 242–256, Feb. 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1053811915010782

[19] D. Vidaurre, A. J. Quinn, A. P. Baker, D. Dupret, A. Tejero-Cantero, and M. W. Woolrich, "Spectrally resolved fast transient brain states in electrophysiological data," *NeuroImage*, vol. 126, pp. 81–95, Feb. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811915010691

[20] E. B. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, "A sticky HDP-HMM with application to speaker diarization," *Ann. Appl. Statist.*, vol. 5, no. 2A, pp. 1020–1056, Jun. 2011. [Online]. Available: https://projecteuclid.org/journals/annals-of-applied-statistics/volume-5/issue-2A/A-sticky-HDP-HMM-with-application-to-speaker-diarization/10.1214/10-AOAS395.full

[21] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989. [Online]. Available: http://ieeexplore.ieee.org/document/18626/

[22] D. Lehmann, H. Ozaki, and I. Pal, "EEG alpha map series: Brain microstates by space-oriented adaptive segmentation," *Electroencephalogr. Clin. Neurophysiol.*, vol. 67, no. 3, pp. 271–288, Sep. 1987. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0013469487900253

[23] J. Cabral, H. Luckhoo, M. Woolrich, M. Joensson, H. Mohseni, A. Baker, M. L. Kringelbach, and G. Deco, "Exploring mechanisms of spontaneous functional connectivity in MEG: How delayed network interactions lead to structured amplitude envelopes of band-pass filtered oscillations," *NeuroImage*, vol. 90, pp. 423–435, Apr. 2014.

[24] J. Britz, D. Van De Ville, and C. M. Michel, "BOLD correlates of EEG topography reveal rapid resting-state network dynamics," *NeuroImage*, vol. 52, no. 4, pp. 1162–1170, Oct. 2010.

[25] M. Breakspear, "Dynamic models of large-scale brain activity," *Nature Neurosci.*, vol. 20, no. 3, pp. 340–352, Mar. 2017. [Online]. Available: http://www.nature.com/articles/nn.4497

[26] F. Freyer, J. A. Roberts, R. Becker, P. A. Robinson, P. Ritter, and M. Breakspear, "Biophysical mechanisms of multistability in resting-state cortical rhythms," *J. Neurosci.*, vol. 31, no. 17, pp. 6353–6361, Apr. 2011, doi: 10.1523/JNEUROSCI.6693-10.2011

[27] L. Cocchi, L. L. Gollo, A. Zalesky, and M. Breakspear, "Criticality in the brain: A synthesis of neurobiology, models and cognition," *Prog. Neurobiol.*, vol. 158, pp. 132–152, Nov. 2017.

[28] M. Gschwind, C. M. Michel, and D. Van De Ville, "Long-range dependencies make the difference—Comment on 'a stochastic model for EEG microstate sequence analysis,'" *NeuroImage*, vol. 117, pp. 449–455, Aug. 2015.

[29] J. A. Roberts, T. W. Boonstra, and M. Breakspear, "The heavy tail of the human brain," *Current Opinion Neurobiol.*, vol. 31, pp. 164–172, Apr. 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0959438814002141

[30] D. Van De Ville, J. Britz, and C. M. Michel, "EEG microstate sequences in healthy humans at rest reveal scale-free dynamics," *Proc. Nat. Acad. Sci. USA*, vol. 107, no. 42, pp. 18179–18184, Oct. 2010, doi: 10.1073/pnas.1007841107.

[31] S. J. Gershman, E. Vul, and J. B. Tenenbaum, "Multistability and perceptual inference," *Neural Comput.*, vol. 24, no. 1, pp. 1–24, Jan. 2012. [Online]. Available: https://direct.mit.edu/neco/article/24/1/1-24/7724

[32] A. Klein and J. Tourville, "101 labeled brain images and a consistent human cortical labeling protocol," *Frontiers Neurosci.*, vol. 6, p. 171, Dec. 2012. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnins.2012.00171

[33] M. Wang, D. Arteaga, and B. J. He, "Brain mechanisms for simple perception and bistable perception," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 35, pp. E3350–E3359, Aug. 2013, doi: 10.1073/pnas.1221945110.

[34] T. Nakamura, K. Kiyono, K. Yoshiuchi, R. Nakahara, Z. R. Struzik, and Y. Yamamoto, "Universal scaling law in human behavioral organization," *Phys. Rev. Lett.*, vol. 99, no. 13, Sep. 2007, Art. no. 138103, doi: 10.1103/PhysRevLett.99.138103.

[35] S.-Z. Yu, "Hidden semi-Markov models," *Artif. Intell.*, vol. 174, no. 2, pp. 215–243, Feb. 2010. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0004370209001416

[36] N. Hudson, "Inference in hybrid systems with applications in neural prosthetics," Ph.D. thesis, California Inst. Technol., Pasadena, CA, USA, 2009. [Online]. Available: http://thesis.library.caltech.edu/5174/

[37] S. Z. Yu, *Hidden Semi-Markov Models: Theory, Algorithms and Applications*. Amsterdam, The Netherlands: Elsevier, 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/C20140025087

[38] K. Murphy, "Hidden semi-Markov models (HSMMs)," Univ. British Columbia, Tech. Rep., Nov. 2002. [Online]. Available: http://www.cs.ubc.ca/~murphyk/Papers/segment.pdf

[39] M. Yaesoubi, R. L. Miller, and V. D. Calhoun, "Mutually temporally independent connectivity patterns: A new framework to study the dynamics of brain connectivity at rest with application to explain group difference based on gender," *NeuroImage*, vol. 107, pp. 85–94, Feb. 2015, doi: 10.1016/j.neuroimage.2014.11.054.

[40] T. Rukat, A. Baker, A. Quinn, and M. Woolrich, "Resting state brain networks from EEG: Hidden Markov states vs. classical microstates," 2016, *arXiv:1606.02344*.

[41] Y. Guédon, "Exploring the state sequence space for hidden Markov and semi-Markov chains," *Comput. Statist. Data Anal.*, vol. 51, no. 5, pp. 2379–2409, Feb. 2007. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167947306002118

[42] M. J. Beal, "Variational algorithms for approximate Bayesian inference," Ph.D. thesis, Gatsby Comput. Neurosci. Unit, Univ. College London, U.K., 2003.

[43] M. Opper and D. Saad, *Advanced Mean Field Methods: Theory and Practice*. Cambridge, MA, USA: MIT Press, Jun. 2001, doi: 10.7551/mitpress/1100.001.0001.

[44] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, Mar. 1951. [Online]. Available: http://projecteuclid.org/euclid.aoms/1177729694

[45] N. J. Trujillo-Barreto, E. Aubert-Vázquez, and P. A. Valdés-Sosa, "Bayesian model averaging in EEG/MEG imaging," *NeuroImage*, vol. 21, no. 4, pp. 1300–1319, Apr. 2004, doi: 10.1016/j.neuroimage.2003.11.008.

[46] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, "Design and construction of a realistic digital brain phantom," *IEEE Trans. Med. Imag.*, vol. 17, no. 3, pp. 463–468, Jun. 1998. [Online]. Available: http://ieeexplore.ieee.org/document/712135/

[47] G. Huiskamp, "Difference formulas for the surface Laplacian on a triangulated surface," *J. Comput. Phys.*, vol. 95, no. 2, pp. 477–496, Aug. 1991. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/002199919190286T

[48] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann, "Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain," *Int. J. Psychophysiol.*, vol. 18, no. 1, pp. 49–65, Oct. 1994.

[49] F. von Wegner, E. Tagliazucchi, and H. Laufs, "Information-theoretical analysis of resting state EEG microstate sequences–non-Markovianity, non-stationarity and periodicities," *NeuroImage*, vol. 158, pp. 99–111, Sep. 2017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1053811917305347

[50] D. M. Endres and J. E. Schindelin, "A new metric for probability distributions," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1858–1860, Jul. 2003. [Online]. Available: http://ieeexplore.ieee.org/document/1207388/

[51] R. E. Kass and A. E. Raftery, "Bayes factors," *J. Amer. Stat. Assoc.*, vol. 90, no. 430, pp. 773–795, Jun. 1995, doi: 10.1080/01621459.1995.10476572.

[52] B.-H. Juang and L. R. Rabiner, "A probabilistic distance measure for hidden Markov models," *AT&T Tech. J.*, vol. 64, no. 2, pp. 391–408, Feb. 1985 doi: 10.1002/j.1538-7305.1985.tb00439.x.

[53] K. J. Friston, K. M. Stephan, J. D. Heather, C. D. Frith, A. A. Ioannides, L. C. Liu, M. D. Rugg, J. Vieth, H. Keber, K. Hunter, and R. S. J. Frackowiak, "A multivariate analysis of evoked responses in EEG and MEG data," *NeuroImage*, vol. 3, no. 3, pp. 167–174, Jun. 1996. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S105381199690018X

[54] F. von Wegner and H. Laufs, "Information-theoretical analysis of EEG microstate sequences in Python," *Frontiers Neuroinform.*, vol. 12, p. 30, Jun. 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/fninf.2018.00030/full

[55] K. Kim, N. T. Duc, M. Choi, and B. Lee, "EEG microstate features for schizophrenia classification," *PLoS ONE*, vol. 16, no. 5, May 2021, Art. no. e0251842.

[56] C. M. Bishop, *Pattern Recognition and Machine Learning* (Information Science and Statistics). Berlin, Germany: Springer-Verlag, 2006.

[57] K. Linkenkaer-Hansen, V. V. Nikouline, J. M. Palva, and R. J. Ilmoniemi, "Long-range temporal correlations and scaling behavior in human brain oscillations," *J. Neurosci.*, vol. 21, no. 4, pp. 1370–1377, Feb. 2001, doi: 10.1523/JNEUROSCI.21-04-01370.2001.

[58] G. Buzsáki and K. Mizuseki, "The log-dynamic brain: How skewed distributions affect network operations," *Nature Rev. Neurosci.*, vol. 15, no. 4, pp. 264–278, Apr. 2014. [Online]. Available: http://www.nature.com/articles/nrn3687

[59] W. J. Freeman, "Deep analysis of perception through dynamic structures that emerge in cortical activity from self-regulated noise," *Cognit. Neurodynamics*, vol. 3, no. 1, pp. 105–116, Mar. 2009. [Online]. Available: http://link.springer.com/10.1007/s11571-009-9075-3

[60] C. T. Kello, G. D. A. Brown, R. Ferrer-i-Cancho, J. G. Holden, K. Linkenkaer-Hansen, T. Rhodes, and G. C. Van Orden, "Scaling laws in cognitive sciences," *Trends Cogn. Sci.*, vol. 14, no. 5, pp. 223–232, May 2010. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S136466131000046X

[61] D. Vidaurre, S. M. Smith, and M. W. Woolrich, "Brain network dynamics are hierarchically organized in time," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 48, pp. 12827–12832, Nov. 2017, doi: 10.1073/pnas.1705120114.

[62] J. P. Borst and J. R. Anderson, "The discovery of processing stages: Analyzing EEG data with hidden semi-Markov models," *NeuroImage*, vol. 108, pp. 60–73, Mar. 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1053811914010210

[63] M. Laganaro, "Inter-study and inter-individual consistency and variability of EEG/ERP microstate sequences in referential word production," *Brain Topography*, vol. 30, no. 6, pp. 785–796, Nov. 2017. [Online]. Available: http://link.springer.com/10.1007/s10548-017-0580-0

[64] G. Navarro, "A guided tour to approximate string matching," *ACM Comput. Surv.*, vol. 33, no. 1, pp. 31–88, Mar. 2001, doi: 10.1145/375360.375365.

ALAND ASTUDILLO received the B.Sc. and M.Sc. degrees in biomedical engineering and the Ph.D. degree in biophysics and computational biology from Universidad de Valparaíso, Valparaíso, Chile, in 2012 and 2023, respectively. Since 2012, he has been working in neuroscience research projects and artificial intelligence applications. His research interests include brain dynamics, cognition, computational neuroscience, artificial intelligence, and machine learning.

NELSON J. TRUJILLO-BARRETO received the B.Sc. degree (Hons.) in nuclear physics from the Higher Institute for Nuclear Science and Technology of Havana, Cuba, in 1995, and the Ph.D. degree in physical sciences from Havana University, in 2006. From 1995 to 2014, he was with the Cuban Neuroscience Center, where he was the Head of the Department for Brain Dynamics. He is currently an Assistant Professor in computational neuroscience with the School of Health Sciences, The University of Manchester, U.K. His research interests include probabilistic and computational modeling of brain function, development of novel methods of analysis of brain imaging data (EEG, MEG, and fMRI), and application of these developments to clinical and non-clinical domains.

DAVID ARAYA GALVEZ received the B.Sc. and M.Sc. degrees in electrical engineering from Universidad de Chile, Santiago, Chile, in 2005, and the Ph.D. degree in biophysics and computational biology from Universidad de Valparaíso, Valparaíso, Chile, in 2023. His Ph.D. research was dedicated to the early detection of psychiatric disorders through brain markers. Currently, he applies artificial intelligence and computational neuroscience with the Institute of Technologies and Innovation for Health and Wellbeing (ITSB), Universidad Andres Bello, Viñal del Mar, Chile.

WAEL EL-DEREDY received the B.Sc. degree in electrical and electronics engineering from Ain Shams University, Cairo, Egypt, the M.Sc. degree in electrical and electronics engineering from the Imperial College London, London, U.K., and the Ph.D. degree in neurocomputing from University College London, London. He is currently a Professor in computational neuroscience with Universidad de Valparaíso and an Associate Researcher with the Advanced Center for Electrical and Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile. His research interests include probabilistic models of brain function, the development of novel methods for neuroimaging and application for neuroengineering, and neuroprostatics. His research is supported by the MRC and EPSRC in the U.K. and CONICYT in Chile.

• • •