Multitaper Infinite Hidden Markov Model for EEG

Andrew H. Song*,[1], *Student Member, IEEE*, Leon Chlon*,[2], Hugo Soulat[3], John Tauber[2], Sandya Subramanian[4], *Student Member, IEEE*, Demba Ba[5], *Member, IEEE*, and Michael J. Prerau[3]

***Abstract*— Electroencephalographam (EEG) monitoring of neural activity is widely used for identifying underlying brain states. For inference of brain states, researchers have often used Hidden Markov Models (HMM) with a fixed number of hidden states and an observation model linking the temporal dynamics embedded in EEG to the hidden states. The use of fixed states may be limiting, in that 1) pre-defined states might not capture the heterogeneous neural dynamics across individuals and 2) the oscillatory dynamics of the neural activity are not directly modeled. To this end, we use a Hierarchical Dirichlet Process Hidden Markov Model (HDP-HMM), which discovers the set of hidden states that best describes the EEG data, without a-priori specification of state number. In addition, we introduce an observation model based on classical asymptotic results of frequency domain properties of stationary time series, along with the description of the conditional distributions for Gibbs sampler inference. We then combine this with multitaper spectral estimation to reduce the variance of the spectral estimates. By applying our method to simulated data inspired by sleep EEG, we arrive at two main results: 1) the algorithm faithfully recovers the spectral characteristics of the true states, as well as the right number of states and 2) the incorporation of the multitaper framework produces a more stable estimate than traditional periodogram spectral estimates.**

***Index Terms*— Infinite Hidden Markov Model; Beam Sampler; Multitaper Spectral Analysis; State-Space Modeling**

I. INTRODUCTION

The electroencephalogram (EEG) is a high temporal-resolution means of observing the oscillatory electrical activity of cortical and sub-cortical brain networks at the scalp. Thus, neuroscientists have long used the EEG as a means of differentiating between different brain states through the characterization of different temporal components within the signal. In particular, spectral analysis of EEG oscillatory dynamics has been a vital tool in the discovery and analysis of neurophysiological activity [1].

Another approach to EEG analysis casts the data as a realization of a generative model, which involves fitting a Hidden Markov Model (HMM). Typically, the number of states in the HMM is defined by historical convention based on features that can be observed by a clinician in the time-domain, rather than by empiricial means. The use of clinical states alone is highly restrictive, thus it is more principled to use a framework that can automatically determine the number of states and their signatures based on the data. Previous observation models for the HMM, which map the observations to the states, range from auto-regressive models [2] to automatic feature extraction by a Deep Belief Network [3]. However, spectral observation models, which leverage the oscillatory nature of the EEG, may be more appropriate, and also possess beneficial theoretical properties [4].

We therefore propose a principled and empirical approach to HMM estimation for EEG data that: 1) has the capacity to discover the optimal number of states 2) has a frequency-domain observation model, 3) has optimal variance reduction in the spectral estimate.

To address 1) and 2), we propose a Bayesian nonparametric framework that integrates the Hierarchical Dirichlet Process Hidden Markov Model (HDP-HMM) [5] for the underlying state dynamics, with an observation model utilizing the asymptotic properties of wide-sense stationary (WSS) time series. To address 3), the multitaper spectral estimation framework [6] is incorporated into the generative model. We show that the multitaper framework, which reduces variance by averaging multiple orthogonal projections of the data, leads to a more stable recovery of the underlying states, compared to the same generative model with no tapers.

For the inference procedure of the proposed generative model, we extend the existing package of the Beam sampler [7], a variant of the Gibbs sampler. Specifically, we derive the conditional probability for the spectral estimates, thus incorporating our observation models into the existing sampler.

II. MODEL

Let $\{y_t\}_t$ be a realization of a wide-sense stationary (WSS) stochastic process. We will use several properties of WSS for parameter estimation. Wide-sense stationarity, or second-order stationarity, states that the mean and the autocovariance of the stochastic process are invariant with respect to time t. In practice, since physiological data are nonstationary, we segment the data into multiple non-overlapping windows, within which we assume WSS.

A. Spectral Representation for WSS Time Series

To analyze the time series sampled at rate F_s in the frequency domain, we use the Discrete Fourier Transform (DFT). With the sampled time series as $\mathbf{y} = [\mathbf{y}_1, \cdots, \mathbf{y}_J] \in \mathbb{R}^J$ and the DFT coefficients as $\mathbf{y}^{(F)} \in \mathbb{C}^J$, we have $\mathbf{y}^{(F)}(w_j) = \frac{1}{J}\sum_{l=1}^{J} \mathbf{y}_l \exp\big(-i2\pi\frac{(j-1)}{J}(l-1)\big)$, where $w_j = \frac{j-1}{J}$ for $j = 1, \cdots, \frac{J}{2}$, is the normalized frequency. [1]

* The authors contributed equally
[1] Electrical Engineering and Computer Science, MIT
[2] Brain and Cognitive Sciences, MIT
[3] MGH Department of Anesthesia, Critical Care and Pain Medicine
[4] Health Sciences and Technology, MIT
[5] School of Engineering and Applied Sciences, Harvard University

This work was supported by National Institute Of Neurological Disorders And Stroke Grant R01 NS-096177 (M.J.P.)

[1] The conversion between f_j, frequency in Hz, and w_j is $f_j = F_s w_j$.

For convenience we use $d_j^{(R)} = Re\{\mathbf{y}^{(F)}(w_j)\}$ and $d_j^{(I)} = Im\{\mathbf{y}^{(F)}(w_j)\}$. We introduce the following proposition [4]:

Proposition 1. If $\mathbf{y}$ is a WSS time series, the DFT coefficients, $d_j^{(R)}$ and $d_j^{(I)}$, are distributed as *asymptotically independent normal*, as $J \to \infty$,

$$d_j^{(R)} \sim \mathcal{N}\left(0, \frac{f(w_j)}{2}\right) \text{ and } d_j^{(I)} \sim \mathcal{N}\left(0, \frac{f(w_j)}{2}\right) \quad (1)$$

where $f(w_j)$ is the true underlying Power Spectral Density (PSD) of the stochastic process at w_j. The independence property is the guiding principle for our observation model.

B. Tapers and Multitaper Spectral Estimation

The estimation of the true PSD (spectral estimation), $f(w_j)$, from the observations (DFT coefficients), has received great attention. The simplest estimate, albeit with high bias and variance, is $\widehat{f}(w_j) = \|\mathbf{y}^{(F)}(w_j)\|^2$, the *periodogram*. For reduction of the bias and variance, a taper function, $\mathbf{h}^{(m)} = [\mathbf{h}_1^{(m)}, \cdots, \mathbf{h}_J^{(m)}] \in \mathbb{R}^J$, with m as a taper index, is usually multiplied by the observation, yielding $\mathbf{y}^{(m,F)}(w_j) = \frac{1}{J}\sum_{l=1}^{J} \mathbf{h}_l^{(m)} \mathbf{y}_l \exp\Big(-i2\pi\frac{(j-1)}{J}(l-1)\Big)$.

We use $d_j^{(m,R)}$ and $d_j^{(m,I)}$ to denote real and imaginary parts of $\mathbf{y}^{(m,F)}(w_j)$, respectively. With tapers, we can extend Proposition 1 to characterize the asymptotic relationships between the tapered coefficients [4]:

Corollary 1. Under regular assumptions for the tapers and the same assumptions as Proposition 1, the following holds for two different tapers $\mathbf{h}^{(m_1)}$ and $\mathbf{h}^{(m_2)}$,

$$\begin{pmatrix} d_j^{(m_1,*)} \\ d_j^{(m_2,*)} \end{pmatrix} \sim \mathcal{N}\left(0, \frac{f(w_j)}{2}\begin{pmatrix} \langle \mathbf{h}^{(m_1)}, \mathbf{h}^{(m_1)} \rangle & \langle \mathbf{h}^{(m_1)}, \mathbf{h}^{(m_2)} \rangle \\ \langle \mathbf{h}^{(m_1)}, \mathbf{h}^{(m_2)} \rangle & \langle \mathbf{h}^{(m_2)}, \mathbf{h}^{(m_2)} \rangle \end{pmatrix}\right) \quad (2)$$

where * denotes either R or I, and the inner product is given as $\langle \mathbf{h}, \mathbf{g} \rangle = \sum_l \mathbf{h}_l \mathbf{g}_l$.

We focus on the discrete prolate spheroidal sequences (DPSS), which are known to be mutually orthogonal with optimal energy concentration properties. In the rest of the paper, we use $\mathbf{h}^{(m)}$ for $m = 1, \cdots, M$, to denote the m^{th} DPSS taper and M the total number of appropriate DPSS tapers, chosen according to other parameters [8]. The DPSS tapered spectral estimates, $\{\|\mathbf{y}^{(m,F)}(w_j)\|^2\}_{m=1}^M$, are approximately uncorrelated and thus can be averaged to obtain spectral estimate $\widehat{f}_{MT}(w_j)$ which can be shown to optimize the reduction in bias and variance, $\widehat{f}_{MT}(w_j) = \frac{1}{M}\sum_{m=1}^{M} \|\mathbf{y}^{(m,F)}(w_j)\|^2$.

C. Generative model

As in the general HMM framework, our generative model consists of two components: 1) the HDP-HMM component, which governs the state transitions and the PSD of each state, and 2) the observation model, denoted as the *(multitaper) spectral emission model*, which governs the generation of time series based on each state.

Let N be the total number of samples in the time series and J the number of samples in each window. Then, T= $\lfloor \frac{N}{J} \rfloor$ denotes the number of windows for the data. We define $\mathbf{y}_t = \{\mathbf{y}_{(t-1)J}, \cdots, \mathbf{y}_{tJ-1}\} \in \mathbb{R}^J$ as the samples in the window $t = 1, \cdots, T$.

We model $\{\mathbf{y}_t^{(F)}(w_j)\}_{t=1}^T$ (or $\{\mathbf{y}_t^{(m,F)}(w_j)\}_{t=1}^T$ for multitaper) as the realization of a generative process driven by latent states $\{s_t\}_{t=1}^T \in \{1, ..., K\}$, where K is the number of discovered states. Finally, $\boldsymbol{\pi}_k = \{\pi_{k,k'}\}_{k'=1}^K$ represents the transition probability from state k to k'.

1) HDP-HMM: The key distinction between the HDP-HMM and the conventional finite state HMM is how the total number of states K is treated. For the HDP-HMM, K is allowed to change throughout the inference procedure, whereas for the finite state HMM, K is fixed to a prespecified number. The full description of the HDP-HMM for our generative model is as follows, for $t = 1, \cdots, T$, $j = 1, \cdots, \frac{J}{2}$, and $k = 1, \cdots, K$.

$$\begin{aligned} &\boldsymbol{\beta} \sim \text{GEM}(\gamma), \quad \boldsymbol{\pi}_k | \beta \sim \text{DP}(\alpha, \boldsymbol{\beta}), \quad b_j^{(s_t)} \sim \Gamma(1,1) \\ &s_t | s_{t-1} \sim \text{Multinomial}(\pi_{s_{t-1}}), \quad f(w_j^{(s_t)}) \sim \text{IG}(a, b_j^{(s_t)}) \end{aligned} \quad (3)$$

where GEM, DP, Γ, and IG denotes the Griffiths-Engen-McCloskey distribution, Dirichlet Process, Gamma, and Inverse Gamma distribution, respectively.

We assume that each state $s_t = 1, \cdots, K$ is parameterized by its PSD, in a diagonal matrix form, $\Sigma_{s_t} = \frac{1}{2}\text{diag}\Big(f(w_1^{(s_t)}), f(w_1^{(s_t)}), \cdots, f(w_{\frac{J}{2}}^{(s_t)}), f(w_{\frac{J}{2}}^{(s_t)})\Big)$, where each element is repeated twice to account for real and the imaginary parts. The PSD for for frequency j of state s_t, $f\big(w_j^{(s_t)}\big)$, is generated from $\text{IG}(a, b_j^{(s_t)})$. The Inverse Gamma is a conjugate prior of the normal likelihood of the observations, further specified in the observation model. To capture different frequency-level characteristics, we place an uninformative Gamma prior over the hyperparameter $b_j^{(s_t)}$.

For the state transitions, we follow the standard HDP-HMM convention. We define $\boldsymbol{\beta} = (\cdots, \beta_k, \cdots)$, where $\sum_{k=1}^{\infty} \beta_k = 1$, as the prior on the state transition distribution, from which $\{\boldsymbol{\pi}_k\}_{k=1}^K$ are sampled. $\boldsymbol{\beta} \sim \text{GEM}(\gamma)$ denotes the hyperprior on $\boldsymbol{\beta}$, referred to as the stick-breaking construction, given by $\beta_k = \beta_k' \prod_{i=1}^{k-1}(1-\beta_i')$ and $\beta_k' \sim \text{Beta}(1, \gamma)$. This provides a systematic approach of assigning values to β_k as $k \to \infty$, while ensuring that $\sum_{k=1}^{\infty} \beta_k = 1$.

In practice, $K = \infty$ is never realized for tractable inference. At each step of the inference procedure, for number of the discovered states K, $\beta_{k \geq K+1}$ are merged together, so that $\beta_{K+1} = \sum_{k=K+1}^{\infty} \beta_k$ and $\sum_{k=1}^{K+1} \beta_k = 1$. The presence of β_{K+1} enables exploring a new state other than previously discovered K states, during the inference step.

2) Spectral Emission Model: We use Eq.(1) as the generative distribution for the observations given the state s_t, denoted as $p\Big(\mathbf{y_t}^{(F)}\Big|\Sigma_{s_t}\Big)$,

$$\begin{aligned} &\big[d_{1,t}^{(R)}, d_{1,t}^{(I)}, \cdots, d_{\frac{J}{2},t}^{(R)}, d_{\frac{J}{2},t}^{(I)}\big] \\ &\sim \mathcal{N}(\bar{\mathbf{0}}, \Sigma_{s_t}) = \prod_{j=1}^{\frac{J}{2}} \left\{\mathcal{N}\left(0, \frac{f(w_j^{(s_t)})}{2}\right)\right\}^2 \end{aligned} \quad (4)$$

Note that the diagonal structure is due to Proposition 1. In practice, we only use $\{w_j\}$ that are of interest.

3) Multitaper Spectral Emission Model: We exploit the mutual orthogonality of the DPSS and **Corollary 1** to develop an extension of the spectral emission model. Without loss of generality, we assume that all the tapers are normalized, such that $\langle \mathbf{h}_l^{(m)}, \mathbf{h}_l^{(m)} \rangle = 1$. Due to orthogonality, the off-diagonal terms in the covariance matrix of Eq.(2) is zero, $\langle \mathbf{h}_l^{(m_1)}, \mathbf{h}_l^{(m_2)} \rangle = 0$, for different DPSS tapers m_1 and m_2. Since *uncorrelatedness* implies *independence* for multivariate normal random variables, $\mathbf{y}_t^{(m,F)}$, can be treated as *independent* observations generated by the same covariance matrix Σ_{s_t}. This can simply be expressed as $p\Big(\mathbf{y}_t^{(m,F)}\Big|\Sigma_{s_t}\Big) = \mathcal{N}(\bar{\mathbf{0}}, \Sigma_{s_t})$, for $m = 1, \cdots, M$. In other words, the multitaper framework essentially produces M independent observations, compared to the conventional approach, with a single observation.

III. Inference via Beam Sampler

The beam sampler is an extension of the Gibbs sampler adapted for sampling the whole state trajectories with possibly infinite number of states. To overcome the challenge of sampling a state trajectory, an auxiliary variable is introduced to effectively upper bound the set of dynamically computed state trajectories (full details outlined in [7]).

To complete the inference procedure, we need to compute the posterior distribution of Σ_{s_t} or equivalently $f(w_j^{(s_t)})$, based on the spectral emission model. We use the following facts: 1) the conditional distribution is also an Inverse Gamma distribution due to conjugacy and 2) there are M tapered observations. Using Bayes' rule, we have

$$
\begin{aligned}
&p\Big(f\big(w_j^{(k)}\big)\Big|\{d_{j,t}^{(m,R)}, d_{j,t}^{(m,I)}\}_{m,t=1}^{M,T}, a, b_j^{(k)}, \{s_t\}_{t=1}^{T}\Big) \\
&= \overbrace{p\Big(\{d_{j,t}^{(m,R)}, d_{j,t}^{(m,I)}\}_{m,t=1}^{M,T}\Big|f(w_j^{(k)}), \{s_t\}_{t=1}^{T}\Big)}^{\text{likelihood}} \\
&\quad \times \underbrace{p\Big(f(w_j^{(k)})\Big|a, b_j^{(k)}\Big)}_{\text{prior}} \\
&= \prod_{\substack{\{s_t=k\},\\ m=1}}^{T,M} p\Big(d_{j,t}^{(m,R)}\Big|f(w_j^{(k)})\Big)p\Big(d_{j,t}^{(m,I)}\Big|f(w_j^{(k)})\Big) IG\Big(a, b_j^{(k)}\Big) \\
&= \frac{1}{2}\mathrm{IG}\Bigg(a + M\sum_{t=1}^{T}\mathbf{1}_{\{s_t=k\}}, \\
&\quad b_j^{(k)} + \frac{1}{2}\sum_{t=1}^{T}\mathbf{1}_{\{s_t=k\}}\sum_{m=1}^{M}\Big\{\|\mathbf{y}_t^{(m,F)}(w_j)\|^2\Big\}\Bigg)
\end{aligned}
\tag{5}
$$

where $\mathbf{1}_{\{s_t=k\}}$ is an indicator function for $\{s_t = k\}$. The likelihood amounts to counting the number of occurrences for state k throughout the entire time series.

We emphasize that the form of the posterior is from the orthogonality of DPSS. If other sets of tapers are used and thus the non-diagonal terms in Eq.(2) are non-zero, the conjugate prior/posterior would have to be inverse-Wishart distribution. Furthermore, the elements of the sampled covariance matrix would need to constrained, such that the diagonal and off-diagonal term are proportional to the ratio of inner products between the tapers, which complicates sampling.

To summarize, the algorithm samples cyclically from the distributions of $\{s_t\}_{t=1}^{T}$, $\{\boldsymbol{\pi}_k\}_{k=1}^{K}$, $\boldsymbol{\beta}$, $\{\Sigma_k\}_{k=1}^{K}$, and the hyperparameters, conditioned on the current values of the other parameters.The empirical distribution for each parameter is constructed after discarding burn-in samples.

IV. Experiments

A. Dataset

We test our model on simulated data that mimics the spectral content of sleep EEG. Using $F_s = 250$Hz, $T = 2000$ windows of 15-second length were simulated and assigned to one of 5 sleep states (specified as S1-S5 in Figure 1). Each state is characterized by its PSD in different frequency bands of interest (FOI : slow 0.1-1Hz, delta 1-4Hz, theta 4-7Hz, alpha 8-12Hz, and sigma 12-15Hz) and by its probability to transition into another state (gathered in a realistic [1] transition matrix). For a given window t, the signal $\mathbf{y}_t \in \mathbb{R}^J$ is the sum of observation noise and independently modelled broad band time series [9] respectively peaking at 0.2, 3, 5, 10 and 14Hz, and accounting for each FOI.

B. Practical details

We run the beam sampler with multiple settings of hyperparameters, $\gamma \in [0.1, 1, 10]$ and $\alpha \in [0.1, 1, 10, 100]$ to test the robustness of the algorithm. For computational efficiency, we take median power of each frequency band (five in total), and use the DFT coefficients of the corresponding frequency as the observations. In addition, we set the hyperparameter for the PSD hyperprior as $a_0 = b_0 = 1$. We *burn-in* the first 2000 samples, and draw a subsequent 500 samples with a step size of 10 to minimize inter-sample autocorrelation. For each setting, we run the sampler twice, one with (no taper) spectral emission model and the other with multitaper spectral emission model.

V. Results

A. Recovery of the states

Across different sets of hyperparameters α and γ, the framework discovers between 3 and 5 states, with the no taper version recovering fewer states than the multitaper version for given hyperparameters, as shown in Table I.

To understand the relations between the recovered states and the true states, especially when fewer states are found, we examine two different aspects. First, the state trajectory in Figure 2 suggests that whenever the S1 (blue) and S2 (orange) occur, the no taper version identifies a single state (blue). Second, one of the recovered states, in Figure 1 top right panel, combines the characteristics of the PSD for the S1 and S2, while other recovered states precisely recover the PSD of S3-S5. Specifically, for the frequency bands where the power difference between S1 and S2 is large, the recovered PSD identifies middle value. Based on this, we conclude that when fewer states are recovered, some states are merged together, rather than one of the states not being discovered. With different hyperparameter settings and when three states are recovered, we consistently observe that S1 and S2 are merged together, as well as S3 and S4.

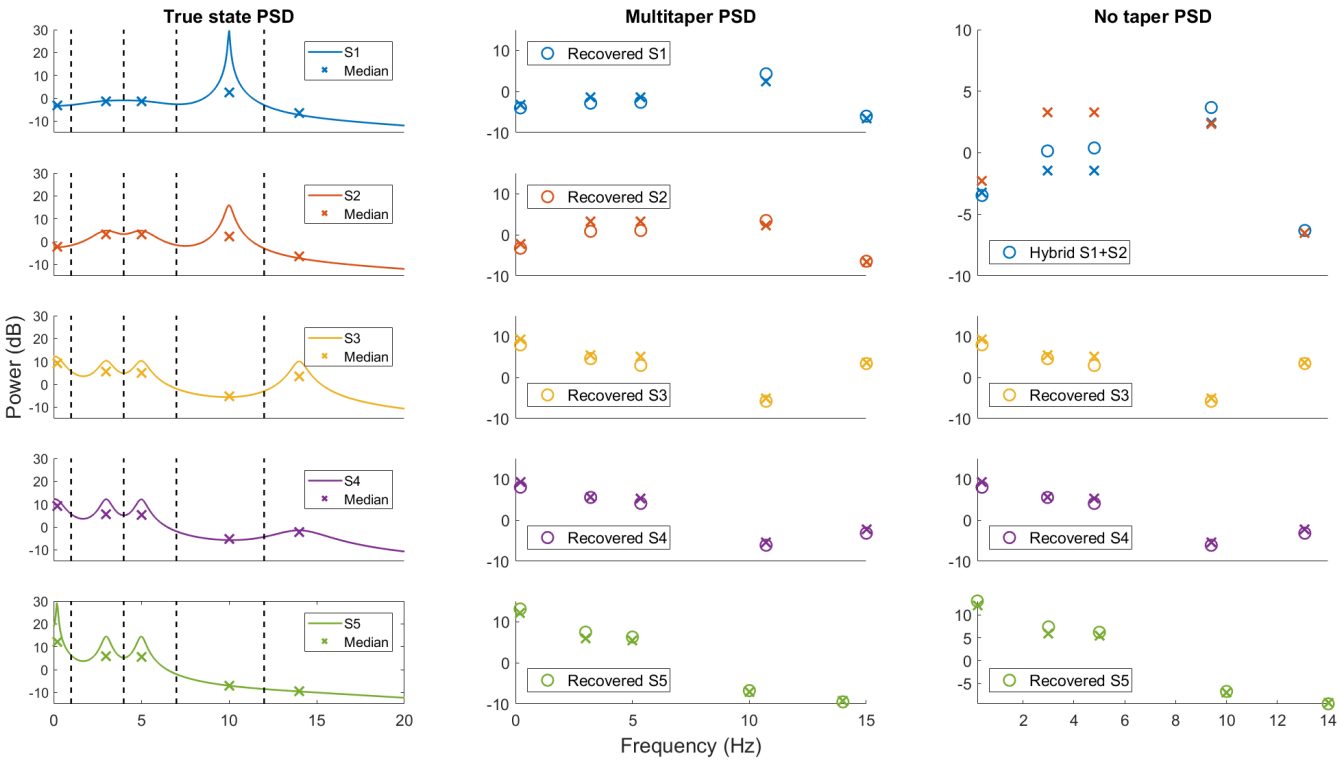

Fig. 1: With five true states, the multitaper version recovers exactly five states, whereas the no taper version recovers only four states, with S1 and S2 merged together. (Left) True PSD (Middle) Recovered PSD with multitaper (Right) Recovered PSD with no taper, both for $\alpha = \gamma = 1$. The markers 'x' and 'o' markers represent the median of true PSD in each frequency band of interest (specified by dotted lines) and the median of the recovered PSD across samples, respectively.

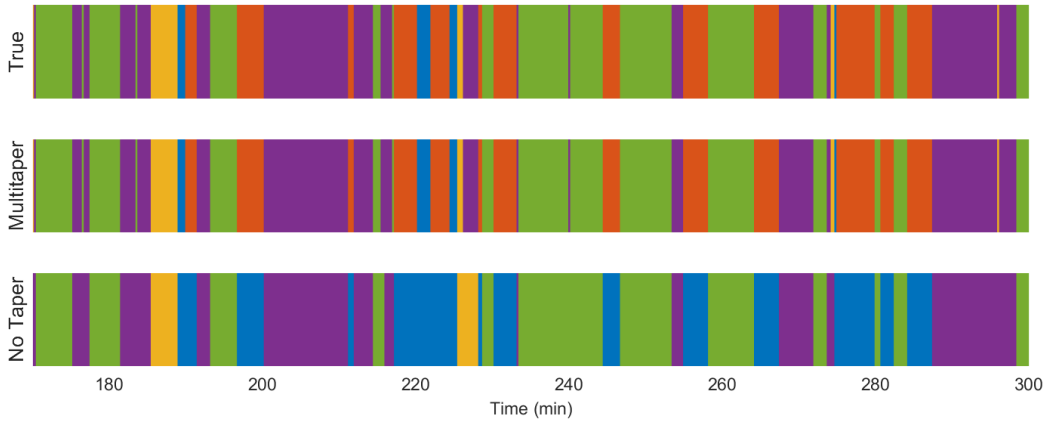

Fig. 2: The multitaper model (five states recovered) produces similar trajectory as the simulated data, whereas the no taper model (four states recovered) does not. Zoomed-in plot for sampled state trajectories ($\alpha = \gamma = 1$) and the simulated data.

Among several possible explanations for the merging phenomenon, such as the structure of the transition matrix and hyperparameters, we focus on the similarity between the PSDs of different states. Figure 3 shows the sum of absolute difference (ℓ_1 distance) between the PSD of the states, $\sum_{w_j} |f(w_j^{k_1}) - f(w_j^{k_2})|$ for $k_1, k_2 = 1, \cdots, 5$. We observe that S1 and S2, as well as S3 and S4 are close to each other in ℓ_1 distance. Indeed, these are the states that are merged together when fewer states are discovered. We also observed that a similar pattern holds for other metrics, such as the ℓ_2 distance and the Cross Spectral Density.

γ \ α	0.1	1	10	100
0.1	5 (5)	5 (5)	5 (5)	4 (3)
1	5 (5)	5 (5)	5 (4)	4 (3)
10	5 (4)	5 (4)	5 (4)	4 (3)

TABLE I: Number of states discovered by multitaper and no taper version (in parenthesis) for different α and γ

B. Comparison between different Spectral Emission models

Based on simulation, we observe two benefits of applying multiple tapers. First, for a given hyperparameter setting, the multitaper spectral emission model always discovers more states compared to the no taper version. This is explained by the lower variance of the spectral estimates, or the lower variance across covariance samples for the multitaper model. With higher variance, it is more likely that the PSD of two different states look similar, which leads to merging.

For further assessment, we examine the spread of the spectral estimates for the states recovered by both models. To also assess the bias, we focus on the hyperparameter settings that yield the recovery of the five states for both approaches. We then examine the difference between the spectral estimates and the true PSD of the corresponding state, shown in the Figure 4 for one of the states (S3).

As expected, the use of multiple tapers results in a smaller variance in the estimated PSD when compared to an approach without tapers and also slight decrease in bias, with the power difference closer to 0. The lower variance for the multitaper model is consistent across other recovered states, and also for the different hyperparameters where fewer states are recovered, although no conclusion can be made about the bias in these scenarios.

VI. CONCLUSIONS

In this work, we explored a generative model that explicitly uses the frequency characteristics of the EEG, with the flexibility of discovering the states without a-priori specification of the number of states. In addition, we show that incorporating the derivation based on asymptotic normality and the multitaper framework results in a stable recovery of the states, as well as lower variance of the spectral estimates compared to the approach with no tapers. This work serves as a proof-of-concept for this multitaper HDP-HMM model, providing the groundwork necessary for future studies involving experimental EEG data, and a comparison to related HDP-HMM [10] and finite state HMM methods. Overall, this novel framework can provide an empirical means of automating and increasing the dimensionality of existing clinical EEG analyses, as well as serve as a platform for EEG biomarker development.

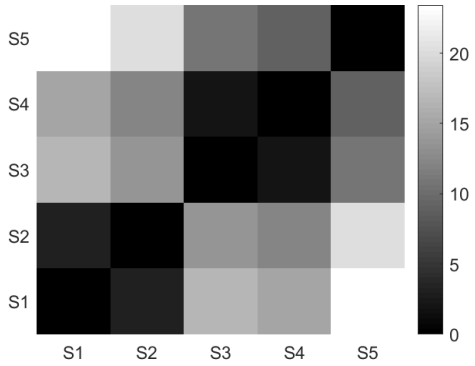

Fig. 3: The ℓ_1 distance between the PSDs of different states indicates that S1 and S2, and S3 and S4 are close to each other (dark color).

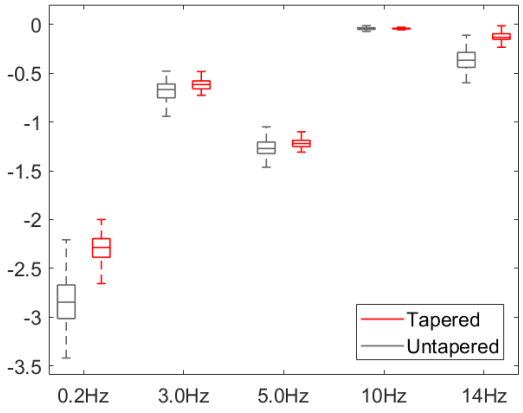

Fig. 4: Distribution of errors for the spectral estimate and the true PSD across samples for S3 shows that the multitapered version greatly reduces the variance.

REFERENCES

[1] Michael J. Prerau, Ritchie E. Brown, Matt T. Bianchi, Jeffrey M. Ellenbogen, and Patrick L. Purdon. Sleep Neurophysiological Dynamics Through the Lens of Multitaper Spectral Analysis. *Physiology*, 32(1):60–92, 1 2017.

[2] Diego Vidaurre, Andrew J. Quinn, Adam P. Baker, David Dupret, Alvaro Tejero-Cantero, and Mark W. Woolrich. Spectrally resolved fast transient brain states in electrophysiological data. *NeuroImage*, 126:81 – 95, 2016.

[3] Martin Längkvist, Lars Karlsson, and Amy Loutfi. Sleep Stage Classification Using Unsupervised Feature Learning. *Advances in Artificial Neural Systems*, 2012:1–9, 7 2012.

[4] David R. Brillinger. *Time series : data analysis and theory*. Society for Industrial and Applied Mathematics, 2001.

[5] Matthew J Beal, Zoubin Ghahramani, and Carl Edward Rasmussen. The Infinite Hidden Markov Model. *Advances in Neural Information Processing Systems 14*, pages 577–584, 2002.

[6] D.J. Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096, 1982.

[7] Jurgen Van Gael, Yunus Saatci, Yee Whye Teh, and Zoubin Ghahramani. Beam Sampling for the Infinite Hidden Markov Model. *Proceedings of the 25th International Conference on Machine Learning*, pages 1088–1095, 2008.

[8] Behtash Babadi and Emery N. Brown. A Review of Multitaper Spectral Analysis. *IEEE Transactions on Biomedical Engineering*, 61(5):1555–1564, 5 2014.

[9] Takeru Matsuda and Fumiyasu Komaki. Multivariate Time Series Decomposition into Oscillation Components. *Neural Computation*, 29(8):2055–2075, 8 2017.

[10] Kyle Ulrich, David E Carlson, Wenzhao Lian, Jana Schaich Borg, Kafui Dzirasa, and Lawrence Carin. Analysis of Brain States from Multi-Region LFP Time-Series. *Advances in Neural Information Processing Systems 27*, pages 2483–2491, 2014.