Bayesian Nonparametric Identification of Frequency-Selective Neural Oscillatory States

A PREPRINT

Shoto Yamada
Faculty of Computer Science, Institute for Neuroinformatics
Ruhr-Universität Bochum
Bochum, Germany
Shoto.Yamada@ruhr-uni-bochum.de

Simon E. Nagel
Faculty of Computer Science, Institute for Neuroinformatics
Ruhr-Universität Bochum
Bochum, Germany
Simon.Nagel-q7k@ruhr-uni-bochum.de

Xenia Kobeleva
Faculty of Medicine, Institute for Neuroinformatics
Ruhr-Universität Bochum
Bochum, Germany
research@xenia-kobeleva.com

Robert Schmidt
Faculty of Computer Science, Institute for Neuroinformatics
Ruhr-Universität Bochum
Bochum, Germany
robert.schmidt@ruhr-uni-bochum.de

December 29, 2025

ABSTRACT

Identifying neural oscillations is essential for linking fast brain dynamics to underlying cognitive states. However, this is challenging because oscillatory events can be brief, embedded in $1/f$-like background activity, and may comprise an unknown number of spectrally distinct states. Conventional approaches often apply narrowband band-pass filters to one or a few predefined frequency bands, followed by fixed power- or amplitude-thresholding to identify oscillatory events. Although recent unsupervised alternatives based on Hidden Markov models (HMM) address these limitations, they still require *a priori* specification of the number of states and can underfit or overfit when the number of states is misspecified. We propose a Bayesian nonparametric method that identifies distinct oscillatory states while inferring an appropriate number of states directly from the data. This method combines time-delay embedding (TDE) with the Dirichlet-process Gaussian mixture model (DP-GMM). TDE augments the signal with time-shifted copies, enabling the DP-GMM to capture frequency-specific local autocovariance structures, while the Dirichlet-process prior adapts model complexity by pruning inactive components. We benchmarked the approach against a filter-based thresholding method and the time-delay embedded HMM using single-channel synthetic data designed to mimic neural time series (e.g., EEG, MEG, and local field potentials), with multiple frequency components embedded in $1/f$-like noise. In this setting, the proposed model reliably

recovered multiple distinct frequency components under noisy conditions while also inferring the correct number of oscillatory states. Applied to a resting-state motor-cortex MEG dataset, the model identified multiple frequency-selective, short-lived oscillatory states that exhibited substantial inter-individual heterogeneity in peak frequency, occurrence rate, and power. Overall, this provides a fully unsupervised route to discovering frequency-selective oscillatory states without predefining frequency bands or selecting the number of states in advance.

Keywords Neural signal processing · Burst detection · Aperiodic activity · Local Field Potential (LFP) · Electroencephalography (EEG) · Magnetoencephalography (MEG) · Time-delay embedding · Dirichlet process Gaussian mixture model

1 Introduction

Identifying neural oscillations is essential for linking rapid brain dynamics to cognitive processes, as oscillations support frequency-specific coordination and large-scale integration of neural activity (Buzsáki & Draguhn, 2004; Fries, 2015; Siegel et al., 2012; Varela et al., 2001). Consistent with this view, oscillatory activity has been associated with a broad range of cognitive functions, including, e.g., executive control (Cavanagh & Shackman, 2015; Schmidt et al., 2019) and memory consolidation (Diekelmann & Born, 2010; Rasch & Jan Born, 2013). Crucially, many task-relevant oscillations manifest as transient, burst-like events rather than sustained rhythms (Muralidharan et al., 2022, 2023; Rayson et al., 2025; Schmidt et al., 2023; Zich et al., 2020), making their reliable identification a nontrivial methodological problem. Such events can be obscured by a 1/f-like background (Donoghue et al., 2022) and occur at variable times relative to task events (Williams et al., 2020), and may appear artificially sustained when brief, temporally jittered episodes are averaged across trials or samples (Donoghue et al., 2022; van Ede et al., 2018).

One conventional approach for identifying oscillatory events is to restrict analysis to one or a few predefined frequency bands via narrowband band-pass filtering, followed by amplitude- or power-thresholding (Duchet et al., 2021; Little et al., 2019; Rayson et al., 2022; Shin et al., 2017). However, this method is highly sensitive to its parameter choices. The selection of filter and threshold settings can drastically affect detection outcomes (Langford & Wilson, 2023; Widmann et al., 2015). Moreover, if the true oscillation frequency deviates from the filter frequency, the method can falsely detect oscillations in the filtered band, even when no narrowband oscillation is present in the signal (Donoghue et al., 2022). In addition, the threshold directly determines the apparent duration of detected oscillations, and in practice threshold selection often follows arbitrary conventions rather than principled criteria (Schmidt et al., 2023). This issue becomes even more severe when multiple oscillatory components in distinct frequency bands are present simultaneously. Threshold-based pipelines effectively operate as single-band event detectors and are therefore poorly suited to characterizing multi-frequency structures.

Although a recent unsupervised Hidden Markov Model (HMM) approach addresses these issues (Gohil et al., 2024; Quinn et al., 2019), it still requires a priori selection of the number of states K. Typically, researchers fit HMMs with multiple candidate values of K and select the preferred model using model-selection criteria such as information criteria or the evidence lower bound (ELBO). This model selection problem remains notoriously challenging (Buckby et al., 2023). Simulation studies show that commonly used information criteria (e.g., Akaike information criterion; Akaike, 1998, Bayesian Information Criterion; Schwarz, 1978) often fail to recover the true number of states, typically favoring overly complex models (Bacci et al., 2012; Buckby et al., 2023; Pohle et al., 2017). For electrophysiological data, the ELBO typically continues to increase with larger K, simply favoring models with larger states (Huang et al., 2024; Quinn et al., 2018). As a result, practitioners often resort to heuristic choices of K guided by interpretability (Gohil et al., 2024; Quinn et al., 2019). Consequently, both formal model-selection criteria and the heuristic leave analyses vulnerable to underfitting (i.e., too few states to capture all the oscillatory states) and overfitting (i.e., fragmenting states into multiple superfluous states).

Motivated by these limitations, we propose a Bayesian nonparametric approach for clustering oscillatory structures while minimizing the need to predefine either target frequency bands or the number of oscillatory states. Our method combines time-delay embedding (TDE) (Quinn et al., 2019) with a Dirichlet-process Gaussian mixture model (DP-GMM) (Gershman & Blei, 2012). The Dirichlet-process prior induces sparsity in the mixture weights, allowing the model to allocate probability mass to a limited set of data-supported oscillatory states while shrinking redundant components toward near-zero weight. Importantly, the approach avoids fixed band definitions and global amplitude thresholds; rather, it finds oscillatory components based on frequency-specific local autocovariance structures.

We evaluated our approach on single-channel synthetic time series designed to mimic key properties of neural recordings (EEG, MEG, and local field potentials), comprising multiple oscillatory components embedded in aperiodic $1/f$ noise. We compared performance against two baselines: (i) a conventional filter-and-threshold pipeline and (ii) the time-delay embedded Hidden Markov model (Quinn et al., 2019). We assessed robustness across (a) noise levels and oscillation-

frequency configurations, (b) model-misspecification settings in which the assumed number of states differed from the ground truth, and (c) an empirical resting-state MEG dataset to test performance under real measurement noise and inter-individual heterogeneity. Across these settings, our method consistently identified multiple frequency-selective oscillatory states without fixing the number of states *a priori* and remained robust across noise levels, model-order uncertainty, and inter-individual heterogeneity in the example MEG dataset. Altogether, the DP-GMM provides an unsupervised framework to discover and characterize multiple frequency-selective oscillatory states in complex signals without predefining frequency bands or the number of states in advance.

2 Methods

2.1 Synthetic data generation

We generated one-dimensional time series that alternate between short-lived oscillatory states and $1/f$ noise states, following Cho and Choi (2023) and Quinn et al. (2019). A categorical latent state $s_n \in \{1, \ldots, K\}$ indexes noise ($s_n = 1$) and $K - 1$ oscillatory states ($s_n \geq 2$) for each time point n. Signals were constructed by concatenating variable-length segments until a total duration of N was reached, truncating the final segment if necessary. At the start of each segment, the next state was drawn from a uniform categorical distribution over states with the constraint $s_{\text{next}} \neq s_{\text{current}}$ (no self-transitions). Unless noted otherwise, we used a sampling rate of 250 Hz.

For each segment of each oscillatory state, we generated a sine wave at a frequency of f_{sim}. We randomly varied the duration of the oscillatory segment by sampling the number of cycles uniformly from $\{3, ..., 7\}$. We also varied the amplitude for each oscillatory state segment based on the product of an absolute Gaussian jitter and a fixed power-law frequency scaling: $A = \left|\mathcal{N}(1,\, 0.1^2)\right| \cdot f_{sim}^{-0.15}$. The Gaussian jitter introduced amplitude variability in each simulated oscillatory segment, while the $f_{sim}^{-0.15}$ factor implemented a simple $1/f$-like attenuation of higher frequencies. We finally tapered each segment using a Tukey window for a smooth transition between distinct states.

To mimic a $1/f^{\beta}$ background noise, we first generated standard Gaussian noise, whose duration was uniformly drawn from 0.5 to 3 seconds. We then filtered the Gaussian noise such that its power spectrum decayed as $1/f^{\beta}$, where β controls the decay rate. For all simulations we set β to 1 (i.e., pink noise).

2.1.1 Signal-to-noise ratio (SNR)

To compute the desired signal-to-noise ratio (SNR), we first high-pass filtered the simulated noise segments at 0.5 Hz to suppress slow drifts, which would otherwise inflate the power of the signal. Target SNR values were specified in decibels (dB) and converted to a linear power ratio with $\gamma = 10^{\text{SNR}/10}$. We then scaled the amplitude of each oscillatory state by

$$a_s = \sqrt{\frac{\gamma\, \sigma^2_{\text{noise}}}{\sigma^2_s}}, \quad s \geq 2,$$

where σ^2_{noise} is the power of the noise state ($s = 1$), and σ^2_s is the power of the oscillatory state ($s \geq 2$). With this scaling, the power of each state matched the desired SNR relative to the noise power. After scaling all oscillatory states, we added *unfiltered* pink noise to generate the final time series. Thus, the SNR was calibrated using drift-suppressed noise, while the simulated data retained the full $1/f$ structure.

2.2 Probabilistic models for clustering oscillatory states

2.2.1 Feature construction

To detect oscillatory states, we first transformed the raw time series into a feature representation that captures short-range temporal structure by applying time-delay embedding (TDE) (Figure 1). TDE augments the signal with lagged copies of itself, such that the covariance matrix of the embedded vectors reflects the autocovariance structure of the underlying signal (Quinn et al., 2019). The original time series is embedded at a lag of 0. The embedding dimension E is a hyperparameter that determines how many lags are included in each embedded vector. In our analyses, we standardized the data and then applied TDE to obtain embedded observations, which served as input to both the Dirichlet–process Gaussian mixture model (DP-GMM) and the Hidden Markov Model (HMM).

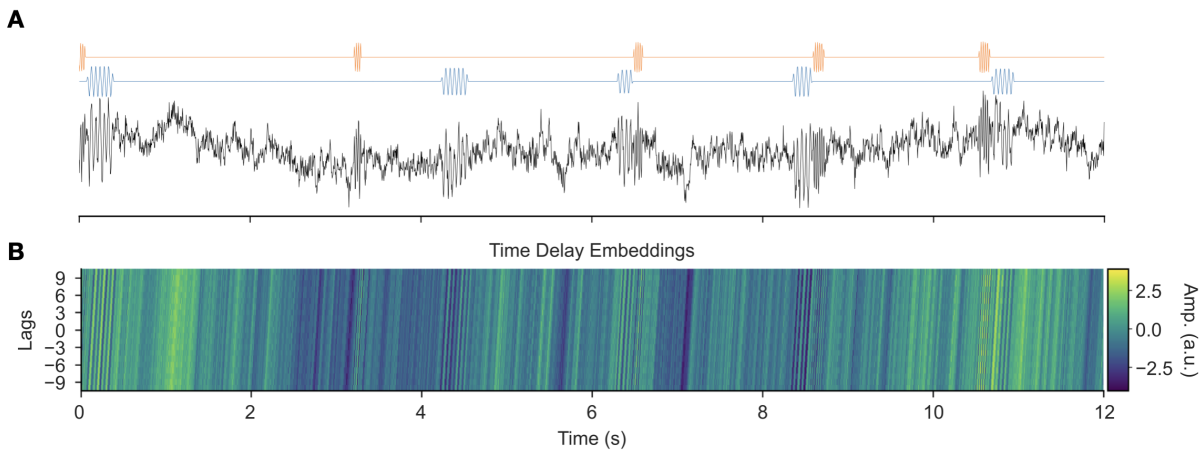

Figure 1: **Time-delay embedding (TDE) stacks lagged copies of the signal so oscillatory states become separable for clustering**(A) Example synthetic signal (black curve) at SNR $= 4$ dB containing brief 20 Hz (blue) and 40 Hz (orange) oscillatory events. Periods without oscillations consist of pink noise. (B) Corresponding TDE representation with embedding dimension $E = 21$, visualized as lagged copies of the black signal over time. The lag 0 corresponds to the original signal.

2.2.2 Dirichlet-process Gaussian mixture model (DP-GMM)

The Dirichlet-process Gaussian mixture model (DP-GMM) is a Bayesian nonparametric extension of the Gaussian mixture model (GMM) that lets the data determine how many mixture components (states) are effectively used (Ferguson, 1973). Informally, the DP prior allows for an unbounded number of potential components, but it typically assigns substantial weight to only a limited number of them (Gershman & Blei, 2012). For computation, we used a truncated stick-breaking approximation with an upper bound K_{upper} on the number of states. This provides a finite representation while retaining the DP prior behavior that unused components receive near-zero posterior weights (Ishwaran & James, 2001).

Stick-breaking generates mixture weights recursively. For each $k = 1, \ldots, K_{\text{upper}}$, we first sample $v_k \sim \text{Beta}(1, \alpha)$. Starting from a stick of length 1, at step k we break off a fraction v_k of the remaining stick and assign that piece as the kth weight w_k (with the remainder passed to the next step). That is, the resulting weights $\mathbf{w} = (w_1, \ldots, w_{K_{\text{upper}}})^\top$ are a deterministic function of $v_{1:K_{\text{upper}}}$. We denote this construction compactly as

$$\mathbf{w} \sim \text{TSB}(\alpha),$$

where $\text{TSB}(\alpha)$ is the truncated stick-breaking prior at K_{upper} components. Smaller values of α imply faster decay of the expected weights and hence sparser mixtures. We placed a $\text{Half-Cauchy}(1)$ prior on the concentration parameter α.

We modeled observations with a zero-mean multivariate normal distribution,

$$\mathbf{x}_n \mid s_n = k \sim \mathcal{N}\big(\mathbf{0}, \Sigma_k\big),$$

where $s_n \in \{1, \ldots, K_{\text{upper}}\}$ is the latent state at a time point n and Σ_k is the state-specific covariance matrix (treated as a random variable). For computation, we parameterized each covariance matrix via a scale-correlation decomposition,

$$\Sigma_k = \text{Diag}(\boldsymbol{\sigma}_k)\, L_k L_k^\top \text{Diag}(\boldsymbol{\sigma}_k),$$

where $\boldsymbol{\sigma}_k \in \mathbb{R}^d$ is the vector of marginal standard deviations, L_k is the Cholesky factor of a correlation matrix, and d is the data dimension (embedding dimension). This decomposition is a flexible alternative to inverse-Wishart priors for covariance matrices, particularly in moderate to high dimensions (Barnard et al., 2000). We placed independent $\text{Half-Cauchy}(1)$ priors on the scale parameters $\sigma_{k,j}$ and an $\text{LKJCholesky}(\eta = 1)$ prior on L_k; for $\eta = 1$, the LKJ prior is uniform over the space of valid correlation matrices (Lewandowski et al., 2009). Together, these priors induce $p(\Sigma_k)$.

Collecting all unobserved random variables into

$$\mathbf{z} \;=\; \big\{\alpha,\ \mathbf{w},\ \Sigma_{1:K_{\text{upper}}},\ \mathbf{s}\big\}\,,$$

the joint distribution factorizes as

$$p(\mathbf{x}, \mathbf{z}) = p(\alpha)\, p(\mathbf{w} \mid \alpha) \prod_{k=1}^{K_{\text{upper}}} p(\Sigma_k) \prod_{n=1}^{N} p(s_n \mid \mathbf{w})\, p(\mathbf{x}_n \mid s_n, \Sigma_{1:K_{\text{upper}}}), \tag{1}$$

where N denotes the total data points, $p(s_n \mid \mathbf{w})$ is categorical with probabilities $\mathbf{w}$, and $p(\mathbf{x}_n \mid s_n = k, \Sigma_{1:K_{\text{upper}}})$ is the Gaussian observation model defined above. Figure 2A provides the graphical representation of the DP-GMM.

Unless otherwise stated, we set the truncation level to $K_{\text{upper}} = \lceil \ln N \rceil$, motivated by the fact that the expected number of occupied clusters under a DP prior grows on the order of $\ln N$ (Korwar & Hollander, 1973). Since K_{upper} is only an upper bound, the DP prior typically assigns negligible posterior weight to many components. We defined the number of active states as the smallest set of components whose posterior mixture weights collectively account for more than 99% of the total mixture weight across the K_{upper} components, treating the remaining components as inactive. We renormalized the weights of the active states to sum to one and assigned each time point to the most likely state via maximum a posteriori (MAP) estimation.

2.2.3 Hidden Markov Model (HMM)

The hidden Markov model (HMM) represents a time series as a sequence of K discrete latent states with Markovian temporal dynamics (Fraser, 2008; Rabiner, 1989). Let $s_n \in \{1, \ldots, K\}$ denote the latent state at time point n, and $\mathbf{x}_n \in \mathbb{R}^d$ the corresponding observation. Conditioned on the current state, observations are generated independently from a state-specific emission model. We used the same zero-mean multivariate normal emission model, and the same covariance prior $p(\Sigma_k)$ as in the DP-GMM (see Section 2.2.2 for details).

The latent state sequence follows a first-order Markov chain governed by an initial state distribution $\boldsymbol{\pi}$ and a transition matrix A, where $A_{ij} = p(s_n = j \mid s_{n-1} = i)$. Specifically,

$$s_1 \sim \text{Categorical}(\boldsymbol{\pi}), \qquad s_n \mid s_{n-1} \sim \text{Categorical}(A_{s_{n-1},:}) \quad (n = 2, \ldots, N).$$

We placed symmetric Dirichlet priors on $\boldsymbol{\pi}$ and on each row of A,

$$\boldsymbol{\pi} \sim \text{Dirichlet}(\mathbf{1}), \qquad A_{i,:} \sim \text{Dirichlet}(\mathbf{1}) \quad (i = 1, \ldots, K),$$

where $\mathbf{1}$ denotes the all-ones vector of length K.

Collecting all latent variables and parameters into

$$\mathbf{z} \;=\; \{\boldsymbol{\pi},\; A,\; \Sigma_{1:K},\; \mathbf{s}\},$$

the joint distribution factorizes as

$$p(\mathbf{x}, \mathbf{z}) = p(\boldsymbol{\pi})\, p(A) \prod_{k=1}^{K} p(\Sigma_k)\, p(s_1 \mid \boldsymbol{\pi}) \prod_{n=2}^{N} p(s_n \mid s_{n-1}, A) \prod_{n=1}^{N} p(\mathbf{x}_n \mid s_n, \Sigma_{1:K}). \tag{2}$$

Figure 2B provides the graphical representation of the HMM.

After fitting the model, we decoded a single state sequence using the Viterbi algorithm, which returns the maximum a posteriori (MAP) state path under the fitted HMM parameters,

$$\hat{\mathbf{s}}_{\text{Viterbi}} \;=\; \arg\max_{\mathbf{s}}\; p(\mathbf{s} \mid \mathbf{x}, \hat{\boldsymbol{\pi}}, \hat{A}, \hat{\Sigma}_{1:K}),$$

where $(\hat{\boldsymbol{\pi}}, \hat{A}, \hat{\Sigma}_{1:K})$ denotes the inferred parameter values (i.e., posterior median).

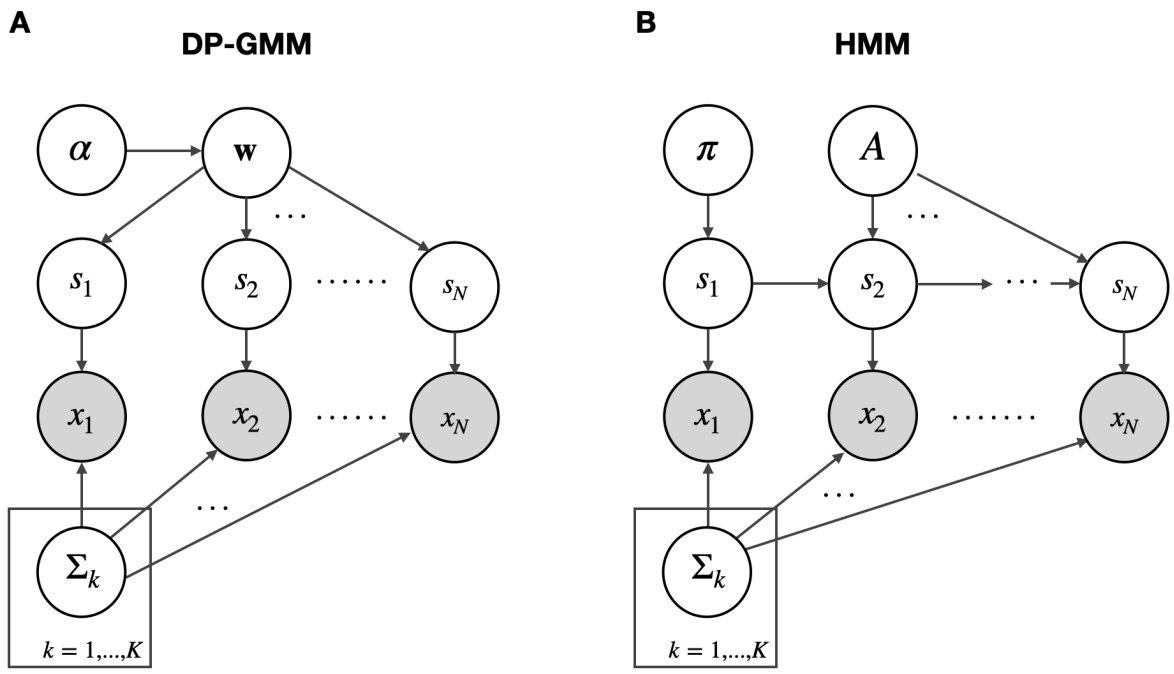

Figure 2: **Model overview: DP-GMM and HMM share the same Gaussian emission model but differ in latent-state dynamics.** (A) DP-GMM: a concentration parameter α induces mixture weights $\mathbf{w}$ (truncated stick-breaking); latent states $s_{1:N}$ are i.i.d. with class probabilities $\mathbf{w}$. Embedded observations follow $\mathbf{x}_n \mid s_n = k \sim \mathcal{N}(\mathbf{0}, \Sigma_k)$, where Σ_k is the state-specific covariance of the time-delay embedded vectors. (B) HMM: an initial distribution $\boldsymbol{\pi}$ and transition matrix A generate a first-order Markov state sequence $s_{1:N}$; observations follow the same Gaussian emission model. Shaded nodes denote observed variables and plates indicate replication over states $k = 1, \ldots, K$.

2.2.4 Stochastic variational inference

Given a probabilistic model specified by a joint distribution $p(x, z)$, Bayesian inference aims to compute the posterior distribution of all unobserved random variables given the observations,

$$p(z \mid x) \;=\; \frac{p(x, z)}{p(x)}. \tag{3}$$

Here, z denotes all latent variables in the model, including continuous parameters and discrete latent states. For the models considered here, exact posterior inference is analytically intractable due to integrals that cannot be evaluated in closed form. Moreover, empirical neural time series are typically large, requiring inference methods that scale to big datasets. We therefore used stochastic variational inference (SVI) to obtain a fast and scalable approximation of the posterior via stochastic gradient optimization (Ranganath et al., 2014; Wingate & Weber, 2013).

SVI approximates $p(z \mid x)$ with a tractable variational distribution $q(z; \phi)$, parameterized by ϕ, by maximizing the evidence lower bound (ELBO) (Blei et al., 2017; Hoffman et al., 2013):

$$\mathrm{ELBO}(\phi) \;=\; \mathbb{E}_{q(z;\phi)}[\log p(x, z) - \log q(z; \phi)]\,. \tag{4}$$

For optimization, we minimized the negative ELBO (NELBO), defined as $\mathrm{NELBO}(\phi) = -\mathrm{ELBO}(\phi)$; minimizing $\mathrm{NELBO}(\phi)$ is equivalent to maximizing $\mathrm{ELBO}(\phi)$. Because this expectation is generally intractable, we optimized a Monte Carlo estimate of the NELBO. Concretely, at each optimization step, we drew L samples (particles) $z^{(1)}, \ldots, z^{(L)} \sim q(z; \phi)$ and approximated the NELBO and its gradients by averaging over these samples; increasing L reduces estimator variance at a higher computational cost.

We implemented our models and inference in Python using the probabilistic programming libraries NumPyro (for the DP-GMM) and Pyro (for the HMM). In preliminary benchmarking of our implementation, NumPyro yielded shorter

runtimes for the DP-GMM; therefore, we used NumPyro for that model. For the continuous latent variables, we used an automatically constructed mean-field Normal variational family (i.e., a diagonal-covariance Gaussian) in both frameworks. Because both models include discrete latent state variables, we marginalized these variables using parallel enumeration, enabling automatic differentiation for the remaining continuous variables and improving optimization stability (Obermeyer et al., 2019; Schulman et al., 2015).

For the DP-GMM, we ran SVI with $L = 10$ Monte Carlo particles, minibatches of size 2^{10} observations (randomly sampled individual data points), a learning rate of $5 \cdot 10^{-2}$, and 3000 training epochs. For the HMM, we optimized on randomly sampled subsequences of length 500 using minibatches of 2^4 subsequences, with $L = 1$ particle, a learning rate of $5 \cdot 10^{-2}$, and 1000 training epochs. After inference, we estimated the final NELBO using 100 Monte Carlo samples from $q(z; \phi)$. We also summarized each continuous latent variable with a single point estimate by taking the median of its approximate posterior distribution under $q(z; \phi)$. To reduce sensitivity to initialization, we fit each model from eight random initializations and retained the run with the lowest final NELBO for subsequent analyses.

2.3 Thresholding method

Among many variants of thresholding methods (Cho & Choi, 2023), we used the one with two amplitude thresholds: a main detection threshold and a lower secondary threshold for defining event boundaries (Cole et al., 2019; Feingold et al., 2015). Given a band-pass filtered signal centered at the target frequency, the detection algorithm proceeds in three steps. First, it computes the analytic signal via the Hilbert transform and takes its magnitude as the amplitude envelope. Second, candidate oscillatory events are identified as periods during which the amplitude envelope exceeds the main detection threshold. Third, for each candidate event, it defines its event onset and offset as the time before and after the amplitude envelope falls below the secondary threshold. These steps yield a binary time series that indicates the presence or absence of oscillatory events at each time point.

In the simulation experiments, the band-pass filter was centered at the known ground-truth frequency f_{sim}. We used a Gaussian filter with a full-width at half maximum of 5 Hz around f_{sim}. Threshold parameters were selected in a data-driven manner. We obtained a binary burst-indicator time series from the detection algorithm and selected the threshold pair by maximizing the Pearson correlation between this indicator and the amplitude envelope of the band-pass filtered signal, a pragmatic criterion used in prior work for tuning burst-detection parameters (Little et al., 2019; Shin et al., 2017).

To extend the thresholding method to scenarios with multiple oscillatory states, each target frequency band was processed individually, with thresholds optimized separately for each band, yielding a binary time series per band. These per-band detections were then combined into a single categorical label per time point. Samples with no detections were labeled as a noise state, whereas samples with one or more detections were assigned to an oscillatory state based on the highest amplitude envelope of the corresponding band.

2.4 Post-hoc k-nearest-neighbor imputation

To suppress brief false-positive state fragments, we applied a simple post-hoc k-nearest-neighbor (k-NN) imputation to all algorithms and report results with imputation by default unless stated otherwise. State segments shorter than two cycles at the target frequency were flagged as spurious, and their state assignments were treated as missing labels. These missing labels were then imputed by assigning the most frequent state among the k nearest temporal neighbors. For multi-frequency cases, we used the highest candidate frequency as the target, yielding the shortest two-cycle duration and, therefore, a conservative relabeling criterion. The two-cycle criterion is a pragmatic heuristic and a common choice in the practice of oscillatory event detection (Cho et al., 2023; Hahn et al., 2022; Liljefors et al., 2024; Liu et al., 2022).

2.5 Performance metric

To measure the performance of clustering oscillatory states, we used the Matthews correlation coefficient (MCC) (Matthews, 1975). A key advantage of the MCC over other commonly used performance metrics (e.g., precision and F1-score) is that the MCC takes into account all four categories in the confusion matrix (i.e., true positives, true negatives, false positives, and false negatives (Yao & Shepperd, 2020)). Additionally, the MCC is robust to class imbalance (e.g., when detecting rare events), while precision and F1-score can overestimate performance in such settings (Boughorbel et al., 2017). The MCC spans from -1 to 1, where 1 indicates perfect classification and -1 indicates completely inverse classification. The value of 0 implies that the clustering performance is the same as random guessing. Since the state time courses are estimated in an unsupervised fashion, their order must be aligned with that of the ground truth state time courses to compute performance. To achieve this, we reordered the estimated states by maximizing the correlation between the ground-truth states and the estimated states before computing the MCC. For multiclass

evaluation with K classes, we computed the MCC directly from the complete $K \times K$ confusion matrix C with entries C_{ij}, where i indexes the true class and j the predicted class. The multiclass Matthews correlation coefficient is defined as

$$\mathrm{MCC} = \frac{cs - \sum_k t_k p_k}{\sqrt{s^2 - \sum_k p_k^2}\,\sqrt{s^2 - \sum_k t_k^2}},$$

where

$$t_k = \sum_j C_{kj} \qquad \text{number of true samples in class } k,$$

$$p_k = \sum_i C_{ik} \qquad \text{number of samples predicted as class } k,$$

$$c = \sum_k C_{kk} \qquad \text{total number of correctly classified samples,}$$

$$s = \sum_{i,j} C_{ij} \qquad \text{total number of samples.}$$

2.6 Simulation scenarios

2.6.1 Two-states scenario

In the two-states scenario, our objective was to establish a practical and empirically grounded decision rule to choose the TDE hyperparameter, the embedding dimension E, in the simplest setting, where both the DP-GMM and the HMM distinguish a single oscillatory state against noise (i.e., two states). Because E sets the effective temporal resolution of the oscillatory structure, we hypothesized that the optimal E should depend on both the oscillatory frequency and the sampling rate. To test this, we simulated five independent 3-min signals per condition at $\mathrm{SNR} = +2$ dB, with simulated target frequencies $f_{sim} \in \{10, 20, 30, 40, 50\}$ Hz, sampling rates $f_s \in \{100, 250, 500\}$ Hz, and embedding dimensions $E \in \{3, 7, \ldots, 51\}$. For both the DP-GMM and the HMM, we used $K = 2$ states, assuming the true number of states was known. The final performance was summarized by the mean MCC across all five samples. The resulting heuristic for how to choose an appropriate E was carried forward to the subsequent benchmarks.

2.6.2 Three-states scenario

In this scenario, we aimed to systematically evaluate whether the DP-GMM remained robust to both noise and various frequency conditions. We benchmarked our model against the HMM and the thresholding method, using a three-state scenario, where two oscillatory states and one noise state were present in a signal with varying noise levels. Here, we generated five independent 3-min signals per condition at $f_s = 250$ Hz with $\mathrm{SNR} \in \{-10, -8, \ldots, 10\}$ dB and six frequency conditions (10,40), (15,40), (20,40), (25,40), (30,40), (35,40) Hz.

To isolate algorithmic performance among algorithms from model-order selection, we set the number of states to the ground truth (i.e., $K = 3$ states) for the DP–GMM and the HMM. The thresholding method additionally assumed knowledge of the true target frequencies and processed them separately with a Gaussian band-pass filter (i.e., the center frequency set to the ground-truth frequency; algorithm details in section 2.3). Notably, in combination with the assumption of the true number of states, this setup placed the thresholding method in a highly optimistic, best-case configuration. We also reported the performance of all three methods with post-hoc k-NN imputation, which suppresses short false-positive fragments. Because simulated oscillatory events were constrained to last at least three cycles by design, any detected segment shorter than two cycles at the highest inserted frequency, $f_{\max}$, must be considered spurious. Such segments were reassigned to the most frequent neighboring state via k-NN imputation, using a window length of $k = \lceil 2.5\, f_s / f_{\max} \rceil = \lceil 2.5 \cdot 250/40 \rceil = 13$ samples. Performance was summarized by the mean MCC across the five simulated samples for each condition.

2.6.3 Four-states scenario

It is usually unknown how many oscillatory states are present in a dataset, making it difficult to determine the number of states a priori. However, the primary advantage of the DP-GMM is that the model retains only the subset of active states while treating the redundant states as inactive. The goal of this scenario was to assess clustering accuracy under model-order uncertainty, where we estimated the number of states rather than setting it to the ground truth. In this scenario, we generated a 3-min synthetic signal at $\mathrm{SNR} = 2$ dB containing oscillatory components at 20, 30, and 40 Hz, as well as a noise state (i.e., four states).

For the DP-GMM, we set the upper bound K_{upper} to $K_{\text{upper}} = \lceil \ln N \rceil = 11$ (N = 149980), where N denotes the total time points of a signal. For the performance comparison, we used $K = 11$ states for the HMM and reported its performance. For a fair comparison in terms of model selection, we also examined HMM models with a varying number of K from 4 to 21 and assessed whether the final NELBO could provide a usable criterion to select optimal K for the HMM. As the highest simulated frequency was 40 Hz, the window for post-hoc KNN imputation was set to 13 samples.

2.7 CTF rest MEG dataset

We applied the time-delay embedded DP-GMM to an empirical, publicly available dataset to evaluate whether it could extract physiologically plausible oscillatory states. We analyzed data from 64 healthy participants recorded on a 275-channel CTF system (5 min per subject) at Nottingham University, UK, as part of the MEGUK partnership (MEG UK Scientific Research Community, 2023). We used the preprocessed parcel-wise time series released by Gohil et al. (2024). In brief, 0.5–125 Hz band-pass and 50/100 Hz notch filters were applied; the data were downsampled to 250 Hz; ocular and cardiac artifacts were removed via automated ICA using EOG/ECG correlations; and the data were finally downsampled to 100 Hz. We selected a single parcel in the left motor cortex and temporally concatenated the corresponding time series from all 65 participants to form a single continuous signal, following prior work using the same strategy to enable group-level state inference from resting-state MEG (Gohil et al., 2024).

3 Results

3.1 Embedding windows spanning ∼2.5 cycles yielded most accurate oscillatory-state estimates

Time-delay embedding (TDE) exposes oscillatory structure by mapping the original signal into vectors of lagged samples, such that oscillations are reflected in the covariance (i.e., the autocovariance matrix across lags) of the embedded vectors. The embedding dimension E determines the number of lags included and thus the effective temporal window (i.e., the maximum time shift spanned by the embedding). Here, we quantified how E affects oscillatory-state estimation using two-state synthetic signals while varying the sampling frequency f_s and target frequency for both the DP-GMM and the HMM. We summarize the joint effect of E and f_s using the embedding-window duration $w_E \approx E/f_s$ (more precisely, $(E-1)/f_s$ for unit-lag embeddings).

For both models, clustering performance depended on E, f_s, and the target frequency. In general, the mean MCC peaked when w_E spanned ≈ 2.5 cycles of the target frequency (Figure 3). At 10 Hz, the mean-MCC curve around this optimum was relatively flat, indicating a broad range of window lengths that yielded near-maximal performance. On the other hand, at higher frequencies, the mean MCC dropped off more sharply once w_E exceeded the optimal range, reflecting a narrower tolerance for overly long windows.

We therefore set the embedding dimension using the cycle-based heuristic $E^\star = \lceil a\, f_s / f_{\text{target}} \rceil$, with $a = 2.5$, where $\lceil \cdot \rceil$ denotes the ceiling operator and f_{target} is the target frequency. For example, for $f_{\text{target}} = 30$ Hz and $f_s = 250$ Hz, $E^\star = \lceil 2.5 \times 250/30 \rceil = 21$. We used this pragmatic, empirically guided heuristic to set the embedding dimension in subsequent analyses.

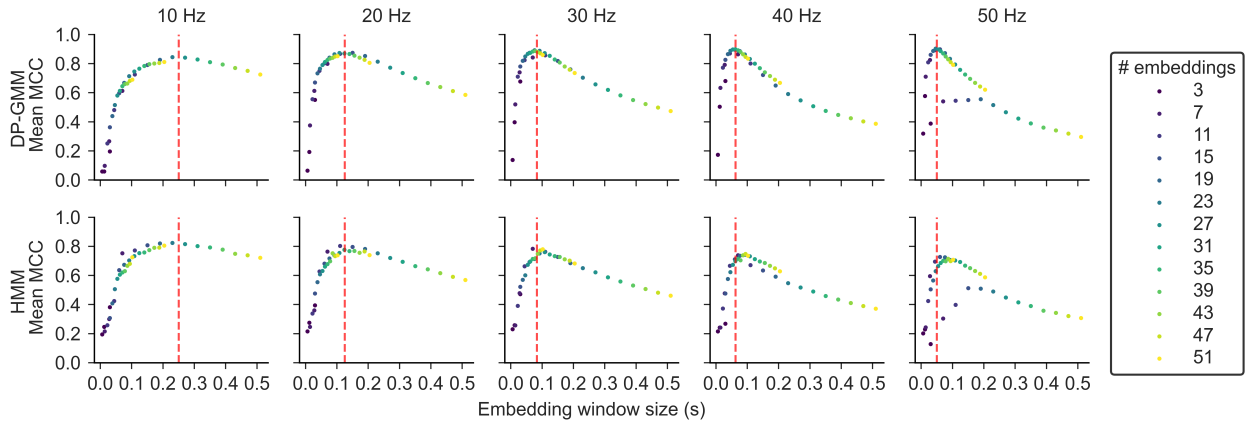

Figure 3: **Clustering performance peaked when the embedding window spanned ∼2.5 cycles of the target frequency.** Mean MCC as a function of embedding-window size w_E for the DP-GMM (top row) and the HMM (bottom row) across target frequencies (columns: 10, 20, 30, 40, 50 Hz). Colors indicate the embedding dimension E. The vertical red dashed line marks 2.5 cycles at the target frequency. For the 50 Hz condition at $f_s = 100$ Hz, the target frequency equals the Nyquist frequency, which likely contributed to the reduced MCC visible for some of the configurations with low number of embeddings.

3.2 The DP-GMM achieved robust oscillatory-state clustering across SNRs and frequencies

Reliable and practically useful clustering of oscillatory states must be robust to noise and applicable to signals containing multiple distinct oscillatory states. We therefore benchmarked the DP-GMM against the HMM and the thresholding method across a range of SNRs and frequency conditions. In this configuration, we set the number of states for both the DP-GMM and the HMM to $K = 3$ (two oscillatory states and one noise state), i.e., assuming that the true number of states was known. The embedding dimension was determined using the empirically guided, cycle-based heuristic described in the previous section. To select the embedding dimension for the two oscillatory states, we set f_{target} to their mean frequency, yielding $E^\star \in \{25\ (10\ \&\ 40\ \text{Hz}),\ 23\ (15\ \&\ 40),\ 21\ (20\ \&\ 40),\ 20\ (25\ \&\ 40),\ 18\ (30\ \&\ 40),\ 17\ (35\ \&\ 40)\}$.

Both the DP-GMM and the HMM occasionally produced brief false-positive fragments under a moderately noisy condition (SNR $= 2$ dB) in the 20 & 40 Hz setting (Figure S1). Although these fragments typically occurred at similar onset times under both models, they tended to have shorter durations under the DP-GMM. This observation motivated a minimal post-hoc k-NN imputation (see section 2.4 for details). After k-NN imputation, spurious state activations were effectively suppressed for the DP-GMM, leading to high clustering accuracy (MCC = 0.89; MCC = 0.79 without imputation), whereas the HMM showed only a marginal improvement (MCC = 0.63; MCC = 0.61 without imputation) (Figure 4A). Notably, the limited improvement for the HMM is consistent with the longer duration of its spurious states, which k-NN imputation corrects only when they are brief. In contrast, the thresholding method exhibited few spurious activations, yielding no measurable improvement with imputation (MCC = 0.86; MCC = 0.86 without imputation).

With imputation, the DP-GMM achieved high clustering accuracy at SNR $= 0$ dB (MCC $= 0.89 \pm 0.02$), outperforming both the HMM (MCC $= 0.62 \pm 0.03$) and the thresholding method (MCC $= 0.85 \pm 0.01$) (Figure 4B). This was only slightly below its performance for a clean signal at SNR $= 10$ dB (MCC $= 0.92 \pm 0.01$). Even at SNR $= -2$ dB, performance remained high (MCC $= 0.80 \pm 0.02$). At lower SNRs (SNR < -4 dB), all methods degraded substantially (mean MCC $\lesssim 0.50$). Without imputation, the DP-GMM performance decreased by $\Delta\text{MCC} \approx 0.08$ across SNR $\in [-4, 4]$ dB, allowing the thresholding method to outperform the DP-GMM for SNR < 2 dB. Nevertheless, the DP-GMM remained more accurate than the HMM across nearly all SNRs (Figure 4B).

The DP-GMM with k-NN imputation achieved consistently high performance across frequency conditions, although accuracy depended on the separation between the simulated frequency pairs (Figure 4C). For the distant 10 & 40 Hz pair, accuracy decreased to MCC $= 0.78 \pm 0.03$ at SNR $= 0$ dB and to MCC $= 0.37 \pm 0.10$ at SNR $= -2$ dB. Relative to the 20 & 40 Hz condition at the same SNRs, this corresponds to $\Delta\text{MCC} = -0.11$ (0 dB) and $\Delta\text{MCC} = -0.43$ (-2 dB). Although performance recovered at higher SNRs, this drop may reflect sensitivity to the fixed embedding duration, which can under-represent one oscillatory state when frequencies are widely separated under noisy conditions. In addition, for the close 35 & 40 Hz pair, performance plateaued at MCC $= 0.75 \pm 0.03$ even at SNR $= 10$ dB, consistent with reduced separability when the two oscillatory states induce highly similar embedded features. Compared with the DP-GMM, the HMM showed greater sensitivity to frequency conditions across pairs (Figure S2). Finally,

although it should be noted that the threshold method was evaluated under a favorable setup, it nevertheless achieved high detection accuracy across many SNR and frequency conditions and, in a subset of settings, matched or exceeded the DP-GMM (Figure S2).

Overall, the DP-GMM achieved high clustering performance also under noisy signals containing multiple oscillatory states. With minimal post hoc k-NN imputation, performance remained robust across SNRs and frequency conditions, despite occasional cases where thresholding achieved higher MCC.

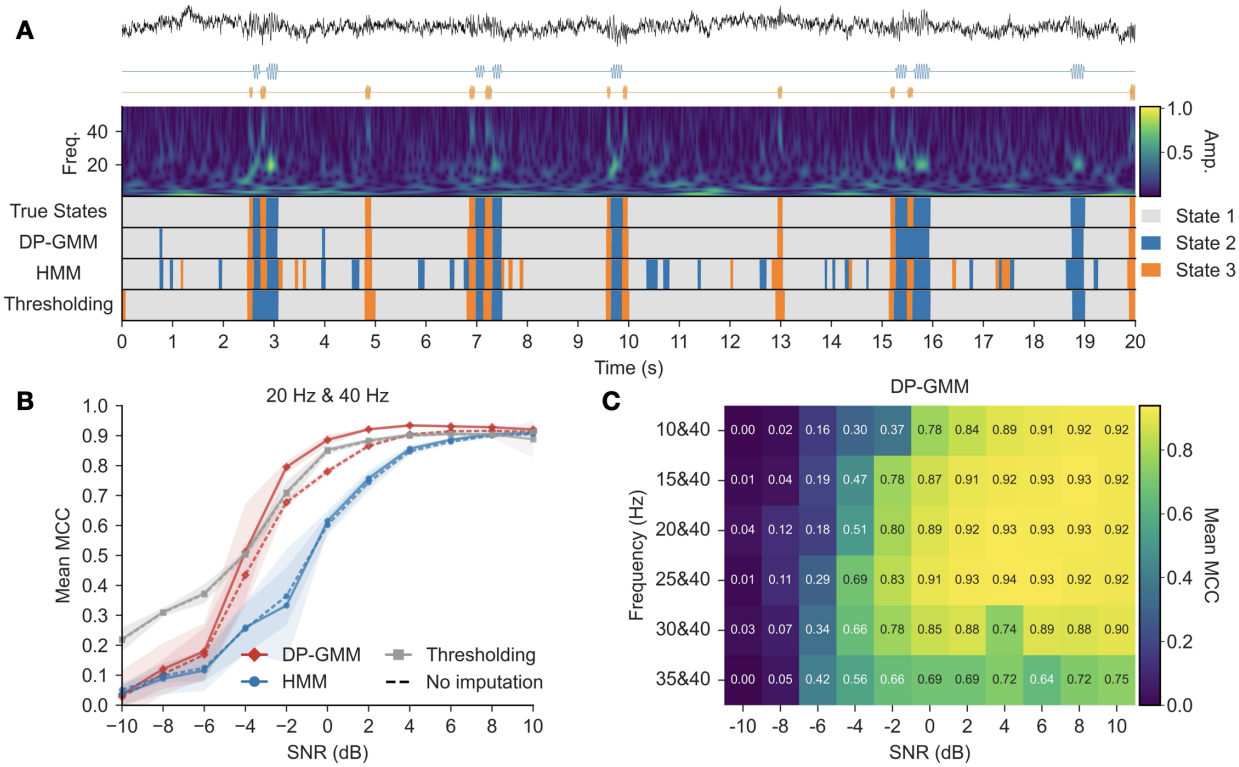

Figure 4: **The DP-GMM accurately identified the oscillatory states under noisy signals.** (A) Example raw trace from three-state synthetic signal containing 20 Hz (blue) and 40 Hz (orange) oscillations at SNR $= 2$ dB, with wavelet amplitude (a.u.), the ground-truth state sequence, and the estimated oscillatory states after k-NN imputation. (B) Mean MCC versus SNR for the 20 and 40 Hz condition (solid: with imputation; dashed: without; shading: ± 1 SD). (C) DP-GMM mean MCC after imputation across SNRs and frequency-pair conditions.

3.3 The DP-GMM recovered the correct number of states while maintaining high clustering accuracy

In the previous sections, we assumed that the true number of oscillatory states was known, allowing us to benchmark the algorithms under a best-case scenario. In practice, however, the true number of states is unknown. Here, we evaluated the DP-GMM clustering performance under model-order uncertainty by specifying only an upper bound on the number of states and allowing the model to prune redundant states. We set the upper bound to $K_{\text{upper}} = 11$ and used the same value for the HMM. We used an embedding dimension of $E = 21$, selected from the mean target frequency of 30 Hz.

Under these more realistic conditions, i.e., without knowing the ground truth about the number of states, the DP-GMM correctly retained four active states and accurately clustered the oscillatory states (MCC $= 0.94$, Figure 5A; 0.87 without imputation, Figure S3). In contrast, when we allowed more states than necessary, the HMM tended to split the noise regime across multiple states, producing many ultra-brief fragments. Because these fragments dominate local neighborhoods, the k-NN imputation could not reliably replace them with a consistent background label, and performance remained low (MCC $= 0.35$; 0.31 without imputation).

The DP-GMM concentrated its posterior mixture weight on a small subset of components. The retained active states accounted for 99.92% of the total mixture mass, leaving only 0.08% assigned to the remaining inactive states (Figure 5B). The active components exhibited distinct oscillatory structures in their estimated autocovariances and showed clear spectral peaks at 20, 30, and 40 Hz, whereas a single noise state primarily captured the $1/f$ structure. In

comparison, the HMM spread its stationary distribution nearly evenly across all 11 states, indicating frequent long-run state occupancy across all 11 states (Figure 5B). Consistent with this, the HMM's autocovariance matrices and state-wise power spectra suggest that the noise state was split into multiple substates.

Moreover, despite an explicit model-selection procedure (selecting the model with the lowest NELBO), the HMM did not identify four states as optimal (Figure S5). Instead, the NELBO decreased nearly monotonically with increasing numbers of states, without a clear elbow.

Taken together, the DP-GMM effectively eliminated redundant components by assigning them negligible mixture weights, thereby recovering the true number of states. Importantly, even under model-order uncertainty and in the four-state setting (three oscillatory states plus one noise state), it maintained high clustering accuracy.

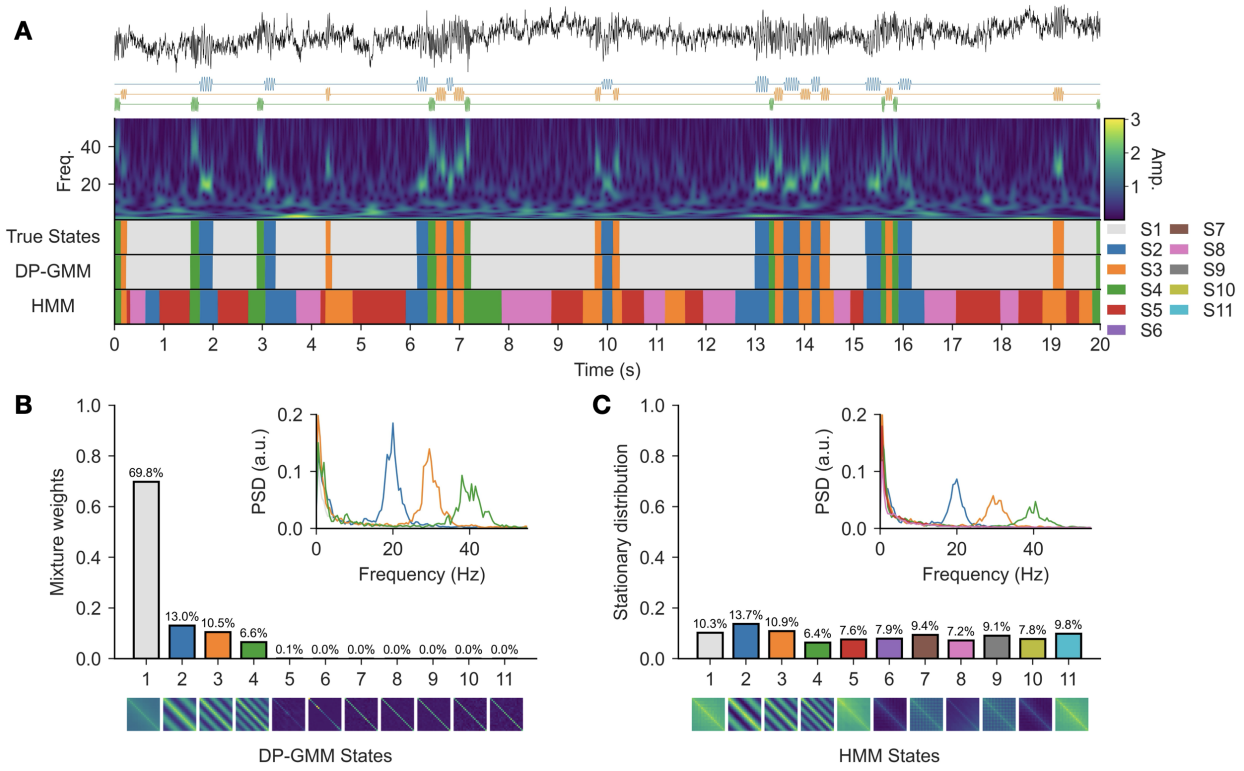

Figure 5: **The DP-GMM recovered the true number of states and achieved high clustering accuracy by eliminating inactive components.** (A) Representative snippet from a four-state synthetic signal containing transient 20 Hz (blue), 30 Hz (orange), and 40 Hz (green) oscillatory states at SNR $= 2$ dB. Inferred state time courses for the DP-GMM and HMM after k-NN imputation. (B) Outputs of the DP-GMM: posterior mixture weights and autocovariance matrices, with state-wise power spectra computed for the active states. (C) Corresponding outputs for the HMM, including the stationary distribution. We computed the stationary distribution $\boldsymbol{\pi}_\infty$ by solving $\boldsymbol{\pi}_\infty^\top = \boldsymbol{\pi}_\infty^\top \mathbf{A}$ subject to $\sum_{k=1}^{K} (\pi_\infty)_k = 1$, where $\mathbf{A}$ is the estimated transition matrix. Note that $\boldsymbol{\pi}_\infty$ does not have the same interpretation as the DP-GMM mixture weights: it reflects the long-run fraction of time the Markov chain spends in each state under $\mathbf{A}$, rather than the probability of a component generating an observation independently of temporal dynamics.

3.4 The DP-GMM revealed multiple transient oscillatory states in resting MEG

Having validated the new approach on synthetic data with known ground truth, we next applied the DP-GMM to a single parcel in the left motor cortex of MEG recordings in a publicly available dataset (MEG UK Scientific Research Community, 2023). To determine the number of embeddings, we first obtained the candidate frequency components. Here, we identified 5 candidate frequencies: 5, 11, 17, 21, and 31.5 Hz based on the peaks of the group-averaged power spectrum density (PSD) (Figure S4). Importantly, these candidate peaks were used only as a pragmatic guide for setting the reasonable embedding and k-NN imputation window. Here, we used 15 embeddings derived from 2.5 cycles of the average frequency of 17.2 Hz, covering a 0.145 s window. We then used the DP-GMM with 15 states ($\lceil \ln 1929818 \rceil$).

Spurious states were corrected by k-NN imputation ($k = 6$, which was based on two cycles of 31.5 Hz, the highest candidate frequency.)

The DP-GMM decomposed resting motor-cortex activity into two aperiodic states (States 1–2) and five short-lived, spectrally distinct oscillatory states (States 3–7). In Figure 6A, the inferred state path was largely occupied by the two aperiodic states, intermittently interrupted by brief activations of oscillatory states that aligned with localized increases in the wavelet representation. These seven active states accounted for 99.67% of the total posterior mixture weight, while the remaining components had negligible weights and were thus treated as inactive (Figure 6B). The estimated covariances for the oscillatory states (States 3–7) exhibited distinct striped patterns consistent with oscillatory structure. State-wise spectra further confirmed that the active oscillatory states were frequency-selective, each exhibiting a narrowband peak at approximately 9.5 Hz (State 6), 11.5 Hz (State 4), 19.5 Hz (State 5), 22.0 Hz (State 7), and 26.5 Hz (State 3) (Figure 6C). In contrast, State 1 showed no clear narrowband peak, whereas State 2 exhibited a pronounced $1/f$-like spectrum with maximal power at 2.5 Hz. Notably, the DP-GMM resolved oscillatory states that were not prominent as local maxima in the group-averaged PSD. At the same time, several inferred states aligned broadly with the prominent PSD candidates ($\approx$ 11–22 Hz), indicating that the candidate-peak heuristic provided a reasonable coarse guide. The 5 Hz and 31.5 Hz candidates were not expressed as distinct oscillatory states. In particular, the ~ 5 Hz candidate was likely driven by the filter cutoff (0.5-125 Hz band-pass; Figure S4), rather than reflecting a distinct oscillatory state.

Having identified a common set of oscillatory states at the group level, we next examined how these states were expressed across individual participants. Fractional occupancy varied across participants for all states (Figure 7A). The two aperiodic states (State 1-2) and the two oscillatory states (State 3-4) were present in all 65 participants but showed substantial inter-individual variability in occupancy (median [IQR]: State 1 0.371 [0.230, 0.517], State 2 0.173 [0.098, 0.280], State 3 0.188 [0.093, 0.272], State 4 0.101 [0.067, 0.162]). In contrast, State 5-7 were sparse and near-zero for many participants (medians 0.01 - 0.03), with State 7 absent in 12/65 participants (present in 81.5%). To further inspect the characteristic of the identified states, we also computed the mean lifetime, state power, occurrence rate, and coefficient of variation (CV) of inter-state intervals for each participant (Figure 7B). The dominant baseline state (State 1) persisted longer than the other states (median mean lifetime 0.370 s [0.299, 0.459]), whereas the remaining states exhibited shorter, burst-like episodes (median lifetimes 0.202–0.263 s). Expressed as equivalent cycles at each state's peak frequency, the oscillatory states lasted only a few cycles when converting the median across-participant mean lifetime to cycles: State 6 ≈ 2.50 cycles (9.5 Hz), State 4 ≈ 2.69 cycles (11.5 Hz), State 5 ≈ 3.93 cycles (19.5 Hz), State 7 ≈ 4.84 cycles (22.0 Hz), and State 3 ≈ 6.47 cycles (26.5 Hz). Occurrence rates mirrored the occupancy structure: State 1-4 occurred frequently (median occurrence rates 0.444–0.921 Hz), while State 5–7 were rare (0.048–0.123 Hz). Inter-event timing showed limited regularity, with CVs of onset intervals near 1 across states (medians 0.966–1.164), suggesting irregular onsets. Notably, State 6-7 expressed comparatively high power (median power: State 6 2.78 a.u., State 7 2.10 a.u.) despite their low occupancy.

For comparison, we also applied the HMM with 7 states to the same dataset to assess whether it could recover states comparable to those obtained with the DP-GMM (Figure S6). The HMM recovered several comparable states (Figure S6C), including (i) a predominantly aperiodic $1/f$-like state (State 3), (ii) an ~ 10 Hz state (State 1), (iii) an ~ 20 Hz state (State 6), and (iv) a broader 10-20 Hz band state (State 7) . However, the HMM also exhibited state fragmentation and mixing: State 2 and State 4 showed similar spectra without distinct peaks and alternated rapidly in the decoded sequence (Figure S6A), indicating over-partitioning of background activity. In addition, State 5 exhibited a peak at ~ 27 Hz while retaining a substantial $1/f$ aperiodic component, yielding a bimodal PSD. Finally, we evaluated whether the negative evidence lower bound (NELBO) could guide model order selection; however, it continued to decrease as K increased, consistent with our synthetic-data results (Figure S7).

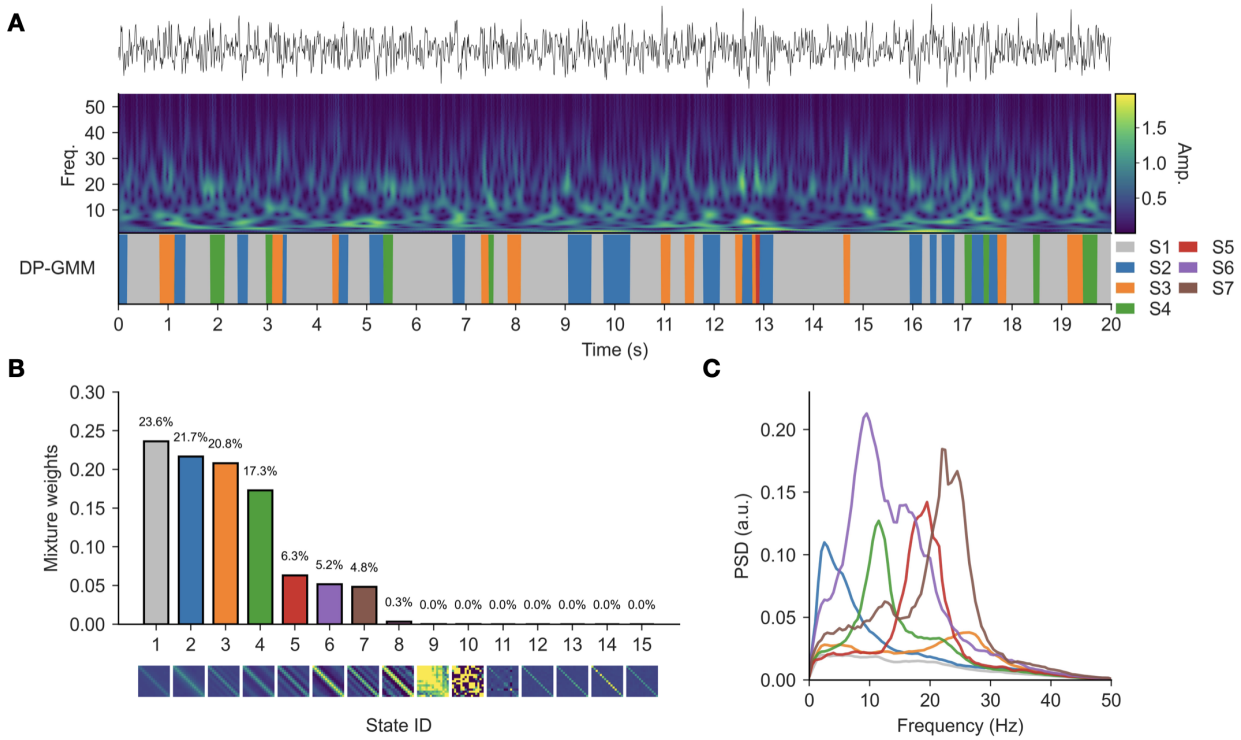

Figure 6: **The DP-GMM identified seven spectrally distinct states in resting motor-cortex MEG activity.** (A) Decoded state sequence for the seven active states after post-hoc k-NN imputation. (B) Posterior mixture weights for all $K_{\text{upper}} = 15$ components, shown alongside the corresponding state autocovariance matrices. (C) State-wise power spectra computed from the seven active states.

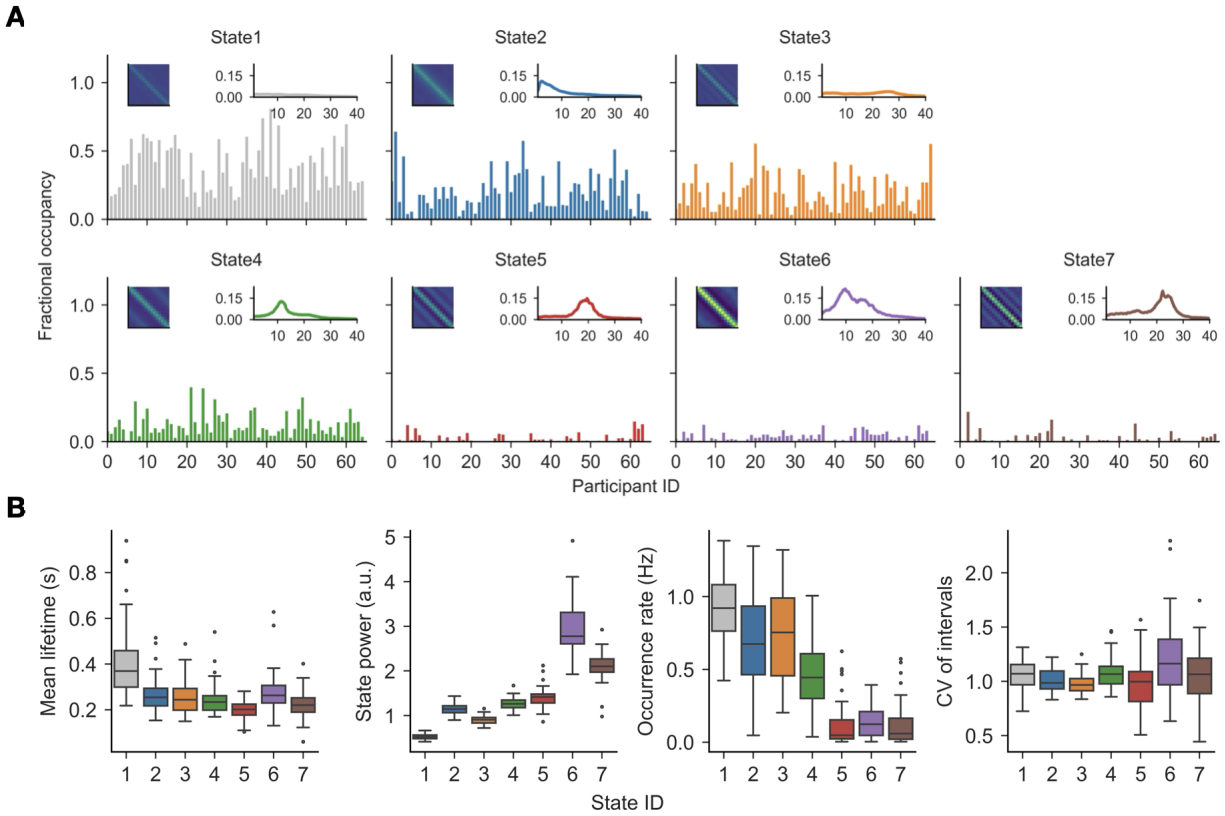

Figure 7: **Identified states showed strong inter-subject variability and burst-like oscillatory dynamics.** (A) Fractional occupancy across participants ($N = 65$) for each of the seven active states. Insets show the state-specific autocovariance and power spectrum. (B) Distributions across participants of mean lifetime, state power, occurrence rate, and coefficient of variation (CV) of inter-event intervals for each state.

4 Discussion

This work demonstrates that combining time-delay embedding (TDE) with a Dirichlet-process Gaussian mixture model (DP-GMM) enables unsupervised identification of multiple spectrally distinct oscillatory states from single-channel time series, without predefining frequency bands or the number of states. In synthetic benchmarks spanning noisy, multi-frequency conditions, the DP-GMM outperformed the HMM and a thresholding approach after applying the same minimal post-hoc k-NN imputation to suppress brief false-positive fragments. Critically, the Dirichlet-process prior pruned inactive components and recovered the true number of oscillatory states, reducing sensitivity to model-order misspecification. Applied to resting-state motor-cortex MEG, the model identified several oscillatory states that emerged as short, burst-like activations lasting only a few cycles, with substantial inter-individual heterogeneity in peak frequency, occurrence rate, and power.

The DP-GMM outperformed the HMM in accurately identifying multiple oscillatory states primarily because the two models exploit temporal structures in fundamentally different ways. The HMM enforces temporal persistence through a transition matrix, which can discourage rapid state switching (Quinn et al., 2019) but can also prolong erroneous state activations once they occur. In contrast, the DP-GMM does not impose an explicit Markovian dynamical prior; instead, local temporal structure enters through TDE, which maps short-range dynamics into feature space. This allows the DP-GMM to differentiate oscillatory events based on the local covariance structure without committing to prolonged state persistence, making spurious activations easier to terminate and, consequently, shorter.

Synthetic benchmarks showed that both the DP–GMM and the HMM produced false-positive state assignments, likely because TDE can cause some non-oscillatory segments to transiently resemble oscillatory structures in the embedded space. For the DP–GMM, these spurious runs were typically brief and could be corrected by minimal post-hoc k-NN imputation. Although increasing the k-NN window could further suppress false positives, it would also raise the risk of reassigning true oscillatory events to neighboring states, thereby increasing false negatives. Moreover, when the

HMM was fitted with an excessive number of states, it tended to split a single state into multiple substates, producing rapidly alternating sub-states. In this regime, post-hoc k-NN imputation is ineffective because the neighboring states are dominated by competing sub-states rather than a single stable state, so k-NN cannot reliably collapse the alternation back into a coherent state.

We also observed that the choice of embedding dimension (i.e., the number of lags) strongly influences clustering performance for both the DP-GMM and the HMM. As a pragmatic heuristic, we set the embedding window to span approximately 2.5 cycles at the frequency of interest, reflecting a trade-off between representing oscillatory structure and preserving temporal precision. This simple cycle-based rule can break down at very slow or very fast frequencies, where it can yield an impractically large or small number of lags. Moreover, in multi-frequency settings, any single embedding choice is necessarily a compromise. This compromise can be suboptimal when the signal contains both low and high frequencies simultaneously, because the resulting embedding dimension may be too short for the lowest band and too long for the highest band. A range of algorithmic approaches for selecting embedding dimensions in state-space reconstruction exists (for a review, see Tan et al., 2023), but their applicability to oscillatory-state clustering remains an open direction.

For the thresholding method, we assumed both the true number of states and the true frequency bands. These assumptions placed thresholding in an unrealistically favorable setting, under which it achieved high segmentation accuracy across a broad range of SNRs and frequency conditions, and in some conditions even exceeded the DP-GMM. We attribute this advantage to the narrowband filtering, which increases effective SNR within the targeted band (de Cheveigné & Nelken, 2019) and facilitates the detection of band-limited events. Thresholding can therefore be effective when a specific target band can be justified *a priori*. However, in practice, selecting and justifying target bands is nontrivial. Peak frequencies and bandwidths can vary across participants (Haegens et al., 2014) and brain regions (Mahjoory et al., 2020), making fixed detection bands difficult to defend. Even when a target band is justified, amplitude thresholds defined globally across the dataset can be poorly calibrated when oscillation amplitude varies substantially across time and recordings, leading to missed low-amplitude events or inflated false positives (Langford & Wilson, 2023).

Our benchmarking simulations had several limitations. First, we restricted our evaluation to the commonly studied range of oscillation frequencies. Higher-frequency activity such as hippocampal ripple-range oscillations (~ 80–250 Hz (Liu et al., 2022)) and high gamma (70–150 Hz (Synigal et al., 2020)), as well as very low frequencies including delta (0.5–4 Hz) and theta (4–8 Hz (Attar, 2022)), remains to be explored. Second, we did not simulate nested or concurrent oscillations (e.g., high-frequency ripples nested within 10–16 Hz spindles during cortical slow oscillations, ~ 0.5–1 Hz) (Antony et al., 2018; Latchoumane et al., 2017; Oyanedel et al., 2020; Staresina et al., 2015). Relatedly, our models assume that each time point belongs to a single state, whereas recent approaches such as DyNaMo allow overlapping states (Gohil et al., 2024); such models are natural candidates for capturing nested and concurrent activity and may complement our Bayesian nonparametric approach. Third, we simulated oscillations as sinusoidal, although neural oscillations are often non-sinusoidal (e.g., asymmetric or sawtooth-like cycles) (Belluscio et al., 2012; Cole & Voytek, 2017; Rayson et al., 2025). Despite these simplifications, we expect the approach to remain applicable when the signal-to-noise ratio is sufficient and the embedding window captures the relevant cycle lengths because it leverages autocovariance structure to identify oscillatory states. For example, nested oscillations should produce multiple off-diagonal ridge patterns in their autocovariance matrix, and non-sinusoidal waveforms should yield sharper, asymmetric ridge structures, both distinguishable from an aperiodic $1/f$ background. Systematic evaluation of these regimes is an important direction for future work.

A natural extension of our framework is to combine nonparametric priors with richer temporal dynamics. Hidden semi-Markov models (HSMMs) explicitly model state durations via duration distributions rather than relying solely on self-transition probabilities (Yu, 2010). This explicit-duration variant of the HMM may reduce brief false positives by penalizing unrealistically short state activations. There are also nonparametric Bayesian HMM/HSMM analogs, such as the hierarchical Dirichlet process HMM (HDP-HMM) (Beal et al., 2001) and HDP-HSMM (Johnson & Willsky, 2013). Like the DP–GMM, these models utilize Dirichlet process priors to reduce unused states to negligible occupancy, thereby adapting the number of states to the data. Integrating time-delay embedding with these duration-regularized, nonparametric dynamical models is another promising direction for future work that may combine the shrinkage benefits of the DP–GMM with explicit control over state persistence.

Taken together, these results indicate that the DP-GMM clustering on TDE representations provides a robust, fully unsupervised framework for characterizing oscillatory events in noisy neural recordings, reducing dependence on narrowband filtering and *a priori* model-order selection. Conceptually, this reframes oscillation analysis from detecting activity within preselected bands to discovering discrete, frequency-selective states whose occupancy, duration, and spectral profile can be quantified and compared. This state-based perspective should facilitate more reproducible analyses of time-varying oscillatory dynamics across participants, brain regions, and task contexts, particularly in

settings where the relevant frequencies and the number of oscillatory states are not known in advance, and where oscillatory dynamics are linked to cognitive processes.

Acknowledgments

Parts of the computations were performed on the HPC cluster *Elysium* at Ruhr University Bochum, subsidized by the DFG (INST 213/1055-1).

Data and code availability

The code used for simulation, model fitting, and analysis is available at https://github.com/schmidt-neural-data-science/TDE-DP-GMM/tree/main.

We analyzed resting-state MEG data from the MEGUK partnership ("CTF rest" dataset; 275-channel CTF system, University of Nottingham). Raw MEG data are available from the official MEGUK database subject to their data-access process. We used the preprocessed parcel-wise time series provided via the `osl-dynamics` workflow (Gohil et al., 2024), which includes scripts to obtain and preprocess the dataset. We do not redistribute raw MEG data in this repository.

References

Akaike, H. (1998). Information Theory and an Extension of the Maximum Likelihood Principle [Series Title: Springer Series in Statistics]. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected Papers of Hirotugu Akaike* (pp. 199–213). Springer New York. https://doi.org/10.1007/978-1-4612-1694-0_15

Antony, J. W., Piloto, L., Wang, M., Pacheco, P., Norman, K. A., & Paller, K. A. (2018). Sleep Spindle Refractoriness Segregates Periods of Memory Reactivation. *Current biology: CB*, *28*(11), 1736–1743.e4. https://doi.org/10.1016/j.cub.2018.04.020

Attar, E. T. (2022). Review of electroencephalography signals approaches for mental stress assessment. *Neurosciences*, *27*(4), 209–215. https://doi.org/10.17712/nsj.2022.4.20220025

Bacci, S., Pandolfi, S., & Pennoni, F. (2012, December). A comparison of some criteria for states selection in the latent Markov model for longitudinal data [arXiv:1212.0352 [stat]]. https://doi.org/10.48550/arXiv.1212.0352

Barnard, J., McCulloch, R., & Meng, X. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*. Retrieved November 21, 2025, from https://www.semanticscholar.org/paper/Modeling-covariance-matrices-in-terms-of-standard-Barnard-McCulloch/d922535f117db63b46ff9d09421ecd0e277d873b

Beal, M., Ghahramani, Z., & Rasmussen, C. (2001). The Infinite Hidden Markov Model. *Advances in Neural Information Processing Systems*, *14*. Retrieved August 24, 2024, from https://papers.nips.cc/paper_files/paper/2001/hash/e3408432c1a48a52fb6c74d926b38886-Abstract.html

Belluscio, M. A., Mizuseki, K., Schmidt, R., Kempter, R., & Buzsáki, G. (2012). Cross-Frequency Phase–Phase Coupling between Theta and Gamma Oscillations in the Hippocampus [Publisher: Society for Neuroscience Section: Articles]. *Journal of Neuroscience*, *32*(2), 423–435. https://doi.org/10.1523/JNEUROSCI.4122-11.2012

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association*, *112*(518), 859–877. https://doi.org/10.1080/01621459.2017.1285773

Boughorbel, S., Jarray, F., & El-Anbari, M. (2017). Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric (Q. Zou, Ed.). *PLOS ONE*, *12*(6), e0177678. https://doi.org/10.1371/journal.pone.0177678

Buckby, J., Wang, T., Fletcher, D., Zhuang, J., Takeo, A., & Obara, K. (2023). Finding the number of latent states in hidden Markov models using information criteria. *Environmental and Ecological Statistics*, *30*(4), 797–825. https://doi.org/10.1007/s10651-023-00584-5

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science (New York, N.Y.)*, *304*(5679), 1926–1929. https://doi.org/10.1126/science.1099745

Cavanagh, J. F., & Shackman, A. J. (2015). Frontal midline theta reflects anxiety and cognitive control: Meta-analytic evidence. *Journal of Physiology, Paris*, *109*(1-3), 3–15. https://doi.org/10.1016/j.jphysparis.2014.04.003

Cho, H., Adamek, M., Willie, J., & Brunner, P. (2023). Novel Cyclic Homogeneous Oscillation Detection Method for High Accuracy and Specific Characterization of Neural Dynamics [Journal Abbreviation: eLife Publication Title: eLife], *12*. https://doi.org/10.7554/eLife.91605

Cho, S., & Choi, J. H. (2023). A guide towards optimal detection of transient oscillatory bursts with unknown parameters. *Journal of Neural Engineering*, *20*(4), 046007. https://doi.org/10.1088/1741-2552/acdffd

Cole, S., Donoghue, T., Gao, R., & Voytek, B. (2019). NeuroDSP: A package for neural digital signal processing. *Journal of Open Source Software*, *4*(36), 1272. https://doi.org/10.21105/joss.01272

Cole, S. R., & Voytek, B. (2017). Brain Oscillations and the Importance of Waveform Shape [MAG ID: 2568766977]. *Trends in Cognitive Sciences*, *21*(2), 137–149. https://doi.org/10.1016/j.tics.2016.12.008

de Cheveigné, A., & Nelken, I. (2019). Filters: When, Why, and How (Not) to Use Them. *Neuron*, *102*(2), 280–293. https://doi.org/10.1016/j.neuron.2019.02.039

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews. Neuroscience*, *11*(2), 114–126. https://doi.org/10.1038/nrn2762

Donoghue, T., Schaworonkow, N., & Voytek, B. (2022). Methodological considerations for studying neural oscillations [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejn.15361]. *European Journal of Neuroscience*, *55*(11-12), 3502–3527. https://doi.org/10.1111/ejn.15361

Duchet, B., Ghezzi, F., Weerasinghe, G., Tinkhauser, G., Kühn, A. A., Brown, P., Bick, C., & Bogacz, R. (2021). Average beta burst duration profiles provide a signature of dynamical changes between the ON and OFF medication states in Parkinson's disease [Publisher: Public Library of Science]. *PLOS Computational Biology*, *17*(7), e1009116. https://doi.org/10.1371/journal.pcbi.1009116

Feingold, J., Gibson, D. J., DePasquale, B., & Graybiel, A. M. (2015). Bursts of beta oscillation differentiate post-performance activity in the striatum and motor cortex of monkeys performing movement tasks [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *112*(44), 13687–13692. https://doi.org/10.1073/pnas.1517629112

Ferguson, T. S. (1973). A Bayesian Analysis of Some Nonparametric Problems [Publisher: Institute of Mathematical Statistics]. *The Annals of Statistics*, *1*(2), 209–230. https://doi.org/10.1214/aos/1176342360

Fraser, A. M. (2008). *Hidden Markov models and dynamical systems* (Nachdr.). Soc. for Industrial; Applied Math.

Fries, P. (2015). Rhythms For Cognition: Communication Through Coherence. *Neuron*, *88*(1), 220–235. https://doi.org/10.1016/j.neuron.2015.09.034

Gershman, S. J., & Blei, D. M. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, *56*(1), 1–12. https://doi.org/10.1016/j.jmp.2011.08.004

Gohil, C., Huang, R., Roberts, E., Van Es, M. W., Quinn, A. J., Vidaurre, D., & Woolrich, M. W. (2024). Osl-dynamics, a toolbox for modeling fast dynamic brain activity. *eLife*, *12*, RP91949. https://doi.org/10.7554/eLife.91949

Haegens, S., Cousijn, H., Wallis, G., Harrison, P. J., & Nobre, A. C. (2014). Inter- and intra-individual variability in alpha peak frequency. *NeuroImage*, *92*, 46–55. https://doi.org/10.1016/j.neuroimage.2014.01.049

Hahn, L. A., Balakhonov, D., Lundqvist, M., Nieder, A., & Rose, J. (2022). Oscillations without cortex: Working memory modulates brainwaves in the endbrain of crows. *Progress in Neurobiology*, *219*, 102372. https://doi.org/10.1016/j.pneurobio.2022.102372

Hoffman, M., Blei, D. M., Wang, C., & Paisley, J. (2013, April). Stochastic Variational Inference [arXiv:1206.7051 [cs, stat]]. Retrieved August 2, 2024, from http://arxiv.org/abs/1206.7051

Huang, R., Gohil, C., & Woolrich, M. W. (2024, January). Modelling variability in functional brain networks using embeddings. https://doi.org/10.1101/2024.01.29.577718

Ishwaran, H., & James, L. F. (2001). Gibbs Sampling Methods for Stick-Breaking Priors. *Journal of the American Statistical Association*, *96*(453), 161–173. https://doi.org/10.1198/016214501750332758

Johnson, M. J., & Willsky, A. S. (2013). Bayesian nonparametric hidden semi-Markov models. *J. Mach. Learn. Res.*, *14*(1), 673–701.

Korwar, R. M., & Hollander, M. (1973). Contributions to the Theory of Dirichlet Processes [Publisher: Institute of Mathematical Statistics]. *The Annals of Probability*, *1*(4), 705–711. https://doi.org/10.1214/aop/1176996898

Langford, Z. D., & Wilson, C. R. E. (2023, December). Simulations reveal that beta burst detection may inappropriately characterize the beta band. https://doi.org/10.1101/2023.12.15.571838

Latchoumane, C.-F. V., Ngo, H.-V. V., Born, J., & Shin, H.-S. (2017). Thalamic Spindles Promote Memory Formation during Sleep through Triple Phase-Locking of Cortical, Thalamic, and Hippocampal Rhythms. *Neuron*, *95*(2), 424–435.e6. https://doi.org/10.1016/j.neuron.2017.06.025

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *J. Multivar. Anal.*, *100*(9), 1989–2001. https://doi.org/10.1016/j.jmva.2009.04.008

Liljefors, J., Almeida, R., Rane, G., Lundström, J. N., Herman, P., & Lundqvist, M. (2024). Distinct functions for beta and alpha bursts in gating of human working memory [Publisher: Nature Publishing Group]. *Nature Communications*, *15*(1), 8950. https://doi.org/10.1038/s41467-024-53257-7

Little, S., Bonaiuto, J., Barnes, G., & Bestmann, S. (2019). Human motor cortical beta bursts relate to movement planning and response errors (K. Ganguly, Ed.). *PLOS Biology*, *17*(10), e3000479. https://doi.org/10.1371/journal.pbio.3000479

Liu, A. A., Henin, S., Abbaspoor, S., Bragin, A., Buffalo, E. A., Farrell, J. S., Foster, D. J., Frank, L. M., Gedankien, T., Gotman, J., Guidera, J. A., Hoffman, K. L., Jacobs, J., Kahana, M. J., Li, L., Liao, Z., Lin, J. J., Losonczy, A., Malach, R., … Buzsáki, G. (2022). A consensus statement on detection of hippocampal sharp wave ripples

and differentiation from other fast oscillations [Publisher: Nature Publishing Group]. *Nature Communications*, *13*(1), 6000. https://doi.org/10.1038/s41467-022-33536-x

Mahjoory, K., Schoffelen, J.-M., Keitel, A., & Gross, J. (2020). The frequency gradient of human resting-state brain oscillations follows cortical hierarchies (L. Dugué, L. L. Colgin, & L. Dugué, Eds.) [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *9*, e53715. https://doi.org/10.7554/eLife.53715

Matthews, B. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, *405*(2), 442–451. https://doi.org/10.1016/0005-2795(75)90109-9

MEG UK Scientific Research Community. (2023). The scientific community for {MEG} research in the United Kingdom and Ireland. Retrieved March 15, 2025, from https://meguk.ac.uk

Muralidharan, V., Aron, A. R., Cohen, M. X., & Schmidt, R. (2023). Two modes of midfrontal theta suggest a role in conflict and error processing. *NeuroImage*, *273*, 120107. https://doi.org/10.1016/j.neuroimage.2023.120107

Muralidharan, V., Vignesh Muralidharan, Adam R Aron, Adam R. Aron, Robert Schmidt, & Robert Schmidt. (2022). Transient beta modulates decision thresholds during human action-stopping [MAG ID: 4220971693 S2ID: e42a56e8f2598cde6c31f768f09854b1721d6914]. *NeuroImage*, 119145–119145. https://doi.org/10.1016/j.neuroimage.2022.119145

Obermeyer, F., Bingham, E., Jankowiak, M., Chiu, J. T., Pradhan, N., Rush, A. M., & Goodman, N. D. (2019). Tensor Variable Elimination for Plated Factor Graphs. *ArXiv*. Retrieved November 4, 2025, from https://www.semanticscholar.org/paper/Tensor-Variable-Elimination-for-Plated-Factor-Obermeyer-Bingham/94639c9ebdee773d6224504fab5c218a6149bd53

Oyanedel, C. N., Durán, E., Niethard, N., Inostroza, M., & Born, J. (2020). Temporal associations between sleep slow oscillations, spindles and ripples [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejn.14906]. *European Journal of Neuroscience*, *52*(12), 4762–4778. https://doi.org/10.1111/ejn.14906

Pohle, J., Langrock, R., Beest, F. v., & Schmidt, N. M. (2017, April). Selecting the Number of States in Hidden Markov Models - Pitfalls, Practical Challenges and Pragmatic Solutions [arXiv:1701.08673 [stat]]. https://doi.org/10.48550/arXiv.1701.08673

Quinn, A., van Ede, F., Brookes, M., Heideman, S., Nowak, M., Seedat, Z., Vidaurre, D., Zich, C., Nobre, A., & Woolrich, M. (2019). Unpacking Transient Event Dynamics in Electrophysiological Power Spectra. *Brain Topography*, *32*. https://doi.org/10.1007/s10548-019-00745-5

Quinn, A. J., Vidaurre, D., Abeysuriya, R., Becker, R., Nobre, A. C., & Woolrich, M. W. (2018). Task-Evoked Dynamic Network Analysis Through Hidden Markov Modeling. *Frontiers in Neuroscience*, *12*, 603. https://doi.org/10.3389/fnins.2018.00603

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286. https://doi.org/10.1109/5.18626

Ranganath, R., Gerrish, S., & Blei, D. M. (2014). Black Box Variational Inference. *Artificial Intelligence and Statistics*, 814–822.

Rasch, B., & Jan Born. (2013). About Sleep's Role in Memory. *Physiological Reviews*, *93*(2), 681–766. https://doi.org/10.1152/physrev.00032.2012

Rayson, H., Debnath, R., Alavizadeh, S., Fox, N., Ferrari, P. F., & Bonaiuto, J. J. (2022). Detection and analysis of cortical beta bursts in developmental EEG Data. *Developmental Cognitive Neuroscience*, *54*, 101069. https://doi.org/10.1016/j.dcn.2022.101069

Rayson, H., Moreau, Q., Gailhard, S., Szul, M. J., & Bonaiuto, J. J. (2025). Beta Burst Waveform Diversity: A Window onto Cortical Computation [Publisher: SAGE PublicationsSage CA: Los Angeles, CA]. *The Neuroscientist*. https://doi.org/10.1177/10738584251390779

Schmidt, R., Herrojo Ruiz, M., Kilavik, B. E., Lundqvist, M., Starr, P. A., & Aron, A. R. (2019). Beta oscillations in working memory, executive control of movement and thought, and sensorimotor function. *The Journal of Neuroscience*, *39*(42), 8231–8238. https://doi.org/10.1523/JNEUROSCI.1163-19.2019

Schmidt, R., Rose, J., & Muralidharan, V. (2023). Transient oscillations as computations for cognition: Analysis, modeling and function. *Current Opinion in Neurobiology*, *83*, 102796. https://doi.org/10.1016/j.conb.2023.102796

Schulman, J., Heess, N., Weber, T., & Abbeel, P. (2015). Gradient Estimation Using Stochastic Computation Graphs. *Advances in Neural Information Processing Systems*, *28*. Retrieved November 21, 2025, from https://proceedings.neurips.cc/paper_files/paper/2015/hash/de03beffeed9da5f3639a621bcab5dd4-Abstract.html

Schwarz, G. (1978). Estimating the Dimension of a Model [Publisher: Institute of Mathematical Statistics]. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Shin, H., Law, R., Tsutsui, S., Moore, C. I., & Jones, S. R. (2017). The rate of transient beta frequency events predicts behavior across tasks and species. *eLife*, *6*, e29086. https://doi.org/10.7554/eLife.29086

Siegel, M., Donner, T. H., & Engel, A. K. (2012). Spectral fingerprints of large-scale neuronal interactions. *Nature Reviews Neuroscience*, *13*(2), 121–134. https://doi.org/10.1038/nrn3137

Staresina, B. P., Bergmann, T. O., Bonnefond, M., van der Meij, R., Jensen, O., Deuker, L., Elger, C. E., Axmacher, N., & Fell, J. (2015). Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. *Nature Neuroscience*, *18*(11), 1679–1686. https://doi.org/10.1038/nn.4119

Synigal, S. R., Teoh, E. S., & Lalor, E. C. (2020). Including Measures of High Gamma Power Can Improve the Decoding of Natural Speech From EEG. *Frontiers in Human Neuroscience*, *14*, 130. https://doi.org/10.3389/fnhum.2020.00130

Tan, E., Algar, S., Corrêa, D., Small, M., Stemler, T., & Walker, D. (2023). Selecting embedding delays: An overview of embedding techniques and a new method using persistent homology. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *33*(3), 032101. https://doi.org/10.1063/5.0137223

van Ede, F., Quinn, A. J., Woolrich, M. W., & Nobre, A. C. (2018). Neural Oscillations: Sustained Rhythms or Transient Burst-Events? *Trends in Neurosciences*, *41*(7), 415–417. https://doi.org/10.1016/j.tins.2018.04.004

Varela, F., Lachaux, J.-P., Rodriguez, E., & Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration [Publisher: Nature Publishing Group]. *Nature Reviews Neuroscience*, *2*(4), 229–239. https://doi.org/10.1038/35067550

Widmann, A., Schröger, E., & Maess, B. (2015). Digital filter design for electrophysiological data – a practical approach. *Journal of Neuroscience Methods*, *250*, 34–46. https://doi.org/10.1016/j.jneumeth.2014.08.002

Williams, A. H., Poole, B., Maheswaranathan, N., Dhawale, A. K., Fisher, T., Wilson, C. D., Brann, D. H., Trautmann, E. M., Ryu, S., Shusterman, R., Rinberg, D., Ölveczky, B. P., Shenoy, K. V., & Ganguli, S. (2020). Discovering precise temporal patterns in large-scale neural recordings through robust and interpretable time warping. *Neuron*, *105*(2), 246–259.e8. https://doi.org/10.1016/j.neuron.2019.10.020

Wingate, D., & Weber, T. (2013, January). Automated Variational Inference in Probabilistic Programming [arXiv:1301.1299 [stat]]. https://doi.org/10.48550/arXiv.1301.1299

Yao, J., & Shepperd, M. (2020). Assessing software defection prediction performance: Why using the Matthews correlation coefficient matters. *Proceedings of the Evaluation and Assessment in Software Engineering*, 120–129. https://doi.org/10.1145/3383219.3383232

Yu, S.-Z. (2010). Hidden semi-Markov models. *Artificial Intelligence*, *174*(2), 215–243. https://doi.org/10.1016/j.artint.2009.11.011

Zich, C., Quinn, A. J., Mardell, L. C., Ward, N. S., & Bestmann, S. (2020). Dissecting Transient Burst Events. *Trends in Cognitive Sciences*, *24*(10), 784–788. https://doi.org/10.1016/j.tics.2020.07.004

5 Supplements

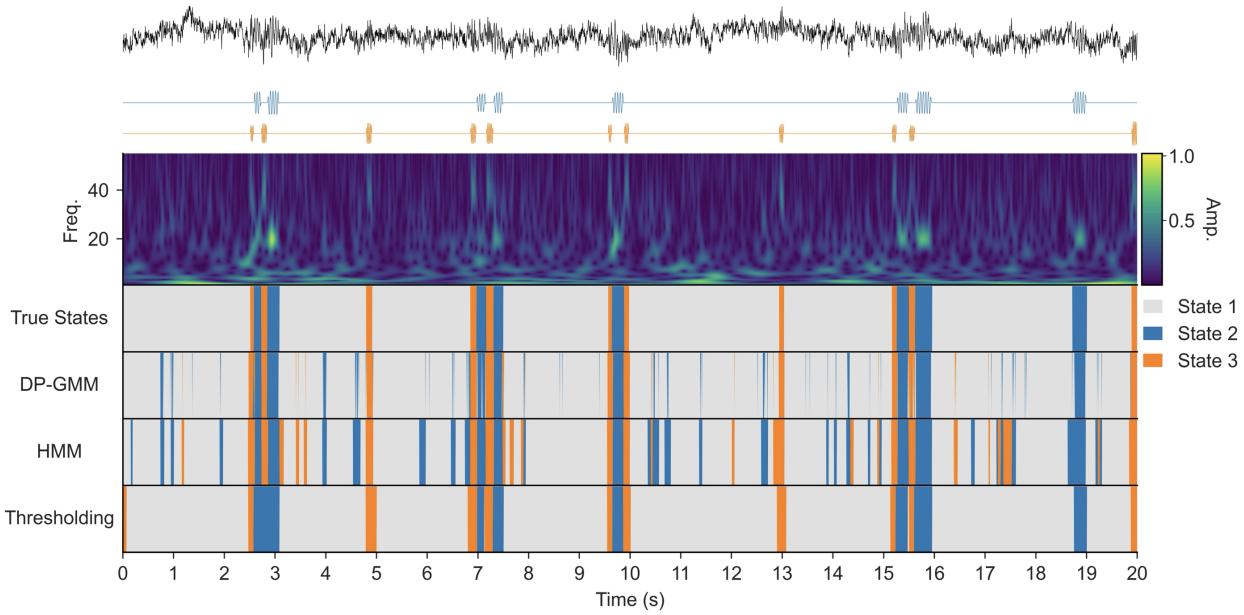

Figure S1: **Brief state fragments are visible in the TDE-based model outputs.** Rows show the three-state raw trace with 20Hz (blue) and 40 Hz (orange) oscillations, wavelet amplitude, ground-truth states, and estimated states from DP-GMM, HMM, and the thresholding method example *before* k-NN imputation.

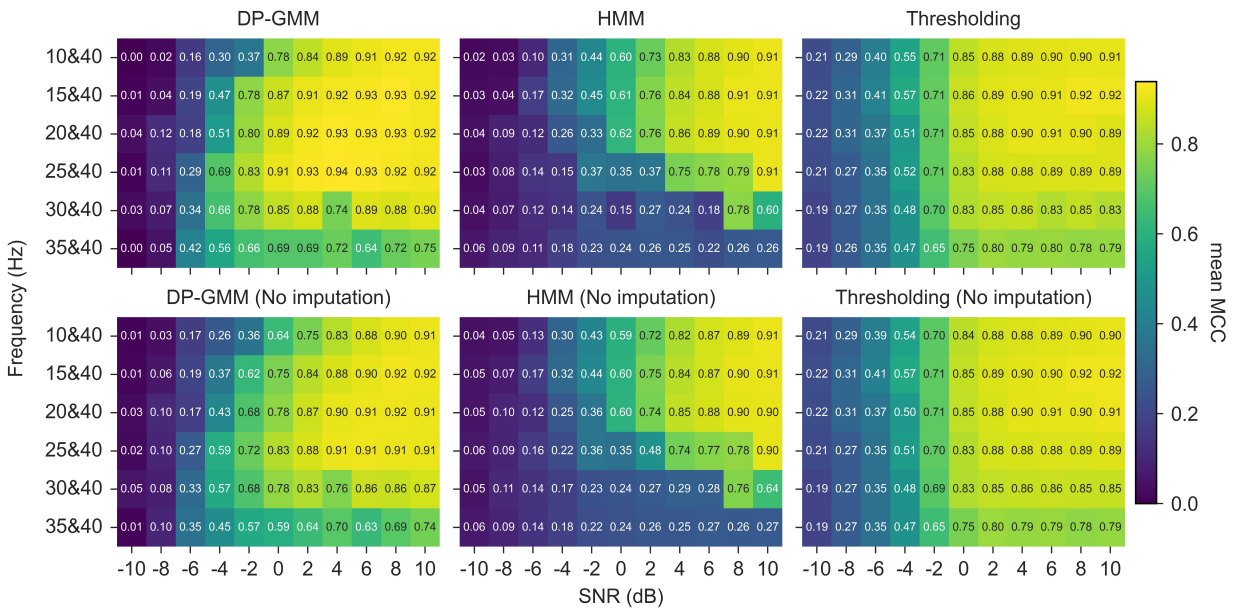

Figure S2: **Performance heatmaps across SNR and frequency pairs with and without k-NN imputation.** Mean MCC heatmaps for the three-state scenario as a function of SNR (x-axis) and oscillatory frequency pair (y-axis). Columns compare methods (DP-GMM, HMM, and thresholding). The top row shows results after post-hoc k-NN imputation of brief fragments, whereas the bottom row shows the corresponding results without imputation.

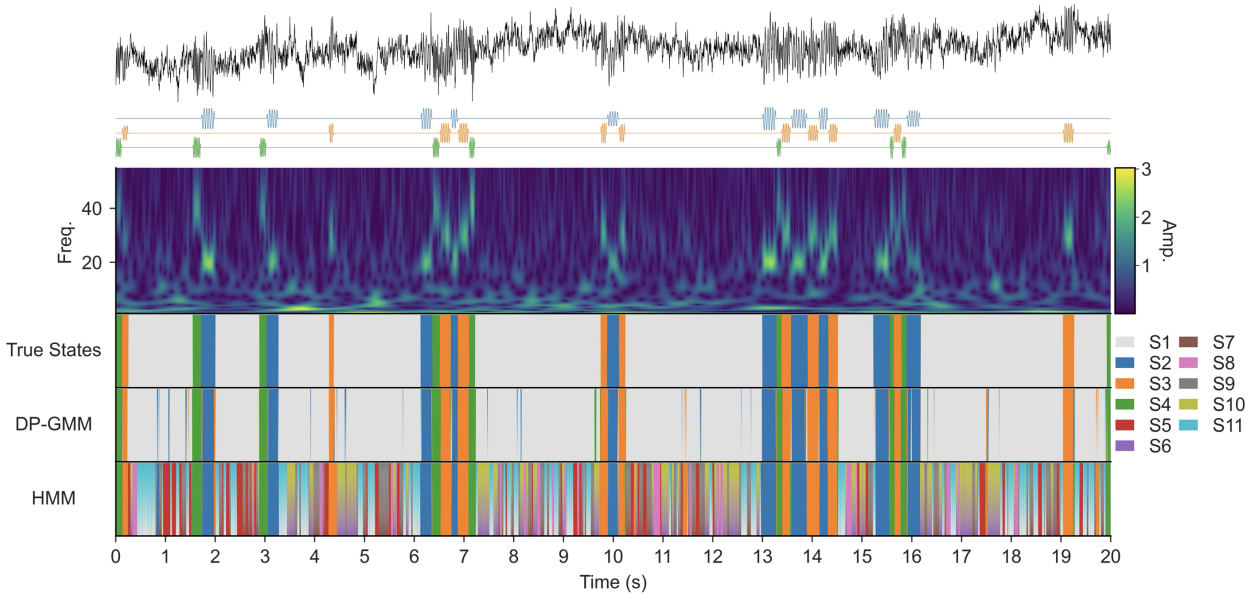

Figure S3: **In the over-specified HMM, noise is fragmented into ultra-brief substates, limiting the effectiveness of post-hoc k-NN imputation.** From top to bottom: example trace from the four-state simulation at SNR $= 2$ dB, its time-frequency representation, and inferred state time courses for the DP-GMM and the HMM *before* applying post-hoc k-NN imputation. Although both models were specified with 11 states, the DP-GMM reduced the extra states to four active states, whereas the HMM distributed occupancy across many short-lived noise-like substates.

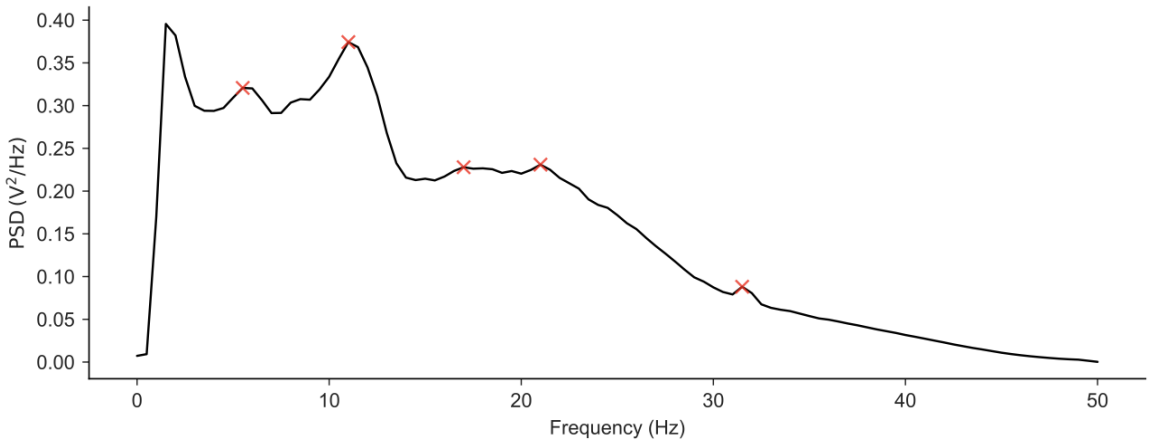

Figure S4: **Candidate peak frequencies from the group-averaged resting-state PSD used to set the TDE and post-processing windows.** Group-averaged power spectral density (PSD; V^2/Hz) of CTF resting-state MEG, estimated with Welch's method at $f_s = 100$ Hz. Local maxima (red crosses) were detected using a minimum peak separation of 4 Hz (via `scipy.signal.find_peaks` with the `distance` parameter). The lowest-frequency peak at ~ 2 Hz was excluded as likely reflecting the aperiodic $1/f$ background. The remaining candidate frequencies were 5, 11, 17, 21, and 31.5 Hz. Their mean (≈ 17.2 Hz) set the TDE window to 2.5 cycles ($E = 15$ lags; ≈ 0.145 s), and the highest candidate (31.5 Hz) defined the two-cycle minimum-duration threshold used for k-NN fragment imputation ($k = 6$).

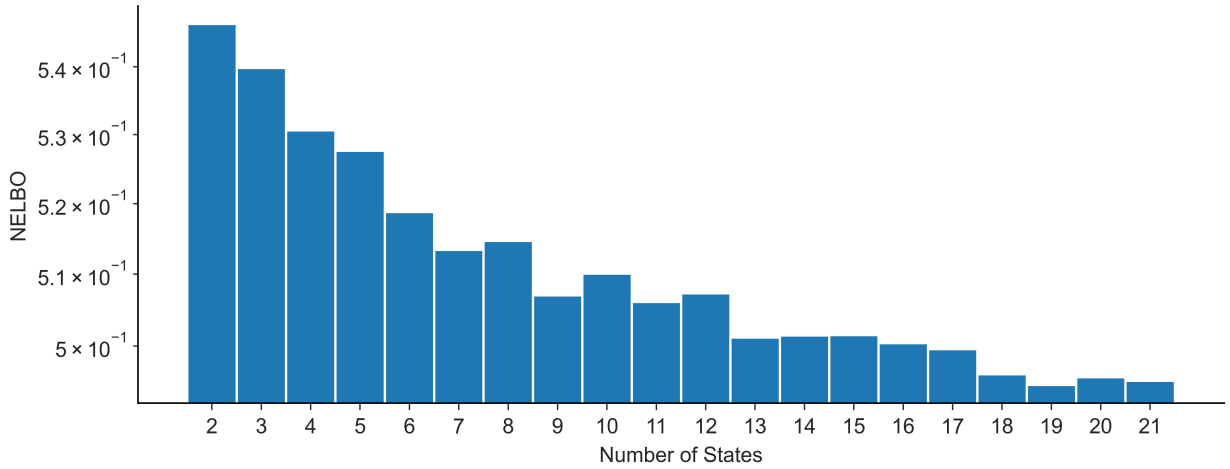

Figure S5: **(N)ELBO does not provide a clear criterion for selecting the true number of HMM states.** HMM model selection by negative ELBO (NELBO) as a function of the number of states $K \in \{4, \ldots, 21\}$ on the four-state synthetic dataset ($K_{\text{true}} = 4$). The NELBO decreases nearly monotonically with K, yielding no clear elbow or optimum at $K = 4$ and instead favoring increasingly large models.

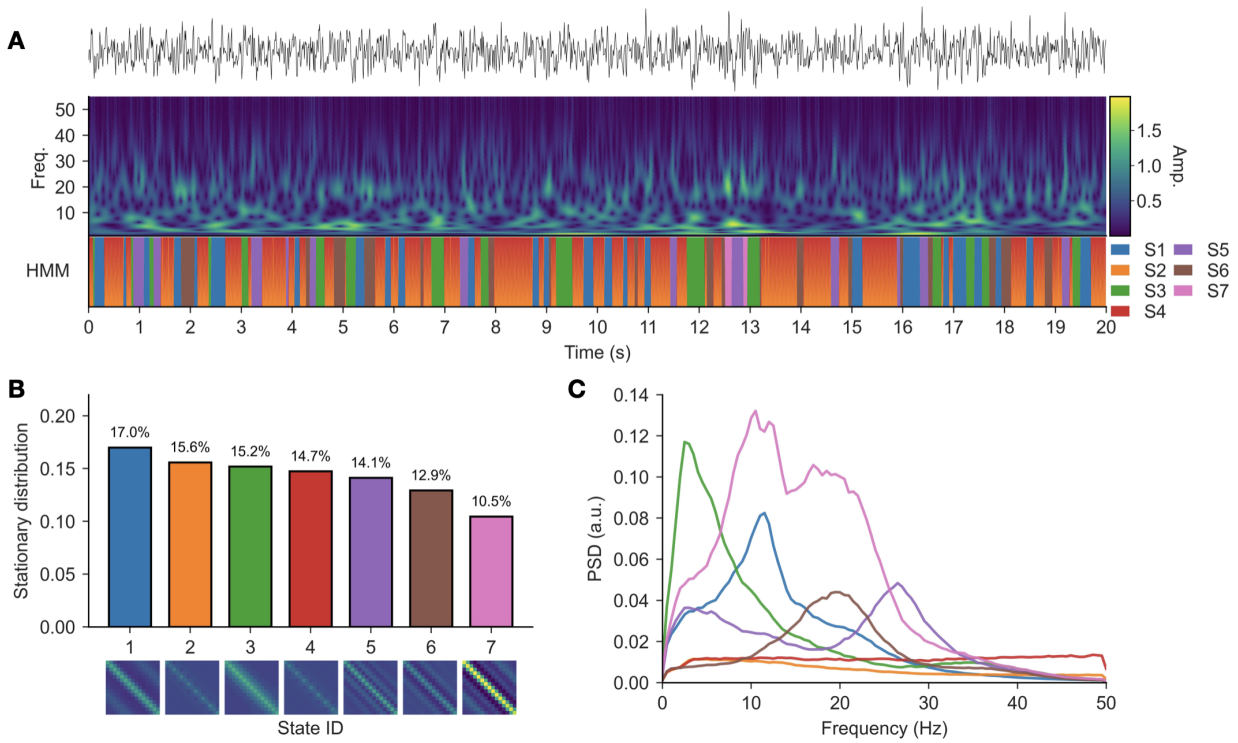

Figure S6: **Seven-state HMM without imputation recovers several DP-GMM-like states but shows rapid alternation between background substates (States 2 and 4).** (A) Example CTF resting-state trace, time-frequency representation, and decoded HMM state sequence (Viterbi path), highlighting rapid switching between State 2 (orange) and State 4 (red). (B) Estimated stationary state occupancies for $K = 7$ with state-specific autocovariance matrices. (C) State-wise power spectra. No post-hoc k-NN imputation was applied.

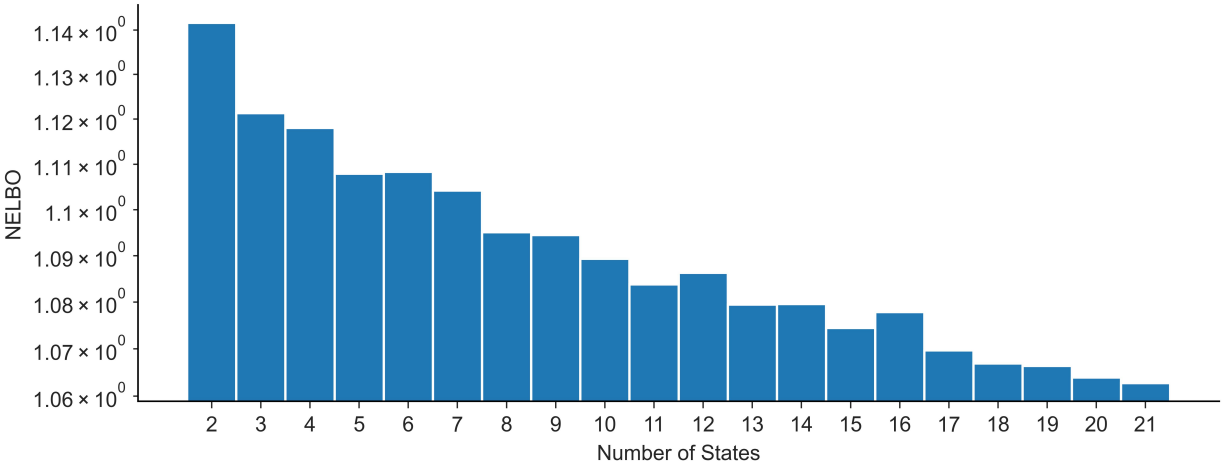

Figure S7: **NELBO favors models with more states on real MEG data and does not yield a clear choice of K.** Negative ELBO (NELBO) as a function of the number of states $K \in \{2, \ldots, 21\}$ for HMMs fit to the CTF resting-state dataset. The NELBO decreases overall as K increases, providing no clear optimum at a small K.