EVOLUTIONARY STATE-SPACE MODEL AND ITS APPLICATION TO TIME-FREQUENCY ANALYSIS OF LOCAL FIELD POTENTIALS

Xu Gao[1], Weining Shen[1], Babak Shahbaba[1], Norbert J. Fortin[1]
and Hernando Ombao[2]

[1] *University of California, Irvine and*
[2]*King Abdullah University of Science and Technology*

Abstract: We propose an evolutionary state-space model (E-SSM) for analyzing high-dimensional brain signals, the statistical properties of which evolve over the course of a nonspatial memory experiment. Under the E-SSM, brain signals are modeled as mixtures of components (e.g., an AR(2) process) with oscillatory activity at predefined frequency bands. To account for the potential nonstationarity of these components (because brain responses can vary throughout an experiment), the parameters are allowed to vary over epochs. Compared with classical approaches, such as independent component analyses and filtering, the proposed method accounts for the entire temporal correlation of the components and accommodates nonstationarity. For inference purposes, we propose a novel computational algorithm based on a Kalman smoother, maximum likelihood, and blocked resampling. The E-SSM model is applied in simulation studies and applied to multi-epoch local field potential (LFP) signal data, collected from a nonspatial (olfactory) sequence memory task study. The results confirm that our method captures the evolution of the power of the components across different phases in the experiment, and identifies clusters of electrodes that behave similarly with respect to the decomposition of different sources. These findings suggest that the activity of electrodes does change over the course of an experiment in practice. Thus, treating these epoch recordings as realizations of an identical process could lead to misleading results. In summary, the proposed method underscores the importance of capturing the evolution in brain responses over the study period.

Key words and phrases: Auto-regressive model, brain signals, state-space models, spectral analysis, time-frequency analysis.

1. Introduction

The goal of this study is to develop a novel statistical model for investigating the evolution of a brain process during a learning experiment. To infer brain neuronal activity, electrophysiological recordings such as local field poten-

tials (LFPs) and electroencephalograms (EEGs) are commonly used to indirectly measure the electrical activity of neurons. In this study, we consider LFPs from multiple electrodes that capture the integration of membrane currents in a local region of the cortex (Mitzdorf (1985)).

In practice, LFPs are observed spatio-temporal signals at different tetrodes. In a motivating example, an olfactory (nonspatial) sequence memory experiment was performed in a memory laboratory to study how neurons learn the sequential ordering of presented odors (Allen et al. (2016)). In this study, LFP recordings for a rat are obtained from an implanted plate with 12 electrodes. One epoch corresponds to about one second in physical time. We further study the behavior of these LFPs by examining their spectra. In Figure 1, we plot box plots of the log periodograms across all epochs from one electrode. These plots reveal that LFPs contain power at distinct bands: delta (0–4 Hz), alpha (8–12 Hz), and the high-beta, low-gamma (30–35 Hz) bands. As an exploratory step, we divide the entire experiment into three phases: early, middle, and late. In each phase, we compute the average periodogram (averaged across epochs); see the left side of Figure 2. On the right side of the figure, we plot the relative periodogram (obtained by rescaling the periodogram so that the relative periodogram for each frequency sums to one), and find that the spectral power evolves during the course of the experiment. During the early phase, power has a broad (rather than concentrated) spread across bands. However, in the late phase, power seems to be more concentrated at the lower beta band.

In summary, our preliminary results suggest there exists a strong similarity in the LFP waveforms across many electrodes. Moreover, the spectra of the LFPs appear to change across the epochs in the experiment. In a recent study, Gao et al. (2019) proposed a matrix data-clustering approach that also indicated the existence of spectra heterogeneity (Gao, Shen and Ombao (2018)). Therefore, we need a statistical model capable of describing the evolution of LFP signals over the course of epochs to help understand how the rat learns the sequence of the odor presentation.

In the literature, LFPs and other electrophysiological signals are commonly characterized as mixtures of different underlying brain oscillatory processes, and several approaches have been used to estimate these latent independent sources (Whitmore and Lin (2016); Einevoll et al. (2007); Prado and Lopes (2013)). For example, a data-adaptive method such as an independent component analysis (ICA) or a principal component analysis (PCA) can provide estimates for the unobserved cortical sources. However, they usually do not take into account the

spectral structure within the underlying sources that could evolve over the course of the experiment, given multiple epochs. Moreover, without any constraint on the structure of the sources, it is extremely difficult to pool information across epochs in the experiment. Recently, Fiecas and Ombao (2016) studied the dynamics of LFPs during the course of an experiment using Cramér representations. Their approach does not consider low-dimensional representations, which are indispensable to modeling high-dimensional multi-electrode LFPs.

To overcome the aforementioned limitations, we develop an evolutionary state-space model (E-SSM) that explicitly captures the evolutionary behavior in high-dimensional time series. The E-SSM shares a similar form to the classical state-space model (SSM; see Shumway and Stoffer (2013)), but differs in that the parameters vary across epochs; furthermore, the mixing matrix is unknown, and therefore has to be estimated. Moreover, the E-SSM manages to capture the temporal correlation of each of the latent sources by characterizing them using second-order autoregressive [AR(2)] processes. The reason for choosing AR(2) is its ability to capture the precise oscillatory behavior of these latent sources. In particular, by parameterizing these sources as AR(2), we can easily constrain the power of each source to center at prespecified frequency bands, such as delta (0–4 Hz), alpha (8–12 Hz), and high-beta gamma (> 30 Hz). These frequency bands are chosen according to standard convention in neuroscience, based upon previous electrophysiological data analyses (Deuschl (1999)). The use of an AR(2) mixture here is analogous to using Gaussian mixture models in classical density estimation problems. Compared with classical methods such as the ICA and PCA, the sources produced by the E-SSM are more directly interpretable in terms of their oscillatory properties.

The main contributions of this study are as follows. First, the proposed E-SSM model provides a rigorous framework for modeling brain activity and connectivity, as well as the dynamic behavior of the brain during an experiment. In particular, our model accounts for the temporal evolution/dependence of the spectrum power for particular frequency bands throughout the experiment, as well as the temporal structure among the latent sources. Second, the E-SSM yields interpretable results by modeling predominant frequency bands associated with brain functional states using AR(2) processes. Third, we show theoretically that the spectrum of an arbitrary weakly stationary time series can be approximated by the spectrum of an AR(2) mixture, which gives a theoretical justification for the use of AR(2) mixtures. We also provide a strong consistency result for the MLE of the E-SSM. Fourth, by applying the E-SSM model, one
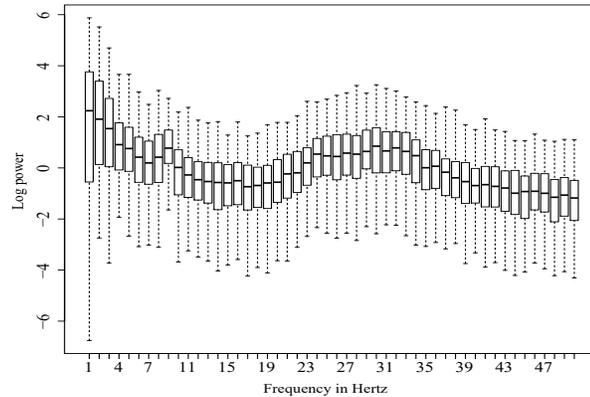
Figure 1. Log periodogram box plots for each frequency, obtained for all 247 epochs at electrode T22.
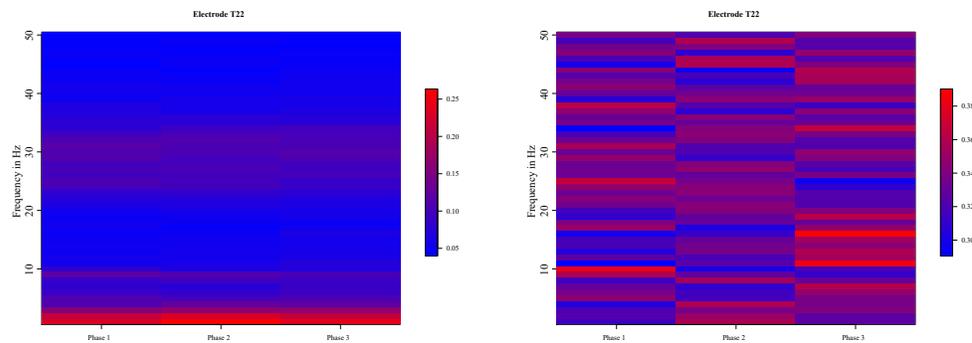

Figure 2. Left: The heatmap of the averaged periodogram among Phase 1 (epochs 1–80), Phase 2 (81–160) and Phase 3 (161–247) at electrode T22. The original signals are rescaled to unit variance. Right: The heatmap of the relative periodogram (summing to one for each frequency). The spectral power (decomposition of the waveform) evolves across phases of the experiment. Please refer to the online version for colored figures.

can easily analyze both the time and the frequency domains and, thus, provide a complete characterization of the underlying brain process. Fifth, the E-SSM model and the proposed estimation method are, in general, intuitive, and can be implemented easily owing to the existing theory and algorithms for SSMs. However, the key difference is the generalization of the multiple epochs setting, which allows us to pool information across epochs, and provides a flexible mixing matrix-estimation step.

2. The E-SSM

In this section, we discuss the models used to infer the latent structures in LFPs and their evolution across epochs over an entire experiment. We first describe the model for a single epoch, and then discuss the extension to multiple epochs.

2.1. SSM for a single epoch

Denote $t = 1, \ldots, T$ as the time points in a single epoch, and $\boldsymbol{Y}_t = (Y_t(1), \ldots, Y_t(p))'$ as the observed LFPs, where p is the number of electrodes. For any fixed time point t, we assume that $\boldsymbol{Y}_t$ is a *mixture* of q latent independent source signals $\boldsymbol{S}_t = (S_t(1), \ldots, S_t(q))'$, where q is the number of spatial source signals. We assume $p \geq q$. Then, the model can be expressed as $\boldsymbol{Y}_t = M\boldsymbol{S}_t + \boldsymbol{\epsilon}_t$, where M is the mixing matrix, $\boldsymbol{\epsilon}_t = (\epsilon_t(1), \ldots, \epsilon_t(p))'$ is noise that follows $N(\boldsymbol{0}, \tau^2 \boldsymbol{I_p})$, and $\boldsymbol{I_p}$ is an identity matrix of dimension p. Each of the independent latent signals $S_t(l)$, for $l = 1, \ldots, q$, models the source that represents the oscillatory activity at a set of prespecified frequency bands (e.g., delta, alpha, and gamma).

Modeling the source signals $\boldsymbol{S}_t$

One important parameterization in our model is to constrain the sources to have an AR(2) structure, such that each represents a particular oscillator: delta (δ: 0–4 Hz), theta (θ: 4–8 Hz), alpha (α: 8–12 Hz), lower beta (β: 12–18 Hz) and gamma (γ: > 30 Hz). Recall that an autoregressive operator of order 2 is defined by

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2, \tag{2.1}$$

where B is a backshift operator, defined by $B^\ell S_t = S_{t-\ell}$, and ϕ_1, ϕ_2 are the corresponding coefficients. It can be shown that the spectrum of an AR(2) process with noise level σ_ω is $f_S(\omega) = \sigma_w^2/(|1 - \phi_1 \exp(-2\pi i\omega) - \phi_2 \exp(-4\pi i\omega)|^2)$. To illustrate its use in practice, we plot the spectrum of an AR(2) process with $\phi_1 = 1.976, \phi_2 = -0.980$, and $\sigma_w = 0.1$ in Figure 20 in the online Supplementary Material. It can be seen that there is a peak at frequency $\omega = 10$ Hz, which means the frequency band around $\omega = 10$ Hz dominates the process and, thus, produces the most power. This property of the AR(2) model makes it potentially useful for characterizing brain signals (such as LFPs) with oscillations at either broad or narrow frequency bands.

We now explain the connection between the AR(2) coefficients and the spectrum (i.e., the location and the spread of the peak). First, the process is causal

when the roots of the polynomial in Equation (2.1) have magnitudes greater than one. Furthermore, under causality, Jiru (2008) and Shumway and Stoffer (2013) demonstrate that when the roots of the polynomial in Equation (2.1) are complex-valued and have magnitude greater than one, the spectrum attains a peak centered around the phase of the roots. Moreover, when the magnitude of the roots becomes larger than one, the peak becomes less concentrated around the phase.

Motivated by this result, we fix the phase (or argument) of each AR(2) polynomial root in order to model each of the bands obtained from previous study results. As noted, fixing the phase is consistent with neuroscience standards and, thus, will not be a constraint in practice. In the field of neuroscience, neural oscillations are widely captured at all levels, including LFPs, EEGs and neuro spike trains (Busch, Dubois and VanRullen (2009)). To characterize these oscillatory patterns, one typical approach is to convert the original electrophysiological signals to the spectrum domain using a Fourier transformation. In this way, oscillations can be represented by moduli and phases. The first interesting frequency band (alpha band) was introduced by Gerrard and Malcolm (2007). Subsequently, other bands have been studied including the delta, theta, beta, and gamma bands. As shown in the motivating example in Section 1, our collaborators from neuroscience studied the frequency domain behavior, concluding that "low-gamma oscillations were more strongly modulated by temporal context and performance than theta oscillations" (Allen et al. (2016)). Thus, by fixing the phase in our model, we are able to maintain consistency with neuroscience standards, which means there is not a constraint in practice. To model the evolution across epochs, we allow the modulus of the AR(2) polynomial roots to change among epochs. As a result, because the phases of the roots for each of the latent independent source signals are fixed, the AR(2) process is uniquely determined by the modulus and the variance. In practice, the value of modulus controls the spread of the spectrum curves. For an AR(2) process $S_t = \phi_1 S_{t-1} + \phi_2 S_{t-2} + w_t$, the modulus ρ and phase ψ of the roots of the polynomial have the relationship that $\phi_1 = 2\rho^{-1} cos(\psi)$ and $\phi_2 = -\rho^{-2}$. This is analogous to using a Gaussian mixture model (or any location-scale mixture, in general) for density estimation.

Generalized SSM

Following the previous discussion, the latent independent spatial source signals are modeled as multivariate AR(2)s, $\boldsymbol{S}_t = \Phi_1 \boldsymbol{S}_{t-1} + \Phi_2 \boldsymbol{S}_{t-2} + \boldsymbol{\eta}_t$, where $\Phi_1 = \text{diag}(\phi_{11}, \dots, \phi_{q1})$, and $\Phi_2 = \text{diag}(\phi_{12}, \dots, \phi_{q2}) \in \mathbb{R}^{q \times q}$ are diagonal matrices,

and the noise $\boldsymbol{\eta}_t = (\eta_1(t), \ldots, \eta_q(t))' \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I_q})$. The final model can hence be viewed as a generalized SSM:

$$
\begin{aligned}
\boldsymbol{Y}_t &= \widetilde{M} \boldsymbol{X}_t + \boldsymbol{\epsilon}_t, \\
\boldsymbol{X}_t &= \widetilde{\Phi} \boldsymbol{X}_{t-1} + \widetilde{\boldsymbol{\eta}}_t,
\end{aligned}
\tag{2.2}
$$

where $\boldsymbol{X}_t = (\boldsymbol{S}_t', \boldsymbol{S}_{t-1}')'$, $\widetilde{M} = (M, \mathbf{0}) \in \mathbb{R}^{p*2q}$, $\widetilde{\Phi} = \begin{bmatrix} \Phi_1 & \Phi_2 \\ \boldsymbol{I_q} & \mathbf{0} \end{bmatrix}$, and $\widetilde{\boldsymbol{\eta}}_t = (\boldsymbol{\eta}_t', \mathbf{0})'$.
Note that the residual $\boldsymbol{\epsilon_t}$ is assumed to be independent over time t. This implies that all temporal correlations are characterized by the underlying latent signals $\boldsymbol{S}_t$. The model in (2.2) is not a regular SSM, because the mixing matrix $\widetilde{M}$ is unknown. Moreover, following the aforementioned discussion, the coefficients of the autoregressive processes are determined by the modulus $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_q)$ and phase $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_q)$ of the autoregressive polynomial roots. Because we are interested in particular frequency bands, we fix the phase $\boldsymbol{\psi}$, and the state equation in (2.2) is parameterized by $\boldsymbol{\rho}$ and σ^2.

2.2. E-SSM for multiple epochs

Next, we generalize the model in Section 2.1 to accommodate multiple epochs. We assume that across epochs, the mixing matrix M is fixed, and that the latent independent autoregressive processes evolve through the modulus $\boldsymbol{\rho}$. This assumption implies that the cortical structure remains unchanged across epochs for each individual. We denote the epochs in the experiment as $r = 1, \ldots, R$. Then the model is given by

$$
\begin{aligned}
\boldsymbol{Y}_t^{(r)} &= \widetilde{M} \boldsymbol{X}_t^{(r)} + \boldsymbol{\epsilon}_t^{(r)}, \\
\boldsymbol{X}_t^{(r)} &= \widetilde{\Phi}^{(r)} \boldsymbol{X}_{t-1}^{(r)} + \widetilde{\boldsymbol{\eta}}_t^{(r)},
\end{aligned}
\tag{2.3}
$$

where the definitions of $\boldsymbol{Y}_t^{(r)}, \widetilde{M}, \boldsymbol{X}_t^{(r)}, \widetilde{\Phi}^{(r)}, \boldsymbol{\epsilon}_t^{(r)}$, and $\widetilde{\boldsymbol{\eta}}_t^{(r)}$ are similar to those in Equation (2.2), except for the additional superscript for epoch r.

In the proposed model, we assume an autoregressive structure that evolves across epochs. This assumption is inspired by the preliminary analysis in Section 1, which showed that the power spectrum evolves during the experiment. Accordingly, the evolutionary spectrum of each latent source will be captured easily in explicit form as $f^{(r)}(\omega) = \sigma_w^{2(r)} / (|1 - \phi_1^{(r)} \exp(-2\pi i\omega) - \phi_2^{(r)} \exp(-4\pi i\omega)|^2)$. We also assume that the mixing matrix is invariant to epochs. This is because the network structure of subjects does not change across the phases of the exper-

iments. To reiterate, the nonstationarity is captured by the AR(2) coefficients.

The literature includes numerous discussions on identifiability issues related to SSMs (Hamilton (1994)). Indeed, for a general SSM, the same representation can be obtained by applying an orthogonal transformation on matrices. Zhang and Hyvärinen (2011) proposed a nonGaussian constraint to avoid the identifiability issue. In this study, to ensure the uniqueness of the solution, we require that each component of the latent independent source signals $\boldsymbol{S}(t)$ have unit variance, and that the entries of $\widetilde{M}$ be positive.

3. Inferences for the E-SSM

3.1. Estimating the E-SSM

For the E-SSM with a single epoch, we propose an iterative algorithm that incorporates a Kalman filter and the least squares method for parameter estimation purposes. Further details are given in Section 3 of the Supplementary Material.

Next, we extend the previous method to a multiple epoch setting in Equation (2.3). Here, the major challenge lies in pooling information from different epochs in order to estimate the epoch-invariant mixing matrix. Inspired by the resampling approach used to model time series with a Gaussian process (Gao, Shahbaba and Ombao (2018)) and the linear mixed model (Cheng et al. (2014)), we propose a blocked resampling-based approach. The key idea of the approach is as follows. First, we divide the epochs into blocks; then, for each block, we estimate the corresponding mixing matrix and the epoch-specific AR(2) parameters. These blocks retain the temporal sequence of the epochs, and the final estimate at a previous epoch serves as the initial estimate of the mixing matrix at the current epoch. The final estimates of the mixing matrix obtained from each block are averaged to produce the estimate for the common mixing matrix. For the next step, given the estimated mixing matrix, we follow Algorithm 1 to obtain estimates of the epoch-specific AR(2) parameters. The approach is summarized below.

II.A We fix the length of the blocked resampling sampler as l. We draw the starting epoch index s from the set $\{1, \ldots, R - l + 1\}$. Then, at the current iteration, the blocked resampling sampler is $(\{\boldsymbol{Y}_t^{(s)}\}_{t=1}^T, \ldots, \{\boldsymbol{Y}_t^{(s+l-1)}\}_{t=1}^T)$.

A.1. Starting with epoch s, we implement the approach for a single epoch in Section 2.1 on $\{\boldsymbol{Y}_t^{(s)}\}_{t=1}^T$ to obtain the estimates $\widetilde{M}^{(s)}$.

A.2. Starting with epoch $s+1$ and the initial value $\widetilde{M}^{(s)}$, we repeat A.1 to obtain the estimates $\widetilde{M}^{(s+1)}$.

A.3. We repeat A.2 until the last epoch $s+l-1$. We denote the final estimates $\widetilde{M}^{(s+l-1)}$ as the ultimate estimates of the resampling sampler $(\{\boldsymbol{Y}_t^{(s)}\}_{t=1}^T, \ldots, \{\boldsymbol{Y}_t^{(s+l-1)}\}_{t=1}^T)$. The pipeline of the procedure is summarized below:

$$
\begin{bmatrix} \boldsymbol{Y}_1^{(s)} \\ \boldsymbol{Y}_2^{(s)} \\ \cdots \\ \boldsymbol{Y}_T^{(s)} \end{bmatrix} \rightarrow \widetilde{M}^{(s)} \rightarrow \begin{bmatrix} \boldsymbol{Y}_1^{(s+1)} \\ \boldsymbol{Y}_2^{(s+1)} \\ \cdots \\ \boldsymbol{Y}_T^{(s+1)} \end{bmatrix} \rightarrow \widetilde{M}^{(s+1)} \cdots \rightarrow \begin{bmatrix} \boldsymbol{Y}_1^{(s+l-1)} \\ \boldsymbol{Y}_2^{(s+l-1)} \\ \cdots \\ \boldsymbol{Y}_T^{(s+l-1)} \end{bmatrix} \rightarrow \widetilde{M}^{(s+l-1)}.
$$

II.B. Repeat II.A until a sufficient number of resampling estimates is obtained. Compute the average of those estimates, defined by $\widetilde{M}_g$, as the global estimate of $\widetilde{M}$.

II.C. Substitute the global estimate $\widetilde{M}_g$ into each epoch. Following Algorithm 1 for each epoch, we obtain the estimates of $\boldsymbol{\rho}^{(r)}$, $\sigma^{2(r)}$, and $\tau^{2(r)}$, for $r = 1, \ldots, R$.

The overall work flow is given in Figure 10 of the Supplementary Material. Note that because the mixing matrix $\widetilde{M}$ is the same across epochs, we use the blocked resampling strategy to obtain the global estimates sequentially. Given that estimate, we proceed to make an inference on each epoch. For the choice of the length l, we recommend starting with $l = CR^{1/k}$, and then increasing l until a stable result is obtained, where $k = 3, 4$, C is a constant, and R is the number of epochs.

3.2. Testing for differences across epochs

Inspired by the preliminary results shown in Figure 3, we assume that all epochs can be divided into different phases, among which there exist discrepancies in $\widetilde{\Phi}^{(r)}$. In order to test whether such differences in $\widetilde{\Phi}^{(r)}$ are significant across phases, we propose a permutation test in which we shuffle the epochs between phases, and then implement the E-SSM to obtain the parameter estimates and their reference distributions. We provide a simulation example to demonstrate its use in Section 6.

4. Theory

We start with a strong consistency result for the MLE of the proposed E-SSM model. Denote $\Theta = (M, \widetilde{\Phi}^{(1)}, \ldots, \widetilde{\Phi}^{(R)})$ as the collection of parameters in the multiple epoch model in (2.3). Let $\hat{\Theta}$ and $\Theta_0 = (M_0, \Phi_0^{(1)}, \ldots, \Phi_0^{(R)})$ be the MLE and the true value of Θ, respectively. Theorem 1 below states that, under mild conditions, $\hat{\Theta}$ is a strongly consistent estimator for Θ_0.

Theorem 1. *Suppose that the AR(2) process in the definition of $\boldsymbol{S}_t^{(k)}$ (2.3) is causal for every epoch $k = 1, \ldots, R$. Assume M_0 is of full column rank, $(\Phi_0^{(1)}, \ldots, \Phi_0^{(R)})$ is of full row rank, and every parameter (matrix) in Θ_0 belongs to a known compact support. Then, $\hat{\Theta}$ converges to Θ_0, almost surely.*

The assumptions in Theorem 1 are easily satisfied in many situations. For example, in our case, $\Phi_0^{(i)}$, for $i = 1, \ldots, R$, are diagonal matrices in which the elements are AR(2) coefficients centered at prespecified frequency bands. As long as these bands are different, this assumption is satisfied. The consistency result also applies for the single-epoch model (2.2) by letting the number of epochs $R = 1$. The proof of the theorem, which we defer to the Supplementary Material, is based on the consistency results for the general hidden Markov model in Douc et al. (2011). Next, we give an AR(2) decomposition theorem, stating that the spectrum of any weakly stationary process can be approximated by that of a linear mixture of AR(2) processes. This result provides a theoretical justification for representing individual sources by AR(2) models, owing to their ability to present each source signal at prespecified frequency bands.

Theorem 2. *Let Y_t be a weakly stationary time series with zero mean and continuous spectrum $f_Y(\omega)$. Let $0 = \omega_0 < \omega_1 < \cdots < \omega_J = 1/2$, and $\xi = \max_{j=1}^J |\omega_j - \omega_{j-1}|$. Denote $S_t^{(j)}$, for $j = 1, \ldots, J$, as independent AR(2) processes with unit variance and spectrum $f_{S^{(j)}}(\omega)$, such that the phase of its AR polynomial roots, denoted by $\psi^{(j)}$, satisfies $\psi^{(j)} \in [\omega_{j-1}, \omega_j)$. Consider a family of processes $\{Q_{t,J}\}_{J=1}^\infty$ defined by $Q_{t,J} = \sum_{j=1}^J a_j S_t^{(j)}$ with nonnegative coefficients $\{a_j\}_{j=1}^J$, and let $\mathcal{F}_J$ be the collection of spectra of $\{Q_{t,J}\}$. Assume that $\xi \to 0$ as $J \to \infty$. Then,*

$$\inf_{f \in \mathcal{F}_J} \|f_Y - f\|_2 \to 0 \quad as \quad J \to \infty. \tag{4.1}$$

Moreover, if f_Y is Lipschitz continuous and $\omega_k = k/(2J)$, for $k = 0, \ldots, J$, then for any sufficiently large J and some positive constant C,

$$\inf_{f \in \mathcal{F}_J} \|f_Y - f\|_\infty < CJ^{-1}. \tag{4.2}$$

Theorem 2 states that the minimum approximation error of the spectrum from a class of finite mixture AR(2) models is negligible, given that the number of terms J goes to infinity. In other words, the AR(2) mixture gives a consistent estimate of the spectral density, given that J is chosen to be sufficiently large. Moreover, if we assume that the frequencies ω_k are equally spaced, then the convergence rate is essentially equivalent to that of the equally spaced Fourier series based on a Jackson-type inequality. The convergence rate for a finite Fourier series with nonuniformly spaced frequency bands is still unknown, to the best of our knowledge (Epstein (2005)). The proof of the theorem is given in Section 2 of the Supplementary Material.

5. A Comparison with Existing Methods

Here, we discuss a few major differences between our method and existing approaches, including the ICA and classical SSMs.

The ICA has been widely used in single/between-subject electrophysiological exploratory analyses. For example, Makarova et al. (2011) proposed an ICA method to segregate pathways using partially overlapped synaptic territories from hippocampal LFPs. To investigate the variability across different subjects or subgroups, Guo (2011) proposed a general group probabilistic ICA (pICA) framework, which was later extended (Wang and Guo (2018); Lukemire et al. (2018)), to accommodate cross-subject structures in multi-subject spatial-temporal brain signals. Although these methods work well under certain settings, there is still much room for improvement in modeling electrophysiological signals. First, they do not have a mechanism for capturing how the parameters (and spectral properties) of the latent source signals evolve across epochs over an entire experiment. Most existing methods concatenate the signals from different epochs, and then estimate the parameters as though these signals are realizations of the *same* underlying process. However, because the "reconstructed" latent sources vary across epochs, there is no rigorous framework for modeling how these parameters may change across epochs. Second, existing methods do not take into account the temporal structure of the latent sources. In fact, these sources are estimated for each time point, independently of the other time points. Third, current ICA methods for source modeling may not produce interpretable results from a spectral analysis of electrophysiological signals. In fact, brain researchers have observed an association between the power at different frequency bands and brain functional states (Michel et al. (1992)). Thus, a framework is necessary

that accounts for the evolution of the power at these frequency bands over many epochs. Lastly, there are limitations to the connections between time and frequency domain analyses. Methods for time and frequency domains are developed almost exclusively from each other. However, this is counter-intuitive, because these two approaches ought to be used concurrently in order to give a complete characterization of brain processes.

6. Simulation Studies

6.1. Results of a single epoch analysis

We first evaluate the proposed E-SSM under the single epoch setting. We simulate data from three independent AR(2) processes that correspond to delta (δ: 0–4 Hz), alpha (α: 8–12 Hz), and lower beta (β: 12–18 Hz). We randomly generate a positive "mixing" matrix M, and fix the number of electrodes of the observational brain signals to 20. In summary, following the notation in Section 2.1, we have the following: $p = 20$, $T = 1,000$, $q = 3$, $\tau^2 = 1$, $\sigma^2 = 0.1$, $(\rho_1, \psi_1) = (1.0012, 2)$, $(\rho_2, \psi_2) = (1.0012, 8)$, $(\rho_3, \psi_3) = (1.0012, 15)$.

We implement the method proposed in Section 2.1, and plot the periodograms of the true and reconstructed signals in Figure 1 in the Supplementary Material. The estimated source signals share the same shape as the true signals. We also compare the results with those of the ICA in Figure 11 in the Supplementary Material. It is clear that the ICA is unable to recover the three latent bands, whereas our method manages to separate the spectral components very accurately. These results are consistent with our discussion on the possible drawbacks of the ICA.

6.2. Results of the multiple-epoch analysis

Here, we evaluate the performance of the proposed method for multiple epochs. We choose 20 electrodes and three latent independent AR(2) processes. To model the evolution across epochs, we allow the modulus $(\rho_1^{(r)}, \rho_2^{(r)}, \rho_3^{(r)})$ to increase from $(1.001, 1.001, 1.001)$, with an increment of 0.00005, as the epoch r propagates. All remaining parameters are the same as those in Section 6.1. Figure 2 in the Supplementary Material shows a heatmap of the periodogram from electrode 1 as the epochs evolve. The results look satisfactory. Figure 3 in the Supplementary Material shows the periodograms of the true and estimated signals from the three underlying AR(2) processes. For the delta, alpha, and lower-beta bands, we can see the peaks at the corresponding dominating frequencies from the true and estimated signals. As the epochs evolve, we find that

the true and estimated periodograms both spread out around the dominating frequency, indicating that the pattern of the periodograms from the reconstructed AR(2) process is consistent with that of the true AR(2) process.

We also applied an ICA to the simulated data set; the results are shown Figures 12 and 13 in the Supplemenart Material. As expected, the ICA barely separates the three underlying latent sources, and rarely captures the spread of power as the epoch evolves. This supports our previous discussion that the ICA neglects the dynamics across epochs.

6.3. Results for settings derived from the data

Here, we simulate data using the parameter setting from the motivating example from the sequence memory study. We use the estimated modulus $(\hat{\rho}_1^{(r)}, \hat{\rho}_2^{(r)}, \hat{\rho}_3^{(r)})$, variances $(\hat{\sigma}^{2(r)}, \hat{\tau}^{2(r)})$, and mixing matrix $\widetilde{M}$ to generate signals across 12 electrodes and 247 epochs. To evaluate the performance of the E-SSM, we also apply the classical SSM estimation methods as a benchmark. Specifically, we fit an SSM for each epoch, and obtain the epoch-specific parameter estimates. Note that this is the approach that most existing methods use to analyze signals with multiple epochs. As an alternative, we also compute the average of the epoch-specific estimates.

We compare the mean squared errors (MSEs) of the parameters obtained from the E-SSM and SSM. In Table 1, it is clear that the E-SSM successfully captures the evolution of the parameters, in contrast to the classical SSM. Across all frequency bands, the benefits are significant. These results highlight the advantages of using the E-SSM when signals comprise multiple epochs. At the same time, it also indicates the potential loss of information if we naively average over all epochs when conducting an analysis.

As a comparison, we also applied the ICA to the simulated data set. The results in Figures 14 and 15 in the Supplementary Material suggest that our method manages to estimate the mixing matrix very accurately, whereas the ICA misses most of the patterns across the electrodes. Figures 16 and 17 in the Supplementary Material show the periodograms obtained from our method and the ICA, respectively. Once again, the ICA is unable to recover the true signals or identify the dynamics across epochs.

6.4. Results of the permutation test

Following similar strategies to those in the previous simulations, we generated five latent AR(2) processes, corresponding to delta, theta, alpha, lower-beta and

Table 1. MSE obtained from the E-SSM and SSM (benchmark)

Parameters	E-SSM	SSM (average)	SSM (single)
$\widetilde{\Phi}$ (delta band)	$\mathbf{3.33 \times 10^{-5}}$	7.27×10^{-5}	5.53×10^{-5}
$\widetilde{\Phi}$ (alpha band)	$\mathbf{1.41 \times 10^{-5}}$	3.23×10^{-5}	2.89×10^{-5}
$\widetilde{\Phi}$ (gamma band)	$\mathbf{1.69 \times 10^{-5}}$	8.07×10^{-5}	2.00×10^{-5}
τ^2	$\mathbf{9.31 \times 10^{-6}}$	2.03×10^{-4}	1.91×10^{-4}
σ^2	$\mathbf{1.93 \times 10^{-1}}$	1.93×10^{-1}	1.91×10^{-1}

Table 2. Type-I error/power table of the proposed permutation test.

		$\delta(\rho) = 0$	$\delta(\rho) = 1$	$\delta(\rho) = 2$	$\delta(\rho) = 3$	$\delta(\rho) = 4$
Scenario A	ρ_2	0.056	0.110	0.586	0.966	0.924
	ρ_5	0.046	0.112	0.600	0.992	0.955
Scenario B	ρ_2	0.050	0.108	0.418	0.844	1.000
	ρ_5	0.047	0.268	0.586	1.000	1.000

gamma bands. We assumed there are two phases, with 40 epochs in total. In Scenario A, we fixed modulus $\rho_i^{(r)} = 1.001$, for $i = 1, \ldots, 5$, $r = 1, \ldots, 20$, in Phase 1. We then changed the values of $\rho_2^{(r)}, \rho_5^{(r)}$ in Phase 2, and denoted $\delta(\rho)$ as the modulus difference between phases. In Scenario B, we allowed the modulus to slowly increase by 5×10^{-5} starting from 1.001 in Phase 1, and from various values in Phase 2. The other parameters remained the same as those in the previous simulation settings. Table 2 summarizes the proportion of rejecting the null hypothesis, based on $1,500$ replications. It can be seen that the Type-I error rates are close to the nominal level of .05, and that the power increases to one rapidly under both scenarios.

6.5. Sensitivity analysis

We conducted an extensive sensitivity analysis to investigate the performance of the proposed E-SSM when the underlying model assumption is violated, including when the number of AR(2) mixture components is mis-specified, and when the underlying signal deviates from an AR(2) process. For a detailed discussion, see Section 4 in the online Supplementary Material.

7. Analysis of LFP Data from Olfactory Sequence Memory Study

7.1. Data description

The LFP data set was obtained from an experiment that searched for direct evidence of coding for the memory of sequential relationships between nonspatial events (Allen et al. (2016)). During the experiment, rats were provided with a series of five odors, multiple times. Then, as the rats performed tasks, LFPs were recorded in the CA1 pyramidal layer of the dorsal hippocampus. The LFP data set in this study incorporates 12 electrodes and 247 epochs. Each epoch is recorded over one second, aligned to port entry, and sampled at 1,000 Hz and, thus, has $T = 1,000$ time points.

7.2. Exploratory analysis

We are interested in addressing two questions. First, we wish to determine how the original high-dimensional signals can be sufficiently represented by lower-dimensional summary signals. Second, we wish to assess whether and how the spectral properties of the LFP signals evolve across epochs during the experiment.

To address the first question, we note the assertion in other studies (e.g., Makarova et al. (2014)) that the natural geometry of these neuronal assemblies gives rise to possible spatial segregation. This suggests that it is plausible to represent LFP data using lower-dimensional summaries. In this nonspatial sequence memory study, we observe a similar pattern across all 12 electrodes. In Figure 21 in the Supplementary Material, although the power varies within each electrode, the synchrony of the pattern across electrodes is still critical. For example, electrodes T13 and T14 behave almost identically. Electrodes T7, T8, and T9 also follow the same pattern during the experiment. Moreover, as part of this exploratory analysis, we conducted a spectral PCA (Brillinger (1964)), which is widely used in exploratory analyses of brain imaging data (Wang, Ting and Ombao (2016); Wang et al. (2019)). Figure 7 in the Supplementary Material presents box plots of the percentage of variability accounted for by the first one and the first three components. It can be shown that three components (a mixture of delta, alpha, and gamma bands) account for roughly 92% of the variability, with the first component accounting for 70%. These findings validate the assumption that the original LFPs can be projected into low-dimensional source signals, without substantial loss of information. In this paper, we build on this preliminary analysis by giving a more specific characterization of these signal summaries or components using the AR(2) process.
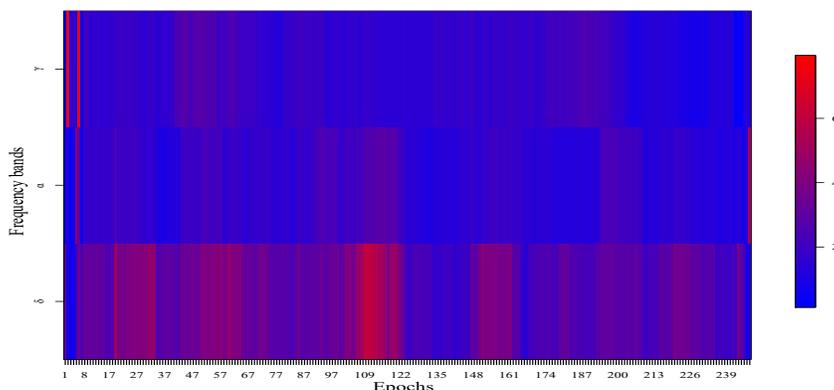
Figure 3. The evolution of the power spectrum among delta (0–4 Hz), alpha (8–12 Hz) and gamma (30–35 Hz) bands. Each band is averaged over all electrodes. Please refer to the online version for colored figures.

To address the second question, we examined the log periodogram box plots in Figure 1 across all frequencies. Here, we find that the powers are quite spread out, especially at lower frequencies and at the two peaks around the delta and slow gamma bands. The heatmap in Figure 2 illustrates the dynamics of the early, middle, and late stages of the overall session. Figure 21 in the Supplementary Material shows how the power ecolves across all electrodes, particularly on the delta, alpha, and gamma bands. It shows that higher-frequency bands dominate in the early stage, but that lower- frequency bands capture more power during the experiment. In Figure 3, an interesting pattern emerges: the burst of gamma activity in Phase 1 of the epochs is not replicated in the other phases. One possible interpretation is that the odor sequence (on which the animals have had extensive training) is re-encoded early in each session, which requires high-frequency (gamma) activity. However, later in the session, the gamma activity is regulated, and other lower frequencies (delta and alpha) become more prominent. Further research is necessary to identify the latent lower-dimensional source signals that drive the observed LFPs.

7.3. Results and discussion

We applied our proposed E-SSM method to this study. Figure 8 in the Supplementary Material shows time series plots of the modulus (root magnitudes) corresponding to each of the three frequency bands as the epochs evolve. In this plot, we can clearly identify the evolution of each individual module and a strong
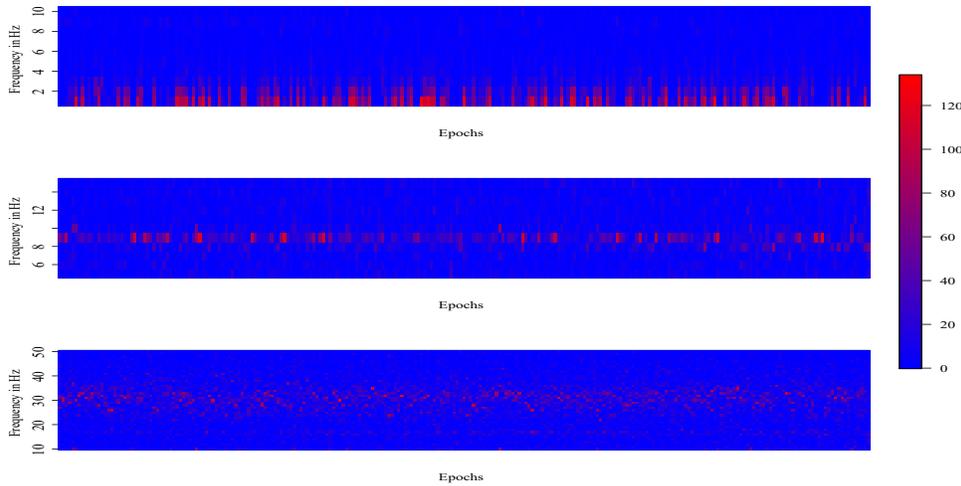
Figure 4. The periodograms of the estimated latent AR(2) processes corresponding to delta (top), alpha (middle), and gamma (bottom) bands. Please refer to the online version for colored figures.

temporal dependence. Figure 4 displays the power of the three latent source signals evolving during the experiment. We observe that the delta band captures the most power among the bands, and is persistent across all phases. The alpha band attains its maximum power during the early phase, diminishes quickly in the middle stage, and obtains more power in the final stage. There appear to be discontinuities in the delta, alpha, and gamma power across the entire experiment. One interpretation of these results is that these on-off patterns could simply be random variation. Another is that they denote the resetting of neuronal responses. This phenomenon of phase resetting in neurons is also observed in many biological oscillators. In fact, it is believed that phase resetting plays a role in promoting neural synchrony in various brain pathways. Nevertheless, it is imperative to be cautious about blindly assuming that a neuronal process behaves identically across epochs, because doing so could produce misleading results.

We also study the mixing matrix to investigate how electrodes are associated across the three frequency bands. From Figure 5, at the delta band, electrodes T13, T14, T16, T19, T22, and T23 are likely to be linked in terms of large power. Electrodes T15, T2, T7, T8 , and T9 share the lowest power. At the alpha band, electrodes T16, T22, and T23 maintain the most power; in contrast, electrodes T15, T2, and T7–9 obtain the lowest power. This pattern of association may result from the anatomical connections. Similarly, at the gamma band, electrodes
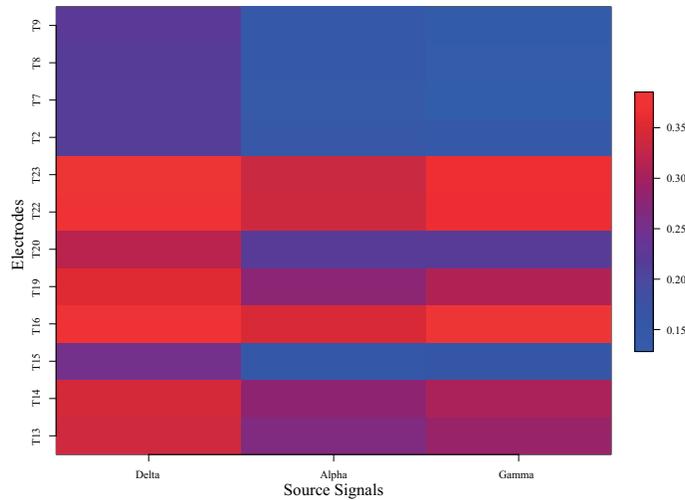
Figure 5. The estimated mixing matrix. Darker colors represent heavier weights given by the latent processes (delta, alpha, gamma) on the LFPs. Please refer to the online version for colored figures.

are connected in the same way as at the alpha band. We also used a cluster analysis on the entries of the "mixing" matrix to understand the connections between the electrodes. Similarly to the results shown in Figure 5, we are able to identify the same pattern in Figure 6 using a visualization of the cluster analysis. At the delta band, electrodes T13, T14, T16, T19, T20, T22, and T23 share the same pattern, while T3, and T7–9 are in the same cluster. Clusters at the alpha and gamma bands are roughly identical, which supports the results in Figure 5. To the best of our knowledge, this approach (i.e., clustering of electrodes or nodes) has not been used previously for this kind of analysis. This has potential for future explorations on synchrony among neuronal populations. Finally, note that the specific parametric AR(2) structure in our E-SSM facilitates ease of interpretation of the oscillatory activity of these sources.

Model validation and diagnostics were performed using sample auto-correlations (ACF) and partial auto-correlations (PACF), calculated from the residuals. Figure 9 in the Supplementary Material shows an example of values obtained from a representative electrode. We can easily observe the uncorrelated structure among the residuals. A p-value of 0.75, based on the Ljung-Box test, provides further evidence of white noise residuals and, thus, we conclude that the proposed E-SSM fits this LFP data well.

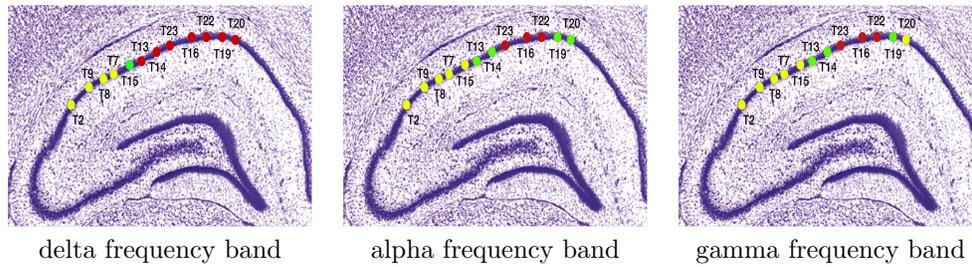| delta frequency band | alpha frequency band | gamma frequency band |

Figure 6. Cluster analysis results among all three frequency bands. The same color indicates the same cluster. Please refer to the online version for colored figures.

8. Conclusion

We have proposed an E-SSM that allows the latent source signals to evolve across epochs. Although the results reported in this paper are quite promising, nevertheless, modeling the evolution/dynamics across epochs remains a challenge in general. For example, we have ignored subject-specific random effects, which should be taken into account in future work. In the simulation studies, we set the phase parameter to the true values used in simulating the data. We also tried using estimates from the true data. This was done by first calculating the periodograms for each channel and trial, and then choosing the frequencies with peaks located as the phase parameters. Although these estimated phase parameters deviated slightly from the true values in some cases, the estimated periodograms were able to capture the structure of the true latent sources. These findings are consistent with the results from the sensitivity analysis.

In addition, we chose to fix the locations of the spectrum peaks at pre-determined values. In a frequency domain analysis, there are basically two approaches to obtaining the power for particular bands: average and integral (Delorme and Makeig (2004)). Specifically, using the average is more straight-forward, but it neglects the range of frequency bands (e.g., theta: 4–8 Hz vs. lower-beta: 12–18 Hz). Using the integral is more complicated and more sensitive to the choice of the range of frequency bands. A common observation in practice is that lower-beta bands are usually more "flattened out" than theta bands when using the average approach. The key point to both approaches is to find a "center" for each frequency band. In our model, because each latent AR(2) process corresponds to a particular band, we choose the center of the range as the exact phase parameter for each frequency band. For example, we fix the phase at 10 Hz for the alpha band. There are a few reasons to do so. First, from the existing literature, the power spectra of particular frequency bands most of-

ten achieve their peaks at the center of the range (Buzsaki (2006)). Second, by fixing the peak beforehand, we can avoid identifiability issues. If we let the data drive the estimates of the location, we could run into the identifiability problem easily; that is, we can change the columns of the mixing matrix and their corresponding AR(2) sources to get the same observed signals. Third, we conducted sensitivity analyses on different peaks within each band. The results show that the "constructed" signals are quite similar to the original ones. Fourth, recent studies (e.g., Allen et al. (2016)) report that the approach of using the center of a range produces consistent and interpretable results. In future research, it will be of interest to develop a more flexible methodology that takes into account data uncertainty when determining the location of spectrum peaks.

Note that the spectrum of a weakly stationary process being approximated by the spectrum of an AR(2) mixture does not necessarily imply that the original process is approximated by the AR(2) mixture. The focus of this study is motivated by a frequency-domain analysis of the imaging data, where the actual LFP values are not as important as their frequency-domain implications. Thus, further research should develop new models based on AR(2) mixture processes (or any other meaningful basis in practice) within a time-domain framework.

Supplementary Material

Technical proofs, an additional simulation, and data analysis results are provided in the online Supplementary Material.

Acknowledgments

References

Allen, T. A., Salz, D. M., McKenzie, S. and Fortin, N. J. (2016). Nonspatial sequence coding in ca1 neurons. *The Journal of Neuroscience* **36**, 1547–1563.

Brillinger, D. (1964). A frequency approach to the techniques of principal components, factor analysis and canonical variates in the case of stationary time series. In *Invited Paper, Royal Statistical Society Conference, Cardiff Wales.*

Busch, N. A., Dubois, J. and VanRullen, R. (2009). The phase of ongoing eeg oscillations predicts visual perception. *Journal of Neuroscience* **29**, 7869–7876.

Buzsaki, G. (2006). *Rhythms of the Brain.* Oxford University Press.

Cheng, Q., Gao, X., Martin, R. et al. (2014). Exact prior-free probabilistic inference on the heritability coefficient in a linear mixed model. *Electronic Journal of Statistics* **8**, 3062–3076.

Delorme, A. and Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of Neuroscience Methods* **134**, 9–21.

Deuschl, G. (1999). *Recommendations for the Practice of Clinical Neurophysiology: Guidelines of the International Federation of Clinical Neurophysiology.* Elsevier.

Douc, R., Moulines, E., Olsson, J. and van Handel, R. (2011). Consistency of the maximum likelihood estimator for general hidden markov models. *The Annals of Statistics* **39**, 474–513.

Einevoll, G. T., Pettersen, K. H., Devor, A., Ulbert, I., Halgren, E. and Dale, A. M. (2007). Laminar population analysis: estimating firing rates and evoked synaptic activity from multielectrode recordings in rat barrel cortex. *Journal of Neurophysiology* **97**, 2174–2190.

Epstein, C. L. (2005). How well does the finite fourier transform approximate the fourier transform? *Communications on Pure and Applied Mathematics* **58**, 1421–1435.

Fiecas, M. and Ombao, H. (2016). Modeling the evolution of dynamic brain processes during an associative learning experiment. *Journal of the American Statistical Association* **111**, 1440–1453.

Gao, X., Shahbaba, B. and Ombao, H. (2018). Modeling binary time series using gaussian processes with application to predicting sleep states. *Journal of Classification* **35**, 549–579.

Gao, X., Shen, W., Hu, J., Fortin, N. and Ombao, H. (2019). Modeling local field potentials with regularized matrix data clustering. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, 597–602. IEEE.

Gao, X., Shen, W. and Ombao, H. (2018). Regularized matrix data clustering and its application to image analysis. *arXiv preprint arXiv:1808.01749*.

Gerrard, P. and Malcolm, R. (2007). Mechanisms of modafinil: a review of current research. *Neuropsychiatric Disease and Treatment* **3**, 349.

Guo, Y. (2011). A general probabilistic model for group independent component analysis and its estimation methods. *Biometrics* **67**, 1532–1542.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton university press, Princeton.

Jiru, A. R. (2008). *Relationships between Spectral Peak Frequencies of a Causal AR (P) Process and Arguments of Roots of the Associated ar Polynomial.* PhD thesis, San Jose State University.

Lukemire, J., Wang, Y., Verma, A. and Guo, Y. (2018). Hint: A toolbox for hierarchical modeling of neuroimaging data. *arXiv preprint arXiv:1803.07587*.

Makarova, J., Ibarz, J. M., Makarov, V. A., Benito, N. and Herreras, O. (2011). Parallel read-out of pathway-specific inputs to laminated brain structures. *Frontiers in Systems Neuroscience*, **5**, 77.

Makarova, J., Ortuño, T., Korovaichuk, A., Cudeiro, J., Makarov, V. A., Rivadulla, C. and Herreras, O. (2014). Can pathway-specific lfps be obtained in cytoarchitectonically complex

structures? *Frontiers in Systems Neuroscience* **8**, 66.

Michel, C., Lehmann, D., Henggeler, B. and Brandeis, D. (1992). Localization of the sources of eeg delta, theta, alpha and beta frequency bands using the fft dipole approximation. *Electroencephalography and Clinical Neurophysiology* **82**, 38–44.

Mitzdorf, U. (1985). *Current Source-Density Method and Application in Cat Cerebral Cortex: Investigation of Evoked Potentials and EEG Phenomena*. American Physiological Society.

Prado, R. and Lopes, H. F. (2013). Sequential parameter learning and filtering in structured autoregressive state-space models. *Statistics and Computing* **23**, 1–15.

Shumway, R. H. and Stoffer, D. S. (2013). *Time Series Analysis and its Applications*. Springer Science & Business Media.

Wang, Y. and Guo, Y. (2018). A hierarchical independent component analysis model for longitudinal neuroimaging studies. *arXiv preprint arXiv:1808.01557*.

Wang, Y., Ting, C.-M., Gao, X. and Ombao, H. (2019). Exploratory analysis of brain signals through low dimensional embedding. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, 997–1002. IEEE.

Wang, Y., Ting, C.-M. and Ombao, H. (2016). Exploratory analysis of high dimensional time series with applications to multichannel electroencephalograms. *arXiv preprint arXiv:1610.07684*.

Whitmore, N. W. and Lin, S.-C. (2016). Unmasking local activity within local field potentials (lfps) by removing distal electrical signals using independent component analysis. *NeuroImage* **132**, 79–92.

Zhang, K. and Hyvärinen, A. (2011). A general linear non-gaussian state-space model: Identifiability, identification, and applications. In *JMLR Workshop and Conference Proc., Asian Conf. on Machine Learning*, 113–128.

Department of Statistics, University of California, Irvine, 2226 Donald Bren Hall, Irvine, CA 92697-1250, USA.

E-mail: xgao2@uci.edu

Department of Statistics, University of California, Irvine, 2206 Donald Bren Hall, Irvine, CA 92697-1250, USA.

E-mail: weinings@uci.edu

Department of Statistics, University of California, Irvine, 2224 Donald Bren Hall, Irvine, CA 92697-1250, USA.

E-mail: babaks@uci.edu

University of California, Irvine, 106 Bonney Research Laboratory Building Department of Neurobiology and Behavior, Irvine, CA 92697, USA.

E-mail: norbert.fortin@uci.edu

Statistics Program, 4700 King Abdullah University of Science and Technology (KAUST) Thuwal, 23955-6900, Kingdom of Saudi Arabia, USA.

E-mail: hernando.ombao@kaust.edu.sa