Variational Covariance Smoothing for Dynamic Functional Connectivity Analysis

Ruwanthi Abeysekara[1,2], Jonathan Z. Simon[1,2,3], and Behtash Babadi[1,2]
[1]Department of Electrical & Computer Engineering, University of Maryland, College Park, MD
[2]Institute for Systems Research, University of Maryland, College Park, MD
[3]Department of Biology, University of Maryland, College Park, MD
E-mails: ruwanthi@umd.edu, jzsimon@umd.edu, behtash@umd.edu

***Abstract*—Functional connectivity is among the widely used metrics to assess the network-level attributes of brain function. While most existing analysis frameworks assume static functional connectivity during the course of an experiment, to capture neural dynamics over short time scales, a time-varying notion of functional connectivity is required. By revealing how neural networks reconfigure in response to changing external stimuli, internal states, and task demands, time-varying functional connectivity can be leveraged to study flexible cognition, such as working memory, attention, and decision-making. A major challenge in estimating time-varying functional connectivity from high-dimensional neural is the associated computational complexity. Existing methods trade off accuracy for computational efficiency, especially in applications that require real-time or near real-time processing. Here, we build on existing work using covariance-domain state-space models and introduce a framework based on variational inference that allows low-complexity estimation of time-varying functional connectivity and construction of confidence intervals. We validate the performance of the proposed method using simulation studies. Our results reveal significant gains in computational complexity compared to existing methods, while maintaining high accuracy.**

***Index Terms*—Functional connectivity, Covariance-domain state-space model, Variational inference, Filtering, Smoothing.**

I. Introduction

Brain networks undergo rapid modulation and evolution due to internal states and external stimuli [1]. Consequently, they can be viewed as dynamic systems with numerous states and transient connectivity patterns that emerge across various time scales and contexts [2], [3]. Comprehending the complex nature of these multi-scale neural networks, which are interconnected via short and long-range pathways, heavily relies on the analytical tools' capability to capture the stochastic and dynamic behaviors of such data.

Functional connectivity, often described as the temporal correlation among the activities of different neural assemblies indicating significant dependence between distant brain regions [4], [5] can be analyzed through the covariance of neurophysiological signals obtained from Electroencephalography (EEG), Magnetoencephalography (MEG), Positron Emission Tomography (PET), and functional Magnetic Resonance Imaging (fMRI).

This work has been supported in part by the National Science Foundation awards no. ECCS2032649 and OISE2020624 and the National Institutes of Health award no. R01-DC019394.

The high dimensionality of modern-day neural data severely limits current analysis techniques, compelling researchers to choose between lower accuracy with faster processing or higher accuracy at a significant computational cost. While researchers have made significant progress in developing statistical and signal processing techniques for analyzing brain connectivity patterns within complex, multi-channel neural recordings, computational complexity remains a persistent challenge that requires more focused attention.

The Gaussian assumption is a staple of neural data analysis. Assuming Gaussian-distributed time series, the functional connectivity (represented by the covariance of multi-sensor time series) follows a Wishart distribution. Wishart distributions are inherently constrained to a subset of the $\mathbb{R}^{n\times n}$ space, complicating the use of standard Bayesian state-space analysis methods. To address this, the authors in [6] introduced a state-space model [7] where observations and hidden states are restricted to a subset of symmetric positive-definite matrices as an extension to [8]–[11]. This model allows for posterior distribution computation of hidden states and system parameters, as well as filtered and smoothed densities, which entail drawing sample paths from said densities. However, in high-dimensional settings, the volume of data space grows exponentially with the number of dimensions. As a result, sampling methods require an impractically large number of samples to accurately estimate quantities of interest leading to a phenomenon known as 'the curse of dimensionality'. Due to this, functional connectivity analysis often uses regions of interest, or employs sliding-window based analysis assuming independence of adjacent windows. As such, most existing methods trade off computational demands with information loss or biases [12], [13].

In this work, we address this issue by proposing a novel smoother via Variational Inference (VI). VI offers a powerful alternative to sampling methods by presenting closed-form solutions for the latent variables. Rather than sampling, it leverages optimization by defining a family of approximate densities over the latent variables. It then approximates complex probability distributions with simpler, more tractable distributions [14]–[17] by locating the member of this family that minimizes the Kullback-Leibler (KL) divergence from the true posterior.

Building on the state-space model introduced in [6], we construct a *variational smoother* that allows inference with

significantly lower complexity, while preserving accuracy on par with existing sampling-based methods. As an application of the inference procedure, we construct estimators and confidence intervals in closed-form. We evaluate the performance of the proposed variational smoother using simulation studies that resemble dynamic changes in brain functional connectivity. Our results reveal significant computational gains, while achieving accuracy comparable to or better than sampling-based methods.

The outline of this paper is as follows. In Section II we give an overview of existing results and formulate the problem. In Section III, we present our proposed variational smoothing solution. Numerical validations are presented in Section IV, followed by our concluding remarks in Section V.

II. Background and Problem formulation

Consider a p-dimensional time series vector of duration K, representing an observed process. In the context of neural activity, p represents the number of sensors, and K the recording duration. Let $Z_{i,j} \in \mathbb{R}$ denote the observation from the jth sensor at time i, $j = 1, 2, \cdots, p$, $i = 1, 2, \cdots, K$. By segmenting the time series into T non-overlapping windows of length w, such that $K = Tw$, we assume that let $Z_i := [Z_{i,1}\, Z_{i,2}, \cdots, Z_{i,p}] \in \mathbb{R}^p$ follows a zero-mean multivariate Gaussian distribution with covariance C_t, for $(t-1)w + 1 \leq i \leq tw$, where $t = 1, 2, \cdots, T$ denotes the window index. It then follows that the un-normalized sample covariance $Y_t := \sum_{i=(t-1)w+1}^{tw} Z_i Z_i^\top \sim W_p(w, C_t)$, where $W_p(w, C_t)$ is the p-dimensional Wishart distribution with degrees of freedom w and scale matrix C_t [18].

A key challenge in capturing the covariance dynamics is the constraint that for all $t = 1, 2, \cdots, T$, C_t must be a positive semi-definite matrix. As such, common additive state-space models are not suitable, since unlike the Gaussian family, the Wishart family is not distributionally closed under addition. In [6], [8], the following state evolution model is proposed:

$$X_t = \frac{U_{t-1}^\top \Psi_t U_{t-1}}{\lambda}, \qquad \Psi_t \sim \beta_p\left(\frac{n}{2}, \frac{k}{2}\right) \tag{1}$$

where $X_t := C_t^{-1}$ is the precision matrix, U_{t-1} is the upper Cholesky factor of X_{t-1}, $\beta_p(\cdot, \cdot)$ is multivariate beta distribution, and $\lambda > 0$ is a scaling parameter. It can be shown that if X_{t-1} is Wishart distributed, X_t will also be.

In conjunction with the evolution model in Eq. (1), we consider a general Wishart model for the observations, given by [6]:

$$Y_t \sim W_p\left(k, \frac{(kX_t)^{-1}}{1-\lambda}\right) \tag{2}$$

where the degrees of freedom k can be chosen as w, if the data are generated as explained earlier. Note that $C_t = \frac{X_t^{-1}}{k(1-\lambda)}$, which is a modification of the observation model in [6] that assumes $C_t = \frac{X_t^{-1}}{k}$. As we will show later, the role of λ in our model is to provide further smoothness when performing one-step ahead prediction.

A. Inference via Forward Filtering and Backward Sampling

The inference method proposed in [6] is based on the common forward filtering and backward sampling procedure. Let $D_t = \{Y_t\} \cup D_{t-1}$ denote the available observations at time t, for $t = 1, ..., T$ with $D_0 = \{\Sigma_0\}$ where Σ_0 is some initial positive semi-definite matrix. It can be shown that the filtering posterior of X_t is given by [6]:

$$(X_t | D_t) \sim W_p(k + n, (k\Sigma_t)^{-1}) \tag{3}$$

where,

$$\Sigma_t = \lambda \Sigma_{t-1} + (1-\lambda) Y_t. \tag{4}$$

In other words, $\Sigma_\tau := \sum_{i=0}^{\tau} \lambda^i Y_{\tau-i}$. Using the recursion (4), one can compute Σ_t, $t = 1, 2, \cdots, T$ and thereby have access to the filtering density at all times. Note that by restricting $\lambda \leq 1$, the recursion of Σ_t provides a convex combination of the past and present observations, in which the contribution from the past decreases over time [19].

In order to obtain the smoothing density, i.e., $p(\{X_t\}_{t=1}^T | D_T)$ for inference, the authors in [6] leverage the following decomposition:

$$p(\{X_t\}_{t=1}^T | D_T) = p(X_T | D_T) \prod_{t=1}^{T-1} p\left(X_t | X_{t+1}, D_t\right) \tag{5}$$

and show that $(X_t | X_{t+1}, D_t)$ has a *shifted* Wishart distribution, i.e., given X_{t+1} and D_t, X_t can be expressed as

$$X_t = \lambda X_{t+1} + Z_{t+1}, \tag{6}$$

where $Z_{t+1} \sim W_p(k, (k\Sigma_t)^{-1})$. While this decomposition does not provide a convenient closed-form smoothing distribution, it can be used to draw sample paths from $p(\{X_t\}_{t=1}^T | D_T)$. Suppose that a total of L sample paths to be drawn. Starting at $t = T$, we have $(X_T | D_T) \sim W_p(k + n, (k\Sigma_T)^{-1})$. Thus, for the ℓ^{th} sample path, one can draw a sample from $p(X_T | D_T)$ denoted by $X_T^{(\ell)}.\ell = 1, 2, \cdots, L$. Using the backward recursion in Eq. (6) for $t = T-1, T-2, \cdots, 1$, given a sample $X_{t+1}^{(\ell)}$ at time $t+1$, one can draw a sample $Z_{t+1}^{(\ell)} \sim W_p(k, (k\Sigma_t)^{-1})$ and form $X_t^{(\ell)}$ as $\lambda X_{t+1}^{(\ell)} + Z_{t+1}^{(\ell)}$.

The L sample paths can be used for inference and estimation procedures. For example, we can approximate the expectation of a function $f\left(\{X_t\}_{t=1}^T\right)$ using the sample average as:

$$\mathbb{E}\left[f\left(\{X_t\}_{t=1}^T\right)\right] \approx \frac{1}{L} \sum_{\ell=1}^{L} f\left(\{X_t^{(\ell)}\}_{t=1}^T\right), \tag{7}$$

for large enough L.

B. Key Challenge and Our Contribution

The key bottleneck in the usage of this backward sampling solution is it's computational complexity: computing empirical averages using L backward sample paths $(X_t | X_{t+1}, D_t)$ inherently suffers from the 'curse of dimensionality'. Our key contribution to this work is to provide a low-complexity alternative to the computationally demanding backward sampling via variational inference. As we will show in the following

section, our proposed method provides smoothed estimates and their confidence intervals in closed-form. In addition, our numerical simulations show that variational smoothing achieves superior mean square error and significantly lower runtime compared to backward sampling.

III. Proposed Variational Smoothing Solution

Variational inference approximates complex posterior distributions by employing a computationally efficient alternative to demanding procedures that require sampling from the posterior. The objective function of variational inference is known as the Evidence Lower Bound (ELBO):

$$ELBO = E_{(q\{X_t\}_{t=1}^T)}\left[\ln\left[p\left(\{X_t\}_{t=1}^T|\{Y_t\}_{t=1}^T\right)\right]\right] - E_{(q\{X_t\}_{t=1}^T)}\left[\ln\left[q\left(\{X_t\}_{t=1}^T\right)\right]\right], \tag{8}$$

where $q(\cdot)$ is a member of predefined family of distributions. By carefully picking this family, maximizing the ELBO strikes a balance between effectively explaining the data and adhering to prior assumptions.

A. Variational Formulation

Given that our latent variables are precision and/or covariance matrices, we consider a mean-field variational family consisting of product-form Wishart distributions:

$$q\left(\{X_t\}_{t=1}^T\right) = \prod_{t=1}^T q\left(X_t\right) \tag{9}$$

where $q(X_t) \sim W_p(m, V_t)$ is Wishart distributed with m degrees of freedom and covariance V_t. The parameters m and $(V_t)_{t=1}^T$ are unknown variational quantities to be estimated by maximizing the ELBO.

The log-posterior of $(X_t)_{t=1}^T$ is given by:

$$\begin{aligned}\ln p\left(\{X_t\}_{t=1}^T|\{Y_t\}_{t=1}^T\right) = \sum_{t=1}^T & \frac{k-p-1}{2}\ln|X_t - \lambda X_{t+1}| \\ & - \frac{k}{2}\operatorname{tr}(\Sigma_t X_t) + \lambda\frac{k}{2}\operatorname{tr}(\Sigma_t X_{t+1}) \\ & - \frac{kp}{2}\ln 2 - \ln\Gamma_p\Big(\frac{k}{2}\Big) + \frac{k}{2}\ln|k\Sigma_t|.\end{aligned}$$

Given that the expectation of $\ln|X_t - \lambda X_{t+1}|$ does not have a closed-form solution, we expand this term up to the first-order in λ, resulting in $\ln|X_t - \lambda X_{t+1}| = \ln|X_t| - \lambda\operatorname{tr}(X_t^{-1}X_{t+1}) + \mathcal{O}(\lambda^2)$. We can thus use the following identities regarding the entropy and expected log-det of a Wishart random variable:

$$\begin{aligned}H[X_t] = & \frac{p+1}{2}\ln|V_t| + \frac{1}{2}p(p+1)\ln 2 + \ln\Gamma_p\left(\frac{m}{2}\right) \\ & - \frac{m-p-1}{2}\psi_p\left(\frac{m}{2}\right) + \frac{mp}{2}\end{aligned} \tag{10}$$

where $H[\cdot]$ denotes the information entropy of the Wishart distribution, $\Gamma_p(\cdot)$ denotes the multivariate gamma function and $\psi_p(\cdot)$ is the multivariate digamma function (the derivative of the log of the multivariate gamma function), and

$$\mathbb{E}[\ln|X_t|] = \psi_p\left(\frac{n}{2}\right) + p\ln(2) + \ln|V_t|, \tag{11}$$

as well as the first and second-order moments of the Wishart and Inverse Wishart matrices [20]–[22] to compute the ELBO as given by:

$$\begin{aligned}ELBO = & \sum_{t=1}^T \frac{k-p-1}{2}\left\{\psi_p\Big(\frac{m}{2}\Big) + p\ln 2 + \ln|V_t| \right. \\ & \left. - \lambda\frac{m}{m-p-1}\operatorname{tr}(V_t^{-1}V_{t+1})\right\} \\ & - \sum_{t=1}^T\frac{mk}{2}\operatorname{tr}(\Sigma_t V_t) + \sum_{t=1}^T\lambda\frac{mk}{2}\operatorname{tr}(\Sigma_t V_{t+1}) \\ & - \frac{Tkp}{2}\ln 2 - T\ln\Gamma_p(k/2) + \sum_{t=1}^T\Big(\frac{k}{2}\Big)\ln|k\Sigma_t| \\ & + \frac{p+1}{2}\sum_{t=1}^T\ln|V_t| + \frac{Tp(p+1)}{2}\ln 2 + T\ln\Gamma_p\Big(\frac{m}{2}\Big) \\ & - T\frac{m-p-1}{2}\psi_p\Big(\frac{m}{2}\Big) + \frac{mp}{2}.\end{aligned} \tag{12}$$

After taking the derivative of Eq. (12) with respect to V_t, we seek a solution of the form $V_t = V_t^{(0)} + \lambda V_t^{(1)}$, where $V_t^{(0)}$ and $V_t^{(1)}$ are to be found. After simplification, these quantities are given by:

$$V_t^{(0)} = \frac{1}{m}\Sigma_t^{-1}, \tag{13}$$

and

$$\begin{aligned}V_t^{(1)} = & \frac{m(k-p-1)}{k(m-p-1)}V_{t+1} - \frac{(k-p-1)}{km(m-p-1)}\Sigma_t^{-1}V_{t-1}^{-1}\Sigma_t^{-1} \\ & + \frac{1}{m}\Sigma_t^{-1}\Sigma_{t-1}\Sigma_t^{-1},\end{aligned} \tag{14}$$

where $V_{t+1} = V_{t+1}^{(0)} + \lambda V_{t+1}^{(1)}$ and $V_{t-1} = V_{t-1}^{(0)} + \lambda V_{t-1}^{(1)}$. The form of Eq. (14) allows recursive computation, akin to a backward filter. Given the estimated V_t's, m can be found by numerically maximizing the ELBO. Thus, one can iterate between estimating V_t's and m. We have observed empirically that setting m to be a small integer multiple of k provides a near-optimal solution. Thus, the computation of V_t may only be done once.

B. Estimation with Confidence Intervals

As an application of the inference aforementioned procedures, we consider constructing estimators and their respective confidence intervals. For the forward filter, we take the mean of the posterior density as the estimator:

$$\widehat{X}_t^f := \mathbb{E}\{X_t|D_t\} = \frac{k+n}{k}\Sigma_t^{-1}. \tag{15}$$

Given that both the mean and variance of the elements of a Wishart distributed random variable linearly scale with the degree of freedom, in the large degree regime, i.e., $n+k \gg 1$, the standard deviation is much smaller than the mean. Thus, we use Gaussian approximation to construct confidence intervals.

As such, confidence interval for $(\widehat{X}_t^f)_{ij}$ at a significance level α is given by:

$$\left[\frac{k+n}{k}(\Sigma_t^{-1})_{ij} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{n+k}{k^2}\left((\Sigma_t^{-1})_{ij}^2+(\Sigma_t^{-1})_{ii}(\Sigma_t^{-1})_{jj}\right)}\right],$$

where $z_{\frac{\alpha}{2}}$ is the z-score at level $\frac{\alpha}{2}$.

For the backward sampling, one needs to evaluate the mean and variance using the L sample paths. For the (i,j) element of $(X_t|D_T)$, the sample mean forms the estimator:

$$(\widehat{X}_t^{bs})_{ij} := \frac{1}{L}\sum_{\ell=1}^{L}(X_t^{(\ell)})_{ij}. \tag{16}$$

The sample variance is also given by:

$$\sigma_{t,ij}^{bs\,2} := \frac{1}{L}\sum_{\ell=1}^{L}\left((X_t^{(\ell)})_{ij} - (\widehat{X}_t^{bs})_{ij}\right)^2. \tag{17}$$

Thus, the confidence interval for $(\widehat{X}_t^{bs})_{ij}$ at a significance level α is given by $\left[(\widehat{X}_t^{bs})_{ij} \pm z_{\frac{\alpha}{2}}\sigma_{t,ij}^{bs}\right]$.

Finally, for the proposed variational smoother, given the closed-form variational density $X_t|D_T \sim W_p(m, V_t)$, the estimator is given by:

$$\widehat{X}_t^{vs} := mV_t, \tag{18}$$

and the confidence interval for $(\widehat{X}_t^{vs})_{ij}$ at a significance level α is given by:

$$\left[m(V_t)_{ij} \pm z_{\frac{\alpha}{2}}\sqrt{m\left((V_t)_{ij}^2 + (V_t)_{ii}(V_t)_{jj}\right)}\right]. \tag{19}$$

Noting the relation between X_t and C_t as $C_t = \frac{X_t^{-1}}{k(1-\lambda)}$ and the fact that the inverse of a Wishart distributed matrix is distributed according to an Inverse Wishart, estimators and confidence intervals for C_t can be obtained in a similar fashion, but are omitted here for brevity.

IV. Results

We generated a simulation study, inspired by changes in brain functional from one state to another. We thus considered 20×20 covariance structure with the following dynamics: consider a transition window of duration T_{tr} samples in which the system changes from an initial to a final state. For a total duration of $T > T_{tr}$ samples, we considered 3 segments: first, the covariance remains at an initial value C_{ini} for the first segment of length $\frac{1}{2}(T - T_r)$, then changes in a convex linear fashion from C_{ini} to some final value C_{fin} for a total of T_{tr} samples, i.e., $C_\tau = (1 - \frac{\tau}{T_{tr}})C_{fin} + \frac{\tau}{T_{tr}}C_{ini}$, followed by a final segment of staying constant at C_{fin}, totaling T samples.

Fig. 1 shows the structures of C_{ini} and C_{fin} used in this simulation study. Here, $p = 20$, $T = 600$ samples, and $T_{tr} = 200$ samples. For all the algorithms considered, we fixed $\lambda = 0.5$, $k = w = 50$, computed n using the relation $\frac{1}{\lambda} = 1 + \frac{k}{n-p-1}$ [6], and for the variational smoother chose $m = 5k$. For the backward sampling, we considered $L = 10, 100,$ and 1000 sample paths.

Fig. 2 shows sample time traces of the estimated precision and covariance matrices for the $(1,2)$ element. Colored hulls

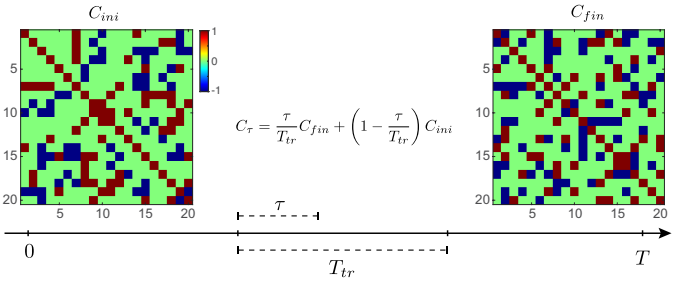

Fig. 1. Initial and final covariance matrices C_{ini} and C_{fin}. The covariance in the transition window of length T_{tr} is a convex linear combination of C_{ini} and C_{fin}.

show the 95% confidence intervals. As it can be observed visually, our proposed method exhibits superior performance in terms of both its variance and size of the confidence intervals.

To quantify this observation, Table I compares three different performance metrics: total normalized mean square error (NMSE), the NMSE of the off-diagonal elements only, and the CPU time. Note that the off-diagonal elements are of higher importance in precision/covariance estimation, as they allow to quantify the coupling between different variables, hence we used the off-diagonal NMSE as an additional metric here.

Boldface numbers show the best performance in each row. In estimating the precision matrix, our proposed variational smoother provides more than 3 dB gain in both NMSE metrics, and has comparable runtime to the forward filter. It is noteworthy that the backward sampling smoother has a significantly higher runtime for a range of sample sizes $L = 10, 100, 1000$ considered here. In estimating the covariance matrices, our proposed method outperforms the others in terms of off-diagonal NMSE, but has about 2 dB loss in terms of total NMSE.

V. Conclusion

The brain's inherent dynamic function demands analytical tools that go beyond traditional static measures of functional connectivity. Time-varying functional connectivity analysis allows us to investigate the flexible reconfiguration of neural networks, giving us crucial insights into fundamental cognitive processes and the disruption of those processes in neurological and psychiatric disorders. Our work addresses a key challenge in this field, often referred to as the 'curse of dimensionality'. By utilizing Bayesian modeling and variational inference, we have developed a computationally efficient technique for analyzing time-varying functional connectivity. This approach demonstrates promising accuracy with very low computational costs in our simulation studies, making it a potentially powerful alternative tool for understanding the complex dynamics of the brain. Future work includes validating our proposed methodology in application to large-scale EEG and MEG recordings.

References

[1] M. Abeles, *Corticonics: Neural circuits of the cerebral cortex*. Cambridge University Press, 1991.

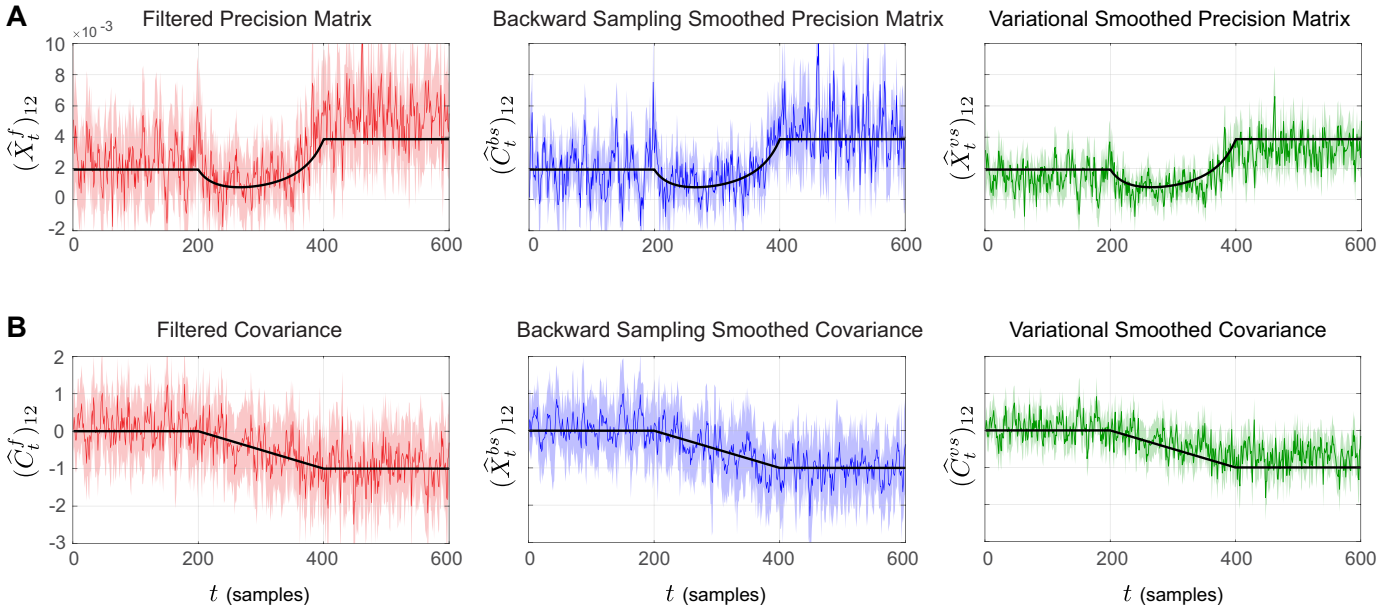

Fig. 2. Time traces of element $(1, 2)$ of the estimated mateices. A) Filtered (left), Backward Sampling Smoothed with $L = 10$ (middle), and Variational Smoother (right) estimates of the precision matrix X_t shown for its $(1, 2)$ element. B) Same as panel A, but for the estimated covariance matrices. Solid black curves show the ground truth time traces. Colored hulls show 95% confidence intervals.

		Forward Filter	Backward Sampling Smoother			Variational Smoother
			$L = 10$	$L = 100$	$L = 1000$	
Precision Matrix Estimation	Total NMSE (dB)	-4.26	-6.66	-6.68	-6.70	$\mathbf{-10.00}$
	Off-Diagonal NMSE (dB)	-2.15	-3.54	-3.54	-3.54	$\mathbf{-7.14}$
Covariance Matrix Estimation	Total NMSE (dB)	-9.20	-10.14	-10.77	$\mathbf{-10.83}$	-8.58
	Off-Diagonal NMSE (dB)	-1.90	-3.07	-3.71	-3.79	$\mathbf{-3.80}$
CPU Time (s)		$\mathbf{0.04}$	0.84	4.66	44.07	0.08

TABLE I
PERFORMANCE METRICS FOR THE EXISTING AND PROPOSED METHODS

[2] G. Deco, V. K. Jirsa, and A. R. McIntosh, "Emerging concepts for the dynamical organization of resting-state activity in the brain," *Nature Reviews Neuroscience*, vol. 12, no. 1, pp. 43–56, 2011.

[3] T. McKenna, T. McMullen, and M. Shlesinger, "The brain as a dynamic physical system," *Neuroscience*, vol. 60, no. 3, pp. 587–605, 1994.

[4] V. Sakkalis, "Review of advanced techniques for the estimation of brain connectivity measured with eeg/meg," *Computers in Biology and Medicine*, vol. 41, no. 12, pp. 1110–1117, 2011.

[5] A. A. Fingelkurts, A. A. Fingelkurts, and S. Kähkönen, "Functional connectivity in the brain—is it an elusive concept?" *Neuroscience & Biobehavioral Reviews*, vol. 28, no. 8, pp. 827–836, 2005.

[6] J. Windle and C. M. Carvalho, "A tractable state-space model for symmetric positive-definite matrices," 2014.

[7] Z. Ghahramani and M. Jordan, "Factorial hidden markov models," *Advances in neural information processing systems*, vol. 8, 1995.

[8] H. Uhlig, "On singular wishart and singular multivariate beta distributions," *The Annals of Statistics*, pp. 395–405, 1994.

[9] ——, "Bayesian vector autoregressions with stochastic volatility," *Econometrica: Journal of the Econometric Society*, pp. 59–73, 1997.

[10] R. Prado and M. West, *Time series: modeling, computation, and inference*. Chapman and Hall/CRC, 2010.

[11] M. K. Pitt and S. G. Walker, "Constructing stationary time series models using auxiliary variables with applications," *Journal of the American Statistical Association*, vol. 100, no. 470, pp. 554–564, 2005.

[12] J. M. Sanchez-Bornot, M. E. Lopez, R. Bruña, F. Maestu, V. Yousofzadeh, S. Yang, D. P. Finn, S. Todd, P. L. McLean, G. Prasad *et al.*, "High-dimensional brain-wide functional connectivity mapping in magnetoencephalography," *Journal of Neuroscience Methods*, vol. 348, p. 108991, 2021.

[13] A. Hillebrand, G. R. Barnes, J. L. Bosboom, H. W. Berendse, and C. J. Stam, "Frequency-dependent functional connectivity within resting-state networks: an atlas-based meg beamformer solution," *Neuroimage*, vol. 59, no. 4, pp. 3909–3921, 2012.

[14] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, "An introduction to variational methods for graphical models," *Machine learning*, vol. 37, pp. 183–233, 1999.

[15] M. D. Hoffman and D. M. Blei, "Structured stochastic variational inference," pp. 361–369, 2015.

[16] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *Journal of the American statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.

[17] C. Wang and D. M. Blei, "Variational inference in nonconjugate models," *Journal of Machine Learning Research*, 2013.

[18] R. J. Muirhead, *Aspects of multivariate statistical theory*. John Wiley & Sons, 2009.

[19] R. G. Brown, "Statistical forecasting for inventory control," 1959.

[20] L. Haff, "An identity for the wishart distribution with applications," *Journal of Multivariate Analysis*, vol. 9, no. 4, pp. 531–544, 1979.

[21] G. Letac and H. Massam, "All invariant moments of the wishart distribution," *Scandinavian Journal of Statistics*, vol. 31, no. 2, pp. 295–318, 2004.

[22] A. N. Bishop, P. Del Moral, A. Niclas *et al.*, "An introduction to wishart matrix moments," *Foundations and Trends® in Machine Learning*, vol. 11, no. 2, pp. 97–218, 2018.