frontiers

Oscillatory Biomedical Signals: Frontiers in Mathematical Models and Statistical Analysis

Hau-Tieng Wu[1], **Tze Leung Lai**[2,*], **Gabriel G. Haddad**[3], **and Alysson Muotri**[4]

[1] *Department of Mathematics, Duke University, Durham, NC 27708-0320, USA*
[2] *Department of Statistics, Stanford University, Stanford, CA 94305-4065, USA*
[3] *Department of Pediatrics and Rady Children's Hospital, University of California at San Diego, La Jolla/San Diego, CA 92093/92123, USA*
[4] *Department of Cellular & Molecular Medicine and Department of Pediatrics, University of California at San Diego, La Jolla, CA 92037, USA*

Correspondence*:
Tze Leung Lai
lait@stanford.edu

ABSTRACT

Herein we describe new frontiers in mathematical modeling and statistical analysis of oscillatory biomedical signals, motivated by our recent studies of network formation in the human brain during the early stages of life and studies forty years ago on cardiorespiratory patterns during sleep in infants and animal models. The frontiers involve new nonlinear-type time-frequency analysis of signals with multiple oscillatory components, and efficient particle filters for joint state and parameter estimators together with uncertainty quantification in hidden Markov models and empirical Bayes inference.

1 INTRODUCTION

The 2017 Nobel Prize in Physiology or Medicine was awarded to Jeffrey Hall and Michael Rosbash of Brandeis University, and Michael Young of Rockefeller University, "for their discoveries of molecular mechanisms controlling the circadian rhythm." In 1984, they succeeded in isolating the "*period* gene" (i.e., the gene that controls the circadian rhythm). Hall and Rosbash then went on to "discover PER, the protein encoded by *period*, accumulated during the night and degraded during the day." In 1994, Young answered a "tantalizing puzzle" concerning how PER produced in the cytoplasm could reach the cell nucleus where genetic material is located. He discovered a second gene *timeless*, encoding the TIM protein so that TIM bound to PER can enter the cell nucleus to block the *period* gene activity. "Such a regulatory feedback mechanism explained how this oscillation of cellular protein levels emerged, but questions lingered", such as what controlled the frequency of the oscillations. Young identified another gene *doubletime* encoding the DBT protein that delayed the accumulation of the PER protein. The three laureates "identified additional proteins required for the activation of the *period* gene, as well as for the mechanisms by which light can synchronize the circadian clock."

One of us (Muotri) was PI of a project on "spontaneous network formation" displaying "periodic and regular oscillatory events that were dependent on glutamatergic and GABAergic signaling" during early brain maturations, for which structural and transcriptional changes "follow fixed developmental programs defined by genetics"; see Trujillo et al. (2019) who also found that "the oscillatory activity transitioned to more spatiotemporally irregular patterns which synchronous network activity resembled features similar to

those observed in preterm human EEG." This project is similar in spirit to the exemplary work of Hall, Rosbash and Young but the "experimental inaccessibility" of the human brain during the early stages of life pushes mathematical modeling and statistical analysis of the oscillatory signals and events to new frontiers that we present in the next section. We describe in the next paragraph the underlying biomedical background of this project.

One of the major recent realizations, especially in the neurosciences, is that while we can obtain important information from animal studies, there are major differences between humans and animals. This is manifested in many ways, especially that major clinical trials that were based on animal findings did not pan out. Therefore if we intend to study pathogenesis of disease, treat them, prevent them or cure diseases of childhood or in the adults, we need to re-focus our scientific approaches and strategies in order to be more efficient and effective. Since embryonic stem cells are often problematic to obtain for ethical reasons, the discovery of being able to re-program somatic cells from humans into induced pluro-potential stem cells (iPSCs, taking these somatic cells back into their "history") and differentiate them into different relatively mature cell types has opened a major avenue for the scientific community, resulting in the 2012 Nobel Prize in Physiology or Medicine to John Gurdon of Cambridge and Shinya Yamanaka of Kyoto. If these iPSCs are exposed to the right growth factors, they would assemble into early human brain (brain organoids) by an amazing process of self-organizing the 3-dimensional cellular elements that recapitulate the network, cellular and membrane properties of neurons and glia. Many types of organoids such as kidney, intestine, liver and lung organoids have been recently developed. These organoids have been particularly useful for studying either normal early human biology or developmental disorders as in neuro-developmental diseases.

2 METHODS

The statistical methods used by Trujillo et al. (2019, pp. 16-19) in their analysis of data on oscillatory signals and events consist of (i) multi-electrode array (MEA) recording and custom analysis, (ii) network event analysis that involves detecting spikes (when at least 80% of the maximum spiking values) over the length of the recording when reached at least 1 sec away from any other network event, (iii) oscillatory spectral power analysis, in which "oscillatory power" is defined as "peaks in the PSD (power spectral density estimated by Peter D. Welch's method) above the aperiodic $1/f$ power law decay", (iv) resampled Pearson's correlation coefficient between neonatal age and each of 12 EEG features. Because of "the inability to interrogate the electrophysiology of intact human brains" and the emergence of induced pluripotent stem cells (iPSC) and organoids as "a scaled-down and three-dimensional model of the human brain, mimicking various developmental features at cellular and molecular levels," Trujillo et al. (2019, pp. 4, 7-9, 11, 19) used oscillatory dynamics of LFP (local field potential) and other mesoscopic brain signals, which manifest "a phenomenon known as cross-frequency phase-amplitude coupling (PAC) wherein the high-frequency content of LFP is entrained to the phase of slow oscillations." Noting that "the pattern of alternating periods of quiescence and network-synchronized events resembles electrophysiological signatures in preterm human EEG," they analyzed "a publicly available dataset of 101 serial EEG recordings from 39 preterm infants ranging from 24 to 38 weeks post-menstral age", containing 23 precomputed features (including spectral power in canonical oscillatory bands, duration and timing of "spontaneous activity transients" or SATs) for each EEG record. To compare the features between cortical organoids and preterm infants, they "trained a regularized regression model (ElasticNet) with cross-validation for hyparameter selection" based on the preterm infants' EEG recordings and applied the model to the organoid dataset to "obtain the predicted developmental time." The results were mixed and they concluded that "given the potential roles of synchronized and oscillatory network dynamics in coordinating information flow between developed

brain regions, these results highlight the potential for cortical organoids to advance our understanding of functional physiology" and to model "cellular interactions and neural circuit dysfunctions related to neurodevelopmental and neuropsychiatric pathologies" that "affect millions of people but otherwise lack an existing animal model." These statistical methods are "custom" (or traditional) methods, as acknowledged by Trujillo et al. (2019). We describe innovative and powerful methods in the next two subsections, first for time-frequency analysis of oscillatory biomedical signals with time-varying features, and then a new hidden Markov model (HMM) which incorporates the key features, of the cortical organoid model and provides uncertainty quantification for empirical Bayes inference based on the model and observed data.

2.1 Time-Frequency Analysis of Signals with Multiple Oscillatory Components

The first author (Wu) has been working on time-frequency analysis (TFA) and its applications to high-frequency biomedical signals in the last ten years. Examples include electrocardiography, electroencephalogram, local field potential, photoplethysmogram (PPG), actinogram, peripheral venous pressure (PVP), arterial blood pressure, phonocardiogram, airflow respiratory signal, to name several. Usually, these signals are composed of multiple components, each of which reflects the dynamics of a physiological system. The analysis is challenged by the physiological variability that appears in form of time-varying frequency and amplitude, or even time-varying oscillatory pattern that is referred to as the "wave-shape function". Furthermore, depending on the signal, the "waxing and waning" effect is sometimes inevitable for its components; see Wu (2020, Figure 1) for an illustration. Take the widely applied PPG signal as an example. Shelley (2007) has given an introduction to photoplethysmogram (PPG) and its applications "beyond the calculation of arterial oxygen saturation and heart rate". In addition to the well-known cardiac component reflecting hemodynamic information, PPG may contain the respiratory dynamics as another component. The frequency of the cardiac component (respiratory component, respectively) is impacted by the heart rate variability (breathing rate variability, respectively). Cicone and Wu (2017) provide an algorithm to "extract both heart and respiratory rates" from the PPG signal and thereby to analyze their interactions. Such information can be used in conjunction with other biomedical signals reflecting hemodynamics. In particular, PVP is ubiquitous in the hospital environment and a rich source of hemodynamic information (Wardhan and Shelley, 2009). But it typically has low signal-to-noise ratio (SNR) and its oscillatory pattern is sensitive to the physiological status, making it is much less used in comparison with PPG. Wu et al. (2020) have developed new signal processing tools to facilitate its use.

Combining time-frequency analysis (TFA) with statistical analysis, the lack of which in previous work "presents an opportunity for much future research", is illustrated in Figure 2 (applied to PPG, fetal ECG and fetal heart rate variability) of Wu (2020) who describes several recent advances in TFA for high-frequency biomedical signals. There are several challenges common to different biomedical signal processing problems. The first is how to estimate the dynamics (e.g., how to quantify the time-varying frequency, amplitude, or wave-shapes) of the signal. The second is to assess signal quality and determine artifacts, distinguishing between physiological and non-physiological ones. The third is to identify oscillatory components and the fourth is to decompose the signal into constituent components. To address these challenges, several TFA tools have been proposed. In addition to the traditional linear-type time-frequency analysis tools like short-time Fourier transform (STFT), continuous wavelet transform (CWT) and bilinear time-frequency analysis tools (Flandrin, 1999), several nonlinear-type tools have been developed and applied, including the reassignment method, empirical mode decomposition (EMD), Blaschke decomposition (BKD), adaptive locally iterative filtering (ALIF), sparse time-frequency representation (STFR),synchrosqueezing transform (SST), scattering transform (ST), concentration of frequency and time (ConcFT), de-shape, dynamic diffusion maps, and manifold learning (Huang et al., 1998; Nahon, 2000; Daubechies et al., 2011, 2016; Mallat, 2012; Lin et al., 2018, 2021; Wang et al.,

2020). The statistical properties of these methods have been relatively unexplored and we are currently investigating them; new methods to handle emerging scientific problems might be developed on the way.

2.2 Efficient Particle Filters for Joint State and Parameter Estimation in HMM

During the past three years the second author (Lai) has been developing a new Markov Chain Monte Carlo (MCMC) procedure called "MCMC with sequential substitutions" (MCMC-SS) for joint state and parameter estimation in hidden Markov models. The basic idea is to approximate an intractable distribution of interest (or target distribution) by the empirical distribution of N representative atoms, chosen sequentially by an MCMC procedure, so that the empirical distribution approximates the target distribution after a large number of iterations as explained below.

Lai's work in this area began with the landmark paper of Gordon et al. (1993) on the development of sequential Monte Carlo (SMC), also called particle filters, for the estimation of latent states in a hidden Markov model (HMM). Liu (2008) contains a collection of techniques that have been developed since then, with examples of applications in computational biology and engineering, and Chan and Lai (2013) provide a general theory of particle filters. Let $= \{X_t, t \geqslant 1\}$ be a Markov chain and let $Y_1, Y_2, \ldots$ be conditionally independent given , such that $X_t \sim p_t(\cdot|X_{t-1}), Y_t \sim g_t(\cdot|X_t)$ in which p_t and g_t are density functions with respect to measure ν_X and ν_Y. The density function p_T of $X_{0:T} = (X_0, \ldots, X_T)$ conditional on $Y_{1:T} = (Y_1, \ldots, Y_T)$ is

$$p_T(x_{0:T}|Y_{1:T}) \propto \prod_{t=1}^{T} [p_t(x_t|x_{t-1}) g_t(Y_t|x_t)].$$

This conditional distribution is often difficult to sample from and the normalizing constant is also difficult to compute for high-dimensional or complicated state spaces, and particle filters use sequential Monte Carlo that involves importance sampling and resampling to circumvent this difficulty. The particle filter computes $\mathbb{E}[\psi(X_{0:T})|Y_1, \ldots, Y_T]$ by the recursive Monte Carlo scheme summarized in Algorithm 1. Let $X^m_{0:t-1}$ denote the sample path of the mth particle (trajectory), $1 \leqslant m \leqslant M$. The scheme uses importance sampling from a proposal density q_t to circumvent this difficulty and updates not only the particles $X^m_{0:t-1}$ but also the associated weights w^m_{t-1} and ancestor A^m_{t-1} of $X^m_{0:t}$. It is initialized with $A^m_0 = m$ and $w^m_0 = 1$. The SMC estimate of $\psi_T := \mathbb{E}[\psi(X_{0:T})|Y_{1:T}]$ is

$$\tilde{\psi}_T = \left(\sum_{m=1}^{M} w^m_T \psi(X^m_{0:T}) \right) \Bigg/ \left(\sum_{m=1}^{M} w^m_T \right).$$

By using martingale theory, Chan and Lai (2013) provide a comprehensive theory of the SMC estimate $\tilde{\psi}_T$, which includes asymptotic normality and consistent standard error estimation as the following:

THEOREM 1. *Under certain integrability conditions,*

$$\sqrt{M}(\tilde{\psi}_t - \psi_T) \Rightarrow N(0, \sigma^2).$$

Moreover, letting $\overline{w}_t = M^{-1} \sum_{i=1}^{M} w^i_t$, σ^2 *can be consistently estimated by*

$$\tilde{\sigma}^2 = \frac{1}{M} \sum_{m=1}^{M} \left(\sum_{i: A^i_{T-1} = m} \frac{w^i_T}{\overline{w}_T} \left[\psi(X^i_{0:T}) - \tilde{\psi}_T \right] \right)^2.$$

Algorithm 1 SMC with M particles

1. Initialization: $A_0^i = i, w_0^i = 1$ for $i = 1, \ldots, M$.
2. Importance sampling at stage $t \in \{1, \ldots, T\}$: Generate conditionally independent X_t^i from $q_t(\cdot|X_{0:t-1}^i)$ and set $X_{0:t}^i = (X_{0:t-1}^i, X_t^i), w_t^i = p_t(X_t^i|X_{0:t-1}^i)g_t(y_t|X_t^i)/q_t(X_t^i|X_{0:t-1}^i), i = 1, \ldots, M$.
3. Bootstrap resampling at stage $t \in \{1, \ldots, T-1\}$: Generate i.i.d. random variables $B_t^1, \ldots, B_t^M$ such that $P(B_t^i = j) = w_t^j / \sum_{i=1^M} w_t^i, j = 1, \ldots, M$. Let $(X_{0:t}^m, A_t^m) = (X_{0:t}^{B_t^m}, A_{t-1}^{B_t^m}), m = 1, \ldots, M$.

Chan and Lai (2013, Lemmas 1 and 4) use the following representation of $\tilde{\psi}_T - \psi_T$ to derive Theorem 1. Let $w_t(x_{0:t}) = p_t(x_t|x_{t-1})g_t(Y_t|x_t)/q_t(x_t|x_{0:t-1})$, in which that Y_t can be treated as constants since the particle filter is the conditional distribution of $X_{0:t}$ given the observations $Y_1, \ldots, Y_T$. Let $H_t^m = (\overline{w}_1, \ldots, \overline{w}_t)/\prod_{j1}^t w_j^m, \eta_t = \mathbb{E}_q\left[\prod_{i=1}^t w_i(X_{0:t})\right]$, where $\mathbb{E}_q$ denotes expectation under which $X_t|X_{0:t-1}$ has the conditional density function $q_t(\cdot|X_{0:t-1})$ for $1 \leqslant t \leqslant T$. Letting $\Psi_0 = \psi_T$ and $\Psi_t(X_{0:t}) = \mathbb{E}_q\left\{\psi(X_{0:T})\prod_{i=1}^T w_i(X_{0:i})\middle|X_{0:t}\right\}$ for $1 \leqslant t \leqslant T$, define

$$\epsilon_{2t-1}^m = \sum_{i:A_{t-1}^m=m} \left\{\psi\left(X_{0:t}^i\right) - \Psi_{t-1}\left(x_{0:t-1}^i\right)\right\} H_{t-1}^i,$$

$$\epsilon_{2t}^m = \sum_{i:A_{t-1}^m=m} \left(\#_t^m - mW_t^i\right)\left\{\Psi_t\left(X_{0:t}^i\right) H_t^{B_t^i} - \Psi_0\right\},$$

in which $W_t^i = w_t^i / \sum_{j=1}^M w_t^j$, $\#_t^i$ is the number of copies of $X_{0:t}^i$ generated by boostrap resampling from $\left\{X_{0:t}^1, \ldots, X_{0:t}^M\right\}$ in Algorithm 1 (where the B_t^i are also defined). Then $\left(\#_t^1, \ldots, \#_t^M\right) \sim$ Multinomial $\left(M; W_t^1, \ldots, W_t^M\right)$ and

$$\tilde{\psi}_T - \psi_T = \left\{(\overline{w}_1 \cdots \overline{w}_T)^{-1}\eta_T\right\} M^{-1} \sum_{m=1}^M \left(\epsilon_1^m + \cdots + \epsilon_{2T-1}^m\right);$$

see Eq.(3.3) and (3.36) of Chan and Lai (2013) who show that $\{\epsilon_t^m, 1 \leqslant t \leqslant 2T-1\}$ is a martingale difference sequence and that $(\overline{w}_1 \cdots \overline{w}_T)^{-1}\eta_T = 1 + o_p(1)$ under the integrability assumptions $\eta_T < \infty$ and $\mathbb{E}_q\left[\prod_{i=1}^T w_i^2(X_{0:t})\right] < \infty$.

Algorithm 2 PMCMC at kth iteration, initialized with $\theta^0 \sim f(\cdot)$

1. $\theta^* \sim f(\cdot|\theta_{k-1})$.
2. Run SMC (Algorithm 1) to generate M particles $X_{0:t}^{m,k}$ with corresponding weights $w_T^{m,k}$. Let $\tilde{p}_T(\theta^*) = \sum_{m=1}^M w_T^{m,k}$.
3. Accept θ^* with probability $1 \wedge \{\tilde{p}_T(\theta^*)f(\theta_{k-1}|\theta^*)\}/\{\tilde{p}_T(\theta_{k-1})f(\theta^*|\theta_{k-1})\}$.
4. If θ^* is accepted, let $\theta^k = \theta^*$ and $\left(X_{0:t}^{m,k}, w_T^{m,k}\right)$ be the corresponding weighted particles.

The assumption of a single fully-specified HMM in particle filter is often too restrictive in applications since the model parameters are usually unknown and also need to be estimated sequentially from the observed data. A standard method to estimate unknown parameters is to assume a prior distribution

for the unknown parameter vector and to use Markov chain Monte Carlo (MCMC) to estimate the posterior distriubtion. Andrieu et al. (2010) carried out this method for time-homogeneous Markov chains $X_t \sim p_\theta(\cdot|X_{t-1})$ for $t \geqslant 1$ and $X_0 \sim p_\theta(\cdot)$, with latent states X_t and observations $Y_t \sim g_\theta(\cdot|X_t)$, in which θ is an unknown parameter with a prior density function $\pi(\cdot)$ with respect to some measure $\nu_{\boldsymbol{\theta}}$ on the parameter space Θ. The posterior density of $(\theta, X_{0:T})$ given $Y_{1:T}$ is proportional to

$$p_T(\vartheta, x_{0:T}) = \pi(\vartheta)p_\vartheta(x_0)\prod_{t=1}^{T}\{p_\vartheta(x_t|x_{t-1})g_\vartheta(Y_t|x_t)\}\,.$$

PMCMC uses SMC involving M particles (each of which consists of a sampled parameter and state trajectory) at every iteration k to construct an approximation $\tilde{p}_T$ to p_T in a Metropolis-Hastings (MH) MCMC scheme that uses a proposal density $f(\cdot|\theta_{k-1})$ with respect to the measure v_θ to sample θ_k at the kth iteration, as summarized in Algorithm 2. Chopin et al. (2013, Section 1.2) point out the difficulties in the asymptotic analysis of PMCMC as k becomes infinite. In particular, although Andrieu et al. (2010) have shown that under some strong assumptions, PMCMC converges to a measure in total variation norm as $k \to \infty$, for fixed value of M, the limiting measure is not the target posterior distribution of $(\theta, X_{0:t})$. On the other hand, allowing M to approach ∞ with k would lead to an analytically intracAlgorithm scheme involving state spaces whose dimensions change with k. Chopin et al. (2013) propose the SMC^2 scheme to target heuristically the posterior distribution of $(\theta, X_{0:t})$ given $Y_{1:t}$ $(1 \leqslant t \leqslant T)$ as follows. It involves N θ-particles, which we will call "atoms", and attaches to each atom θ a particle filter that propagates and resamples M particles (state trajectories $X_{0:t}^m$) generated by SMC (as in Algorithm 1 with the given θ). It carries out the MH iterations to determine if a candidate atom is accepted (as in Step (c) of Algorithm 2). For the N atoms $\theta_t^1, \ldots, \theta_t^N$ and their corresponding importance weights at time t generated in this way, if the degeneracy criterion in Chopin (2002) is satisfied, carry out bootstrap resampling of the weighted parameter-particle set to replace it by an unweighted set, but no convergence theory as $k \to \infty$ is provided.

Although MCMC methods with MH iterations are widely used computational tools in Bayesian inference on $\theta \in \Theta$ that has prior density function with respect to some measure ν_θ, they do not have convergence rate guarantees in terms of the number of iterations to automate termination of the iterations. On the other hand, if the target density p, which is the posterior density of θ given $Y_{1:t}$, were known and easy to sample from, then standard Monte Carlo approximation of $\mu := E_p(\psi(\theta))$ could be carried out by generating i.i.d. $\theta_1, \ldots, \theta_N$ from $p(\cdot)\,\mathrm{d}m$ and using the sample average $\tilde{\mu} = N^{-1}\sum_{n=1}^{N}\psi(\theta_n)$ to estimate μ. Under the assumption $E_p(\psi^2(\theta)) < \infty$, the estimated standard error is $\tilde{\sigma}_N/\sqrt{N}$, and $\tilde{\mu} \pm N^{-1/2}\tilde{\sigma}_N\zeta_{1-\alpha/2}$ is an approximate $(1-\alpha)$-level confidence interval for μ, where $\tilde{\sigma}_N^2 = (N-1)^{-1}\sum_{n=1}^{N}(\psi(\theta_n) - \tilde{\mu})^2$ and ζ_q is the qth quantile of the standard normal distribution. This follows from the classical central limit theorem and is very useful for determining N to ensure $\tilde{\mu}$ to be within some prescribed tolerance limit ϵ of μ: $N^{-1/2}\tilde{\sigma}_N\zeta_{1-\alpha/2} \leqslant \epsilon$, and has inspired Lai to develop, with his current Ph.D. students Huanzhong Xu, Michael Hongyu Zhu, and former Ph.D. student Hock Peng Chan, the following novel MCMC algorithm which is asymptotically equivalent to the oracle procedure that assumes known target density p and which they call MCMC with sequential state substitutions (MCMC-SS).

As in MH, let f be a given function that is proportional to the target density. Let $\{q(\cdot|\gamma) : \gamma \in \Gamma\}$ be a family of positive proposal densities with respect to some measure m, where Γ is a convex subset of $\mathbb{R}^d$. MCMC-SS initializes by choosing $\gamma_0 \in \Gamma^o$ and generating νB i.i.d. $\theta_{1,0}^1, \ldots, \theta_{1,0}^\nu; \ldots; \theta_{B,0}^1, \ldots, \theta_{B,0}^\nu$ from the proposal distribution $q(\cdot|\gamma_0)\,\mathrm{d}m$, thereby forming the B disjoint sets $\Theta_{b,0} = \{\theta_{b,0}^1, \ldots, \theta_{b,0}^\nu\}$. At stage k, it uses the sequential substitution procedure $\mathrm{SS}(\Theta_{b,k}, \boldsymbol{w}_k^b)$ in Algorithm 3 to update the atom set

Algorithm 3 Updating procedure $\mathrm{SS}(\Theta_{b,k}, \mathbf{w}_{\mathrm{k}}^{\mathrm{b}})$ for MCMC-SS

1. Sample $\tilde{\theta}$ from $q(\cdot|\gamma_{b,k-1})$ as candidate atom.
2. Let $\theta^b_{\nu+1,k-1} = \tilde{\theta}$ and compute
$$\lambda^b_{i,k} = q\left(\theta^b_{i,k-1}|\gamma_{b,k-1}\right) \Big/ f\left(\theta^b_{i,k-1}\right), i = 1, \ldots, \nu+1.$$
3. Sample J from $\{1, \ldots, \nu+1\}$ with probability $\pi^b_{i,k} = \lambda^b_{i,k} \Big/ \left(\sum_{j=1}^{\nu+1} \lambda^b_{j,k}\right)$ for i.
4. If $J = \nu+1$, let $\Theta_{b,k} = \Theta_{b,k-1}$. Otherwise let $\Theta_{b,k} = \left(\Theta_{b,k-1} \cup \{\tilde{\theta}\}\right) \setminus \left\{\theta^b_{\tilde{J},k-1}\right\}$.
5. Let $w^b_{i,k} = 1/\pi^b_{i,k}$ for $i = 1, \ldots, \nu$, and $\mathbf{w}^b_k = \left(w^b_{1,k}, \ldots, w^b_{\nu,k}\right)$.

in the bth block and to assign the weight $w^b_{i,k}$ to the ith atom in $\Theta_{b,k}$, $b = 1, \ldots, B$. MCMC-SS estimates $\mu = \mathbb{E}_p\psi(\theta)$ by
$$\hat{\psi} = \frac{1}{B(K-\kappa)} \sum_{b=1}^{B} \sum_{k=\kappa+1}^{K} \hat{\psi}_{b,k}, \text{ with } \hat{\psi}_{b,k} = \frac{\sum_{i=1}^{\nu} w^b_{i,k} \psi(\theta^b_{i,k})}{\sum_{i=1}^{\nu} w^b_{i,k}},$$

in which κ represents an initial burn-in period that is asymptotically negligible as $\kappa = o(K)$. In many applications, the parameter γ of the family of proposal densities is a function $\gamma : \mathcal{P} \to \Gamma$, where $\mathcal{P}$ is the space of probability measures on Θ. Assuming this framework, we now describe the choice of $\gamma_{b,k-1}$ in Algorithm 3. For $k \leqslant \kappa$, let $\gamma_{b,k-1} = \nu^{-1} \sum_{\theta \in \Theta_{b,k-1}} \gamma(\theta)$, which is the mean of the empirical measure of the atoms in the bth block at the end of stage $k-1$. On the other hand, for $k > \kappa$, we pool across blocks by letting $\gamma_{k-1} = B^{-1} \sum_{b=1}^{B} \gamma_{b,k-1}$, which we use as the modified $\gamma_{b,k-1}$ for all blocks. Therefore, after the burn-in period, we can carry out the update $\mathrm{SS}(\Theta_{b,k})$ in the order $b = 1, \ldots, B$, so that if the candidate atom in $\mathrm{SS}(\Theta_{b,k})$ is not used for block b, it can serve as candidate atom for block $b+1$ ($\leqslant B$), which then does not need to generate another random variable from $q(\cdot|\gamma_{k-1})$, an obvious advantage for high-dimensional and complicated states. Lai, Xu, Zhu and Chan have developed a comprehensive asymptotic theory of MCMC-SS showing its asymptotic optimality with respect to computational and statistical criteria, and have also derived consistent estimators of the standard errors for the Monte Carlo state/parameter estimates. Details are given in Lai et al. (2021) whose main results are summarized in the following and who also describe numerically stable implementation of Algorithm 3 that can be vectorized and parallelized and illustrate its applications to latent variable analysis with uncertainty quantification in image reconstruction and brain network development.

THEOREM 2. *Suppose* $\mathbb{E}_p\psi^2(\theta)$ *and there exist* $\beta > \alpha > 0$ *and* $V : \Theta^\nu \to [1, \infty)$ *such that for* $\gamma : \mathcal{P} \to \Gamma$,
$$\int_{\Theta^\nu} V(\boldsymbol{\theta}) q(\theta^1|\gamma_0) \ldots q(\theta^\nu|\gamma_0) \mathrm{d}m^\nu(\boldsymbol{\theta}) < \infty \text{ with } \boldsymbol{\theta} = (\theta^1, \ldots, \theta^\nu), \text{ and}$$
$$\alpha V(\boldsymbol{\theta}) \leqslant \lambda(\tilde{\theta}|\gamma(\boldsymbol{\theta})) \leqslant \beta V(\boldsymbol{\theta}) \text{ for all } \boldsymbol{\theta} \in \Theta^\nu \text{ and } \tilde{\theta} \in \Theta,$$

where $\lambda(\tilde{\theta}|\gamma) = f(\tilde{\theta}|\gamma)/p(\tilde{\theta})$.

(i) Let $G_{b,k}$ be the joint distribution of $\left(\theta^b_{1,k}, \ldots, \theta^b_{\nu,k}\right)$ and let Q^ν be the probability measure on Θ^ν that has the density of ν independent components each of which has density $q(\cdot|\gamma_f)$ with respect to m, where $\gamma_f = \text{argmin}_{\gamma\in\Gamma} I(q_\gamma\|f)$ and $I(q\|f) = \mathbb{E}_f\{\log(q(\theta)/f(\theta))\}$ is the Kullback-Leibler divergence (or relative entropy) of q from the target density f in Algorithm 3. Then there exist positive constants a and c such that $\|G_{b,k} - Q^\nu\|_V \leqslant ce^{-ak}$ for $1 \leqslant k \leqslant K$, where $\|\cdot\|_V$ denotes the weighted total variation norm associated with the weight function V. Hence after $k \succ \log B$ iterations, $\sum_{b\leqslant B}\|G_{b,k} - Q^\nu\|_V \to 0$.

(ii) Let $N = B(K-\kappa)$ be the total number of atoms used to define the MCMC-SS estimate of $\tilde{\psi}$ of $\mu = \mathbb{E}_p(\psi(\theta))$. Then as $K \to \infty$ and $B \to \infty$ such that $B = \mathcal{O}(K)$,

$$\sqrt{N\nu}(\hat{\psi} - \mu) \Rightarrow N(0, \sigma^2),$$

where $\sigma^2 = Var_p(\psi(\theta))$ and can be consistently estimated by

$$\hat{\sigma}^2 = \frac{1}{B(K-\kappa)}\sum_{b=1}^{B}\sum_{k=\kappa+1}^{K}\frac{1}{\nu-1}\sum_{\theta\in\Theta_{b,k}}(\psi(\theta) - \hat{\psi}_{b,k})^2.$$

As shown by Lai et al. (2021), with probability approaching 1 by large k, the candidate atom $\tilde{\theta}$ in Algorithm 3 substitutes some existing atom in $\Theta_{b,k-1}$. Hence, similar to the case of known target density p from which $\tilde{\theta}$ is sampled, the newly sampled atom features in the weighted average $\hat{\psi}_{b,k}$. The reason we need the weighted average, with "importance sampling weights" $w^b_{i,k}$, is that for large k, the conditional distribution of $\Theta_{b,k}$ given $\Theta_{b,k-1}$ behaves like the ν-fold product measure Q^ν on Θ^ν. This shows that importance sampling (likelihood ratio) weights $w^b_{i,k}$ are needed to convert Q to P and suggests the asymptotic optimality of $\hat{\psi}$, which is the overall average of the $B(K-\kappa)$ estimates $\hat{\psi}_{b,k}$, similar to $\hat{\mu}$ that is described for the case of known p. Each random variable generated in the MCMC-SS scheme asymptotically contributes weight $(N\nu)^{-1}$ to (a) the estimate $\hat{\psi}$ of μ and (b) the asymptotic variance of $\hat{\psi}$. Theorem 2 shows that there is in fact considerable flexibility in the choice of the factors K (the number of iterations) and B (the number of blocks) in $N = B(K-\kappa)$ that determines the scaling factor in the central limit theorem, although the theorem highlights the case $B = \mathcal{O}(K)$ to emphasize that K should not be chosen too small relative to B. Lai et al. (2021) give an application to uncertainty quantification in the following image reconstruction problem. Cotter et al. (2013) propose to use MCMC methods "whenever the target measure has density with respect to a Gaussian process or Gaussian random field reference measure". A wide range of applications involving such a framework considers Bayesian inference on a latent random field $\{u(x) : x \in D\} \subset \mathbb{R}^d$ generated by some stochastic partial differential equation (SPDE) in which D is a connected subset of $\mathbb{R}^{d'}$, based on data generated by some nonlinear function of the random field. It is shown that after discretization and truncation to fit into this framework, the Radon-Nikodym derivative of the target measure P with respect to the reference measure Q has the form

$$(dQ/dP)(u) \propto \exp(-l(u))$$

for some real-valued function l, which Cotter et al. (2013) call "potential" in their substantive applications. The advantage of using a zero-mean Gaussian random field reference measure Q is that it is specified by the covariance operator $\mathcal{C}$ whose eigenvalues λ_i and orthonormal eigenfunctions ϕ_i yield the Karhunen-Loève expansion $u(x) = \sum_{i=1}^{\infty}\xi_i\phi_i(x)$, with i.i.d. ξ_i that are $N(0, \lambda_i^2)$ and $\sum_{i=1}^{\infty}\lambda_i^2 < \infty$. Cotter et al. (2013)

use a random truncation τ with a sieve prior to convert the infinite-dimensional expansion to a finite sum $u(x) = \sum_{i=1}^{\tau} \xi_i \phi_i(x)$. In addition, a discrete approximation of the random field $u(x)$ is used, with x taken over a mesh of width δ in each coordinate.

MCMC-SS uses a parametric family of Gaussian proposal measures $Q(\gamma)$ instead of a single one Q by Cotter et al. (2013). Putting $1/L(\theta) = \exp(-l(u(x)))$, we can also incorporate the random truncation τ and possibly also other random effects ρ into the state $\theta = (\tau, \zeta_1, \ldots, \zeta_\tau, \rho)$, where $\zeta_j = \mathcal{G}(u(x_j))$, $j = 1, \ldots, \tau$, and $\mathcal{G}$ is an operator associated with the SPDE and the discretization scheme for which x_j belongs to a discrete subset of D. With this definition of θ, MCMC-SS uses the updating procedure described in Algorithm 3 . Cotter et al. (2013, Sect. 4.2) have argued that simply applying MCMC to a discretized random field leads to a singular reference measure with respect to the target measure. However, the MCMC procedure they consider is the random walk Metropolis algorithm that involves the acceptance probability $a(u, v) = \min\{1, (d\eta^*/d\eta)(u, v)\}$, where η is the measure defined by the transition kernel $q(u, v)$ of the MCMC algorithm (i.e., $v|u \sim q(u, \cdot)$) and η^* is the measure obtained by reversing the roles of u and v in the definition of η. Their Theorem 6.3 shows that after discretization, η^* is singular with respect to η and therefore "all proposal moves are rejected with probability 1" for the random walk Metropolis algorithm, which proposes $v^{(k)} = u^{(k)} + \beta\xi^{(k)}$, with $\xi^{(k)} \sim N(0, C)$, and chooses $u^{(k+1)} = v^{(k)}$ with probability $a(u^{(k)}, v^{(k)})$, setting $u^{(k+1)} = u^{(k)}$ if $v^{(k)}$ is rejected. To get around this difficulty, they introduce a pre-conditioned Crank-Nicolson (pCN) adjustment, which proposes $v^{(k)} = \sqrt{1 - \beta^2}u^{(k)} + \beta\xi^{(k)}$. Here $\beta^2 = 8\delta/(2 + \delta)^2$ and C is the covariance matrix (after truncation and discretization) of the covariance operator $\mathcal{C}$ for the Gaussian proposal measure. Because MCMC-SS does not involve η and η^*, it does not require the pCN adjustments; see Lai et al. (2021) for details and further discussion.

Making use of bounds on a weighted total variation norm of the difference between the target distribution and the empirical measure defined by the sample paths of the MCMC procedure, Lai et al. (2021) have developed an asymptotic theory of the MCMC-SS estimates, as both K and N approach ∞, of functionals of the target distribution. This asymptotic theory includes asymptotic normality of the MCMC-SS estimates, provides consistent estimators of their standard errors, and establishes their asymptotic optimality by deriving certain oracle properties. Implementation via sequential Monte Carlo schemes called "particle filters" and parallelization is also given. In his Ph.D. thesis, Zhu who is a coauthor of Lai et al. (2021) describes a numerically stable implementation of MCMC-SS that can be vectorized and parallelized, using Julia v0.62 (Bezanson et al., 2017) and the ArrayFire GPU library (Yalamanchili et al., 2015). He also develops scalable implementations for high-dimensional states/parameters using differentiation through mixture distributions for stochastic gradient descent; see Zhu (2021).

In the context of cortical organoids described in the first paragraph of Section 2, the target distribution is the posterior distribution of a precomputed feature of the organoid as a scaled-down model of the preterm human brain, conditional on the observations which are the 101 serial EEG recordings from 39 preterm infants. The uncertainty quantification (Lai et al., 2021) of the posterior distribution of a precomputed feature of cortical organoids provides a principled and systematic approach to the comparison of the feature between cortical organoids and the observations from the preterm infants, in contrast to the lack of uncertainty quantification for the approach and results of Trujillo et al. (2019, pp. 8-9 and Fig. 4A, B, C, D on p. 31) mentioned in the first paragraph of Section 2. Moreover, the methods of time-frequency analysis in the preceding subsection can be used to compute the predictive distribution of the feature of the cortical organoids given the observations, which is the same as the target distribution. The predictive distribution typically also involves an unspecified hyperparameter vector $\boldsymbol{\theta}$, as in manifold learning of

Wang et al. (2020). This corresponds to a Bayesian approach with prior densities belonging to a family of proposal densities $q(\boldsymbol{\theta}|\gamma)$, in which $\gamma \in \Gamma$ indexes the family and Γ is a convex subset of $\mathcal{R}^d$. Lai et al. (2021) have shown that MCMC-SS eventually samples from $q(\cdot|\gamma_p)$ that has the smalllest Kullback-Leibler divergence from $p(\cdot)$, and therefore from the target density if it belongs to $\{q(\cdot|\gamma) : \gamma \in \Gamma\}$.

3 DISCUSSION AND CONCLUDING REMARKS

Haddad and Lai actually initiated similar research forty years ago when they worked on cardiorespiratory patterns during sleep in a SIDS (Sudden Infant Death Syndrome) project at Columbia University's Pediatrics Department; see Haddad et al. (1981) who describe the study population consisting of 12 infants "with one or more episodes of aborted SIDS" (four of whom had siblings who died of SIDS), and 19 normal infants, all born full-term except for one aborted SIDS infant born at 37 weeks of gestation. After describing the study design and methods of statistical analysis. Haddad et al. (1981) presented results on total tidal volume (Vt), respiratory cycle time (Ttot), and increase in Vt/Ttot resulting from 2% increase in CO_2 concentration in sleeping chamber, comparing aborted SIDS to normal infants in both REM (rapid eye movement) and quiet sleep. Because of the inability to induce stress such as loaded breathing as in Bazzy and Haddad (1984) and Haddad et al. (1986), animal models involving sheep, puppies and dogs were used; see also Haddad et al. (1984). In particular, Bazzy and Haddad (1984) "studied diaphragmatic muscle function during inspiratory flow resistive loaded breathing" in 6 unanesthetized sheep over periods of 6-8 months. Data were collected (baseline) and after application of the loads that were sustained for up to 90 minutes. Loads were divided into mild (< 50 cm$H_2O\cdot$ l$^{-1}\cdot$s), moderate (50-150 cm$H_2O\cdot$ l$^{-1}\cdot$s) and severe (> 150 cm$H_2O\cdot$ l$^{-1}\cdot$s). They found that "(1) the diaphragm is capable of generating large pressure for prolonged periods with no evidence of fatigue, (2) with very high inspiratory resistive loads mechanical failure of the diaphragm can occur, (3) diaphragmatic fatigue is associated with acute hypercania and therefore failure of the entire respiratory pump, and (4) a decrease in integrated EMG (iEMG) and a concomitant shift in the EMG power spectral density towards lower frequencies precede the mechanical failure of the diaphragm." Thus, similar to the power spectral density of the EEG signal in the first paragraph of Section 2, Bazzy and Haddad use shift of the power spectrum of the EMG towards lower frequencies to identify the onset of diaphragmatic muscle fatigue in adult sheep. The frontier methods of time-frequency analysis in Section 2.1 are therefore also relevant to the problem of diaphragmatic muscle fatigue and rhythmic variations in cardiorespiratory signals studied by Haddad and Lai forty years ago. Pointing out that in the 1980s "investigators from various disciplines focused their efforts on finding out whether SIDS is related to hypoxia or anoxia (acute or chronic) before death and whether this relation is responsible for events leading to death", Haddad (1992) reviewed "studies in the recent past" from various fields — epidemiology, physiology of infant death and SIDS, pathology of the airway, and animal studies. Although "most of the evidence accumulated so far, including that obtained in the past two years, is circumstantial", he concluded that "SIDS was little understood for many years until, over the past few years, its basic underlying genetic defect was better characterized (from recent animal and human studies), and light could finally be seen at the end of the tunnel", again linking genetics and feedback mechanisms to see this light, as in the exemplary work of Hall, Rosbach and Young on the circadian rhythm. Combining various clues and insights from different areas/studies via an empirical Bayes model is the capability of the frontier approach described in Section 2.2; see Chen et al. (2018, Sects 3.6.3, 5.4, 6.2.3, 7.4) for post-marketing monitoring of medical product safety.

A related direction of our ongoing research is to combine several biomedical signals, which form a multivariate time series, thereby providing a more holographic view of a human subject. For example, in a intensive care unit, PPG can be combined with EEG, EMG, respiratory and other signals to evaluate a

patient's health status. Liu et al. (2019, 2021) have applied a combination of ST and EEG channels to study sleep dynamics, and an "interpretable machine learning algorithm" to assess consistency of sleep-stage scoring rules across multiple sleep centers. How to utilize available information from multiple centers is a sensor fusion problem. We are currently combining recent advances in severe fusion with those in TFA to develop integrated statistical analysis of the multivariates time series of multiple biomedical signals.

ACKNOWLEDGMENTS

We thank Lai's research assistant Chenru Liu at Stanford University for her valuable contributions to the preparation, timely submission and revision of the paper. We also thank the reviewers whose innovative comments and suggestions have resulted in substantial improvement of the presentation. Lai's research was supported by the National Science Foundation under DMS-1811818. Haddad's research was supported by the National Institutes of Health under 1R01HL146530 and 1R21NS111270. Moutri's research was supported by the National Institutes of Health under DP2-OD006495.

REFERENCES

Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B*, 72(3):269–342.

Bazzy, A. R. and Haddad, G. G. (1984). Diaphragmatic fatigue in unanesthetized adult sheep. *Journal of Applied Physiology*, **57**(1):182–190.

Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1):65–98.

Chan, H. P. and Lai, T. L. (2013). A general theory of particle filters in hidden Markov models and some applications. *Annals of Statistics*, 41(6):2877–2904.

Chen, J., Heyse, J. F., and Lai, T. L. (2018). *Medical Product Safety Evaluation: Biological Models and Statistical Methods*. Chapman & Hall/CRC, Boca Raton FL.

Chopin, N. (2002). A sequential particle filter method for static models. *Biometrika*, 89(3):539–552.

Chopin, N., Jacob, P. E., and Papaspiliopoulos, O. (2013). SMC2: an efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society: Series B*, 75(3):397–426.

Cotter, S. L., Roberts, G. O., Stuart, A. M., and White, D. (2013). MCMC methods for functions: modifying old algorithms to make them faster. *Statistical Science*, 28:424–446.

Daubechies, I., Lu, J., and Wu, H. T. (2011). Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool. *Appl Comput Harmon Anal*, **30**:243–261.

Daubechies, I., Wang, Y., and Wu, H. T. (2016). ConceFT: concentration of frequency and time via a multitapered synchrosqueezed transform. *Philos Trans Roy Soc A-Math Phys Eng Sci*, **374**:20150193.

Flandrin, P. (1999). Time-frequency/time-scale analysis. *In Wavelet analysis and its applications*, **10**.

Gordon, N. J., Salmond, D. J., and Smith, A. F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings F (Radar and Signal Processing)*, volume 140, pages 107–113.

Haddad, G. G. (1992). The multifaceted sudden infant death syndrome: controversies and hope. *Current Opinion in Pediatrics*, **4**(3):426–430.

Haddad, G. G., Jeng, H. J., Bazzy, A. R., and Lai, T. L. (1986). Within-breath electromyographic changes during loaded breathing in adult sheep. *Journal of Applied Physiology*, **61**(4):1316–1321.

Haddad, G. G., Jeng, H. J., Lee, S. H., and Lai, T. L. (1984). Rhythmic variations in R-R interval during sleep and wakefulness in puppies and dogs. *American Journal of Physiology: Heart and Circulatory Physiology*, **247**(1):H67–H73.

Haddad, G. G., Leistner, H., Lai, T. L., and Mellins, R. B. (1981). Ventilation and ventilatory pattern during sleep in aborted sudden infant death syndrome. *Pediatric Research*, **15**(5):879–883.

Huang, N. E., Shen, Z., R., L. S., C., W. M., Shih, H., Zheng, Q., Yen, N. C., Tung, C. C., and Liu, H. (1998). The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc London Ser A-Math Phys Eng Sci*, **454**:903–995.

Lai, T. L., Xu, H., Zhu, M. H., and Chan, H. P. (2021). MCMC with sequential substitutions for joint state and parameter estimation in hidden Markov models. *Technical Report, Department of Statistics, Stanford University*.

Lin, C. Y., Su, L., and Wu, H. T. (2018). Wave-shape function analysis – when cepstrum meets time-frequency analysis. *J Fourier Anal Appl*, **24**:451–505.

Lin, Y. T., John, M., and Wu, H. T. (2021). Wave-shape oscillatory model for biomedical time series with applications. *Foundations of Data Science*.

Liu, G. R., Lin, T. Y., Wu, H. T., Sheu, Y. C., Liu, C. L., Liu, W. T., Yang, M. C., Ni, Y. L., Chou, K. T., Chen, C. H., Wu, D., Lan, C. C., Chiu, K. L., Chiu, H. Y., and Lo, Y. L. (2021). Large-scale assessment of consistency in sleep stage scoring rules among multiple sleep centers using an interpretable machine learning algorithm. *Journal of Clinical Sleep Medicine*, 17(2):159–166.

Liu, G. R., Lo, Y. L., Malik, J., Sheu, Y. C., and Wu, H. T. (2019). Diffuse to fuse EEG spectra–Intrinsic geometry of sleep dynamics for classification. *Biomedical Signal Processing and Control*, 55(5):101576.

Liu, J. S. (2008). *Monte Carlo Strategies in Scientific Computing, 2nd ed*. Springer, New York.

Mallat, S. (2012). Group invariant scattering. *Pure Appl Math*, **10**:1331–1398.

Nahon, M. (2000). Phase evaluation and segmentation. *PhD thesis, New Haven: Yale University*.

Shelley, K. H. (2007). Photoplethysmography: beyond the calculation of arterial oxygen saturation and heart rate. *Anesth. Analg.*, **105**:531–536.

Trujillo, C. A., Gao, R., Negraes, P. D., Gu, J., Buchanan, J., Preissl, S., Wang, A., Wu, W., Haddad, G. G., Chaim, I. A., Domissy, A., Vandenberghe, M., Devor, A., Yeo, G. W., Voytek, B., and Muotri, A. R. (2019). Complex oscillatory waves emerging from cortical organoids model early human brain network development. *Cell Stem Cell*, **25**(4):558–569.

Wang, S. C., Wu, H. T., Huang, P. H., Chang, C. H., Ting, C. K., and Lin, Y. T. (2020). Novel imaging revealing inner dynamics for cardiovascular waveform analysis via unsupervised manifold learning. *Anesth. Analg.*, **130**:1244–1254.

Wu, H. T. (2020). Current state of nonlinear-type time-frequency analysis and applications to high-frequency biomedical signals. *Current Opinion in Systems Biology*, **23**:8–21.

Yalamanchili, P., Arshad, U., Mohammed, Z., Garigipati, P., Entschev, P., Kloppenborg, B., Malcolm, J., and Melonakos, J. (2015). ArrayFire - A high performance software library for parallel computing with an easy-to-use API.

Zhu, M. (2021). Improving and accelerating particle-based probabilistic inference. *Ph.D. dissertation in Computer Science, Stanford University*.