IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 14, 2020, accepted November 8, 2020, date of publication December 4, 2020,
date of current version December 16, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3042737

Estimation of Time-Varying Spectral Peaks and Decomposition of EEG Spectrograms

PATRICK A. STOKES, (Member, IEEE), AND MICHAEL J. PRERAU, (Member, IEEE)
Division of Sleep and Circadian Disorders, Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA

Corresponding authors: Patrick A. Stokes (patrick.stokes@gmail.com) and Michael J. Prerau (mprerau@bwh.harvard.edu)

This work was supported by the National Institute of Neurological Disorders and Stroke under Grant R01NS096177-05.

ABSTRACT Detection of spectral peaks and estimation of their properties, including frequency and amplitude, are fundamental to many applications of signal processing. Electroencephalography (EEG) of sleep, in particular, displays characteristic oscillations that change continuously throughout the night. Capturing these dynamics is essential to understanding the sleep process and characterizing the heterogeneity observed across individuals. Most sleep EEG analyses rely on either time-averaged spectra or bandpassed amplitude/power. Unfortunately, these approaches obscure the time-variability of peak properties, require specification of a priori criteria, and cannot distinguish power from nearby oscillations. More sophisticated approaches, using various spectral models, have been proposed to better estimate oscillatory properties, but these too have limitations. We present an improved approach to spectrogram decomposition, tracking time-varying parameterized peak functions and dynamically estimating their parameters using a modified form of the iterated extended Kalman filter (IEKF) that incorporates discrete On/Off-switching of peak combinations and a sampling step to draw the initial reference trajectory. We evaluate this approach on two types of simulated examples—one nearly within the model class and one outside. We find excellent performance, in terms of spectral fits and accuracy of estimated states, for both simulation types. We then apply the approach to real EEG data of sleep onset, obtaining quality spectral estimates with estimated peak combinations closely matching the expert-scored sleep stages. This approach offers not only the ability to estimate time-varying parameters of spectral peaks but, moving forward, the potential to estimate the governing dynamics and analyze their variability across nights, subjects, and clinical groups.

INDEX TERMS Electroencephalography, Kalman filter, parameter estimation, sleep, spectral analysis, spectral peaks, spectrogram.

I. INTRODUCTION

Many applications of spectral analysis involve detection and estimation of characteristic peaks or functions within background noise, such as audio and speech signal processing, magnetic resonance spectroscopy (MRS), and analysis of physiological signals. Moreover, in many cases, the composition or properties of the peaks vary in time. For example, electroencephalographic (EEG) recordings of the brain during various states, processes, and conditions display characteristic time-varying oscillations, which organize functional activity and facilitate information transfer between regions [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Liu.

EEG of sleep, in particular, is characterized by patterns of oscillatory activity that change at multiple time scales over the course of the night [2]–[7]. These oscillations are classically defined by frequency ranges and other properties [8], like duration. Some arise and persist over extended periods of time (∼minutes), whereas others are more transient, lasting less than a few seconds.

EEG oscillations are not pure sinusoids but variable phenomena and complex waveforms in aperiodic background noise. The waveform shapes reflect the underlying neuronal dynamics and change with physiological state [9]–[12]. The properties of the resulting broadband spectral peaks (e.g., peak bandwidth and shape) correspond to the waveform morphology. Likewise, the background noise is believed to reflect the relative excitation and inhibition of the underlying neural

populations and appears spectrally as a '$1/f$' decay [13]. When visualized as a spectrogram, the oscillations of EEG thus appear as peaks moving on a background of noise, as the properties of these peaks (e.g., frequency location, amplitude, and bandwidth) and those of the noise background change with time.

In addition to the temporal variability, EEG activity varies spatially across the scalp corresponding to the cortical localization of the underlying processes [14]–[16]. Sleep EEG, specifically, displays a topographic pattern of oscillatory peak dynamics [5], [17]–[21]. Crucially, the peak properties (including frequency location, prominence, and even number) are also known to vary widely between individuals [16], [22], and in sleep, often deviate substantially from the classical definitions [21], [23], [24]. Moreover, the oscillatory properties have been found to change with age [12], [13], [19], [25], [26] and with clinical pathology. Alterations and aberrations in sleep EEG oscillations have been found in schizophrenia [27], [28], autism [29], epilepsy [30], sleep disorders [31]–[34], stroke [35], and neurodegenerative disorders [36], for example.

Sleep is thus a continuous, dynamic process, and modeling and analyzing the full variability of sleep EEG oscillations is imperative for understanding the process of sleep and identifying pathological biomarkers. Such analyses require an approach that 1. captures the spectral shapes and relevant peak properties of the oscillatory activity, 2. tracks the continuous changes of the properties of interest throughout the night, 3. identifies the (relatively) discrete On/Off-switching of the transient oscillations, and 4. flexibly accounts for the intra-subject spatio-temporal variations and inter-subject heterogeneity of spectral peaks. Unfortunately, most sleep EEG analyses do not account for this variability, averaging over the dynamics and/or applying a priori definitions and criteria across subjects.

Traditionally, the process of sleep is discretized into five stages—wake, rapid eye movement (REM), non-REM (NREM) 1, NREM 2, and NREM 3—based on the presence and prominence of strictly defined oscillations. All clinical analyses of sleep EEG are done by having an expert technician manually score the entire night's sleep record into the five discrete stages to produce a hypnogram. This is done by visualizing the EEG signals—as well as other physiological signals, like electrocardiogram (ECG) and electrooculogram (EOG)—in 30 second segments as time-domain traces and assessing the presence/absence of the characteristic oscillations according to the clinical definitions [8]. This process is extremely time consuming and highly subjective [37], and the resulting hypnogram very coarsely discretizes the continuous dynamics of the sleep process.

Quantitative analyses of oscillatory activity in the sleep EEG are most commonly based on either average spectra or bandpassed amplitudes. Average spectra are formed by averaging over spectral estimates from separate time windows [6], [17], [35], usually based on sleep stage. From these spectra, oscillation frequencies, amplitudes, and other properties can be determined, but nearly all of the temporal variations are averaged out. Alternatively, the EEG signal is bandpassed to the frequency range of interest and instantaneous amplitude or power within the band computed [18], [24], [27], [31], [38], e.g., via Hilbert transform. While this provides instantaneous amplitude/power estimates, specific frequency and bandwidth estimates are unavailable, being instead fixed by the specified range. Thus, both approaches obscure, to some degree, the temporal variations of the peak properties, both are susceptible to confounding of power/amplitude estimates due to leakage from nearby or overlapping oscillations, and both are rely on a priori band definitions and other oscillation criteria that do not account for inter-subject heterogeneity. Some analyses, usually those analyzing specific detected oscillations, do tailor detection criteria to individual subjects [4], [20], [24], [35], [38], [39]. But despite these adjustments, such analyses still impose hard criteria on phenomena that do not necessarily follow such demarcations [21]. Automatic oscillation-detection methods have also been shown to disagree substantially with each other and to perform poorly relative to expert scoring [37].

A variety of approaches have been proposed to better estimate the oscillatory properties of interest from EEG or other signals. Olbrich and Achermann [40] use windowed-autoregressive (AR) models to identify oscillatory events and estimate frequencies and amplitudes. Tarvainen *et al.* [41] estimate a state-space autoregressive-moving average (ARMA) model for nonstationary EEG. Dubois *et al.* [42] and Matsuda and Komaki [43] use state-space models of time-varying parameterized oscillations, estimated via unscented particle filter (UPF) and Kalman smoother (KS), respectively. Yet other approaches have been proposed to estimate instantaneous amplitudes and frequencies of signals comprising variable sinusoids [44]–[46]. While these approaches perform well at identifying frequencies and amplitudes, the underlying AR and sinusoids do not fully capture the spectral shape of many EEG oscillations. Moreover, AR parameters are not directly interpretable in terms of frequencies and amplitudes, which must be derived. Haller *et al.* [9] offers a spectral decomposition fitting explicitly parameterized peak functions. However, as with the windowed-AR of [40], there is no temporal continuity between estimates of separate windows. Prerau *et al.* [47] proposed a particle filter approach to estimate time-varying, parameterized peaks of an EEG spectrogram. While this method enables tracking of variable peaks, it encountered several limitations and difficulties. First, the method, like the state-space approaches of [43]–[46], does not handle the phenomena of peaks quickly turning off (disappearing) and on (reappearing), as with sleep spindles in EEG of non-REM sleep. More importantly, the high dimensionality of the observations, i.e., the instantaneous spectral estimates, quickly leads to degenerate particle weightings.

We propose an improved method for the decomposition of spectrograms into sets of time-varying parameterized peak functions. We develop and demonstrate this method in the

context of analyzing sleep EEG, but the method may be applicable to analogous problems in other fields, like spectroscopy. Our method augments the state-space model of [47] to capture the discrete On/Off-switching of peaks and uses a variation of the iterated extended Kalman filter (IEKF) to overcome the degeneracy of the particle filter and estimate the varying peak parameters. Because the peaks are parameterized in terms of directly interpretable properties, the temporal continuity imposed by the state-space model is interpretable as well. The resulting decomposition of the spectrogram provides not only improved peak identification over traditional approaches, like average spectra and bandpass filtering, but also direct estimation of peak properties and their uncertainties. These estimates further enable statistical analyses and comparisons of peak properties over time. Characterization of these dynamics, and their variation within and between individuals and between patient groups, will not only further our understanding of the process of sleep but possibly reveal diagnostic markers.

II. METHODS

We propose using a state-space model to represent a spectrogram as combinations of parameterized peak oscillations, where the peak parameters vary smoothly in time and the combination of On/Off-peaks changes dynamically as well. The peak parameters and On/Off-combination are estimated using a modified version of the IEKF and a decomposition of the spectrogram so obtained. We demonstrate this method and test its relative performance on two types of simulated data. We then apply the method to real EEG data from sleep. (Code is available for download in the supplemental material as well as at http://sleepeeg.org/peaktracking/.)

A. PARAMETERIZED SPECTRAL APPROXIMATION

In this model, at each time t, the instantaneous spectrum y_t of the estimated spectrogram, evaluated at vector of discrete frequency bins ω, is approximated as the sum of a set of peak functions $h^{(i)}$ observed in noise v_t,

$$y_t = \sum_{i \in \mathcal{I}} h^{(i)}\left(\omega; x_t^{(i)}\right) + v_t. \tag{1}$$

Here, $\mathcal{I}$ is the set of available peaks, the $h^{(i)}$ are their respective peak functions, and the $x_t^{(i)}$ are their respective sets of parameters. The observation noise is assumed zero-mean, $E[v_t] = 0$, and temporally uncorrelated with covariance $E\left[v_{t_1} v'_{t_2}\right] = \mathrm{R} \cdot \delta_{t_1 - t_2}$.

B. PEAK FUNCTIONS

The model is flexible in terms of the number and types of peak functions that can be included. The parameters are generally chosen to be interpretable in terms of location (frequency) and shape (amplitude, bandwidth, etc.) of the corresponding peak. Those that we have implemented include a Gaussian (with and without harmonics), a shifted-gamma, a box-exponential, and an exponential-decay. The parameterized equations of the peaks are given in Appendix A.

These peak functions were chosen to represent classically recognized oscillatory peaks in sleep EEG. We use an exponential-decay to represent the background spectral density, shifted-gammas to represent slow (< 1.5 Hz) and delta-theta (1.5–8 Hz), a Gaussian with harmonics to represent alpha (8–12 Hz), a Gaussian to represent sigma (12–16 Hz), and a box-exponential to represent the 60 Hz line noise.

C. PARAMETER EVOLUTION AND BOUNDS

The parameters of the peak functions form the state vector x_t, varying smoothly over time following a random walk,

$$x_t = \mathrm{F} \cdot x_{t-1} + w_t. \tag{2}$$

The state noise w_t is assumed zero-mean, $E[w_t] = 0$, and temporally uncorrelated with covariance $E\left[w_{t_1} w'_{t_2}\right] = \mathrm{Q} \cdot \delta_{t_1 - t_2}$. It is further assumed the state and observation noises are uncorrelated, $E\left[v_{t_1} w'_{t_2}\right] = 0$. We use a decay factor $\mathrm{F} = 0.9 \cdot \mathrm{I}$ for the state transition to enhance stability of the resulting filter estimates.

Many of the peak parameters are required to be positive (e.g., amplitudes, frequencies, and bandwidths). It is often advantageous to further restrict parameters to specified intervals (e.g., a specific frequency range for given peak). To achieve this, we employ exponential and sigmoid link functions, respectively. Their parameterized equations are given in Appendix B. The resulting bounded versions of the dynamic state variables, $\bar{x}_{t,k} = l_k\left(x_{t,k}\right)$, then parameterize the observation peak functions, where k indexes the component state variables and their respective link functions.

D. COMBINATIONS OF ON/Off-PEAKS AND THEIR TRANSITIONS

In many applications, like EEG analysis, the oscillation content of a signal may change rapidly, with sets of peaks appearing or disappearing, often in particular combinations, almost instantaneously. It is, thus, frequently of benefit and interest to allow peaks to turn on and off in certain combinations at various times, and to model the transitions between these combinations. We model the dynamic On/Off-switching of peaks as a probabilistic discrete transition among a set of allowable combinations $\mathcal{J}$. With α_t the vector of probabilities of being in each combination and Φ the matrix of transition probabilities (i.e., Φ_{j_2, j_1} is the probability of transitioning from combination j_1 to j_2), the probabilities vary according to

$$\alpha_t = \Phi \cdot \alpha_{t-1}. \tag{3}$$

For clarity and brevity, we variably refer to the combination of On/Off-peaks as the On/Off-peaks combo, the On/Off-combo, or even simply the combo. The formation of the combo transition matrix Φ is described in Alg. 5 in Appendix C.

E. STATE-SPACE MODEL

Together, the observation and state equations of the state-space model are as follows,

$$y_t = \sum_{i \in \mathcal{I}(j_t)} h^{(i)}\left(\omega; \overline{x}_t^{(i)}\right) + v_t, \tag{4}$$

$$x_{t+1} = \mathrm{F} \cdot x_t + w_{t+1}, \text{ and} \tag{5}$$

$$\alpha_{t+1} = \Phi \cdot \alpha_t, \tag{6}$$

where $\overline{x}_{t,k} = l_k\left(x_{t,k}\right)$ are the bounded peak parameters and $\mathcal{I}(j_t)$ is the set of On-peaks in On/Off-combo j_t. For brevity and clarity moving forward, we denote the overall observation function for each combo j by $h^j(x)$, where $h^j(x) = \sum_{i \in \mathcal{I}(j)} h^{(i)}\left(\omega; l\left(x^{(i)}\right)\right)$ incorporates the sub-selection of parameters, the application of link functions, and the summation of peak functions, and suppresses the dependence on frequency bins.

F. IEKF WITH DISCRETE SWITCHING AND SAMPLED REFERENCE TRAJECTORY

The peak parameters x_t and On/Off-peak combo j_t are estimated via a modified form of the IEKF. The general IEKF is detailed in references such as [48]. We make two modifications to the IEKF—one to handle the discrete switching of the On/Off-combo and one to further handle the nonlinearity and improve the convergence. For real-data applications, we also add checks to handle non-convergence.

The outline of the approach is as follows. Given the filter estimates for the previous time step, the common prediction estimates are obtained. For each On/Off-combo, random state samples are drawn from a multivariate Gaussian with mean and covariance given by the prediction estimates, and the maximum likelihood draw becomes the initial reference trajectory for that combo. Filter estimates are obtained for each combo by the iterated updates, starting at their respective initial references. The posterior probability of each combo is computed from the total likelihoods and transition probabilities. The new filter estimates are then the posterior mode combo and its state and state-error covariance. The general steps for the full algorithm are given in Alg. 1.

1) IEKF STEP AND ESTIMATION OF ON/Off-COMBO

The discrete switching of the On/Off-combination is handled analogously to adaptive parameter estimation as detailed in references such as [49]. At each time t, the IEKF updates are computed separately for each combo j to obtain separate prediction and filter estimates of the state and state-error covariance of the approximating Gaussian density. In our case, the state equation is linear, so the prediction updates of the state and state-error covariance estimates are straightforward, $\hat{x}^j_{t|t-1} = \mathrm{F} \cdot \hat{x}^j_{t-1|t-1}$ and $P^j_{t|t-1} = \mathrm{F}P^j_{t-1|t-1}\mathrm{F}^T + \mathrm{Q}$, respectively. Due to the nonlinear dependence of the observation function on the state, the filter updates of the state and state-error covariance estimates, $\hat{x}^j_{t|t}$ and $P^j_{t|t}$, respectively, are approximated by linearizing the update equations around a reference trajectory η^j_t. These equations are given in Alg. 3.

Algorithm 1: Iterated Extended Kalman Filter With Predictive Sampling and Probabilistic Parameter Switching

Data: $y_{1:T}$, $\mathcal{I}$, $\mathcal{J}$, R, F, Q, N_d, N_{iter}, $\hat{x}_0$, P_0, and $\hat{\alpha}_0$
Result: $\{\hat{x}_{t|t}, P_{t|t}, \hat{j}_{t|t}\}_{t=1}^T$
$\hat{x}_{0|0} \leftarrow \hat{x}_0$; $P_{0|0} \leftarrow P_0$; $\hat{\alpha}_{0|0} \leftarrow \hat{\alpha}_0$
for $t \leftarrow 1$ **to** T **do**
  $\hat{x}_{t|t-1} \leftarrow \mathrm{F} \cdot \hat{x}_{t-1|t-1}$
  $P_{t|t-1} \leftarrow \mathrm{F}P_{t-1|t-1}\mathrm{F}^T + \mathrm{Q}$
  for $j \leftarrow 1$ **to** $N_{\mathcal{J}}$ **do**
    $\varepsilon^j_{t|t-1} \leftarrow y_t - h^j\left(\hat{x}_{t|t-1}\right)$
    $\eta^j_{t,1} \leftarrow$ referenceTrajectory (Alg. 2 or 6)
    $\{\hat{x}^j_{t|t}, P^j_{t|t}, \mathrm{LL}^j_t\} \leftarrow$ filterUpdate (Alg. 3, 7, or 8)
    $\varepsilon^j_{t|t} \leftarrow y_t - h^j\left(\hat{x}^j_{t|t}\right)$
  end
  $\{\hat{x}_{t|t}, P_{t|t}, \hat{\alpha}_{t|t}\} \leftarrow$ modeOnOffCombo (Alg. 4)
end

Algorithm 2: referenceTrajectory—Max Likelihood

Data: $\hat{x}_{t|t-1}$, $P_{t|t-1}$, N_d, j, y_t, h^j, R
Result: $\eta^j_{t,1}$
$d^{j,1} \leftarrow \hat{x}_{t|t-1}$
$\varepsilon^{j,1} \leftarrow y_t - h^j\left(d^{j,1}\right)$
for $i \leftarrow 2$ **to** N_d **do**
  $w^{j,i} \sim \mathcal{N}\left(0, P_{t|t-1}\right)$
  $\overline{w}^{j,i} \leftarrow$ only retain components relevant to j
  $d^{j,i} \leftarrow \hat{x}_{t|t-1} + \overline{w}^{j,i}$
  $\varepsilon^{j,i} \leftarrow y_t - h^j\left(d^{j,i}\right)$
end
$i^* \leftarrow \arg\min_i \left(\varepsilon^{j,i}\right)^T \mathrm{R}^{-1}\varepsilon^{j,i}$
$\eta^j_{t,1} \leftarrow d^{j,i^*}$

In the basic EKF, the prediction estimate serves as the reference trajectory, $\eta^j_t = \hat{x}^j_{t|t-1}$. In the IEKF, the approximation is improved by iteratively using the resulting estimate as subsequent reference and stopping at convergence.

The linearized equations utilize the derivative matrix of the observation function, $M^j(\eta) = \partial_x h^j(\eta)$. The derivatives of the individual peak functions and the link functions with respect to their parameters are straightforward to compute. They are included in Appendixes A and B.

The filter estimates of the combo probabilities $\hat{\alpha}_{t|t}$ are computed by taking the observation likelihoods,

$$p\left(y_t \mid j_t, y_{1:t-1}\right) = \int_{x_t} p\left(y_t \mid x_t, j_t, y_{1:t-1}\right) \cdot p\left(x_t \mid j_t, y_{1:t-1}\right) dx_t, \tag{7}$$

obtained separately from the IEKF for each combo, weighting by the predicted transition probabilities, $p\left(j_t \mid y_{1:t-1}\right) = \Phi_{j,\cdot} \cdot \hat{\alpha}_{t-1|t-1}$, and normalizing,

$$p\left(j_t \mid y_{1:t}\right) = \frac{p\left(y_t \mid j_t, y_{1:t-1}\right) p\left(j_t \mid y_{1:t-1}\right)}{\sum_{j_t} p\left(y_t \mid j_t, y_{1:t-1}\right) p\left(j_t \mid y_{1:t-1}\right)}. \tag{8}$$

Algorithm 3: filterUpdate—EKF/IEKF

Data: $\hat{x}_{t|t-1}$, $P_{t|t-1}$, $\eta^j_{t,1}$, N_{iter}, y_t, h^j, $\partial_x h^j$, R
Result: $\hat{x}^j_{t|t}$, $P^j_{t|t}$, LL^j_t
for $k \leftarrow 1$ **to** N_{iter} **do**
$M \leftarrow \partial_x h^j\left(\eta^j_{t,k}\right)$
$K \leftarrow P_{t|t-1}M^T\left[MP_{t|t-1}M^T + \text{R}\right]^{-1}$
$z \leftarrow y_t - h^j\left(\eta^j_{t,k}\right) - M\left[\hat{x}_{t|t-1} - \eta_k\right]$
$\eta^j_{t,k+1} \leftarrow \hat{x}_{t|t-1} + Kz$
end
$\hat{x}^j_{t|t} \leftarrow \eta^j_{t,N_{\text{iter}+1}}$
$P^j_{t|t} \leftarrow [\text{I} - KM]\,P_{t|t-1}\,[\text{I} - KM]^T + K\text{R}K^T$
$\text{LL}^j_t \leftarrow -\frac{1}{2}z^T\left[MP_{t|t-1}M^T + \text{R}\right]^{-1}z$
$\quad - \frac{1}{2}\ln\det(MP_{t|t-1}M^T + \text{R}) - \frac{N_y}{2}\ln(2\pi)$

Algorithm 4: modeOnOffCombo

Data: $\hat{\alpha}_{t-1|t-1}$, Φ, $\text{LL}_t^{1:N_\mathcal{J}}$, $\hat{x}_{t|t}^{1:N_\mathcal{J}}$, $P_{t|t}^{1:N_\mathcal{J}}$, $\varepsilon_{t|t}^{1:N_\mathcal{J}}$, $\hat{x}_{t|t-1}$, $P_{t|t-1}$, $\varepsilon_{t|t-1}^{1:N_\mathcal{J}}$, R
Result: $\hat{x}_{t|t}$, $P_{t|t}$, $\hat{\alpha}_{t|t}$
if $\min_j\left(\varepsilon^{jT}_{t|t-1}R^{-1}\varepsilon^j_{t|t-1}\right) < \min_j\left(\varepsilon^{jT}_{t|t}R^{-1}\varepsilon^j_{t|t}\right)$ **then**
$j^* \leftarrow \arg\min_j\left(\varepsilon^{jT}_{t|t-1}\text{R}^{-1}\varepsilon^j_{t|t-1}\right)$
$\hat{x}_{t|t} \leftarrow \hat{x}_{t|t-1}$; $P_{t|t} \leftarrow P_{t|t-1}$; $\hat{\alpha}_{t|t} \leftarrow \mathbf{e}_{j^*}$
else
$j^* \leftarrow \arg\max_j \text{LL}^j_t$
for $j \in \mathcal{J}$ **do**
$\text{LR}^j_t \leftarrow \text{LL}^j_t - \text{LL}^{j^*}_t + \log\left(\Phi_{j,\cdot}\cdot\hat{\alpha}_{t-1|t-1}\right)$
$\quad - \log\left(\Phi_{j^*,\cdot}\cdot\hat{\alpha}_{t-1|t-1}\right)$
end
$\hat{j}_{t|t} \leftarrow \arg\max_j \text{LR}^j_t$
$\hat{x}_{t|t} \leftarrow \hat{x}^{\hat{j}_{t|t}}_{t|t}$; $P_{t|t} \leftarrow P^{\hat{j}_{t|t}}_{t|t}$; $\hat{\alpha}_{t|t} \leftarrow \mathbf{e}_{\hat{j}_{t|t}}$
end

The observation likelihoods are computed as logarithms, $\ln p(y_t \mid j_t, y_{1:t-1})$, and denoted LL^j_t in the algorithms.

While in principle the joint filter density $p(x_t, j_t \mid y_{1:t})$ could be computed as a mixture of separate Gaussian state estimates, we have found that this performs poorly. Instead, at each time step, we select the On/Off-combination with the maximum probability, $\hat{j}_{t|t} = \arg\max_{j_t} p(j_t \mid y_{1:t})$, and approximate the combo filter probabilities by the corresponding indicator vector, $\hat{\alpha}_{t|t} \approx \mathbf{e}_{\hat{j}_{t|t}}$. This selected combo and its associated filter estimates then serve as the overall filter estimates, $\hat{x}_{t|t} \approx \hat{x}^{\hat{j}_{t|t}}_{t|t}$ and $P_{t|t} \approx P^{\hat{j}_{t|t}}_{t|t}$, and the starting point for all combos at the next time step. The selection of the modal On/Off-combo is detailed in Alg. 4.

2) IMPROVED REFERENCE TRAJECTORY

To improve the ability of the iterations to converge, we add a step between the prediction and filtering to sample for an improved initial reference trajectory. At each time t, for each combination j, we draw $N_d = 1000$ state samples $d^j_{1:N_d}$ from the common prediction density, $\mathcal{N}\left(\hat{x}_{t|t-1}, P_{t|t-1}\right)$, evaluate the predicted observations, and take the initial reference trajectory $\eta^j_{t,1}$ to be the draw of maximum likelihood, or equivalently that with minimum sum of square prediction errors. The details are given in Alg. 2.

3) MISSING DATA, ARTIFACTS, AND NON-CONVERGENCE

As in any Kalman filter application, instances of missing observations are easily handled by skipping the filter step and propagating the predictions forward. Real EEG data are often corrupted by brief artifacts. These can be handled in off-line analyses by detection and exclusion prior to filtering, treating the corresponding time points as missing data. It would further be possible to incorporate the detection of artifacts in on-line fashion at the expense of increased computation time. At each time, prior to the combo selection, the filter estimates of all combos are compared to the prediction estimates. If a prediction estimate is of greater likelihood than all filter estimates, then the current observation is treated similarly to being an artifact. The corresponding combo is selected and its prediction estimates retained as the filter updates. This is a further safeguard against missed artifacts and iterative divergence and is indicated in Alg. 4.

G. ALTERNATIVE FILTER ESTIMATES

To evaluate the relative importance of the draws and iterations in improving performance, we compare the performance of the above method, referred to as IEKF-d, with that of several other related methods. The basic EKF is obtained by omitting the prediction draws and only performing the first filtering iteration. An EKF with improved reference trajectory, referred to as EKF-d, is obtained by utilizing the draws while only applying the single iteration. The standard IEKF uses iterations without drawing the initial reference trajectory.

We also evaluate the possibility of using higher-order terms in the filtering to improve the quality of the estimates. The iterations in the filtering step of the IEKF can be viewed as a mode optimization via Newton-Raphson with the 2nd-derivatives set to zero. We implemented a version of the IEKF that uses the Hessian in the iteration steps to evaluate whether its inclusion improved the estimates. This method was tested both without particle draws, referred to as IEKF-2, and with particle draws, referred to as IEKF-2d. We also implemented a Gaussian 2nd-order EKF, referred to as EKF-2o, following [48]. Lastly, we tested whether selecting the draw of maximum posterior filter density, referred to as IEKF-dm, as opposed to maximum likelihood, improved convergence. The filter update equations of these variant methods are detailed in the algorithms in Appendix E. The updates for IEKF-2/IEKF-2d are given in Alg. 7. The updates for the EKF-2o are given in Alg. 8. The maximum posterior selection of reference trajectory is included in Alg. 6. All methods share the common filtering steps of Alg. 1.

The set of methods compared are thus:

1) Basic EKF—no prediction draws, single filter update (i.e., one iteration).
2) EKF with draws and one iteration.
3) Standard IEKF—no draws, 10 iterations.
4) IEKF with draws and 10 iterations.
5) IEKF with 2nd-derivative and no draws.
6) IEKF with 2nd-derivative and draws.
7) 2nd-order EKF—no draws, single filter update.
8) EKF with draws evaluated at posterior.

H. SIMULATED DATA

We compared the performance of the variant methods by applying them to two types of simulated data with variable number of peaks present. Every simulation features a set of 1 to 5 peaks on an exponential-decay background. The parameters of all peaks, including the background, vary in time. The non-background peaks also turn on and off probabilistically. The first simulation type is nearly within the generative model class, with the parameters following hard-bounded random walks. The second simulation type is well outside with underlying model class, with the parameters following pseudo-deterministic patterns, including sines, saw-tooths, and steps, of varying periods. One thousand simulations were generated for each simulation type and number of peaks. Each simulation consists of 100 time points and 300 frequency bins, covering a frequency range of 0–100 Hz.

The simulated frequency range was randomly divided into non-overlapping frequency intervals that serve as the frequency bounds of each non-background peak. Each non-background peak is randomly assigned a peak-type as gamma (25%), Gaussian (50%), or box (25%), where the number of harmonics for each Gaussian is determined by the upper bound of its frequency. All amplitudes were bound between 4 and 20. Bandwidth bounds for each peak had a minimum of 3 and a maximum determined by the width of the frequency bounds. See the available code for the complete list of parameter bounds and other simulation details.

I. STATISTICAL MEASURES

The performance of the variant filters when applied to the simulated data was assessed by a set of statistical measures, chosen to evaluate 1. the model fit to the observed spectrogram, 2. the accuracy of the state estimates, and 3. the ability to correctly determine the On/Off-peaks. The model fit to the spectrogram is assessed by three measures computed from the filter observation errors, $\varepsilon_{t|t} = y_t - \hat{h}^{t|t}(\hat{x}_{t|t})$. The mean observation error measures any bias in the model fit. The mean-square of the observation errors reflect the size of the residuals. The Box Q statistic of the observation errors measures two-dimensional correlations in the residuals, capturing aspects of non-whiteness. The mean and mean-square are computed over all time steps and frequency bins, but the Box Q is computed for windows over 4 time lags and ± 4 frequency bins.

The accuracy of estimating the underlying state was assessed by two measures. The mean-square of the state errors, averaged over state variables only for times when the respective parameters are truly On, measures the accuracy of the specific filter estimates. The probability of the true state values lying within the 95%-filter interval of the state, averaged over state variables for all times, measures the more general reliability of filter density in capturing the state. The ability to identify On/Off-peaks is measured by computing the probability of correct On/Off-determination averaged over all peaks at all times.

Several alternatives to the above measures were also computed to support the robustness of the findings. The maximum of the absolute value of the observations errors provides an alternative measure of the size of the residuals. An alternative assessment of the correlations between observation errors is provided by a Wald statistic of the instantaneous correlations across frequency bins (up to 4 neighbors) and the Moran's I statistic over $3\Delta_t \times 4\Delta_\omega$ windows. The Cohen's kappa of the On/Off-combo estimation is an alternative measure of the ability to identify On/Off-peaks.

The last statistic computed was the run time of the filter estimates. We computed each statistic for each filter estimate of each simulation, and compared the distributions (over simulations) of these statistics to evaluate the relative performance of the various forms of the filter. See the available code for the specific computations of the statistics.

J. EXPERIMENTAL DATA: EEG OF SLEEP ONSET

We applied the IEKF-d (and, for comparison, the basic EKF) to sleep EEG data from a single subject from a study previously presented in [50]. We analyzed a single occipital channel (O2) recorded during the second night. The data were recorded at 500 Hz and downsampled to 200 Hz. The analysis was limited to ~55 minutes covering the sleep onset process. The multitaper spectrogram was computed using 15-second windows with a 7-second slide, a time-half-bandwidth product of 15 with 29 tapers, and with linear detrending, resulting in 2049 frequency bins of width 0.0488 Hz. The frequencies of the observed spectrogram were subselected to improve the filter performance. Between 35 Hz and 55 Hz, every 5th frequency bin was retained, and between 55 Hz and 65 Hz, every 2nd bin was retained. The state-space model used in the IEKF-d comprised an exponential-decay background, a gamma slow-oscillation peak, a gamma delta-theta peak, a Gaussian with two harmonics for the alpha peak, a Gaussian sigma peak, and a box peak for the 60 Hz line noise. The details of the model are given in Appendix F.

The observation noises are independent (i.e., R is diagonal) with $R_{ii} = 0.5$, except for frequency bins between 55 and 65 Hz, where $R_{ii} = 3$. The peak parameter bounds, diagonal state noise covariance Q, initial state $\hat{x}_0$, and diagonal initial state covariance P_0 are given in Table 1 in Appendix F. The initial state values for the background parameters, a_0, r_0, and o_0, are determined by an iterative nonlinear fit to the first

20 time points of data. (See supplemental code for details.) The available On/Off-peak combos, $\mathcal{J}$, and initial combo probabilities, α_0, are given in Table 2 in Appendix F. The combo transition matrix Φ is formed via Alg. 5 in Appendix C with $p_{\text{on}} = 0.2$, $p_{\text{off}} = 0.3$, and $p_{\text{stay}} = 0.8$.

The model and its hyperparameters were chosen, largely by trial-and-error, specifically for the example subject and channel. However, in addition to the temporal dynamics, the spectral content of sleep EEG displays spatial variations across the scalp, vast inter-subject heterogeneity, and dramatic changes due to age and pathology. These variations can deviate widely from the classically defined oscillations and are just beginning to be explored.

The example subject was chosen because their oscillatory activity hews closely to the traditional peak definitions. An occipital channel was used because it most prominently displays the eyes-closed alpha oscillation indicative of wake and, hence, is where sleep onset is most easily observed. And the second-night data was used because the first night of sleep studies is for acclimation of the subjects to the study environment.

To further test the robustness of the approach, we applied the IEKF-d, using the same model and hyperparameters, to second-night, O2-channel data for the remaining subjects from the dataset. The results of three of the additional subjects are presented in the Supplemental for comparison. The purpose of this comparison is not a full exploration of the dynamics and variations of sleep EEG oscillatory activity, which is outside the scope of this paper, but to demonstrate the utility of the proposed approach towards addressing such questions.

III. RESULTS

We evaluated the performance of the proposed IEKF with draws (IEKF-d) by applying it to two types of simulated datasets, hard-bounded random walks from (nearly) within the underlying model class and pseudo-deterministic patterns from outside the underlying model class. We assessed the performance by a set of statistical measures and compared the performance to that of a set of related, variant methods. We then applied the IEKF with draws to real EEG data of the sleep onset process.

A. SIMULATED DATA

The results for the random-walk simulations are given in Appendix G. Fig. 6 shows the IEKF-d estimates for a single simulation of three peaks on an exponential-decay background. Fig. 7 shows the distributions of the main statistical measures across the simulations for the eight variant filters. And Fig. 8 shows the distributions of the additional statistical measures, including the computation times. As may be expected when the simulations are essentially within the underlying model class, the filter performance is exceptional, particularly for filters utilizing draws, with almost universally correct determination of On/Off-peaks and completely white observation residuals.

Because the in-class scenario of these simulations is simpler, and because the near-perfect performance impairs finer comparisons of the statistical measures, we forgo further detail of the random-walk results and move to fully elaborate the results of the more challenging pseudo-deterministic simulations. The same patterns of relative filter performance are seen in the statistical measure distributions for both the random-walk and pseudo-deterministic simulations.

Fig. 1 shows the results for a single pseudo-deterministic simulation with three peaks on an exponential-decay background—the true simulated spectrogram Fig. 1(A), the spectrogram estimated from the IEKF with draws Fig. 1(B), the residuals Fig. 1(C), and the true and estimated On-peaks Fig. 1(D). Qualitatively, we see the excellent agreement between the true and estimated spectrograms, as well as the overall whiteness of the residuals. The agreement between the true and estimated spectrograms requires, not simply accurate estimation of the peak parameters, but correct determination of the combination of On/Off-peaks, which is indeed seen in Fig. 1(D). The true and estimated peak parameters are shown in Figs. 1(H)–1(G), where the estimates (dark colored lines) track the true values (solid black lines) very well, with the true states almost always within the 95%-confidence interval (colored regions). The points of discrepancy primarily occur when a peak is off (dashed black lines) and, as would be expected, the estimates revert toward the center of the bounds while the uncertainties grow. When the peak turns back on, the estimate and its uncertainty collapse to the true value.

Fig. 2 shows the distributions of the statistics computed for the variant methods applied to the sets of 1000 pseudo-deterministic simulations with 1–5 peaks on an exponential background. In general, we find the IEKF with draws produces the best fits by each statistic of all the methods tested. We find that the draws are more important than the iterations at improving the quality of the estimation, with all methods utilizing draws producing smaller, whiter errors than those without. The iterations do improve the estimation, even for methods utilizing draws, though less dramatically, so due to the added computation time, in applications where time is critical, the EKF with draws may be preferable. The methods utilizing 2nd-derivatives actually showed mixed improvements/degradations in performance.

The distributions of the mean residuals are shown in Fig. 2(A.i). Nearly all methods appear zero-mean, with the exception of the basic EKF and EKF-2o, which appear to be negatively biased, likely due to the modeling discrepancy of the 0.9 factor in the state transition F and the absence of compensatory draws or iterations. In Fig. 2(A.ii), the distributions of the MSEs reveal the draws of the reference trajectory greatly improve the quality of the estimates. Each filter with draws shows MSE distributions shifted towards zero and with smaller variance relative to the analogous filter without draws, i.e., MSEs of the IEKF with draws are consistently smaller than those of the IEKF without draws. Comparison of the MSE distributions for filters without and

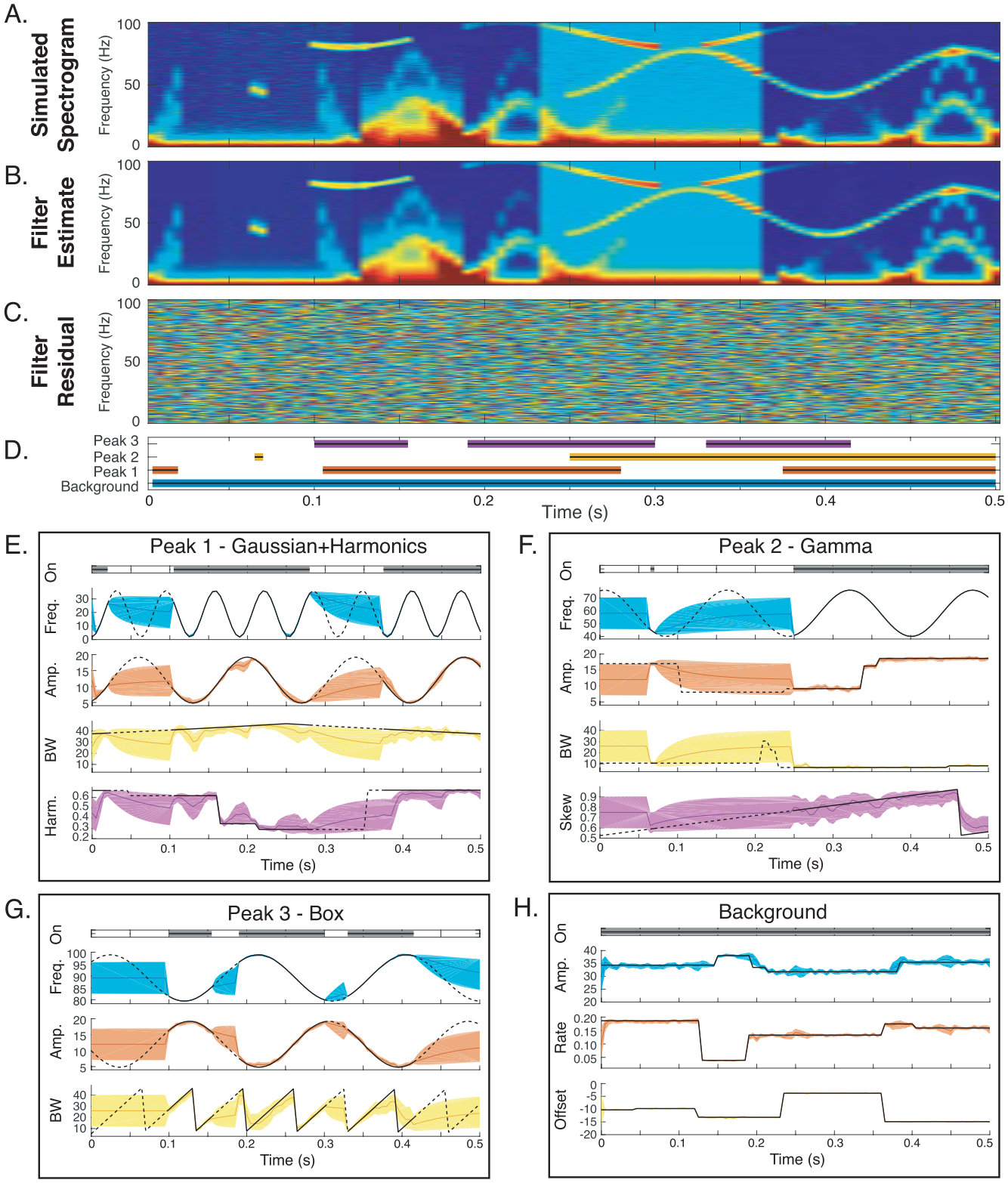

FIGURE 1. Pseudo-deterministic example. (A) Simulated spectrogram of three peaks—one Gaussian with harmonics, one gamma, and one exponential-box—on an exponential-decay background, with peak parameters following pseudo-deterministic patterns. (B) Filter estimate of the spectrogram obtained from IEKF-d. (C) Residual spectrogram, i.e., the difference between the spectrogram and the filter estimate. (D) Indicators of true On-peaks (black lines) and estimated On-peaks (colored lines). (E)–(H) IEKF-d filter state estimates of peak parameters. True parameter values are shown in black (solid when peak is On, dashed when peak is Off). Filter estimates and the 95%-confidence intervals are indicated by the colored lines and regions.

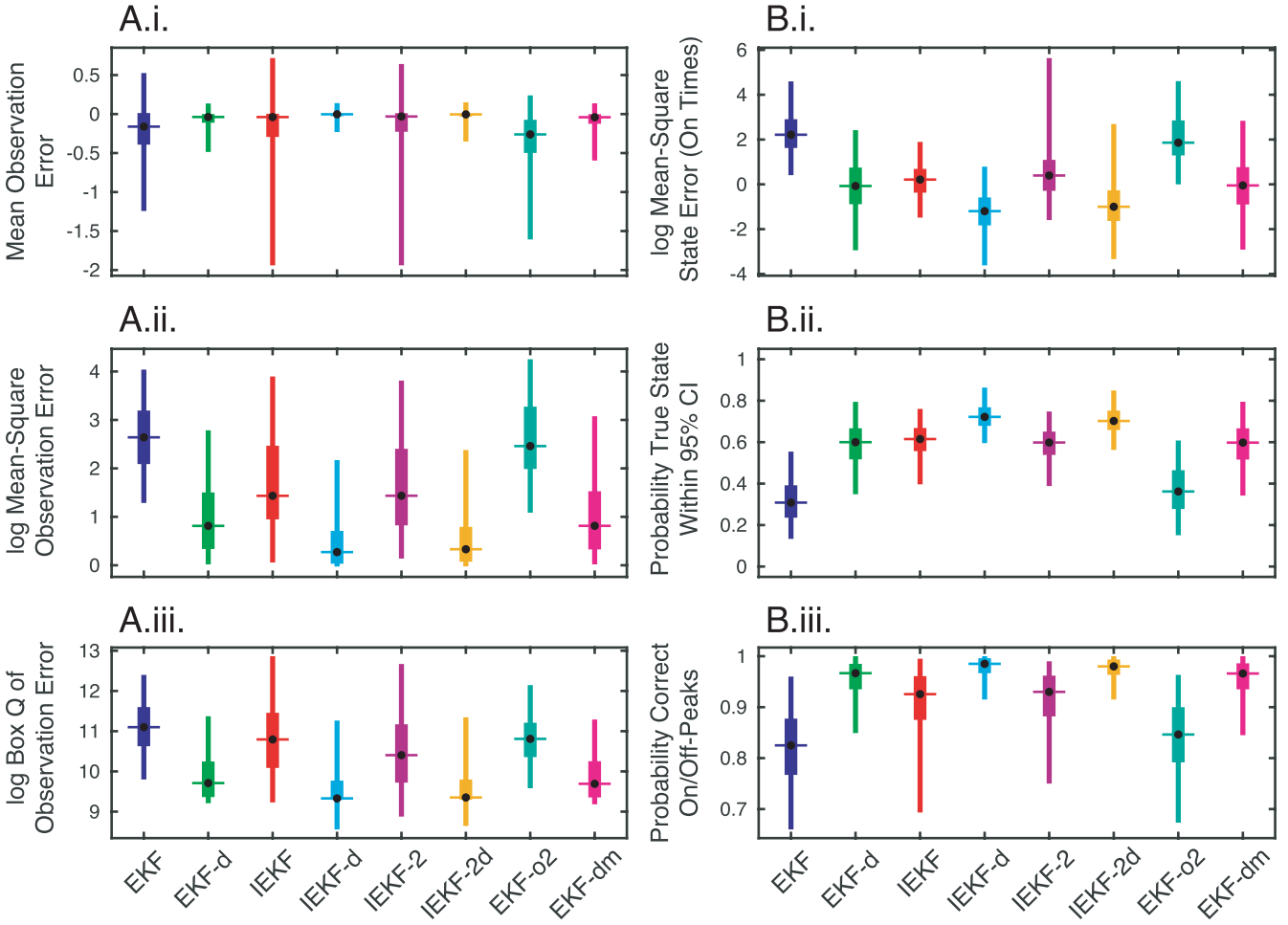

FIGURE 2. Statistical measures of filter performances when applied to 5000 simulated spectrograms of 1–5 peaks on an exponential-decay background with parameters varying pseudo-deterministically. EKF (dark blue), EKF-d (green), IEKF (red), IEKF-d (light blue), IEKF-2 (purple), IEKF-2d (orange), EKF-2o (teal), IEKF-dm (magenta). (A.i) Mean (over time and frequency) of the filter residuals. Most of the filters appear zero-mean, as expected. The exceptions are the EKF and EKF-2o, which do not utilize draws or iterations and show a slight negative bias, likely due to the "model mis-specification" of the 0.9 state transition factor. It is also clear the draws greatly reduce the variance of the filter residuals. (A.ii) Logarithm of the mean (over time and frequency) of the square filter residuals. Again, the draws alone produce a greater reduction in residual variance than the iterations alone (e.g., EKF-d vs. IEKF), but the iterations do add further improvement (e.g., IEKF-d vs. EKF-d). This same pattern of relative performance is also apparent in the other measures. (A.iii) Logarithm of the two-dimensional Box Q statistic of the filter residuals. (B.i) Logarithm of the mean (over time and peak parameters) of the filter state errors, restricted to times when the respective peaks are On. (B.ii) Probability of the true state values lying within the 95%-confidence interval of the filter state estimates. (B.iii) Probability of correctly determined On/Off-status of all peaks at all times.

with iterations—i.e., the EKF with draws and the IEKF with draws, respectively—show the iterations also improve the quality of the estimates, though to a lesser extent than the draws. Further comparisons of EKF with EKF-2o and IEKF with IEKF-2 show the 2nd-derivatives do slightly improve the quality of the estimates without draws, but show negligible improvement over the use of the draws, i.e., IEKF-d vs. IEKF-2d. The same patterns are seen when the maximum absolute value is used as the statistic, shown in Fig. 5(A) in Appendix G.

The distributions of the Box Q statistic, shown in Fig. 2(A.iii), capture two-dimensional correlations in the residuals, a quantitative indicator of non-whiteness. As with the MSE, the Box Q indicates that the draws substantially improve the quality of the estimates, with the distributions being shifted towards zero and of lower variance, while the iterations provide some, but less, improvement. Again, it appears the use of 2nd-derivatives may offer some improvement in the whiteness of the residuals when draws are not involved. The same patterns are seen when instantaneous correlations or Moran's I are used as the statistic, shown in Fig. 5(B) and 5(C), respectively, in Appendix G.

A similar pattern of improvements is seen in the statistical measures of the quality of the state estimates, the mean square state error Fig. 2(B.i), the probability of the true state falling within the 95% confidence estimate Fig. 2(B.ii), and the probability of estimating the On/Off-status of the peaks Fig. 2(B.iii). The slight exception is that the use of second derivatives in the IEKF, i.e., IEKF-2 and IEKF-2d, appears to slightly worsen the state estimates. The same pattern is

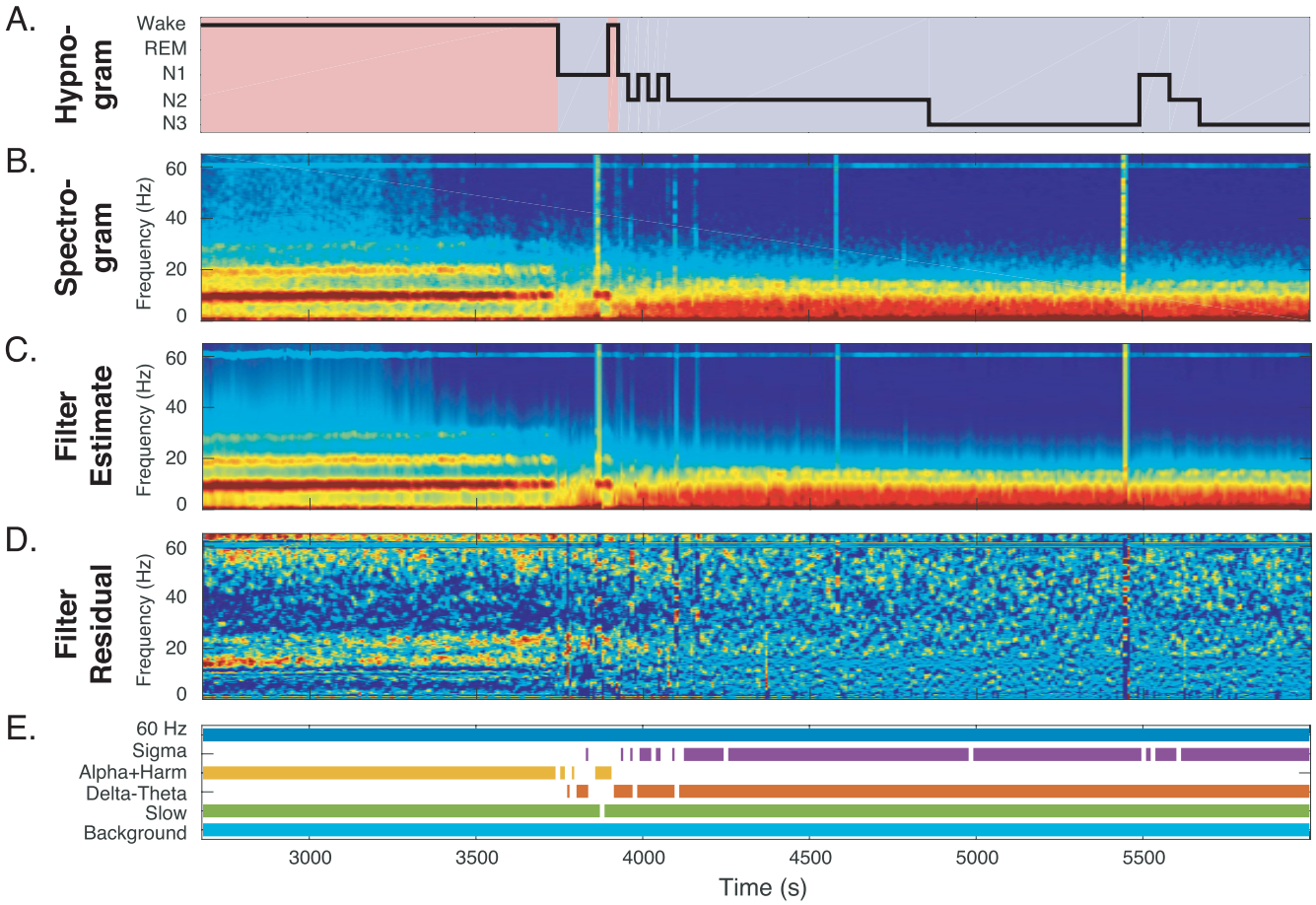

FIGURE 3. Sleep onset process. (A) and (B) Hypnogram and multitaper spectrogram, respectively, of occipital EEG recorded during the transition from wake to sleep. The eyes-closed wake-alpha peak (∼10 Hz) and its two harmonics are initially present before disappearing at around 3750 seconds at the transition to sleep. At that point the slow (< 1.5 Hz) and delta-theta (1.5–8 Hz) power increase, indicative of NREM sleep. The transient increases of sigma (∼14 Hz) power, suggestive of sleep spindles, also occur for the remainder of the spectrogram time. (C) Filter estimate of the spectrogram obtained from IEKF-d, using a model with an exponential-decay background, a gamma slow peak, a gamma delta-theta peak, a Gaussian-with-two-harmonics alpha peak, and a box 60 Hz peak. (D) Residual spectrogram, i.e., the difference between the spectrogram and the filter estimate. (E) Indicators of estimated On-peaks.

seen when the accuracy of the On/Off-combo determination is assessed by Cohen's kappa, which is shown in Fig. 5(D) in Appendix G.

The distributions of computation time are shown in Fig. 5(E) in Appendix G, where the draws are seen to be somewhat computationally more expensive than the iterations. Moreover, the evaluation of draws and selection of reference trajectory based on the filter posterior mode (EKF-dm) as opposed to the maximum likelihood (EKF-d), shows nearly identical performance and computation time, suggesting the filter posterior mode may be an equivalent and more principled means of selecting the reference trajectory.

B. EEG DATA OF SLEEP ONSET

Fig. 3 shows the results of application of the IEKF-d to occipital EEG recorded during the sleep onset process. Figs. 3(A) and 3(B) shows the hyponogram (scored sleep stages) and multitaper spectrogram of EEG data recorded during the sleep onset process, respectively. The transition from wake to NREM is very apparent in the spectrogram, as the alpha (8–12 Hz) oscillation, related to eyes-closed wake, is initially present and disappears and the slow and delta-theta (< 8 Hz) peaks, which are the hallmark of NREM sleep, strengthen in power. This transition is also indicated by the scored stages. Fig. 3(C) shows the estimated spectrogram obtained from the IEKF with draws, Fig. 3(D) shows the residuals, and Fig. 3(E) shows the estimated On-peaks. Qualitatively, the estimated spectrogram is in excellent agreement and the residuals appear white, while the estimated combos are consistent with the scored stages.

The parameter estimates of each peak are shown in Fig. 4. These estimates further indicate the appropriateness of the decomposition of the spectrogram into the relevant peak oscillations, with the estimated alpha amplitude clearly decreasing before the peak disappears at sleep onset, when the delta-theta peak appears and its amplitude begins increasing. These distinct spectral power changes would be conflated in traditional bandpass analysis of alpha amplitude, as the

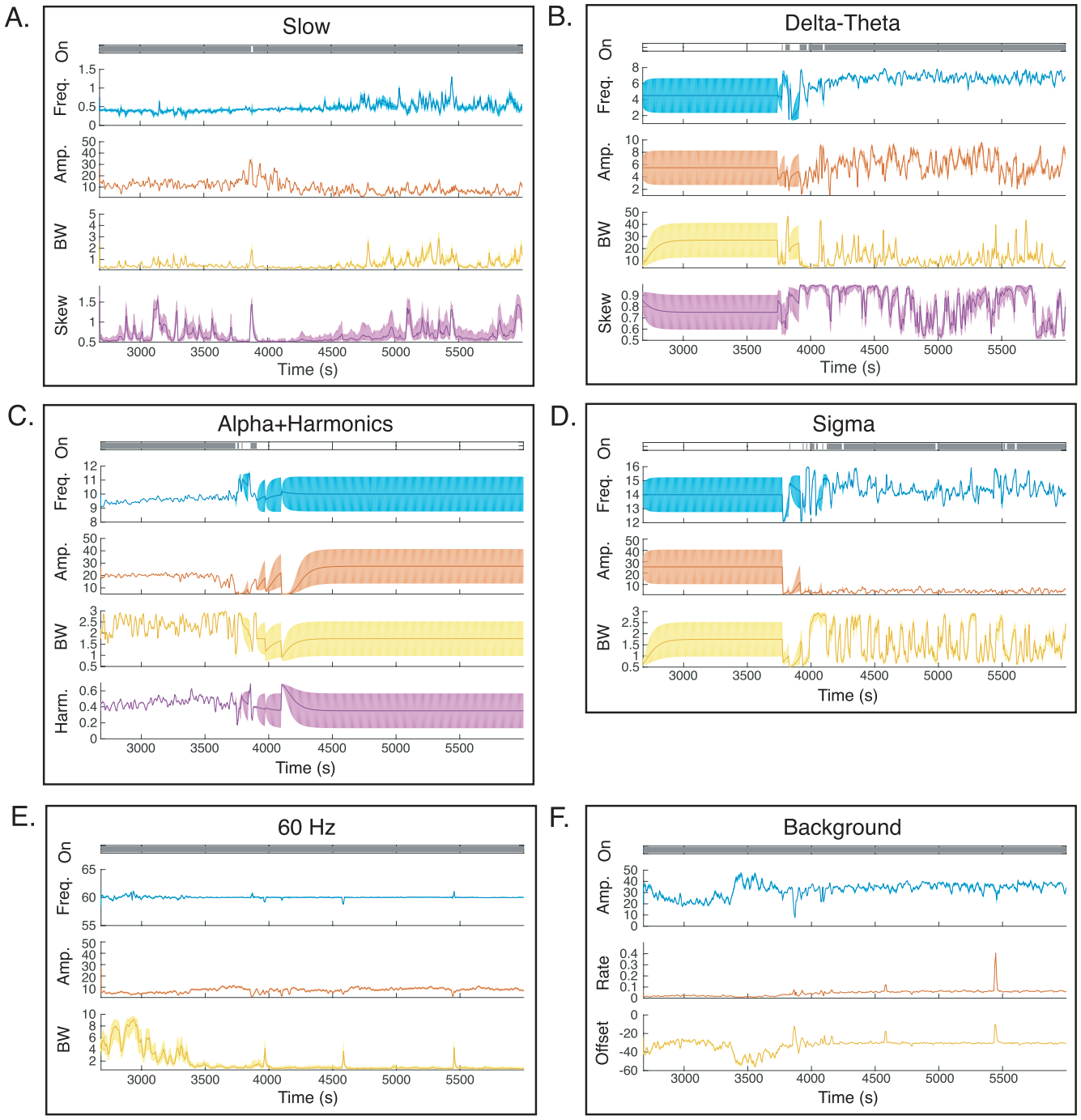

FIGURE 4. Sleep onset process—estimated parameters. (A) Slow gamma peak. (B) Delta-theta gamma peak. (C) Alpha Gaussian peak with two harmonics. (C) Sigma Gaussian peak without harmonics. (D) 60 Hz box peak. (E) Exponential-decay background. Filter estimates and the 95%-confidence intervals are indicated by the colored lines and regions. Grey bars indicate when each peak was estimated to be On.

increasing delta-theta power leaks into the alpha band following sleep onset. Our proposed method not only produces "cleaner" estimates, but, crucially, the direct interpretability of the parameter estimates allows separation of overlapping peaks.

For comparison, the results of applying the basic EKF to the same data are shown in Fig. S1 in the Supplemental. The performance is adequate but does show a few points of impaired tracking, particularly near the quick transitions from ∼3800 s to ∼4200 s, and an extended period of large error, where the filter was unable to recover following the artifact at ∼5500 s.

Spectrograms and IEKF-d filter estimates for three additional subjects are shown in Fig. S2 in the Supplemental.

These estimates were obtained using the same model and hyperparameters as used for the main subject above. Despite the model and hyperparameters not being tuned individually to the additional subjects, the estimates fit the observed spectrograms fairly well.

However, we must stress there is a crucial difference between obtaining passable spectrogram fits and obtaining decompositions with appropriately interpretable parameter estimates. The same model flexibility that enables quality fits can produce estimates with ''incorrect'' peaks, usually due to compensatory changes in parameters. For example, fitting a neighboring alpha peak instead of an actual sigma, or using a large, broad delta-theta to make up for a poor background. Such errors are not present in the examples here, and the state estimates are of reasonable interpretability. But in general, judicious setting of hyperparameters, particularly the parameter bounds, is vital to obtaining correctly interpretable results.

In fact, due to inter-subject heterogeneity in frequency ranges, prominence, and even number of oscillations, one should not expect the hyperparameter settings (nor even the peak functions themselves) determined for one individual to be appropriate for another. Similarly, intra-subject spatial variation means one should not even expect the model to be applicable to other channels from the same individual. The main example subject was chosen because their observed peaks adhere closely to the standard definitions. The additional subjects of the Supplemental show some deviations in peak locations, particularly those of alpha and sigma, but not substantial ones. In this case, the model and hyperparameters are close enough that the fits are acceptable and the interpretability of the decompositions are reasonable as well. But in general, the hyperparameters, and even the model itself, should be uniquely optimized for each dataset.

The full analyses of this variability is, indeed, the motivation and long-term objective of our future work, but is well beyond the immediate scope of this presentation. Here, we have stressed the criticality of these settings in obtaining interpretable spectrogram decompositions and interpretable peak parameter estimates.

IV. DISCUSSION

We have proposed and demonstrated a method to decompose an EEG spectrogram into a set of traditionally recognized oscillatory peaks that vary with time. The parameters of the peak functions form the continuous-valued state of a state-space model, while the combination of On/Off-peaks form a discrete-switching state. The model is estimated using a modified form of the IEKF that draws the initial reference trajectory and at each time step determines the discrete On/Off-combo.

The simulation results show the method tracks both the peak parameters and the On/Off-combo exceptionally well, both within and without the model class. The resulting spectrogram estimates show white residuals with little two-dimensional correlations. Comparisons with variant methods reveal that the step of drawing the initial trajectory for linearization is most responsible for the improvement in the quality of the estimates over a basic EKF. The iterations of the IEKF do add further improvement. The second order EKF did generally improve the estimates over the basic EKF, but less so than the draws or iterations. The use of the second derivative in the iterations of the IEKF, slightly improved the quality of the observation estimates, but at the expense of slightly worse state estimates, and there is no noticeable improvement when the draws are utilized.

Application of the method to an EEG spectrogram of the sleep onset process found great agreement between the observed and estimated spectrograms and estimated On/Off-peak combos consistent with the transition from wake to NREM sleep.

A. ADVANTAGES OVER OTHER APPROACHES TO EEG SPECTRAL ANALYSIS

Our method offers several advantages over standard EEG analyses and other previously proposed methods. The state-space estimation captures temporal changes lost in time-averaged spectral estimates and, to some extent, bandpassed amplitudes. Moreover, the temporal continuity imposed, not present in windowed methods [9], [40], allows for smoother estimates of time-varying properties. Our expanded state-space model even includes discrete On/Off-switching of spectral peaks, a feature not captured by other state-space approaches [43]–[47]. The functional forms of the peaks used in our method result in estimates that more closely match the shapes of the spectral peaks observed in the EEG than sinusoidal or AR(MA) model estimates [40], [42]–[46].

Crucially, our peak functions are parameterized in terms of variables directly interpretable as peak properties, such as location, amplitude, bandwidth, and other shape parameters. Coupled with the temporal continuity of the state-space, this means observed changes in spectral power are interpretable as well. This is particularly valuable in frequency ranges with overlapping bands. Whereas traditional average power and bandpassed amplitude estimates require a priori frequency specification and cannot distinguish oscillatory power from nearby oscillation leakage, our method allows separability of adjacent peaks.

The resulting representation of the spectrogram is both parsimonious and meaningful. Hundreds of frequency bin estimates per time point covering the Nyquist range are captured by a couple dozen directly interpretable state variables. Most importantly, our approach enables quantitative analyses of the peak parameters and their dynamics. The joint uncertainty estimates of the state variables obtained by the Kalman filter allow inference over any peak parameters or related quantities. Estimation of the state transition matrices is of particular interest and, as discussed below, an aim of future work.

More broadly, our approach offers a means towards more fully capturing the variability of locations and shapes of oscillatory peaks in sleep EEG spectrograms. Beyond the

spatio-temporal dynamics, there is wide heterogeneity, across subjects, in frequency location, and even number, of spectral peaks. Only recently have analyses begun exploration of this heterogeneity, and it is rarely addressed in analyses of sleep dynamics and comparisons of clinical groups. Application of the same peak definitions (i.e., frequency ranges) and other criteria across subjects in such analyses can introduce confounds into the estimates and undermine comparisons. Our approach, on the other hand, possesses an inherent flexibility to analyze each subject and channel individually through the setting of parameter bounds and other hyperparameters, and even through selection of the number and type of peaks in the model.

Fundamentally, the innovation of this work is the overall structure of the approach. The specific model choice and hyperparameter settings can, of course, be adapted to each subject and channel. New parameterized peak functions can be utilized, if needed. And any of the filter variants, or even more advanced alternatives, can be employed for estimation, depending on computational constraints. But the general approach—of forming a low-dimensional, interpretably parameterized representation of the spectrogram (or other high-dimensional observations) with a state-space model and estimating the model with a slightly modified EKF/IEKF—provides a foundation for future sleep EEG analyses and analogous applications in other fields.

B. ESTIMATION CHALLENGES AND IMPLEMENTATION CHOICES

Several aspects of the proposed approach were required to overcome unusual, challenging circumstances posed by the problem. The primary difficulty is that faced by the direct particle filter of [47]. Because of the large dimension of the observations, using the likelihood for reweighting of prediction samples results in immediate collapse to a single particle. This challenge is similar to that faced in meteorological and geoscience applications [51] with very large dimensional states and observations. In such applications, the state dynamics are typically complex and nonlinear, and the finer details of the state density estimates of critical import. In our present application, the state dynamics of our current model are simple, linear, and the precise characterization of the state density are not of immediate necessity. By forgoing a precise density, allowing some ''extra uncertainty'' in a sense, and simply using the IEKF with sampled reference trajectory and Gaussian assumption, we are able to not only track the states, but still obtain very precise estimates. More sophisticated ensembling, sampling, or reweighting schemes like those discussed in [51], [52], or [53] may be workable, possibly offering more robust estimates, but likely at the cost of increased computation time.

In real data applications, we have found initialization and setting of hyperparameters, particularly the peak parameter bounds, are of great importance. We determined the settings used here by trial and error, though, in principle, they could be estimated by EM algorithm or sampling methods. These settings were determined specifically for the main example subject of Figs. 3 and 4 but applied to all subjects of the dataset, including the additional examples shown in Fig. S2 of the Supplemental, for which the estimated fits to the spectrograms appear decent. We re-iterate, however, that fitting a spectrogram is not equivalent to obtaining a reasonable decomposition, and the selection of an appropriate model and its hyperparameters is crucial to obtaining the desired decomposition with reasonably interpretable parameter estimates. One should not expect the same model or hyperparameters to be applicable across subjects or different channels from the same subject, due to inter-subject heterogeneity and intra-subject spatial variability. The full exploration of this variation across subjects and across clinical groups is the motivating objective of the current presentation but is far beyond the scope of this work and the subject of future analyses.

C. FUTURE IMPROVEMENTS, EXTENSIONS, AND APPLICATIONS

Moving forward, the ultimate aim is, indeed, the exploration and analysis of sleep dynamics and its variations across subjects and clinical groups through application of the approach to full-night EEGs of larger data sets, i.e., identification and estimation of individual subject models and comparison of the estimated states and governing dynamics. Identification of hyperparameters will be critical in such endeavors, and the imminent steps will focus on implementing methods (e.g., an EM algorithm) to systematically determine appropriate values. Computation time is currently somewhat prohibitive for such estimations, especially for a full night of sleep EEG. Though we have made some efforts to reduce the computation time, there are likely still opportunities for further improvement, such as optimizing the parallelization and implementing the filter in the information form. Alternatively, finding a workable sampling or ensemble scheme, may also enable simultaneous estimation of hyperparameters.

If the computation time is sufficiently improved, other extensions and applications may become more practical. For example, it may be possible to include time-varying state and observation noise covariances in the model and estimate the covariances adaptively. Or real-time processing and artifact handling may be viable.

The approach we have proposed allows tracking and estimation of a set of time-varying, parameterized peaks. While the motivating application is to sleep EEG spectrograms, and the specific parameterized observation functions were chosen to capture the peak properties of primary interest in the tracking of sleep dynamics, the approach may be useful for other EEG applications (e.g., cross-channel coherograms) or other properties of interest (e.g., phase or cross-frequency coupling) through appropriately defined observation functions. Similarly, other applications where high-dimensional observations comprise a set of possible functions with varying parameters of interest (e.g., spectrograms from

MR spectroscopy or audio or speech signal processing) may also be possible with properly chosen observation functions.

Sleep is a complex, dynamic process, and the underlying neural activity presents in the EEG as characteristic oscillations that change over the course of the night. Our approach provides a sparse representation of these spectral dynamics and offers the potential to further model and analyze of the governing dynamics. Future work will focus on estimation of the parameter and combo transition matrices and the analysis of the variability of the dynamics across nights, across individuals, and across different patient groups. Such analyses would not only offer a better understanding of the sleep process but possibly reveal diagnostic markers of pathologies as well.

APPENDIX A
PEAK FUNCTIONS AND THEIR DERIVATIVES
GAUSSIAN PEAK WITH HARMONICS

The spectral peaks of many classically recognized oscillations can be represented by a simple Gaussian,

$$h^{(Gauss)}(\omega; F, A, B) = A \exp\left(\frac{-(\omega - F)^2}{2B}\right), \tag{9}$$

with parameters for central frequency F, maximum amplitude A, and bandwidth B.

Some oscillations, such as eyes-closed occipital alpha, will exhibit harmonics in the spectral domain. We represent this phenomenon by forming the peak at the fundamental frequency and adding similar peaks of decreasing power at subsequent harmonic frequencies.

For the case of a Gaussian peak, the fundamental and its harmonics are of the form

$$h_n^{(Gauss)} = \beta^n A \exp\left(\frac{-(\omega - (n+1)F)^2}{2B}\right), \tag{10}$$

where the power decrease between subsequent harmonics is a multiplicative fraction $\beta \in [0, 1]$, such that β^n is the power of nth harmonic peak as a proportion of the power of the fundamental. With N harmonics and $n = 0$ the fundamental, the overall peak function is

$$h^{(Gauss)}(\omega; F, A, B, \beta) = \sum_{n=0}^{N} h_n^{(Gauss)}. \tag{11}$$

The partial derivatives of the Gaussian peak with harmonics w.r.t the parameters are

$$\frac{\partial h^{(Gauss)}}{\partial F} = \sum_{n=0}^{N} h_n^{(Gauss)} \left[\frac{(n+1)(\omega - (n+1)F)}{B}\right] \tag{12}$$

$$\frac{\partial h^{(Gauss)}}{\partial A} = h^{(Gauss)} \left[A^{-1}\right] \tag{13}$$

$$\frac{\partial h^{(Gauss)}}{\partial B} = \sum_{n=0}^{N} h_n^{(Gauss)} \left[\frac{(\omega - (n+1)F)^2}{2B^2}\right] \tag{14}$$

$$\frac{\partial h^{(Gauss)}}{\partial \beta} = \sum_{n=1}^{N} h_n^{(Gauss)} \left[\frac{n}{\beta}\right] \tag{15}$$

SHIFTED-GAMMA PEAK FUNCTION

Broadband peaks with asymmetric tails are observed in the EEG spectrum, e.g., the low-frequency slow and delta oscillations. Such peaks can be approximated in the form of a gamma distribution,

$$h^{(\Gamma)}(\omega; \alpha, \beta, O, A) = A\left(\frac{\beta(\omega - O)}{(\alpha - 1)}\right)^{\alpha - 1} \times \exp\{(-\beta(\omega - O) + (\alpha - 1)\},$$

where α and β are the shape and rate parameters of the gamma distribution, respectively, O is an additional frequency offset parameter to translate the peak away from the origin, A is the maximum amplitude, and the normalization is obtained by evaluating the gamma distribution function at the mode, $(\alpha - 1)/\beta$. To avoid numerical overflow errors, the function is computed as the log of the gamma distribution and then exponentiated.

The α, β, and O parameters are not directly interpretable in terms of the location and shape of the peak, but more meaningful ones can be computed from them. We take the frequency location F of the peak to be the location of the mode of the shifted gamma distribution, $F = (\alpha - 1)/\beta + O$, the bandwidth B to be the variance of the gamma distribution, $B = \alpha/\beta^2$, and the skewness S to be that of the gamma distribution, $S = 2/\sqrt{\alpha}$. Writing the α, β, and O parameters in terms of the more meaningful parameters,

$$\alpha = \frac{4}{S^2}, \beta = \frac{2}{S\sqrt{B}}, \quad \text{and } O = F - \frac{2\sqrt{B}}{S} + \frac{S\sqrt{B}}{2},$$

we obtain a directly interpretable parameterization of the shifted gamma peak $h^{(\Gamma)}(\omega; F, A, B, S)$.

For ease of computation and derivation of the derivatives, we implement the shifted gamma peak in terms of the original parameters, (α, β, A, and O). The partial derivatives of the peak function w.r.t. the parameters of interest (F, A, B, and S) are obtained by the chain rule

$$\begin{aligned}
\frac{\partial h^{(\Gamma)}}{\partial F} &= \frac{\partial h^{(\Gamma)}}{\partial O}\frac{\partial O}{\partial F} \\
&= h^{(\Gamma)}\left[\frac{-(\alpha - 1)}{\omega - O} + \beta\right] \\
\frac{\partial h^{(\Gamma)}}{\partial A} &= h^{(\Gamma)}\left[A^{-1}\right] \\
\frac{\partial h^{(\Gamma)}}{\partial B} &= \frac{\partial h^{(\Gamma)}}{\partial \beta}\frac{\partial \beta}{\partial B} + \frac{\partial h^{(\Gamma)}}{\partial O}\frac{\partial O}{\partial B} \\
&= h^{(\Gamma)}\left[\frac{(\omega - O)\beta^3}{2\alpha} - \frac{(\alpha - 1)\beta^2}{\alpha} + \frac{(\alpha - 1)^2\beta}{2\alpha(\omega - O)}\right] \\
\frac{\partial h^{(\Gamma)}}{\partial S} &= \frac{\partial h^{(\Gamma)}}{\partial \alpha}\frac{\partial \alpha}{\partial S} + \frac{\partial h^{(\Gamma)}}{\partial \beta}\frac{\partial \beta}{\partial S} + \frac{\partial h^{(\Gamma)}}{\partial O}\frac{\partial O}{\partial S} \\
&= h^{(\Gamma)}\left[-\alpha^{3/2}\log\left(\frac{\beta(\omega - O)}{\alpha - 1}\right) + \sqrt{\alpha}\right.
\end{aligned}$$

$$+\frac{\beta\sqrt{\alpha}\,(\omega-O)}{2}-\frac{\sqrt{\alpha}\left(\alpha^2-1\right)}{2\beta\,(\omega-O)}\Bigg]$$

BOX-EXPONENTIAL PEAK FUNCTION

Increasing the order of the quadratic term in the exponential of the Gaussian peak produces a more rectangular, box-shaped peak,

$$h^{(box)}\,(\omega;F,A,B,P)=A\exp\left(\frac{-(\omega-F)^P}{2B}\right), \quad (16)$$

where, as for the Gaussian, the parameters F, A, and B, represent the central frequency, maximum amplitude, and bandwidth, respectively. The order P is even and determines the sharpness of the rectangularity of the peak. It is fixed and not estimated. This peak closely approximates the spectral peaks formed of machine-generated sources, such as 60 Hz line noise. In the models of 60 Hz noise, we set $P=6$. The partial derivatives w.r.t to the parameters are

$$\frac{\partial h^{(box)}}{\partial F}=h^{(box)}\left[\frac{P\,(\omega-F)^{(P-1)}}{2B}\right] \quad (17)$$

$$\frac{\partial h^{(box)}}{\partial A}=h^{(box)}\left[A^{-1}\right] \quad (18)$$

$$\frac{\partial h^{(box)}}{\partial B}=h^{(box)}\left[\frac{(\omega-F)^P}{2B^2}\right] \quad (19)$$

EXPONENTIAL-DECAY FUNCTION

We model background spectral activity using an exponential-decay function,

$$h^{(decay)}\,(\omega;a,r,o)=a\,(1-r)^{\omega}+o,$$

where the parameters a, r, and o represent the power scaling, decay rate, and baseline offset, respectively. The partial derivatives w.r.t. the parameters are

$$\frac{\partial h^{(decay)}}{\partial a}=(1-r)^{\omega}$$

$$\frac{\partial h^{(decay)}}{\partial r}=-\omega a\,(1-r)^{\omega-1}$$

$$\frac{\partial h^{(decay)}}{\partial o}=1$$

APPENDIX B
LINK FUNCTIONS AND THEIR DERIVATIVES

Some of the peak parameters are mathematically restricted to a certain range of values, e.g., greater than zero. Additionally, it often improves estimation, enhancing the consistency and interpretability of the peaks, to place physiologically-principled bounds on some parameters. We use link functions to impose these restrictions.

ONE-SIDED EXPONENTIAL FUNCTION

A one-sided exponential link function,

$$l_{exp}(x;s,o)=s\exp(x)+o, \quad (20)$$

Algorithm 5: comboTransitionMatrix

Data: $\mathcal{I}$, $\mathcal{J}$, p_{on}, p_{off}, p_{stay}
Result: Φ
for $i\in\mathcal{J}$ **do**
  for $j\neq i\in\mathcal{J}$ **do**
    $\Phi_{j,i}\leftarrow 1$
    for $k\in\mathcal{I}$ **do**
      if $k\in\mathcal{I}(j)$ & $k\notin\mathcal{I}(i)$ **then**
        $\Phi_{j,i}\leftarrow p_{\text{on}}\cdot\Phi_{j,i}$
      else if $k\notin\mathcal{I}(j)$ & $k\notin\mathcal{I}(i)$ **then**
        $\Phi_{j,i}\leftarrow(1-p_{\text{on}})\cdot\Phi_{j,i}$
      else if $k\notin\mathcal{I}(j)$ & $k\in\mathcal{I}(i)$ **then**
        $\Phi_{j,i}\leftarrow p_{\text{off}}\cdot\Phi_{j,i}$
      else
        $\Phi_{j,i}\leftarrow(1-p_{\text{off}})\cdot\Phi_{j,i}$
      end
    end
  end
  $\Phi_{\cdot,i}\leftarrow\left(1-p_{\text{stay}}\right)\cdot\Phi_{\cdot,i}/\sum_{j\neq i}\Phi_{j,i}$
  $\Phi_{i,i}\leftarrow p_{\text{stay}}$
end

Algorithm 6: referenceTrajectory—Max Filter Posterior

Data: $\hat{x}_{t|t-1}$, $P_{t|t-1}$, N_d, j, y_t, h^j, R
Result: $\eta^j_{t,1}$
$d^{j,1}\leftarrow\hat{x}_{t|t-1}$
$\varepsilon^{j,1}\leftarrow y_t-h^j\left(d^{j,1}\right)$
for $i\leftarrow 2$ **to** N_d **do**
  $w^{j,i}\sim\mathcal{N}\left(0,P_{t|t-1}\right)$
  $\overline{w}^{j,i}\leftarrow$ only retain components relevant to j
  $d^{j,i}\leftarrow\hat{x}_{t|t-1}+\overline{w}^{j,i}$
  $\varepsilon^{j,i}\leftarrow y_t-h^j\left(d^{j,i}\right)$
end
$i^*\leftarrow\arg\min_i\left(\varepsilon^{j,i}\right)^T\mathrm{R}^{-1}\varepsilon^{j,i}+\left(\overline{w}^{j,i}\right)^T P^{-1}_{t|t-1}\overline{w}^{j,i}$
$\eta^j_{t,1}\leftarrow d^{j,i^*}$

restricts the value of a variable x to be strictly greater than or less than a given offset value o, where the sign $s\in\{-1,1\}$ determines the direction of the restriction. The first and second derivatives are given by

$$\frac{dl_{exp}}{dx}=\frac{d^2l_{exp}}{dx^2}=s\exp(x). \quad (21)$$

BOUNDED SIGMOID FUNCTION

A sigmoid link function,

$$l_{sig}(x;a,b)=(b-a)\,\frac{1}{1+\exp(-x)}+a, \quad (22)$$

restricts the value of a variable x between upper and lower asymptotic bounds, b and a, respectively. Denoting the standard sigmoid function by $\sigma(x)=(1+\exp(-x))^{-1}$, the first

Algorithm 7: filterUpdate—IEKF With 2nd Derivative

Data: $\hat{x}_{t|t-1}$, $P_{t|t-1}$, $\eta^j_{t,1}$, N_{iter}, y_t, h^j, $\partial_x h^j$, $\partial_{xx} h^j$, R
Result: $\hat{x}^j_{t|t}$, $P^j_{t|t}$, LL^j_t
For ease of notation, $\eta_k \doteq \eta^j_{t,k}$ and $P \doteq P_{t|t-1}$
for $k \leftarrow 1$ **to** N_{iter} **do**
  $M \leftarrow \partial_x h^j(\eta_k)$
  $\partial_x \phi \leftarrow M^T \text{R}^{-1}\left[y_t - h^j(\eta_k)\right] - P^{-1}\left[\eta_k - \hat{x}_{t|t-1}\right]$
  $\partial_{xx}\phi \leftarrow -P^{-1} - M^T \text{R}^{-1} M + \sum_i \left[\partial_{xx} h^j(\eta_k)\right]_i$
  $z \leftarrow y_t - h^j(\eta_k) - M\left[\hat{x}_{t|t-1} - \eta_k\right]$
  $\eta_{k+1} \leftarrow \eta_k - [\partial_{xx}\phi]^{-1}\partial\phi$
end
$\text{LL}^j_t \leftarrow -\frac{1}{2} z^T \left[MPM^T + \text{R}\right]^{-1} z$
$\quad - \frac{1}{2}\ln\det(MPM^T + \text{R}) - \frac{N_y}{2}\ln(2\pi)$
$\hat{x}^j_{t|t} \leftarrow \eta_{N_{\text{iter}}+1}$
$M \leftarrow \partial_x h^j(\eta_{\text{iter}+1})$
$\partial_{xx}\phi \leftarrow -P^{-1} - M^T \text{R}^{-1} M + \sum_i \left[\partial_{xx} h^j(\eta_{\text{iter}+1})\right]_i$
$P^j_{t|t} \leftarrow -[\partial_{xx}\phi]^{-1}$

Algorithm 8: filterUpdate—2nd Order EKF

Data: $\hat{x}_{t|t-1}$, $P_{t|t-1}$, y_t, h^j, $\partial_x h^j$, $\partial_{xx} h^j$, R
Result: $\hat{x}^j_{t|t}$, $P^j_{t|t}$, LL^j_t
For ease of notation, $\eta \doteq \hat{x}_{t|t-1}$ and $P \doteq P_{t|t-1}$
$M \leftarrow \partial_x h^j(\eta)$
$\Theta \doteq [\Omega_m]$, a $N_y \times 1$ vector, where
$\Theta_m \leftarrow \sum_{k,l}^{N_x} P_{k,l}\left[\partial_{x_k x_l} h^j_m(\eta)\right]$
$\Omega \doteq \left[\Omega_{m,n}\right]$, a $N_y \times N_y$ matrix, where
$\Omega_{m,n} \leftarrow \sum_{k,l,p,q}^{N_x}\left[\partial_{x_k x_l} h^j_m(\eta)\right] P_{l,p} P_{k,q}\left[\partial_{x_p x_q} h^j_n(\eta)\right]$
$Y^G \leftarrow MP_{t|t-1}M^T + \text{R} + \frac{1}{2}\Omega$
$z \leftarrow y_t - h^j\left(\hat{x}_{t|t-1}\right) - \frac{1}{2}\Theta$
$\hat{x}^j_{t|t} \leftarrow \hat{x}_{t|t-1} + P_{t|t-1}M^T\left[Y^G\right]^{-1} z$
$P^j_{t|t} \leftarrow P_{t|t-1} - P_{t|t-1}M^T\left[Y^G\right]^{-1} MP_{t|t-1}$
$\text{LL}^j_t \leftarrow -\frac{1}{2} z^T\left[Y^G\right]^{-1} z - \frac{1}{2}\ln\det(Y^G) - \frac{N_y}{2}\ln(2\pi)$

derivative is

$$\frac{dl_{sig}}{dx} = (b-a)\,\sigma(x)\,(1-\sigma(x))\,, \tag{23}$$

while second derivative is

$$\frac{d^2 l_{sig}}{dx^2} = (b-a)\left[\sigma(x) - 3\sigma^2(x) + 2\sigma^3(x)\right]. \tag{24}$$

APPENDIX C
TRANSITION MATRIX FOR COMBINATION OF ON/Off-PEAKS

The transition matrix Φ for the On/Off-combo $j_t \in \mathcal{J}$ with probabilities α_t can be any viable discrete transition matrix. In principle, Φ could be estimated, but here we use a fixed value. It is reasonable for the combo transition probabilities to be related to the probabilities of individual peaks to turn on and off, p_{on} and p_{off}. Additionally, it is helpful to be able to specify the overall degree of "stickiness" in the combos p_{stay}. So the diagonal elements of Φ are set as p_{stay}, while the remaining probabilities are determined by p_{on} and p_{off}, but normalized to equal $1 - p_{\text{stay}}$. Pseudo-code is given below. For the filter estimates and the random walk simulations we use values $p_{\text{on}} = 0.2$, $p_{\text{off}} = 0.2$, and $p_{\text{stay}} = 0.9$.

TABLE 1. Model and parameter bounds used for sleep EEG analysis.

Peak	Type	Dyn.	Param.	Min	Max	Q	$\hat{x}_0$	P_0
Bkgrd.	Decay	0	a	0	50	1	a_0	0.5
			r	0	0.5	1	r_0	0.5
			o	-	-	1	o_0	0.5
Slow	Gamma	1	F	0	1.5	0.1	0	0.1
			A	1	50	0.1	0	0.1
			B	0.1	5	0.1	0	0.1
			S	0.5	1.9	0.1	1	0.1
δ-θ	Gamma	1	F	1	8	0.1	0	0.1
			A	1	10	0.1	0	0.1
			B	4	50	0.1	-3	0.1
			S	0.5	1	0.1	1	0.1
α	Gauss+2	1	F	8	12	0.1	0	0.1
			A	5	50	0.1	0	0.1
			B	0.5	3	0.1	0	0.1
			β	0	0.7	0.1	0	0.1
σ	Gauss	1	F	12	16	0.1	0	0.1
			A	1	50	0.1	0	0.1
			B	0.5	3	0.1	-3	0.1
60 Hz	Box(6)	0	F	55	65	0.1	0	0.1
			A	1	50	0.1	0	0.1
			B	0.5	10	0.1	0	0.1

TABLE 2. Available On/Off-peak combos, $\mathcal{J}$, and initial combo probabilities, $\hat{\alpha}_0$, used for sleep EEG analysis.

	Wake		N1		N2/N3	
Background	1	1	1	1	1	1
Slow	0	1	0	1	1	1
Delta-Theta	0	0	0	0	1	1
Alpha	1	1	0	0	0	0
Sigma	0	0	0	0	0	1
60 Hz	1	1	1	1	1	1
$\hat{\alpha}_0$	0	1	0	0	0	0

The pseudo-code is given below. For the filter estimates and the random walk simulations we use values $p_{\text{on}} = 0.2$, $p_{\text{off}} = 0.2$, and $p_{\text{stay}} = 0.9$.

APPENDIX D
ALTERNATIVE INITIAL REFERENCE TRAJECTORIES

Alg. 6 is the alternative initialization of the reference trajectory, used in the IEKF-dm, that selects the draw of maximum filter posterior density. For those filter variants without draws—the EKF, IEKF, and IEKF-2—the reference trajectory is initialized by the direct assignment of the prediction estimate. This is achieved in either Alg. 2 or 6 by setting $N_d = 1$.

APPENDIX E
ALTERNATIVE FILTER UPDATES

The general filtering algorithm of the proposed approach is given in Alg. 1 in Sec. II-F, along with the EKF/IEKF filter updates in Alg. 3. Two alternative filter updates were were also implemented and tested: an IEKF that treats the filter step as a mode optimization utilizing the 2nd-derivatives, with filter updates given in Alg. 7; and a 2nd-order EKF with Gaussian approximation, following [48], with filter updates given in Alg. 8.

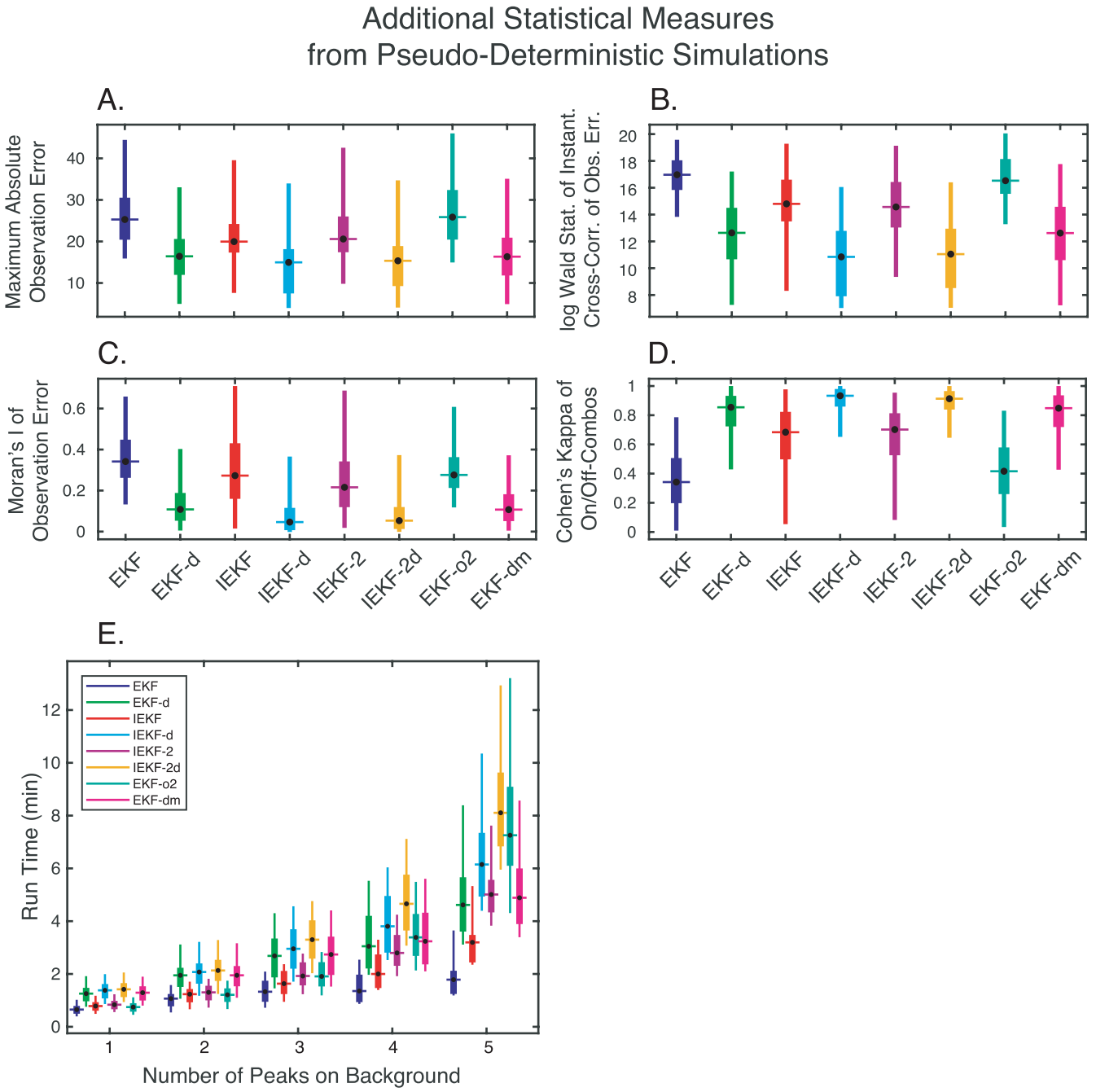

FIGURE 5. Additional statistical measures of filter performances when applied to 5000 simulated spectrograms of 1–5 peaks on an exponential-decay background with parameters varying pseudo-deterministically. EKF (dark blue), EKF-d (green), IEKF (red), IEKF-d (light blue), IEKF-2 (purple), IEKF-2d (orange), EKF-2o (teal), IEKF-dm (magenta). (A) Maximum (over time and frequency) of the absolute value of the filter residuals. This is an alternative to mean-square as a measure of the variability of the residuals. (B) Logarithm of the Wald statistic of the cross-frequency correlations of the filter residuals. (C) Moran's I statistic of the two-dimensional correlations of the filter residuals. The Wald and Moran's I statistics are alternatives to the Box Q as a measure of correlations in the residuals. (D) Cohen's kappa of estimates of the On/Off-combos. This is an alternative to the probability of correct On/Off-peak determination as a measure of the accurate identification of On/Off-peak combos. (E) Computation time of filter estimates.

APPENDIX F
MODEL, PARAMETERS, BOUNDS, AND INITIALIZATIONS FOR SLEEP EEG ANALYSIS

Table 1 details the model used for the analysis of real EEG data from sleep onset. This includes the peaks, their type, whether they are allowed to dynamically switch on and off (Dyn.), their parameters (Param.), the parameter bounds (Min and Max), the parameter state variances (Q), and the initial parameter expected values ($\hat{x}_0$) and variances (P_0). Table 2 lists the set of available On/Off-peak combinations, indicates their approximate sleep stage representation, and specifies the initial On/Off-combo probabilities ($\hat{\alpha}_0$).

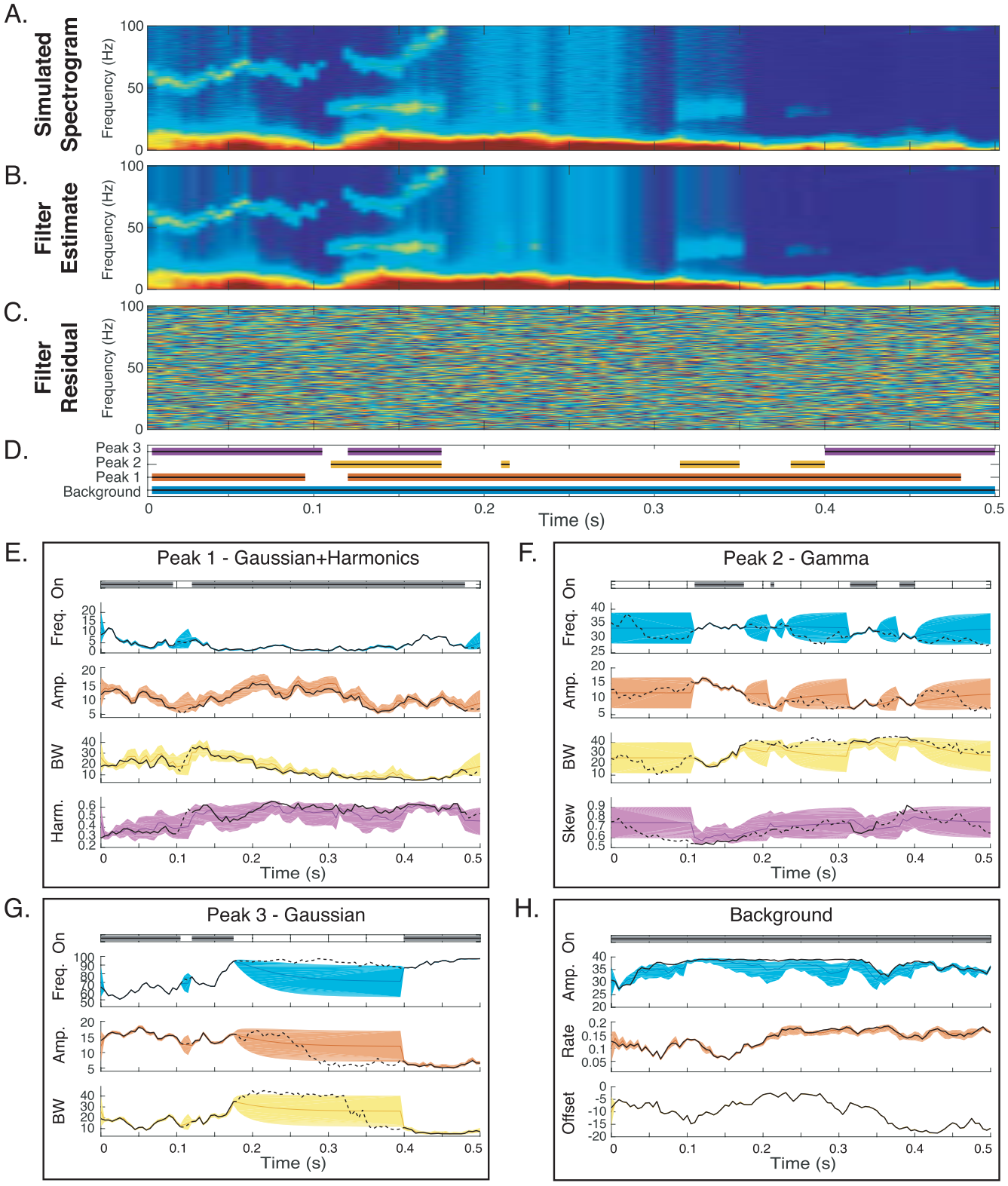

FIGURE 6. Random walk example. (A) Simulated spectrogram of three peaks—two Gaussian (one with harmonics) and one gamma—on an exponential-decay background, with peak parameters following hard-bound random walks. (B) Filter estimate of the spectrogram obtained from IEKF-d. (C) Residual spectrogram, i.e., the difference between the spectrogram and the filter estimate. (D) Indicators of true On-peaks (black lines) and estimated On-peaks (colored lines). (E)–(H) IEKF-d filter state estimates of peak parameters. True parameter values are shown in black (solid when peak is On, dashed when peak is Off). Filter estimates and the 95%-confidence intervals are indicated by the colored lines and regions.

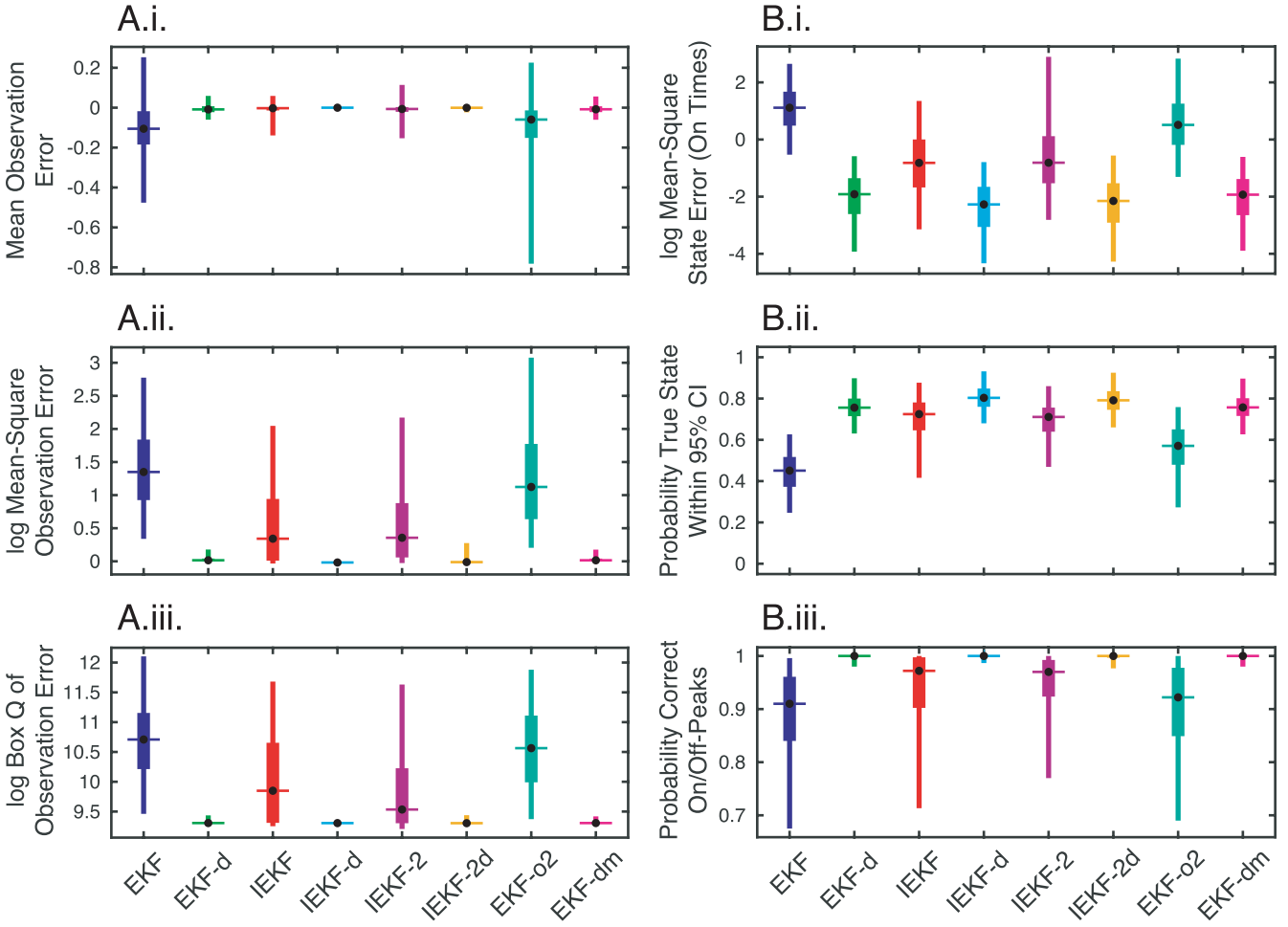

FIGURE 7. Statistical measures of filter performances when applied to 5000 simulated spectrograms of 1–5 peaks on an exponential-decay background with parameters following random walks. EKF (dark blue), EKF-d (green), IEKF (red), IEKF-d (light blue), IEKF-2 (purple), IEKF-2d (orange), EKF-2o (teal), IEKF-dm (magenta). (A.i) Mean (over time and frequency) of the filter residuals. Most of the filters appear zero-mean, as expected. The exceptions are the EKF and EKF-2o, which do not utilize draws or iterations and show a slight negative bias, likely due to the "model mis-specification" of the 0.9 state transition factor. It is also clear the draws greatly reduce the variance of the filter residuals. (A.ii) Logarithm of the mean (over time and frequency) of the square filter residuals. Again, the draws alone produce a greater reduction in residual variance than the iterations alone (e.g., EKF-d vs. IEKF), but the iterations do add further improvement (e.g., IEKF-d vs. EKF-d) (A.iii) Logarithm of the two-dimensional Box Q statistic of the filter residuals. (B.i) Logarithm of the mean (over time and peak parameters) of the filter state errors, restricted to times when the respective peaks are On. (B.ii) Probability of the true state values lying within the 95%-confidence interval of the filter state estimates. (B.iii) Probability of correctly determined On/Off-status of all peaks at all times.

APPENDIX G
ADDITIONAL RESULTS AND FIGURES

This subsection contains additional results and figures.

Fig. 5 shows alternative statistical measures for the filter estimates of the sets of pseudo-deterministic simulations. Like the mean-square of the observation residuals in Fig. 2(A.ii), the maximum of the absolute values in Fig. 5(A) is a measure of the extent of the variability of the observation errors. Like the two-dimensional Box-Q statistic of the observation residuals in Fig. 2(A.iii), the Wald statistic of the instantaneous correlation across neighboring frequency bins in Fig. 5(B) and the Moran's I statistic in Fig. 5(C) are measures of correlations in the observation errors. And like the probability of correct On/Off-peak identification in Fig. 2(B.iii), the Cohen's kappa statistic of the On/Off-combos selection in Fig. 5(D) is a measure of successful determination of the discrete state.

The patterns of relative filter performance are comparable to the analogous statistics, with the draws providing the most improvement in estimation accuracy and the iterations offering further improvement. The use of second derivatives provides a small improvement in observation errors, but with greater variability and somewhat worse performance on state and combo estimation.

Fig. 6 shows the results for a single random walk simulation with three peaks on an exponential-decay background—the true simulated spectrogram Fig. 6(A), the spectrogram estimated from the IEKF with draws Fig. 6(B), the residuals Fig. 6(C), the true and estimated On-peaks Fig. 6(D). Because the simulation is nearly within the underlying

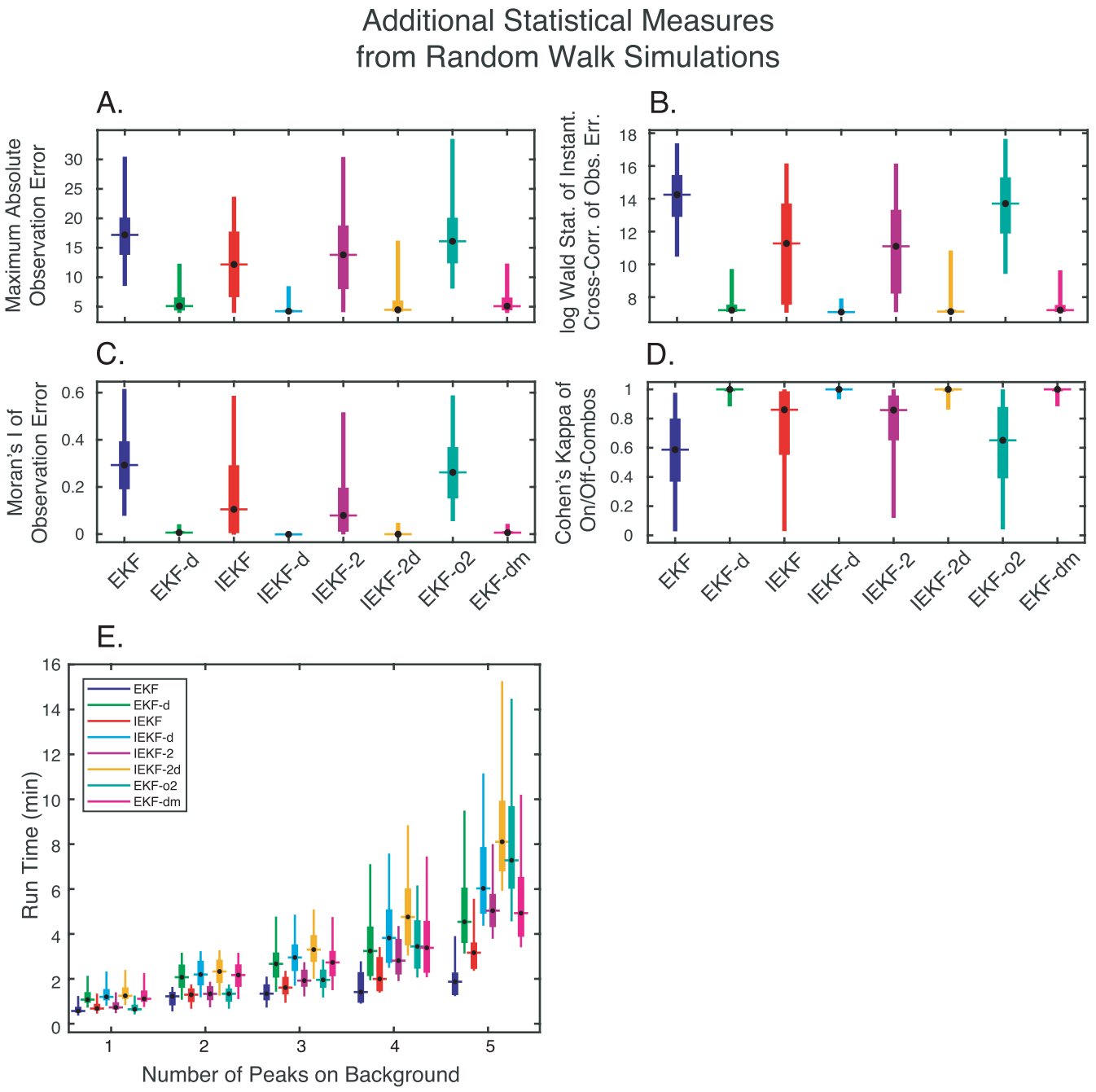

FIGURE 8. Additional statistical measures of filter performances when applied to 5000 simulated spectrograms of 1–5 peaks on an exponential-decay background with parameters following a random walk. EKF (dark blue), EKF-d (green), IEKF (red), IEKF-d (light blue), IEKF-2 (purple), IEKF-2d (orange), EKF-2o (teal), IEKF-dm (magenta). (A) Maximum (over time and frequency) of the absolute value of the filter residuals. This is an alternative to mean-square as a measure of the variability of the residuals. (B) Logarithm of the Wald statistic of the cross-frequency correlations of the filter residuals. (C) Moran's I statistic of the two-dimensional correlations of the filter residuals. The Wald and Moran's I statistics are alternatives to the Box Q as a measure of correlations in the residuals. (D) Cohen's kappa of estimates of the On/Off-combos. This is an alternative to the probability of correct On/Off-peak determination as a measure of the accurate identification of On/Off-peak combos. (E) Computation time of filter estimates.

model class, the performance is even better than for the pseudo-deterministic simulation, with excellent agreement between the true and estimated spectrograms, white residuals, and the correct determination of the combination of On/Off-peaks at each time. The true and estimated peak parameters are shown in Figs. 6(H)–6(G), where the estimated states (dark colored lines) track the true states (solid black lines) almost exactly. The true states are almost always within the very narrow 95%-confidence intervals (colored regions), and the points of discrepancy occur when a peak is off (dashed black lines) and the estimates revert toward the center of the bounds.

Figs. 7 and 8 show the statistical measures for the filter estimates of the sets of random walk simulations. The patterns of relative filter performance are comparable to those for the pseudo-deterministic simulations, with the draws providing the most improvement in estimation accuracy and the iterations offering further improvement. The use of second derivatives provides a small improvement in observation errors, but with greater variability and somewhat worse performance on state and combo estimation.

REFERENCES

[1] G. Buzsáki, *Rhythms of the Brain*. Oxford, U.K.: Oxford Univ. Press, 2006.

[2] E. Olbrich, J. C. Claussen, and P. Achermann, "The multiple time scales of sleep dynamics as a challenge for modelling the sleeping brain," *Philos. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 369, no. 1952, pp. 3884–3901, 2011.

[3] M. J. Prerau, R. E. Brown, M. T. Bianchi, J. M. Ellenbogen, and P. L. Purdon, "Sleep neurophysiological dynamics through the lens of multitaper spectral analysis," *Physiology*, vol. 32, no. 1, pp. 60–92, Jan. 2017.

[4] M. Mölle, T. O. Bergmann, L. Marshall, and J. Born, "Fast and slow spindles during the sleep slow oscillation: Disparate coalescence and engagement in memory processing," *Sleep*, vol. 34, no. 10, pp. 1411–1421, Oct. 2011.

[5] V. Kokkinos and G. K. Kostopoulos, "Human non-rapid eye movement stage II sleep spindles are blocked upon spontaneous K-complex coincidence and resume as higher frequency spindles afterwards," *J. Sleep Res.*, vol. 20, no. 1, pp. 57–72, Mar. 2011.

[6] P. Achermann and A. Borbély, "Temporal evolution of coherence and power in the human sleep electroencephalogram," *J. Sleep Res.*, vol. 7, no. S1, pp. 36–41, 1998.

[7] M. McKeown, C. Humphries, P. Achermann, A. Borbély, and T. Sejnowsk, "A new method for detecting state changes in the EEG: Exploratory application to sleep data," *J. Sleep Res.*, vol. 7, no. S1, pp. 48–56, Jun. 1998.

[8] R. B. Berry, C. L. Albertario, S. M. Harding, R. M. Lloyd, D. T. Plante, S. F. Quan, M. M. Troester, and B. V. Vaughn, *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. Darien, IL, USA: American Academy of Sleep Medicine, 2018.

[9] M. Haller, T. Donoghue, E. Peterson, P. Varma, P. Sebastian, R. Gao, T. Noto, R. T. Knight, A. Shestyuk, and B. Voytek, "Parameterizing neural power spectra," *bioRxiv*, Apr. 2018, Art. no. 299859.

[10] S. R. Cole and B. Voytek, "Brain oscillations and the importance of waveform shape," *Trends Cognit. Sci.*, vol. 21, no. 2, pp. 137–149, Feb. 2017.

[11] R. Cox, W. F. Hofman, M. de Boer, and L. M. Talamini, "Local sleep spindle modulations in relation to specific memory cues," *NeuroImage*, vol. 99, pp. 103–110, Oct. 2014.

[12] A. Mierau, W. Klimesch, and J. Lefebvre, "State-dependent alpha peak frequency shifts: Experimental evidence, potential mechanisms and functional implications," *Neuroscience*, vol. 360, pp. 146–154, Sep. 2017.

[13] B. Voytek, M. A. Kramer, J. Case, K. Q. Lepage, Z. R. Tempesta, R. T. Knight, and A. Gazzaley, "Age-related changes in 1/F neural electrophysiological noise," *J. Neurosci.*, vol. 35, pp. 13257–13265, Sep. 2015.

[14] A. Khanna, A. Pascual-Leone, C. M. Michel, and F. Farzan, "Microstates in resting-state EEG: Current status and future directions," *Neurosci. Biobehavioral Rev.*, vol. 49, pp. 105–113, Feb. 2015.

[15] D. Brunet, M. M. Murray, and C. M. Michel, "Spatiotemporal analysis of multichannel EEG: CARTOOL," *Comput. Intell. Neurosci.*, vol. 2011, pp. 1–15, Jan2011.

[16] S. Haegens, H. Cousijn, G. Wallis, P. J. Harrison, and A. C. Nobre, "Inter- and intra-individual variability in alpha peak frequency," *NeuroImage*, vol. 92, pp. 46–55, May 2014.

[17] E. Werth, P. Achermann, and A. Borbély, "Fronto-occipital EEG power gradients in human sleep," *J. Sleep Res.*, vol. 6, no. 2, pp. 102–112, Jun. 1997.

[18] J. Zeitlhofer, G. Gruber, P. Anderer, S. Asenbaum, P. Schimicek, and B. Saletu, "Topographic distribution of sleep spindles in young healthy subjects," *J. Sleep Res.*, vol. 6, no. 3, pp. 149–155, 1997.

[19] N. Martin, M. Lafortune, J. Godbout, M. Barakat, R. Robillard, G. Poirier, C. Bastien, and J. Carrier, "Topography of age-related changes in sleep spindles," *Neurobiol. Aging*, vol. 34, no. 2, pp. 468–476, Feb. 2013.

[20] C. O'Reilly, J. Godbout, J. Carrier, and J.-M. Lina, "Combining time-frequency and spatial information for the detection of sleep spindles," *Frontiers Hum. Neurosci.*, vol. 9, p. 70, Feb. 2015.

[21] R. Cox, A. C. Schapiro, D. S. Manoach, and R. Stickgold, "Individual differences in frequency and topography of slow and fast sleep spindles," *Frontiers Hum. Neurosci.*, vol. 11, p. 433, Sep. 2017.

[22] D. L. Rowe, P. A. Robinson, and C. J. Rennie, "Estimation of neurophysiological parameters from the waking EEG using a biophysical model of brain dynamics," *J. Theor. Biol.*, vol. 231, no. 3, pp. 413–433, Dec. 2004.

[23] L. De Gennaro, M. Ferrara, F. Vecchio, G. Curcio, and M. Bertini, "An electroencephalographic fingerprint of human sleep," *NeuroImage*, vol. 26, no. 1, pp. 114–122, May 2005.

[24] P. P. Ujma, F. Gombos, L. Genzel, B. N. Konrad, P. Simor, A. Steiger, M. Dresler, and R. Bódizs, "A comparison of two sleep spindle detection methods based on all night averages: Individually adjusted vs. Fixed frequencies," *Frontiers Hum. Neurosci.*, vol. 9, p. 52, Feb. 2015.

[25] K. Crowley, "The effects of normal aging on sleep spindle and K-complex production," *Clin. Neurophysiol.*, vol. 113, no. 10, pp. 1615–1622, Oct. 2002.

[26] A. Nicolas, D. Petit, S. Rompré, and J. Montplaisir, "Sleep spindle characteristics in healthy subjects of different age groups," *Clin. Neurophysiol.*, vol. 112, pp. 521–527, Mar. 2001.

[27] F. Ferrarelli, R. Huber, M. J. Peterson, M. Massimini, M. Murphy, B. A. Riedner, A. Watson, P. Bria, and G. Tononi, "Reduced sleep spindle activity in schizophrenia patients," *Amer. J. Psychiatry*, vol. 164, no. 3, pp. 483–492, Mar. 2007.

[28] E. J. Wamsley, M. A. Tucker, A. K. Shinn, K. E. Ono, S. K. McKinley, A. V. Ely, D. C. Goff, R. Stickgold, and D. S. Manoach, "Reduced sleep spindles and spindle coherence in schizophrenia: Mechanisms of impaired memory consolidation?" *Biol. Psychiatry*, vol. 71, no. 2, pp. 154–161, Jan. 2012.

[29] É. Limoges, L. Mottron, C. Bolduc, C. Berthiaume, and R. Godbout, "Atypical sleep architecture and the autism phenotype," *Brain*, vol. 128, no. 5, pp. 1049–1061, May 2005.

[30] I. Myatchin and L. Lagae, "Sleep spindle abnormalities in children with generalized spike-wave discharges," *Pediatric Neurol.*, vol. 36, no. 2, pp. 106–111, Feb. 2007.

[31] C. O'Reilly, I. Godin, J. Montplaisir, and T. Nielsen, "REM sleep behaviour disorder is associated with lower fast and higher slow sleep spindle densities," *J. Sleep Res.*, vol. 24, no. 6, pp. 593–601, Dec. 2015.

[32] P. Montagna, P. Gambetti, P. Cortelli, and E. Lugaresi, "Familial and sporadic fatal insomnia," *Lancet Neurol.*, vol. 2, no. 3, pp. 167–176, Mar. 2003.

[33] F. Espa, B. Ondze, P. Deglise, M. Billiard, and A. Besset, "Sleep architecture, slow wave activity, and sleep spindles in adult patients with sleepwalking and sleep terrors," *Clin. Neurophysiol.*, vol. 111, no. 5, pp. 929–939, May 2000.

[34] S. L. Himanen, J. Virkkala, E. Huupponen, and J. Hasan, "Spindle frequency remains slow in sleep apnea patients throughout the night," *Sleep Med.*, vol. 4, pp. 229–234, May 2003.

[35] J. M. Gottselig, C. L. Bassetti, and P. Achermann, "Power and coherence of sleep spindle frequency activity following hemispheric stroke," *Brain*, vol. 125, no. 2, pp. 373–383, Feb. 2002.

[36] D. Petit, J.-F. Gagnon, M. L. Fantini, L. Ferini-Strambi, and J. Montplaisir, "Sleep and quantitative EEG in neurodegenerative disorders," *J. Psychosomatic Res.*, vol. 56, no. 5, pp. 487–496, May 2004.

[37] S. C. Warby, S. L. Wendt, P. Welinder, E. G. S. Munk, O. Carrillo, H. B. D. Sorensen, P. Jennum, P. E. Peppard, P. Perona, and E. Mignot, "Sleep-spindle detection: Crowdsourcing and evaluating performance of experts, non-experts and automated methods," *Nature Methods*, vol. 11, no. 4, pp. 385–392, Apr. 2014.

[38] R. Bódizs, J. Körmendi, P. Rigó, and A. S. Lázár, "The individual adjustment method of sleep spindle analysis: Methodological improvements and roots in the fingerprint paradigm," *J. Neurosci. Methods*, vol. 178, no. 1, pp. 205–213, Mar. 2009.

[39] M. Adamczyk, L. Genzel, M. Dresler, A. Steiger, and E. Friess, "Automatic sleep spindle detection and genetic influence estimation using continuous wavelet transform," *Frontiers Hum. Neurosci.*, vol. 9, p. 624, Nov. 2015.

[40] E. Olbrich and P. Achermann, "Analysis of oscillatory patterns in the human sleep EEG using a novel detection algorithm," *J. Sleep Res.*, vol. 14, no. 4, pp. 337–346, Dec. 2005.

[41] M. P. Tarvainen, J. K. Hiltunen, P. O. Ranta-aho, and P. A. Karjalainen, "Estimation of nonstationary EEG with Kalman smoother approach: An application to event-related synchronization (ERS)," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 3, pp. 516–524, Mar. 2004.
[42] C. Dubois, M. Davy, and J. Idier, "Tracking of time-frequency components using particle filtering," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, vol. 4, Mar. 2005, pp. iv/9–iv12.
[43] T. Matsuda and F. Komaki, "Time series decomposition into oscillation components and phase estimation," *Neural Comput.*, vol. 29, pp. 332–367, Feb. 2017.
[44] M. Jabloun, N. Martin, F. Leonard, and M. Vieira, "Estimation of the instantaneous amplitude and frequency of non-stationary short-time signals," *Signal Process.*, vol. 88, pp. 1636–1655, Jul. 2008.
[45] J. Van Zaen, L. Uldry, C. Duchêne, Y. Prudat, R. A. Meuli, M. M. Murray, and J.-M. Vesin, "Adaptive tracking of EEG oscillations," *J. Neurosci. Methods*, vol. 186, no. 1, pp. 97–106, Jan. 2010.
[46] Z. Deprem, A. E. Çetin, and O. Arıkan, "AM/FM signal estimation with micro-segmentation and polynomial fit," *Signal, Image Video Process.*, vol. 8, no. 3, pp. 399–413, Mar. 2014.
[47] M. J. Prerau, P. L. Purdon, and U. T. Eden, "Tracking non-stationary spectral peak structure in EEG data," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 417–420.
[48] A. H. Jazwinski, *Stochastic Processes and Filtering Theory* (Mathematics in Science & Engineering). New York, NY, USA: Academic, 1970.
[49] B. D. O. Anderson and J. B. Moore, *Optimal Filtering* (Information and System Sciences Series). Upper Saddle River, NJ, USA: Prentice-Hall, 1979.
[50] M. J. Prerau, K. E. Hartnack, G. Obregon-Henao, A. Sampson, M. Merlino, K. Gannon, M. T. Bianchi, J. M. Ellenbogen, and P. L. Purdon, "Tracking the sleep onset process: An empirical model of behavioral and physiological dynamics," *PLoS Comput. Biol.*, vol. 10, no. 10, Oct. 2014, Art. no. e1003866.
[51] P. J. Leeuwen, H. R. Künsch, L. Nerger, R. Potthast, and S. Reich, "Particle filters for high-dimensional geoscience applications: A review," *Quart. J. Roy. Meteorol. Soc.*, vol. 145, no. 723, pp. 2335–2365, Jul. 2019.
[52] A. Doucet, N. De Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. (Statistics for Engineering and Information Science). New York, NY, USA: Springer, 2001.
[53] N. Kantas, A. Doucet, S. S. Singh, J. Maciejowski, and N. Chopin, "On particle methods for parameter estimation in state-space models," *Stat. Sci.*, vol. 30, no. 3, pp. 328–351, Aug. 2015.

PATRICK A. STOKES (Member, IEEE) received the B.S. degree in biomedical engineering and the B.A. degree in economics from The University of Texas at Austin, in 2006, and the Ph.D. degree in health sciences and technology from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2015.

From 2015 to 2019, he was a Postdoctoral Research Fellow with the Department of Anesthesia, Critical Care, and Pain Medicine, Massachusetts General Hospital, Boston, MA, USA. From 2019 to 2020, he was a Postdoctoral Research Fellow with the Division of Sleep and Circadian Disorders, Department of Medicine, Brigham and Women's Hospital, Boston. His research interests include methods for system identification and modeling of complex, high-dimensional dynamical systems, and their application to neurophysiological processes and functions.

MICHAEL J. PRERAU (Member, IEEE) received the B.S. degree in computer science and the M.S. degree in biomedical engineering from Columbia University, New York, NY, USA, in 2002 and 2003, respectively, and the Ph.D. degree in computational neuroscience from Boston University, Boston, MA, USA, in 2010.

Since 2016, he has been an Assistant Professor of medicine with the Harvard Medical School, a Faculty Member with the Division of Sleep Medicine, Harvard Medical School, and an Associate Neuroscientist and the Director of the Sleep Neurophysiology Signal Processing Core, Division of Sleep and Circadian Disorders, Brigham and Women's Hospital. His research interests include the linkage between changes in brain dynamics during sleep and the pathophysiology of neurological dysfunction. By applying state-of-the-art quantitative methods to the analysis of the sleep electroencephalogram, it is possible to use sleep for biomarker discovery, facilitating the diagnosis and treatment of neurological, psychiatric, and sleep disorders. His laboratory focuses on developing novel approaches to statistical modeling and signal processing algorithms for the analysis of neural data during sleep, with direct applications to basic science, biomarker discovery, and medical device development. He was a recipient of the Research Image Award for data visualization by the Sleep Research Society, as well as the Young Investigator Oral Presentation Award by the World Sleep Congress.