Hybrid Adaptive Parametric Frequency Analysis: An application to EEG under general anesthesia

Kriton Konstantinidis and Emery N. Brown, *Fellow, IEEE*

***Abstract*— High quality spectra are crucial in anesthesia related procedures where Electroencephalogram (EEG) frequency content can drastically help track different brain states. To this end, an adaptive autoregressive model framework to fit non-stationary EEG data using hybrid Kalman Filtering (HKF) is developed. In this setup, a state-space formulation is adopted. Hybridity arises from the fact that the state vector, which includes the autoregressive parameters, evolves in continuous time, while the observation equation is discrete, to account for the fact that the observations, i.e. EEG data, are discrete points in time. As a smoothing constraint, the parameters are modeled to follow a continuous multivariate random walk. As shown in this work, their adaptive estimation by means of HKF and expectation-maximization (EM) algorithm yields smoother estimation of frequency spectra, outperforming other current purely discrete parametric methods as well as various non-parametric approaches. A variant of the model using non-Gaussian state noise and Sequential Monte Carlo Markov Chain (SMCMC) filtering on a self-organizing state-space model is also presented. Examples of dynamic EEG data were taken from patients under gradually varying doses of propofol and ketamine. The suitability of our method for online use in combination with its ability to smoothly track frequency changes in human EEG signals suggests that it can be used for real time brain state tracking under general anesthesia, facilitating the design of closed loop systems for automatic and precise control of brain states.**

I. INTRODUCTION

Non-stationary time-series frequency analysis has been traditionally a thematic that attracted the attention of the scientific community. This is due to its enormous practical benefits that arise from the fact that most signals with a temporal representation encountered in real-life applications are characterized by time-varying statistics. An example of a real-life application is the use of EEG to track brain states under anesthesia. EEG has been extensively used to characterize brain states under various anesthetics [1], [2] and to automatically adjust drug infusion rates using closed loop control systems [3]. The issue of non-stationarity was usually overcome by various assumptions about local stationarity properties over short time intervals and application of techniques suitable only for stationary signals. Even though such assumptions can have practical benefits and often provide satisfying performances, they are not always sufficient and can be a source of noise. Current techniques used to estimate frequency spectra of non-stationary signals include various non-parametric methods, such as the multitaper approach developed in [4]. Non-parametric approaches suffer from a specific limitation: They cannot yield both high temporal and frequency resolutions simultaneously. Improving time resolution by using a shorter, assumed stationary, data segment results in lower frequency resolution and vice versa. An alternative approach are the so called parametric methods. Such methods are usually based on time-varying linear predictive models, such as autoregressive (AR), moving average (MA) and autoregressive moving average (ARMA) models. The use of such models has led to higher resolution of time-varying frequency spectra. The first to use a state-space framework to fit adaptive autoregressive models for EEG analysis was Schlogl et al. [5]. A recursive least squares method was used to fit the model. Fitting of adaptive autoregressive models using the Kalman filter to estimate time-varying spectra of EEG under propofol is done in [6]. In [5] and [6] the observation and process covariance matrices are manually set before the filtering procedure. Khan et al. [7] uses an expectation-maximization approach to estimate the components of the state-space model in subsequent data windows and then runs a Kalman smoother to finely estimate the autoregressive coefficients. In [8], Cauchy state noise is used. The scale of the Cauchy distribution is assumed identical for all parameters and is jointly estimated with the state using a sequential importance resampling algorithm. In all aforementioned approaches, observations and states are modeled as discrete variables. Inspired by the work of [9] and [10], where continuous, non-adaptive AR models were fit to discrete samples of EEG recordings, this work adopts an intermediate between discrete and continuous approaches and coupled with an EM routine, develops an algorithm for efficient and smooth spectral estimation of non-stationary signals. It also develops an algorithm to fit hybrid autoregressive models using Cauchy state noise by jointly estimating a different scale for each autoregressive parameter's distribution. Finally, it explores the performance of the two hybrid algorithms in different EEG datasets and compares it to that obtained by other widely used non-parametric and discrete parametric methods.

*This work was supported by the The Bertarelli Program in Translational Neuroscience and Neuroengineering (https://bertarelli.hms.harvard.edu/)

K.K. is with the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139 and with Harvard Medical School, Boston, MA 02115 (kritonkonstantinidis@gmail.com)

E.N.B. is with the MIT Department of Brain and Cognitive Sciences, with Harvard Medical School, and with the Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, MA 02114

II. METHODS

A. Hybrid Normal Model

1) State-Space Formulation: In this model, the autoregressive parameters are modeled to follow a

continuous multivariate random walk. They are updated in discrete time steps when a new observation is available, as in the common Kalman Filtering approaches.

State Equation: $\frac{d\mathbf{a}(t)}{dt}$=$\mathbf{w}(t)$ (1)
Observation Equation: $z_k = \mathbf{H}_k\mathbf{a}_k$+$v_k$ (2)
$\mathbf{w}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}(t))$: Continuous multivariate Gaussian process with $\mathbf{0}$ mean and covariance $\mathbf{Q}(t)$ (symmetric, positive-definite)
$v_k \sim \mathcal{N}(0, R(t))$: Discrete scalar white Gaussian observation noise with 0 mean and variance $R(t)$
$\mathbf{a}(t) = [a_1(t), a_2(t), \ldots, a_p(t)]^T$: State vector with the continuous time-varying autoregressive coefficients and covariance matrix $\mathbf{P}(t)$
$\mathbf{H}_k = [z_{k-1}, \ldots, z_{k-p}]$: Observation matrix at time t_k comprising the p past discrete observations
$\mathbf{a}_k = \mathbf{a}(t_k)$: Sampled state vector at time t_k
p: Autoregressive model order
z_k: Observation at time t_k
Let $Z = \{z_k\}_{k=1}^N$ denote the set of the observations, $A = \{\mathbf{a}_k\}_{k=1}^N$ denote the state vector at all times t_k of the observations z_k and N denote the total number of observations.

2) Hybrid Kalman Filtering: With the above model, the HKF equations become:

Initialization: $\mathbf{a}_{0|0} = aryule(Z, p)$ (3), $\mathbf{P}_{0|0} = \mathbf{I}_p$ (4)
Prediction: $\int_{t_{k-1}}^{t_k} \frac{d\mathbf{a}(t)}{dt} d\tau = \mathbf{a}_{k|k-1}$ (5)

$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{Q}\Big|_{\mathbf{P}_{k-1|k-1}} \quad (6) \implies \int_{t_{k-1}}^{t_k} \frac{d\mathbf{P}(t)}{dt} d\tau = \mathbf{P}_{k|k-1}$$

In this design of the hybrid Kalman filter, the prediction equations do not involve update from measurements. The update from the measurements is performed in the discrete part of the filter. Therefore, the Ricatti variance equation that is usually required to be solved in the purely continuous version of the Kalman filter is avoided here.

Filtering:
$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + R)^{-1} \quad (7)$$
$$\mathbf{a}_{k|k} = \mathbf{a}_{k|k-1} + \mathbf{K}_k(z_k - \mathbf{H}_k\mathbf{a}_{k|k-1}) \quad (8)$$
$$\mathbf{P}_{k|k} = (\mathbf{I}_p - \mathbf{K}_k\mathbf{H}_k)\,\mathbf{P}_{k|k-1} \quad (9)$$
$$\mathbf{S}_k = \mathbf{P}_{k|k}(\mathbf{P}_{k|k} + \mathbf{Q})^{-1} \quad (10)$$

Smoothing:
$$\mathbf{a}_{k|N} = \mathbf{a}_{k|k} + \mathbf{S}_k(\mathbf{a}_{k+1|N} - \mathbf{a}_{k+1|k}) \quad (11)$$
$$\mathbf{P}_{k|N} = \mathbf{P}_{k|k} + \mathbf{S}_k(\mathbf{P}_{k+1|N} - \mathbf{P}_{k+1|k})\mathbf{S}_k^T \quad (12)$$

In case of online use of the model, smoothing can be ignored without significant effects on spectrogram quality.

3) EM & Model Selection: In order to estimate the process noise covariance $\mathbf{Q}$ and the observation noise variance R in a maximum likelihood approach, the following EM algorithm can be applied. The procedure is derived by adapting the approach of [11] for our model in an initial EEG sample of 10 seconds. The filter can then be run forward until the end of the recording.

Let $\Theta = \{\mathbf{Q}, R\}$ be the set of parameters whose maximum likelihood values are searched. The goal is to maximize the log-likelihood $log[P(Z|\Theta)]$. Since access to the states is not provided, what is feasible is to calculate the expected complete log-likelihood $\mathcal{Q} = \mathcal{E}_{A|Z}[logP[(Z, A|\Theta)]]$ in the E-Step of the algorithm and maximize it with respect to Θ in the M-Step in an iterative manner of alternating the 2 aforementioned steps.

Initialization: $\mathbf{Q}_{start} = \mathbf{I}_p$ (13), $R_{start} = R(0)$ (14)
E-Step: $\mathcal{Q} = \mathcal{E}_{A|Z}[log[\prod_{k=2}^{N} P(\mathbf{a}_k|\mathbf{a}_{k-1}) \prod_{k=1}^{N} P(z_k|\mathbf{a}_k, \mathbf{H}_k)]]$ (15)
M-Step: $\mathbf{Q}_{ML} = \frac{1}{N-1}[\sum_{k=2}^{N}[\mathbf{P}_{k|N} + \mathbf{a}_{k|N}\mathbf{a}_{k|N}^T - \mathbf{S}_{k-1}\mathbf{P}_{k|N} - \mathbf{a}_{k|N}\mathbf{a}_{k-1|N}^T]]$ (16)
$R_{ML} = \frac{1}{N}\sum_{k=1}^{N}[{z_k}^2 - 2z_k\mathbf{H}_k\mathbf{a}_{k|N} + \mathbf{H}_k[\mathbf{P}_{k|N} + \mathbf{a}_{k|N}\mathbf{a}_{k|N}^T]\mathbf{H}_k^T]$ (17)

The above 2 steps are repeated until convergence. Convergence is established if the relative increase of the marginal log likelihood $\mathcal{L}$ between 2 iterations is below a certain threshold. $\mathcal{L}$ is obtained as follows:

$$\mathcal{L} = -\frac{N}{2}log(2\pi) - \frac{1}{2}\sum_{k=1}^{N} log(R + \mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T) - \frac{1}{2}\sum_{k=1}^{N}\frac{(z_k - \mathbf{H}_k\mathbf{a}_{k|k-1})^2}{R + \mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T} \quad (18)$$

Model selection is then done using the well-known Akaike Information Criterion (AIC), defined as $AIC(p) = 2p - 2\mathcal{L}_{ML}$, where p is the order of the autoregressive model and $\mathcal{L}_{ML}$ is the maximized marginal log-likelihood for that model. The order that yields the lowest AIC will be chosen.

B. Hybrid Cauchy Model

As suggested in [8], the use of non-Gaussian state noise can be advantageous in decomposing EEG spectra. More specifically, replacement of Gaussian by Cauchy noise can provide significant improvements in tracking of instantaneous frequency. In [8], the performance of the filter is better than the Gaussian one in identifying abrupt frequency changes but performs worse in stationary parts of simulated data. Motivated by the approach in [8], a bootstrap hybrid algorithm is developed, that attempts to improve the tracking performance of the Gaussian filter while mitigating outliers in more stationary parts, by taking advantage of hybridity.

1) Self-Organizing State-Space Formulation:

State Equation: $d\mathbf{x}(t) = \mathbf{w}(t)dt$ (19)
Observation Equation: $z_k = \mathbf{H}_k\mathbf{x}_k$+$v_k$ (20)
$\mathbf{w}(t) = [w_1(t), w_2(t), \ldots, w_p(t), \epsilon_1(t), \epsilon_2(t), \ldots, \epsilon_p(t)]^T$, where $w_j(t) \sim \mathcal{C}(0, q_j(t))$ is a continuous Cauchy process

with 0 mean and scale parameter $q_j(t)$ and $\epsilon_j(t) \sim \mathcal{N}(0, R_\epsilon)$ is a continuous Gaussian process with 0 mean and variance R_ϵ

$v_k \sim \mathcal{N}(0, R(t))$ is a discrete scalar white Gaussian observation noise with 0 mean and variance $R(t)$

$\mathbf{x}(t) = [a_1(t), a_2(t), \ldots, a_p(t), ln(q_1(t)), ln(q_2(t)),$

$\ldots, ln(q_p(t))^T = [\mathbf{a}(t), \mathbf{q}(t)]^T$ is the augmented state vector with the continuous time-varying autoregressive coefficients and their respective scale parameter. To ensure positivity of the scale parameters, their logarithm is estimated.

$\mathbf{H}_k = [z_{k-1}, \ldots, z_{k-p}, 0, \ldots, 0]$ is the observation matrix at time t_k comprising the p past discrete observations and p zeros

$\mathbf{x}_k = \mathbf{x}(t_k)$ is the sampled state vector at time t_k

Let T denote the total number of observations in the data, indexed by k, N denote the total number of particles used for the sequential importance resampling, indexed by i, j index the autoregressive parameters $[a_1(t), a_2(t), \ldots, a_p(t)]$ and f_s denote the sampling frequency

2) Discretization of the state equation using the Euler-Maruyama scheme: To start the algorithm, it is necessary to discretize equation (19). To do so, the Euler-Maruyama scheme is deployed, briefly described here.

For a discretization step Δ, (19) can be written as $\mathbf{x}(t+\Delta) = \mathbf{x}(t) + \int_t^{t+\Delta} d\mathbf{w}(t)$ (21). According to Euler-Maruyama, the Ito integral in equation (21) can be discretized as $\Delta\mathbf{x}(t) = \Delta\mathbf{w}(t)$, where $\Delta\mathbf{x}(t) \approx \mathbf{x}(t+\Delta) - \mathbf{x}(t), \Delta\mathbf{w}(t) \sim \mathcal{C}(\mathbf{0}, \mathbf{Q}\Delta)$, for the first p dimensions where $\mathbf{Q}$ is the diagonal matrix with entries $q_j(t)$ and $\Delta\mathbf{w}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_q\Delta)$, where $\mathbf{Q}_q$ is the diagonal matrix with entries $\epsilon_j(t)$ for the next p dimensions.

The Self-Organizing state-space equations can now be written as:

State Equation: $\Delta\mathbf{x}(t) = \Delta\mathbf{w}(t)$ (22)
Observation Equation $z_k = \mathbf{H}_k\mathbf{x}_k + v_k$ (23)

3) Sequential Markov Chain Monte Carlo: Let B denote the number of bootstrap samples and $w_k^{(i),b}$ denote the unnormalized weight of the particle i at time k for the bootstrap sample b and $W_k^{(i),b}$ the normalized one respectively. Let $a_{k,j}^{(i),b}$ denote the value of the particle i for the coefficient j at time k for the bootstrap sample b. Let ESS denote efficient sample size and $[\alpha, \beta]$ denote the interval for uniform initialization of the Cauchy scale parameters. Let $\mathbf{Q}_q$ denote the covariance matrix of the normal distribution of the scale parameters of the Cauchy distributions of the coefficients. The posterior density for a given bootstrap sample is approximated by $\hat{p}(\mathbf{a}_k|z_{1:k}) = \sum_{i=1}^{N} W_k^{(i)} \delta_{\mathbf{a}_k^{(i)}}(\mathbf{a}_k)$ and the autoregressive coefficients are estimated by performing Metropolis-Hastings updates using the approximated posterior as proposal density.

for *b=1:B* **do**
  Initialization: $\mathbf{a}_0^{(i),b} = aryule(Z, p), W_0^{(i),b} = \frac{1}{N}$,
  $\mathbf{q}_0^{(i),b} = diag(\mathcal{U}(\alpha, \beta))$
  Filtering: **for** *k=1:T* **do**
    Sample $\mathbf{a}_{k-1+\Delta}^{(i),b} \sim \mathcal{C}(\mathbf{a}_{k-1}^{(i),b}, \mathbf{Q}_{k-1}\Delta)$
    $\mathbf{q}_{k-1+\Delta}^{(i),b} \sim \mathcal{N}(\mathbf{q}_{k-1}^{(i),b}, \mathbf{Q}_q\Delta)$
    $\mathbf{a}_{k-1+2\Delta}^{(i),b} \sim \mathcal{C}(\mathbf{a}_{k-1+\Delta}^{(i),b}, \mathbf{Q}_{k-1}\Delta)$
    $\mathbf{q}_{k-1+2\Delta}^{(i),b} \sim \mathcal{N}(\mathbf{q}_{k-1+\Delta}^{(i),b}, \mathbf{Q}_q\Delta)$
    $\vdots$
    $\mathbf{a}_k^{(i),b} \sim \mathcal{C}(\mathbf{a}_{k-1+(\frac{1}{\Delta f_s}-1)\Delta}^{(i),b}, \mathbf{Q}_{k-1}\Delta)$
    $\mathbf{q}_k^{(i),b} \sim \mathcal{N}(\mathbf{q}_{k-1+(\frac{1}{\Delta f_s}-1)\Delta}^{(i),b}, \mathbf{Q}_q\Delta)$
    $w_k^{(i),b} = W_{k-1}^{(i),b} p(z_k|\mathbf{H}_k, \mathbf{a}_k^{(i),b}, R) =$
    $W_{k-1}^{(i),b} \mathcal{N}(z_k|\mathbf{H}_k, \mathbf{a}_k^{(i),b}, R)$
    $W_k^{(i),b} = \frac{w_k^{(i),b}}{\sum_{i=1}^{N} w_k^{(i),b}}$
    $ESS_k^b = \frac{1}{\sum_{i=1}^{N} (W_k^{(i),b})^2}$
    if $ESS_k^b < \frac{N}{3}$ **then**
      Resample with replacement N particles from $\{\mathbf{x}_k^{(i),b}\}_{k=1}^N$ according to probabilities $W_k^{(i),b}$
      Reset $W_k^{(i),b} = \frac{1}{N}$
    end
    Approximate $p(\mathbf{a}_k^b, \mathbf{q}_k^b|z_{1:k})$ by
    $\hat{p}(\mathbf{a}_k^b, \mathbf{q}_k^b|z_{1:k}) = \sum_{i=1}^{N} W_k^{(i)} \delta_{\mathbf{a}_k^{(i)}}(\mathbf{a}_k)$
    Sample $\tilde{\mathbf{q}_k}^{(i),b}(0) \sim \hat{p}(\mathbf{a}_k^b, \mathbf{q}_k^b|z_{1:k})$,
    $\tilde{\mathbf{a}_k}^{(i),b}(0) \sim \hat{p}(\mathbf{a}_k^b, \mathbf{q}_k^b|z_{1:k})$,
    $\tilde{\mathcal{L}}(0) = \sum_{t=1}^{k} log(\sum_{i=1}^{N} w_t^{(i),b})$, $\tilde{w_k}^{(i),b}(0) = w_k{}^{(i),b}$
    for *l=1:L* **do**
      $\tilde{\mathbf{q}_k}^{(i),b}(l) \sim \hat{p}(\mathbf{a}_k^b, \mathbf{q}_k^b|z_{1:k})$,
      $\tilde{\mathbf{a}_k}^{(i),b}(l) \sim \hat{p}(\mathbf{a}_k^b, \mathbf{q}_k^b|z_{1:k})$
      $\tilde{w}_{temp}^{(i),b}(l) = \tilde{w}_k^{(i),b}(l-1) \times$
      $\times p(z_k|\mathbf{H}_k, \tilde{\mathbf{a}}_k^{(i),b}(l), \tilde{\mathbf{q}_k}^{(i),b}(l), R) / p(z_k|\mathbf{H}_k, \tilde{\mathbf{a}}_k^{(i),b}(l-1),$
      $\tilde{\mathbf{q}_k}^{(i),b}(l-1), R)$
      $\tilde{\mathcal{L}}_{temp}(l) = \tilde{\mathcal{L}}(l-1)$
      $-log(\sum_{i=1}^{N} \tilde{w}_k^{(i),b}(l-1)) + log(\sum_{i=1}^{N} \tilde{w}_{temp}^{(i),b}(l))$
      With probability $\min(1, \tilde{\mathcal{L}}_{temp}(l) \backslash \tilde{\mathcal{L}}(l-1))$,
      set $\hat{a}_{k,j}^b = \tilde{a_{k,j}}^b(l)$,
      $\tilde{w_k}^{(i),b}(l) = \tilde{w}_{temp}^{(i),b}(l)$
      $W_k^{(i),b} = \frac{\tilde{w}_{temp}^{(i),b}(l)}{\sum_{i=1}^{N} \tilde{w}_{temp}^{(i),b}(l)}$,
      $\tilde{\mathcal{L}}(l) = \tilde{\mathcal{L}}_{temp}(l)$
      Otherwise set $\tilde{a_{k,j}}^{(i),b}(l) = \tilde{a}_{k,j}^{(i),b}(l-1)$,
      $\tilde{\mathcal{L}}(l) = \tilde{\mathcal{L}}(l-1)$
    end
  end
end
$\hat{a}_{k,j} = \frac{1}{B} \sum_{b=1}^{B} \hat{a}_{k,j}^b$

Algorithm 1: Sequential Markov Chain Monte Carlo

4) Log-likelihood of the model: Let θ denote the scale parameters q_j. The log-likelihood that the observed data are generated by the model is defined as $\mathcal{L}(\theta) = log \prod_{k=1}^{N} p\left(z_k \mid z_{k-1}, \ldots, z_0, \theta\right). \quad (24)$

(24) can be approximated using $p\left(z_k \mid z_{k-1}, \ldots, z_0, \theta\right) = \int p(z_k|\mathbf{x}_k)p(x_k|\left(z_{k-1}, \ldots, z_0, \theta\right))d\mathbf{x}_k \approx \frac{1}{N}\sum_{i=1}^{N} p(z_k|t_k^{(i)}) = \frac{1}{N}\sum_{i=1}^{N} w_k^{(i)}$ (25), where $t_k^{(i)}$ denotes the particle i at time k.

Using (25), (24) can be approximated as $\mathcal{L}(\theta) = \sum_{k=1}^{T} log(\sum_{i=1}^{N} w_k^{(i)})$ (26), where $w_k^{(i)}$ is the unnormalized weight of particle i at time k.

The approximated likelihood (26) will be used for model selection with the AIC and for model comparison between Normal and Cauchy models based on likelihood values.

5) Computational Complexity and alternative strategies: For a large number L of steps of Monte Carlo search at each time point the above algorithm can lead the particles to areas of very high likelihood and thus provide very accurate estimates of the coefficients. An issue that arises with this approach is its impracticability for online use due to its high computational complexity of $O(BNTL)$, where each variable has been defined previously. To circumvent this issue, two alternative strategies are presented:

Parameter update based on weighted mean of posterior density: In this case, no Metropolis-Hastings updates are performed and after the sequential importance resampling step, the coefficients are estimated using the weighted mean of the filtered density. This method, even though the estimates of the coefficients may not correspond to particles of the highest likelihood, provides a simple reasonable choice that alleviates the computational burden of performing L MCMC steps every time a new observation arrives. For a given bootstrap sample b, the autoregressive parameters are approximated by $\mathbf{a}_k^b = \sum_{i=1}^{N} a_{k,j}^{(i),b} W_k^{(i),b}$.

Algorithm Parallelization via Virtual Particle Resampling: The resampling step is done to avoid sample degeneracy that happens after some steps of the sampling procedure, where the importance will be only on few particles. Resampling is performed according to the weights the particles have at the time of the resampling and leads the particles to regions that have high likelihood. Nevertheless, resampling causes another issue: Many particles that have low weights end up not being resampled at all. To solve this problem, resampling is performed only when a metric called Effective Sample Size (ESS) is below a certain threshold. The necessity to calculate ESS at every iteration of the filtering procedure renders the parallelization infeasible, as it is a collective operation across all particles that cannot be done in parallel. Another potential strategy is to perform Virtual Particle Resampling [12]. In this approach, a set of virtual particles is generated for each particle and the one that gave the highest importance factor is chosen. Importance factor for a given particle is calculated by evaluating $\hat{p}(z_k|\hat{a}_k)$. Let W denote the number of virtual particles to be generated for each particle and σ the standard deviation which determines how much the generated virtual particles will differ from the original one. Instead of calculating ESS at every time step, resampling proceeds as follows.

for *w=1:W* **do**
  Sample r$\sim \mathcal{U}(-1, 1)$
  $\hat{\mathbf{a}}_{k,w}^{(i),b} = \mathbf{a}_k^{(i),b} + \sigma r$
  $\hat{w}_{k,w}^{(i),b} = \hat{p}(z_k|\hat{a}_{k,w}^{(i),b})$
  if $\hat{w}_{k,w}^{(i),b} > w_k^{(i),b}$ **then**
    $\mathbf{a}_k^{(i),b} = \hat{\mathbf{a}}_{k,w}^{(i),b}$
    $w_k^{(i),b} = \hat{w}_{k,w}^{(i),b}$
  end
end

Algorithm 2: Virtual Particle Resampling

Even though W extra steps have to be performed for each particle at each time point, this is a separate operation for each particle and can be readily parallelized, facilitating the online use of the algorithm.

C. Spectral Estimation & Roughness Metric

Having calculated the optimal estimates for the autoregressive parameters by the Normal or Cauchy model, the spectral density at frequency f and time t is calculated using:

$$S(f,t) = \frac{R}{|1 - \sum_{k=1}^{p} a_k(t) \mathrm{e}^{-\mathrm{i}2\pi k \frac{f}{f_s}}|^2}, \quad 0 \le f \le \frac{f_s}{2} \quad (27)$$

Where $a_k(t)$ is the k-th estimated autoregressive coefficient at time t.

Finally, we define a *roughness* metric of a coefficient $a_k(t)$ between 2 time points t_1 and t_2, $R_k = \int_{t_1}^{t_2} [\frac{d^2 a_k(t)}{dt^2}]^2 dt$ (28) [13]. The lower the metric, the smoother the temporal evolution of the autoregressive coefficients.

III. RESULTS

Results on non-parametric, discrete and hybrid Normal and hybrid Cauchy models will be discussed. Discrete Cauchy models lead to even noisier spectrograms than the Normal models, as expected due to the nature of the heavy-tailed Cauchy distribution.

A. Simulation Results

To validate our hybrid models and to obtain an insight of how hybridity enhances smoothness without restricting adaptability, they were fit along with their discrete counterparts to a simulated noisy sinusoidal wave with stepwise frequency modulation, in order to simulate both abrupt and stationary regimes.

An artificial sinusoidal signal of $T = 60sec$, $z_k = A_k sin(\omega_k \frac{k}{f_s}) + v_k$ was generated at a typical EEG sampling rate $f_s = 250Hz$. v_k is added white Gaussian noise with zero mean and unit variance. Amplitude evolves as $A_k = 1 + \frac{k}{Tf_s}$, $0 \leq \frac{k}{f_s} \leq T$. Let ω_k evolve following the equation below.

$$\omega_k = \begin{cases} 30, & \text{if } 0 \leq \frac{k}{f_s} < 10 \\ 70, & \text{if } 10 \leq \frac{k}{f_s} < 20 \\ 50, & \text{if } 20 \leq \frac{k}{f_s} < 30 \\ 80, & \text{if } 30 \leq \frac{k}{f_s} < 40 \\ 60, & \text{if } 40 \leq \frac{k}{f_s} < \text{T} \end{cases}$$

Let $c_{k|N}$ denote the roots of the characteristic polynomial of the autoregression, which can be written as $c_{k|N} = r_{k|N} \mathrm{e}^{-\mathrm{i}\omega_{k|N}}$, where $r_{k|N}$ is the modulus and $\omega_{k|N}$ is the phase of each root $c_{k|N}$. The dominant frequency is proportional to the phase of the complex roots: $f_{k|N} = \frac{f_s |\omega_{k|N}|}{2\pi}$. Applying this procedure for a discrete and hybrid models on the simulated data, using an autoregressive model of order $p = 2$, $R = 1$ and $\mathbf{Q} = 10^{-3}\mathbf{I}$, $B = 100, \mathbf{Q}_q = 10^{-4}\mathbf{I}$ the following estimations are obtained (Figure 1). A low autoregressive order was used for increased tractability of the poles. The values for the remaining parameters were set empirically.

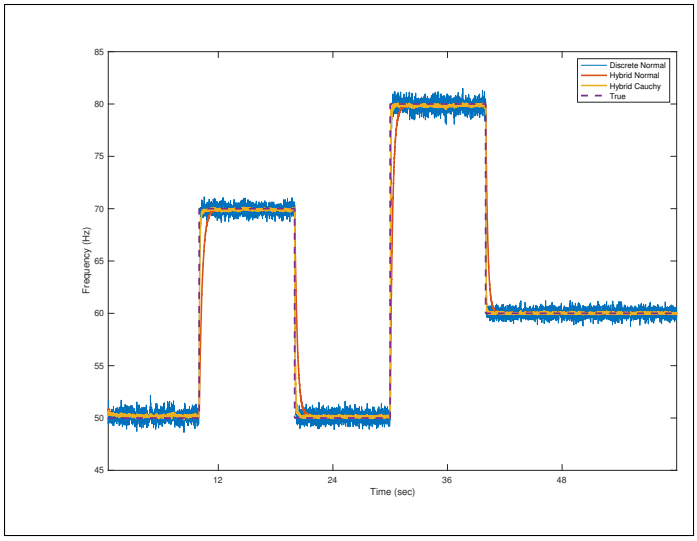

Fig. 1. Instantaneous frequency tracking

All models are able to track the true frequency evolution, but the hybrid ones result in much less fluctuation around the true instantaneous value. Hybrid Cauchy Model does not suffer from the lag of the Hybrid Normal model at the times of abrupt frequency changes. Mean Square Errors (MSE) are shown in Table 1. Values shown are averages and standard deviations for 100 runs.

TABLE I
MEAN SQUARE ERRORS ON SIMULATED DATA

Discrete Normal	Hybrid Normal	Hybrid Cauchy
(3.1836 ± 0.0117)	(2.1229 ± 0.0075)	(1.2607 ± 0.0094)

A lower covariance matrix could potentially be used for the discrete model, but as it will become apparent in the next section of the paper, such a strategy does not help in efficiently smoothing EEG spectrograms.

B. EEG Spectrograms under propofol & ketamine

Spectrograms for propofol and ketamine were calculated using the hybrid and discrete parametric methods, as well as widely used non-parametric methods including the periodogram, the multitaper spectrogram, the state-space periodogram and the recently developed state-space multitaper spectrogram [14]. For the multitapers, 3 tapers were used at a 2Hz spectral resolution. The time window of assumed stationarity was 2 seconds. The performance of our models was compared to that of purely discrete methods that have been used before for EEG spectrogram generation under anesthesia [6] and to that of the non-parametric approaches mentioned above. The EEG recordings are part of de-identified data collected from patients at Massachusetts General Hospital (MGH), as a part of a MGH Human Research Committee approved protocol.

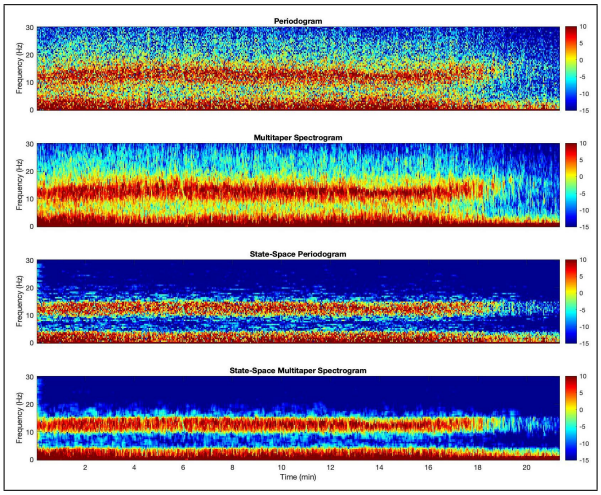

Fig. 2. Non parametric spectrograms of human EEG under propofol

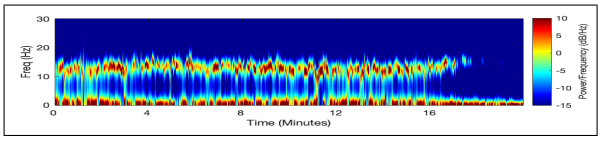

Fig. 3. Discrete Normal parametric spectrogram of human EEG under propofol

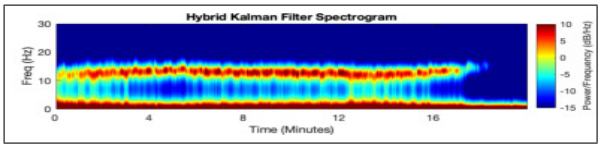

Fig. 4. Hybrid Normal parametric spectrogram of human EEG under propofol

In the case of propofol, non-parametric spectrograms (Figure 2) clearly show an α frequency band in addition

to the slow oscillations, as expected [1]. The periodogram and the multitaper spectrogram are quite noisy. On the other hand, their state-space counterparts seem to underfit as it is very implausible that the α band is evolving as a completely straight line in an actual EEG experiment.

Autoregressive models of order $p = 14$ were used to calculate the parametric spectrograms. In the purely discrete case (Figure 3), the resulting spectrogram seems able to grasp the subtle changes in frequency but is not able to reduce the underlying noise to the same extent as the hybrid spectrogram. The hybrid Normal model (Figure 4) performs better in identifying the slight frequency changes across the α band without introducing extra noise. Since changes between Normal and Cauchy models are not easily detected by the naked eye, instead of showing the Cauchy spectrogram, the spectrogram of differences in frequency power between the two methods is shown (Figure 5). Cauchy model yields similar spectrogram, with slight differences across the α frequency band.

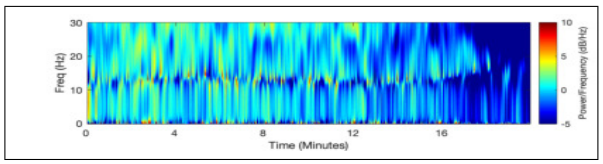

Fig. 5. Difference in power estimated by Normal and Cauchy models for propofol

The performance of the discrete model with a low covariance matrix (in the order of $10^{-5}\mathbf{I}$) as a measure to reduce coefficients' variability and spectrogram noise was tested on the propofol dataset. Propofol was chosen as the temporal evolution of the prominent slow and α frequency bands are smoother than ketamine's $\beta-\gamma$ band and thus it will be easier for the discrete model to try and capture data smoothness. As shown in Figure 6, its performance deteriorated. This is due to a low covariance matrix that prevents the model from correctly estimating the changes in signal frequency and introduces artifacts in the spectrogram. Exponential smoothing was also tested as a potential smoothing method but led to an even lower quality spectrogram than the one in Figure 6.

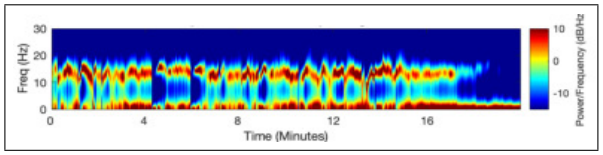

Fig. 6. Discrete Normal spectrogram of human EEG under propofol with lowered $\mathbf{Q}$

For ketamine, a β-γ frequency band is apparent in all non-parametric spectrograms (Figure 7) with peaks at around 30Hz and 50Hz, as expected [1]. The periodogram and multitaper spectrogram seem to identify changes across the bands but they suffer from high noise. On the other hand, as it was the case with propofol, state-space periodogram and multitaper spectrogram oversimplify the frequency evolution and yield a constantly evolving β-γ band. Order $p = 11$ was selected for the parametric spectrograms. The discrete one (Figure 8) is quite noisy, while the hybrid (Figure 9) provides a better result. Differences in power between the two hybrid models is shown in Figure 10. After minute 5, differences start to be more prominent: An indication that when power at the higher frequency bands increases, the use of Cauchy noise allows the model to better identify the change without introducing extra noise.

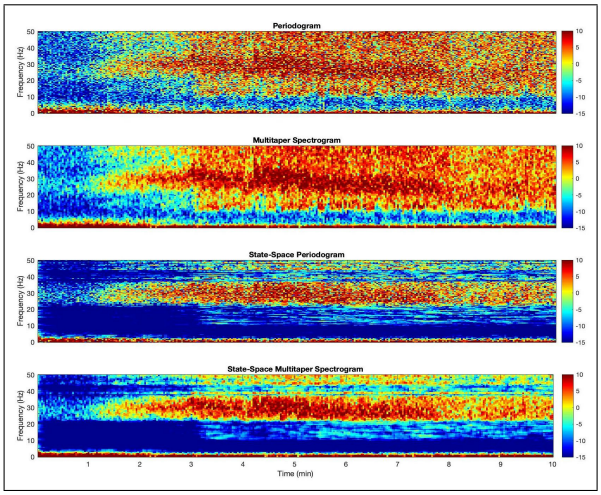

Fig. 7. Non-parametric spectrograms of human EEG under ketamine

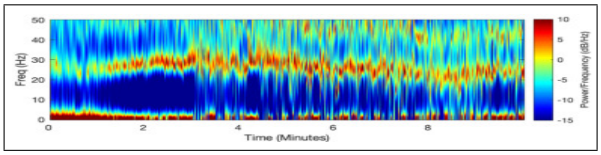

Fig. 8. Discrete Normal parametric spectrogram of human EEG under ketamine

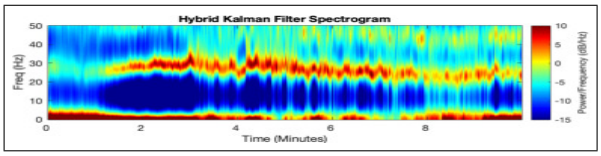

Fig. 9. Hybrid Normal parametric spectrogram of human EEG under ketamine

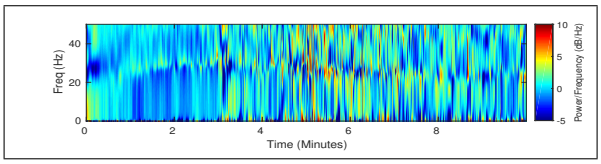

Fig. 10. Difference in power estimated by Normal and Cauchy models for ketamine

The conclusion drawn by visual inspection for smoother spectrograms of the hybrid models is corroborated by the roughness metrics of Table II, as defined in (28). The reported final value for each model is the average of the

metrics for each individual coefficient taking part in the model.

TABLE II
ROUGHNESS METRICS

	Discrete Normal	Hybrid Normal	Hybrid Cauchy
Propofol	$3.3380{\cdot}10^{-5}$	$6.5674{\cdot}10^{-7}$	$2.3795{\cdot}10^{-6}$
Ketamine	$3{\cdot}10^{-3}$	$9.4747{\cdot}10^{-5}$	$2.2733{\cdot}10^{-4}$

As expected, the hybrid Cauchy model yields less smooth coefficients than the hybrid Normal but much smoother than the discrete Normal. We believe that the hybrid Cauchy model estimates the instantaneous frequency with higher accuracy, balancing the tendency for oversmoothing of the hybrid approach by the heavier tales of the Cauchy distribution and the proposal of larger variations of the autoregressive parameters.

Apart from the better performance in simulated data, log-likelihood values of Table III corroborate the aforementioned hypothesis. Averages and standard deviations are shown for 100 runs.

TABLE III
LOG-LIKELIHOOD VALUES FOR HYBRID MODELS

	Propofol	Ketamine
Hybrid Normal	$(4.0854 \pm 0.03485) \cdot 10^5$	$(8.3613 \pm 0.18) \cdot 10^4$
Hybrid Cauchy	$(4.7468 \pm 0.1472) \cdot 10^5$	$(1.0306 \pm 0.083) \cdot 10^5$

IV. Discussion

The hybrid filter is able to compromise efficiently the underfitting regime of the state-space multitaper periodogram with the overfitting regime of the multitaper spectrogram. In relation to the purely discrete version, the hybrid filter is better able to capture the smoothness in the data, while at the same time this ability does not prevent it to identify sharp frequency changes.

In this hybrid approach, no prior knowledge about the evolution of the parameters in the time intervals between the observations is assumed. As a result, the autoregressive coefficients are assumed to follow a random walk. The proposed model can become even more potent in calculating frequency spectra if some prior knowledge about the parameters is available that will guide a reasonable choice of the state transition matrix. On the other hand, predicting a priori how the parameters are expected to evolve is not a trivial issue and demands a fair amount of research and experience with EEG datasets and effects of anesthetics on the brain. Nevertheless, it is expected that incorporating such prior information can enhance the model's performance even more, since by integrating the state equation, the model will be able to predict values for the parameters even in the time intervals between the observations where no new information is available.

Concerning the two different hybrid models, our results agree with [8] and indicate that the use of Cauchy noise is also advantageous when calculating spectra of EEG under anesthesia, in addition to EEG spectral analysis for Event Related Desynchronization that is discussed in [8]. On the other hand, fitting of Cauchy models is more computationally expensive due to the need for Monte Carlo simulations. Gaussian assumptions lead to convenient closed form solutions and allow the use of Kalman Filter that is perfectly suited for online use and do not impose the high computational burden of Monte Carlo simulations, which, even though can be alleviated by parallelization, requires substantial hardware. For this reason both models are presented in this paper to allow for optimal trade-off between desirable accuracy and use of resources one wants to allocate.

Overall, our method provides an accurate tool for frequency analysis of non-stationary data. The suitability of our method for online use, coupled with its ability to provide smoother spectrograms, suggests that it can be used as a monitoring tool for real time brain state tracking under general anesthesia that will further potentiate the design of closed loop systems for automatic and precise control of brain states.

ACKNOWLEDGMENT

K.K. wishes to thank Mr. Andrew Song for his valuable advice as well as Dr. Sourish Chakravarty and Dr. Oluwaseun Johnson-Akeju for the EEG datasets.

References

[1] Purdon PL, Sampson A, Pavone KJ, Brown EN. "Clinical Electroencephalography for Anesthesiologists: Part I: Background and Basic Signatures." Anesthesiology. 2015 Oct;123(4):937-60

[2] Purdon et. al. "Electroencephalogram signatures of loss and recovery of consciousness from propofol." Proc Natl Acad Sci U S A. 2013 Mar 19; 110(12): E1142E1151. Published online 2013 Mar 4. doi: 10.1073/pnas.1221180110

[3] Yang et al. "Developing a personalized closed-loop controller of medically-induced coma in a rodent model." J Neural Eng. 2019 Mar 11. doi:10.1088/1741-2552/ab0ea4

[4] Thomson, D. J. (1982) "Spectrum estimation and harmonic analysis." Proceedings of the IEEE, 70, 10551096

[5] Schlogl A, Flotzinger D, Pfurtscheller G. (1997) "Adaptive autoregressive modeling used for single-trial EEG classification." Biomed Tech (Berl). 1997 Jun;42(6):162-7. EEE Trans Biomed Eng. 1998 May;45(5):553-62.

[6] Georgiadis et al. (2009) "Kalman smoother based time-varying spectrum estimation of EEG during single agent propofol anesthesia.

[7] Mohammad Emtiyaz Khan, Deshpande Narayana Dutt (2007) "An Expectation-Maximization Algorithm Based Kalman Smoother Approach for Event-Related Desynchronization estimation from EEG." IEEE transactions on biomedical engineering, VOL. 54, NO. 7, July 2007 1191

[8] Ting et al. (2010) Spectral estimation of nonstationary EEG using particle filtering with application to event-related desynchronization (ERD). IEEE Transaction on Biomedical Engineering, vol. 58. no. 3, pp. 321-331, 2011. DOI 10.1109/TBME.2010.2088396

[9] Jones RH (1981). Fitting a Continuous Time Autoregression to Discrete Data. Applied Time Series Analysis II, pp. 651682.

[10] A.C. Harvey & James H. Stock (1985). The estimation of higher order continuous time autoregressive models Econometric Theory, 1, 97-112.

[11] Z. Ghahramani and G. E. Hinton, Parameter estimation for linear dynamical systems, Univ. Toronto, Dept. Comput. Sci, Toronto, ON, Canada, Tech. Rep., 1996.

[12] Schwiegelshohn et al. (2016). "A resampling method for parallel particle filter architectures." Microprocessors and Microsystems 47 (2016) 314320

[13] James Ramsay, Giles Hooker, Spencer Graves (2009). "Functional Data Analysis with R and MATLAB." Springer, ISBN:978-0387981840

[14] Kim et al (2018). State-space multitaper time-frequency analysis. Proc Natl Acad Sci U S A. 2018 Jan 2;115(1):E5-E14