Spectral estimation of nonstationary EEG using particle filtering with application to event-related desynchronization (ERD)

Chee-Ming Ting, PhD
Sh-Hussain Salleh
Z. M. Zainuddin
Arifah Bahar

Available at: http://works.bepress.com/chee-ming_ting/4/
Spectral Estimation of Non-Stationary EEG using Particle Filtering with Application to Event-Related Desynchronization (ERD)

Chee-Ming Ting, Sh-Hussain Salleh, Z. M. Zainuddin, and Arifah Bahar

Abstract—This paper proposes non-Gaussian models for parametric spectral estimation with application to event-related desynchronization (ERD) estimation of non-stationary EEG. Existing approaches for time-varying spectral estimation use time-varying autoregressive (TVAR) state-space models with Gaussian state noise. The parameter estimation is solved by conventional Kalman filtering. This study uses non-Gaussian state noise to model AR parameter variation with estimation by Monte Carlo particle filter (PF). Use of non-Gaussian noise such as heavy-tailed distribution is motivated by its ability to track abrupt and smooth AR parameter changes which are inadequately modeled by Gaussian models. Thus, more accurate spectral estimates and better ERD tracking can be obtained. This paper further proposes a non-Gaussian state space formulation of time-varying autoregressive moving average (TVARMA) models to improve the spectral estimation. Simulation on TVAR process with abrupt parameter variation shows superior tracking performance of non-Gaussian models. Evaluation on motor-imagery EEG data shows that the non-Gaussian models provide more accurate detection of abrupt changes in alpha rhythm ERD. Among the proposed non-Gaussian models, TVARMA shows better spectral representations while maintaining reasonable good ERD tracking performance.

Index Terms—Time-varying autoregressive models, particle filters, event-related desynchronization

I. INTRODUCTION

Event-related desynchronization (ERD) and synchronization (ERS) are used to represent frequency-specific changes of on-going EEG activity, induced by specific stimulus, which consist either of decrease or increase of power in specific frequency band. Left and right-hand motor imagery shows alpha rhythm ERD. Conventional Fourier transform based spectral analysis is limited by stationary assumptions and suffers tradeoff between time and frequency resolution. The frequency resolution can be improved by using parametric spectral analysis. Common parametric models used are autoregressive (AR) and autoregressive moving average (ARMA) models which are able to represent appropriately many kinds of natural signals such as speech and EEG. To better model non-stationary signals, time varying AR (TVAR) and ARMA (TVARMA) models have been used, where the AR parameters vary instantaneously with time. The TVAR models were proposed for EEG modeling by [1] and applied for EEG analysis [2]. The TVAR models enable non-stationary spectral analysis by generating instantaneous estimates of power spectrum, thus provide high time-frequency resolution. This model is also adopted here. The TVAR model is formulated into state-space form to enable estimation of TVAR coefficients by using Kalman filter (KF), which is optimal in mean-square sense. TVAR models with KF have been used extensively in power spectrum estimation for biomedical signals in general [2], [3], [4], [5], and in EEG spectral estimation for ERD and ERS [6], [7], [8]. Tarvainen [7] proposed Kalman smoother for TVARMA for ERS and showed faster tracking of ERS and better time-frequency resolution compared with recursive least square (RLS) method and STFT. The work was extended in [6] using the expectation-maximization (EM) for parameter estimation. Results on ERD showed better performance.

Most of the above mentioned studies for non-stationary biomedical signal analysis use TVAR models in linear and Gaussian state-space form where the state and observation noise are assumed to be Gaussian. However, there are processes for which linear Gaussian modeling is inappropriate, for instances time series with abrupt and smooth changes of means. The simple linear Gaussian models with small Gaussian noise variance cannot detect rapidly the abrupt changes while use of large variance produces noisy estimates. The poor AR parameter estimates will result in inaccurate spectrum. This Gaussian model problem was addressed by Kitagawa in an early study [9]. This problem is also inherent in the ERD/ERS problem, where the underlying AR coefficients of EEG process exhibit abrupt changes (perhaps event-related) [7], which cannot be tracked rapidly by Gaussian models and thus suffer tracking lag. To overcome the Gaussian model problem, Kitagawa [9] proposed the non-Gaussian state-space approach as alternative to model nonstationary time series when Gaussian modeling is inappropriate. Monte Carlo filtering and smoothing were used. The state and observation noise distributions are not

necessarily Gaussian. The study shows that in modeling time series with abruptly changing means, use of heavy-tailed distribution such as Cauchy density for state noise gives better detection than the Gaussian noise.

The AR parameter estimation for TVAR model is a problem of online recursive estimation of posterior distributions of states, and its filtered distributions, given observations which arrive sequentially in time. The posterior distributions can be evaluated in closed form analytical solution for only a few cases including the linear Gaussian models using KF. For nonlinear non-Gaussian models, of which the TVAR model with non-Gaussian noise is an example, the analytical solution is intractable. Classical approximations include extended Kalman filter [10] which relies on linearity approximation and Gaussian sum filter [11] constrained by Gaussian mixture approximation of posterior distribution. Another alternative is the sequential Monte Carlo (SMC) method which is not constrained by any linearity and Gaussianity assumptions on the models. SMC methods are simulation-based methods where the posterior distributions are represented by a large numbers of samples or particles with associated weights (refer to [12] for introduction and [13], [14] for recent advances). The first operational SMC methods is bootstrap filter [15], followed by several variants under names of Monte Carlo filter [16], Particle filter (PF) [17], and others. The advantages of SMC method are not only the inference of full exact posterior distributions for complex nonlinear and non-Gaussian estimation problems, but its implementation is efficient, can be in parallel and scalable [13]. The SMC methods have been successfully applied in various areas [18].

There are limited studies of nonlinear and non-Gaussian state-space models estimated using particle filters (PFs) for biomedical signal processing. The PF has been applied recently for event related potential estimation in [19] which however, used Gaussian noises and has not addressed AR modeling. Simulation studies by recent papers [20], [21] which used PFs for AR parameter estimation of simulated non-Gaussian TVAR process with abrupt parameter changes, showed better detection performance of non-Gaussian models than Gaussian models. To the best of authors’ knowledge, there are no studies on applying non-Gaussian models for spectral estimation of ERD and other nonstationary biomedical signal analysis in general. This paper proposes non-Gaussian TVAR state-space models with inference using particle filter for spectral estimation of nonstationary real EEG on ERD problems. We use non-Gaussian distributions on state noise for direct modeling of abrupt parameter changes. The proposed models are better in tracking abrupt and smooth AR parameter variation and give more accurate spectral estimates than Gaussian models. The underlying AR processes during ERD are hypothesized to exhibit abrupt changes. The non-Gaussian heavy-tailed models are expected to be able to track these changes more accurately, which give better detection of the highly non-stationary spectral components and hence more accurate ERD estimates. The non-Gaussian modeling of AR parameters in this work is distinct from [20], [21] which used non-Gaussian observation noise while maintaining Gaussian state noise, which is appropriate to capture observation outliers, but not for modeling parameter variations. Besides, [20], [21] lack of comparisons with more advanced techniques such as Kalman filter, and evaluation of model goodness of fit.

The AR models used to represent the spectral peaks may not be appropriate when both peaks and valleys are present in the true spectrums of real-world processes. This motivates the use of ARMA models capable of modeling both spectral peaks and valleys to provide better time-frequency resolution, and hence more accurate ERD estimation. Besides, the incorporation of MA coefficients may smooth the occasional spurious peaks induced by the proposed heavy-tailed TVAR model. Constrained by Gaussian and linearity assumptions, the existing approaches formulate the TVARMA models into linear-Gaussian form with estimation by KF [7]. To incorporate the MA coefficients in the non-Gaussian TVAR model, we further propose a non-Gaussian state-space formulation of TVARMA models, with coefficient estimation by the flexible PF. The state noise can be modeled by non-Gaussian heavy-tailed process. The driving noise in the observation equation can be viewed as linear combination of distributions weighted by MA coefficients. For model parameter estimation, we use Bayesian based joint state and parameter filtering approach with artificial parameter evolution [22], [23], [24].

The remaining of the paper is organized as follows: Section II describes methodology of ERD estimation using PF. The TVARMA process is formulated into non-Gaussian state-space form. The PF used for TVARMA coefficient estimation, static parameter estimation and ERD estimation are described. Section III presents simulation results and comparisons of the Gaussian and non-Gaussian models on ERD estimation and model fitness evaluation. Conclusion is given in final section.

II. METHODS

A. State-Space Formulation

The TVARMA \((p, q)\) model for signal at time \(t\) is given as

\[
\begin{align*}
y_t &= \sum_{j=1}^{p} a_{i,j} y_{t-j} + \sum_{k=0}^{q} h_{i,k} v_{t-k} \\
&= C_t x_t + G_t v_t
\end{align*}
\]

(1)

where \((a_{1,1}, a_{1,2}, \ldots, a_{1,p})\) and \((h_{1,0}, h_{1,1}, \ldots, h_{1,q})\) are time-varying AR and MA coefficients at time \(t\). \(y_t\) is the observation data which here is EEG signal. \(p\) and \(q\) are model orders of AR and MA parts respectively. In this work, the driving noise \(v_t\) is assumed to be independent and identically distributed (i.i.d.) Gaussian process with zero mean and variance \(\sigma_v^2\). \(v_t \sim N(0, \sigma_v^2)\). The TVAR and TVMA models are special cases of TVARMA models when \(q = 0\) and \(p = 0\), respectively. Here, we propose a non-Gaussian state-space formulation of TVARMA model to include non-Gaussian state noise. The flexible particle filter allows the estimation of ARMA coefficients to be solved in non-Gaussian state-space form. The state-space formulation of TVARMA process consists of observation equation (2) and state equation (3) as

\[
\begin{align*}
y_t &= C_t x_t + x_t^T G_t v_t \\
x_t &= x_{t-1} + w_t
\end{align*}
\]

(2)

(3)

The TVARMA process is written in compact form of (2) as a linear mapping of system state vector of TVAR and TVMA...
coefficients \( \mathbf{x}_t = [a_{t,1}, a_{t,2}, \ldots, a_{t,p}, b_{t,0}, \ldots, b_{t,q}]^T \) to the current observation \( y_t \), with \( \mathbf{C}_t = [y_{t-1}, y_{t-2}, \ldots, y_{t-p}, 0, \ldots, 0] \), vector of past observations. \( \mathbf{v}_t \) is a \((p + q + 1) \times 1\) vector of zero mean Gaussian driving noises \( \mathbf{v}_t \sim N(0, \sigma^2) \), scaled by \( \mathbf{x}_t^T \) and
\[
\mathbf{G} = \begin{bmatrix}
0_{p \times p} & 0 \\
0 & I_{(q+1) \times (q+1)}
\end{bmatrix}.
\]

(4)

Note that the driving noise process in (2) of the proposed models is modeled by linear combination of zero mean Gaussian distribution weighted by MA coefficients. The TVARMA coefficients are assumed to follow a simple first order Markov process with the evolution of states modeled by random walk model as in (3), where \( \mathbf{w}_t \) is i.i.d. zero mean Gaussian state noise with diagonal covariance matrix \( \mathbf{Q} = \sigma^2 \mathbf{I} \) with variance \( \sigma^2 \), \( \mathbf{w}_t \sim N(0, \mathbf{Q}) \). \( \mathbf{w}_t \) and \( \mathbf{v}_t \) are mutually independent for all \( t > 0 \). \( b_{t,0,q} \) is assumed to be stationary and fixed to 1. The observation density \( p(y_t | \mathbf{x}_t) \) and transition density \( p(\mathbf{x}_t | \mathbf{x}_{t-1}) \) of the model are given by

\[
p(y_t | \mathbf{x}_t) = N(y_t | \mathbf{C}_t, \gamma_t^2 \sigma^2)
\]

(5)

\[
p(\mathbf{x}_t | \mathbf{x}_{t-1}) = N(\mathbf{x}_t | \mathbf{x}_{t-1}, \sigma^2 \mathbf{I}).
\]

(6)

The term \( \gamma_t^2 \sigma^2 \) is the variance of sum of Gaussian noises \( \mathbf{v}_{t-1} \) scaled by MA coefficients \( b_{t,0,q} \), where

\[
\gamma_t^2 = \sum_{i=0}^{q} b_{t,i}^2
\]

based on the independency assumption of \( \mathbf{v}_t \).

Besides, the state noise can be non-Gaussian. This study models the state noise as a heavy-tailed non-Gaussian distribution instead of Gaussian one. The Cauchy density has larger deviation at both tails to predict larger parameter variations and tight spread at the central to avoid large fluctuating prediction for smooth changes. We use the typical heavy-tailed distribution i.e. Cauchy distribution. The Cauchy state noise used is characterized by

\[
\mathbf{w}_t \sim C(0, \mathbf{Q}_c), \quad \mathbf{Q}_c = q_c^2 \mathbf{I}
\]

(7)

where \( C(0, \mathbf{Q}_c) \) denotes Cauchy distribution with location 0 and diagonal dispersion matrix \( \mathbf{Q}_c \) with diagonal element \( q_c^2 \) assumed to be identical. The 1-dimensional Cauchy pdf is given as \( p_c(x) = q_c / (\pi(x^2 + q_c^2)) \). For Cauchy state noise, the transition density in (6) is re-written as

\[
p(\mathbf{x}_t | \mathbf{x}_{t-1}) = C(\mathbf{x}_t | \mathbf{x}_{t-1}, q_c^2 \mathbf{I}).
\]

(8)

The proposed non-Gaussian state-space model is now fully specified. The model parameters \( \sigma^2, \sigma_x^2 \) and \( q_c^2 \) are assumed to be fixed and do not vary with time. The observation noise \( \sigma^2 \) is assumed pre-specified. The unknown time-varying state vectors of ARMA coefficients \( \mathbf{x}_t \) are estimated using PF.

B. Particle Filtering

The estimation problem involves estimating sequentially in time the posterior density \( p(\mathbf{x}_t | y_{1:t}) \) where \( \mathbf{x}_{1:t} = \{x_1, \ldots, x_t\} \) and \( y_{1:t} = \{y_1, \ldots, y_t\} \) denote the sequence of EEG observations and state vector of TVARMA coefficients up to time \( t \) respectively. This study focuses on estimation of its marginal, filtered density \( p(\mathbf{x}_t | y_{1:t}) \). Mean of the filtered density is used as TVARMA coefficient estimates. The standard filtering recursion [18] used to estimate the filtered densities can be evaluated in closed form solution only for linear Gaussian models using KF and finite-state hidden Markov models. For nonlinear and non-Gaussian models, alternative simulation-based Monte Carlo methods, such as importance sampling are used where the filtered density can be approximated by empirical density formed by samples or particles and their associated weights [14].

\[
\hat{p}(\mathbf{x}_t | y_{1:t}) = \sum_{i=1}^{N} W^{(i)}_t \delta(x_t - x^{(i)}_t), \quad W^{(i)}_t = \tilde{w}^{(i)}_t / \sum_{j=1}^{N} \tilde{w}^{(j)}_t \quad (9)
\]

where \( \delta(.) \) denotes the Dirac delta mass located at \( x, x^{(i)}_t \) is \( i^{th} \) particle with attached normalized and un-normalized weights, \( W^{(i)}_t \) and \( \tilde{w}^{(i)}_t \). \( N \) is number of particles. To enable recursive estimation, sequential importance sampling (SIS) [12]-[14], a SMC method is used where particles at time \( t \) are sampled recursively using particles at time \( t-1 \) and weights can be updated recursively, by factorizing importance function in recursive form \( \pi(\mathbf{x}_{t-1} | y_{1:t-1}) = \pi(\mathbf{x}_{t-1} | y_{1:t-1}) \pi(\mathbf{x}_t | \mathbf{x}_{t-1}, y_{t}) \).

\[
\pi(\mathbf{x}_t | \mathbf{x}_{t-1}, y_{t}) = \pi(\mathbf{x}_t | \mathbf{x}_{t-1}, y_{t}) \quad \text{is used for filtered density.}
\]

To estimate \( p(\mathbf{x}_t | y_{1:t}) \), sample \( N \) particles of \( \mathbf{x}^{(i)}_t \) from \( \pi(\mathbf{x}_t | \mathbf{x}_{t-1}, y_{t}) \) with weights updated recursively as

\[
\tilde{w}^{(i)}_t = \frac{p(y_t | x^{(i)}_t) p(\mathbf{x}^{(i)}_t | \mathbf{x}_{t-1}^{(i)})}{\pi(x^{(i)}_t | \mathbf{x}_{t-1}^{(i)}, y_{t})}.
\]

(10)

Optimal importance function \( \pi(\mathbf{x}_t | \mathbf{x}_{t-1}^{(i)}, y_{t}) = p(\mathbf{x}_t | \mathbf{x}_{t-1}^{(i)}, y_{t}) \) is used to minimize variance of the importance weights. This study selects a simple choice of importance function i.e. the prior importance function \( \pi(\mathbf{x}_t | \mathbf{x}_{t-1}^{(i)}, \mathbf{x}_{t-1}) = p(\mathbf{x}_t | \mathbf{x}_{t-1}^{(i)}) \) with update of weights reduced to \( \tilde{w}^{(i)}_t = \tilde{w}^{(i)}_t p(y_t | \mathbf{x}_{t-1}^{(i)}) \) [12].

To overcome the weight degeneracy problem of SIS, re-sampling step is added where particles with negligible weights are removed and particles with significant weights are multiplied [13]. Re-sampling step is applied only when the effective sample size \( ESS = \sum_{i=1}^{N} (W^{(i)}_t)^2 \) is less than a specified threshold \( ESS < N_{T} \). The SIS-Resampling (SIR) used for this paper is summarized as in Algorithm 1.

The conditional densities for sampling and computation of weights for our models are given in (5) and (8). The sampling using Cauchy noise will generate a large number of small value samples by its central region and small number of large value samples by its tails, to predict smooth and abrupt changes respectively. These samples are later weighted by the observation conditional probabilities to determine the type of changes. Gaussian noise, however, generates either small or large samples, depends on its variance size and thus is unable
Algorithm 1: SIS - Resampling (SIR)

At time $t = 0$,

**Step 0: Initialization**
- For $i = 1, \ldots, N$, sample $x_i^{(i)} \sim N(0, P_0)$ and set normalized initial weights $W_i^{(0)} = 1/N$.

For time $t \geq 1$,

**Step 1: Importance Sampling**
- For $i = 1, \ldots, N$, sample $x_i^{(t)} \sim p(x_t | x_{t-1}^{(t)}) = C(x_t | x_{t-1}^{(t)}, q_t^2 I)$.
- For $i = 1, \ldots, N$, evaluate the importance weights:
  \[
  \tilde{w}_i^{(t)} = W_{t-1}^{(i)} p(y_t | x_i^{(t)}) = W_{t-1}^{(i)} N(y_t | C(x_i^{(t)}, x_{t-1}^{(t)}), \sigma^2_t).
  \]
- For $i = 1, \ldots, N$, normalize the importance weights:
  \[
  W_t^{(i)} = \frac{\tilde{w}_i^{(t)}}{\sum_{i=1}^N \tilde{w}_i^{(t)}}.
  \]

**Step 2: Resampling**
- If $ESS < N_T$, resample with replacement $N$ particles from the set $\{x_i^{(t)}\}_{i=1}^N$ according to probabilities proportional to weights $\{W_t^{(i)}\}_{i=1}^N$ and reset $W_t^{(i)} = 1/N$, for $i = 1, \ldots, N$.

to provide proper detection of both smooth and abrupt changes simultaneously. Small variance cannot detect rapid changes while large variance produces noisy estimates.

### C. Estimation of Static Parameters

The static model parameters, i.e. variance $\sigma^2_u$ and dispersion $q_v$, for the respective Gaussian and Cauchy state noise assumed unknown and need to estimate, are denoted as $\theta$. Many SMC techniques have been proposed to solve static parameter estimation for general state-space models [25]. The maximum likelihood (ML) methods based on either gradient or expectation-maximization (EM) approach are computationally very expensive and suffer sampling error due to use of biased approximated likelihoods. Here, we consider a more pragmatic Bayesian estimation where the unknown parameters $\theta$ are augmented with the state vector $x_t$ as a single vector $z_t = [x_t^T, \theta_t^T]^T$ with $\theta_t = \theta$, and the posterior density for both state and parameters $p(z_t | y_{1:t})$ is estimated using standard PF [22], [23]. To overcome the sample degeneracy problem, an artificial evolution is introduced to the static parameter as $\theta_t = \theta_{t-1} + \varepsilon_t$, where $\varepsilon_t$ is a small noise term, pretending the static parameters are indeed time-varying. This method, however, implies “loss of information”. To compensate this problem, Liu and West [24] improved the artificial evolution method by sampling the static parameters from a kernel smoothed density with shrinkage modification and with the smoothing parameter controlled by a discounting factor $\delta$ suggested around 0.95–0.99. The conditional evolution density of parameters is now given as

\[
p(\theta_t | \theta_{t-1}) \sim N(\theta_t | a\theta_{t-1} + (1-a)\tilde{\theta}_{t-1}, h^2 V_{s-1})
\]

where $a = (3\delta - 1)/2\delta$ and $h^2 = 1 - a^2$. $\tilde{\theta}$ and $V_s$ are the mean and variance matrix of posterior density $p(\theta_t | y_{1:t})$. This method is sensitive to the setting of its initial density $p(\theta_0)$. Uniform distribution over an interval of possible values of an optimal $\theta$ is recommended [23]. The initial distribution can be tailored by prior information.

### D. Estimation of ERD

The mean of the filtered density of TVARMA coefficients estimated using SIR is used to compute the TVARMA coefficients $\{a_{i,j}\}_{j=1}^p$ and $\{b_{j,k}\}_{k=1}^q$ respectively as $\hat{a}_{i,j} = \sum_{i=1}^N \tilde{x}_{i,j}^{(t)} W_t^{(i)}$ for $1 \leq j \leq p$ and $\hat{b}_{j,k} = \sum_{i=1}^N \tilde{x}_{i,j}^{(t)} W_t^{(i)}$ for $1 \leq k \leq q$, where $\tilde{x}_{i,j}^{(t)}$ is the $j$th element of the particle vector $x_i^{(t)}$. Given the estimated TVARMA coefficients, the time varying power spectrum is computed as

\[
PSD(t, f) = \frac{\sigma^2_v}{f_s} \left| 1 + \sum_{k=1}^p \hat{b}_{k,j} e^{-2\pi if/k} ight|^2, \quad 0 \leq f \leq f_s / 2
\]

where $f_s$ is sampling frequency and $\sigma^2_v$ is error variance prediction. To show ERD at specific frequency band between $f_1$ and $f_2$, we calculate band power as average of power within the band $P_{b}(f) = \sum_{j=f_1}^{f_2} PSD(t, f)^2$ and ERD as relative band power $ERD(t) = (P_{b}(f) - P_{ref}) / P_{ref}$, where

\[
P_{ref} = \sum_{t=1}^T P_{b}(t) / (T_2 - T_1)
\]

is reference power calculated by averaging band power over a reference window from time $T_1$ to $T_2$ when ERD is expected to be absent (probably at the start of experiments) [6]. The ERD estimates are further smoothed by moving averaging technique.

### III. Experimental Results

#### A. Simulation Results

Consider an observation series artificially generated by an AR(4) process with Gaussian driving noise $v_t$ with zero mean and variance $\sigma^2_v = 1$, modified from [20] as follows

\[
y_{1:t} = -0.7348 y_{1:t-1} - 1.882 y_{1:t-2} - 0.7057 y_{1:t-3} - 0.885 y_{1:t-4} + v_t, \quad t = 1, \ldots, 500
\]

\[
y_{1:t} = 1.352 y_{1:t-1} - 1.338 y_{1:t-2} - 0.662 y_{1:t-3} - 0.24 y_{1:t-4} + v_t, \quad t = 501, \ldots, 1000
\]

\[
y_{1:t} = 0.37 y_{1:t-1} + 0.56 y_{1:t-2} + v_t, \quad t = 1001, \ldots, 1500
\]

The TVAR coefficients are set constant for 500-sample duration, with abrupt changes at every 500 sample interval to simulate state parameters with abrupt and smooth changes.

The ‘true’ coefficients given in (13) are estimated by filtered density means for TVAR models with Gaussian and Cauchy state noise, calculated by KF and PF respectively. For this simulation, all model parameters are not estimated,
assumed known and pre-specified. The model order is \( p = 4 \). The initial state distribution is \( N(0, I) \). For Gaussian state noise, \( \sigma^2 = 0.0005 \) is used. The variance is empirically set to satisfactorily detect abrupt AR parameter changes but with less noisy estimates during the stationary part. Additional experiments with large variance showed highly wiggly estimates for the stationary segments despite the ability to detect sudden jumps. The scale of Cauchy distribution is set to \( q_c = 0.005 \). Number of particles \( N = 8000 \) and re-sampling threshold \( N_t = N / 3 \) is set. The tracking of the four AR coefficients using Gaussian and Cauchy model from 10 realizations are shown in Fig. 1. The heavy-tailed model gives better tracking of abrupt changes, compared to the Gaussian model with gradual detection. The Gaussian model with small variance noise is unable to detect abrupt jumps. The heavy-tailed distribution allows larger deviation to predict sudden large changes. Although it induces noisier estimates in stationary part, the estimates are still acceptable. The heavy-tailed model is able to respond rapidly to jumps, however, occasionally produces large deviation from the true values. This may be due to that the heavy-tailed noise produces large deviation samples in a random walk manner using prior importance function to predict abrupt changes; however, these samples may occasionally not be weighted properly and thus produces noisy estimates. The use of prior importance function is justified by its simplicity and less computation. The use of optimal importance function which includes current observation \( y_t \) information in sampling and computation of sample weights may solve this problem, and can be addressed in future work.

Instantaneous log mean square error (LMSE) [20] in function of time is used as objective evaluation for the tracking performance of Gaussian and Cauchy model, as shown in Fig. 2. The instantaneous LMSE at time \( t \) for \( R \) realizations and model order \( p \) is given by

\[
LMSE_t = \log \frac{1}{R} \sum_{r=1}^{R} \left( \frac{1}{p} \sum_{i=1}^{p} (\hat{a}_{i,r} - a_i)^2 \right)
\]

(14)

where \( \hat{a}_{i,r} \) is \( i^{th} \) AR coefficient estimate at time \( t \), and \( a_i \) is the true values. It is shown that Cauchy model gives lower MSE than Gaussian model especially at abrupt change regions, but occasionally induce higher MSE at stationary parts. The average mean squared error (MSE) for Gaussian and Cauchy model are 0.0579 and 0.0386 respectively.

Motivated by analysis in [9], estimated filtered densities using Gaussian and non-Gaussian model are compared. Fig. 3 shows the estimated filtered density \( p(a_{1,3} | y_{1:t}) \) of the first AR coefficient by the Gaussian model using KF. The estimated filtered densities are identical Gaussians. The Gaussian model gives slow detection, reflected by gradual shifts of filtered densities to follow the abrupt changes. Fig. 4 shows the estimated means and 5th, 25th, 75th, 95th percentiles of the filtered density by Gaussian model. The percentile deviations are small and follow the mean values. Fig. 5 shows one realization from kernel smoothed filtered densities for first AR coefficient by Cauchy model using PF. The estimated densities are non-Gaussian. During the sudden shift of AR parameters, the densities become heavy-tailed in one side, even multimodal, to follow abruptly the sudden transients. This is followed immediately, in next prediction, by abrupt shift of estimated density to the new AR mean values. Fig. 6 shows means and respective percentiles of the filtered densities by Cauchy model. Compared with Fig. 4, it is clearly seen that the Cauchy model predicts smoother estimate at stationary part while maintaining faster tracking of jumps. The densities only occasionally induce large, even spurious deviation from the mean, however it affects outermost percentiles and the mean values remain relatively stable. The results in this simulation are consistent with [9]. The non-Gaussian models give better detection of abrupt change by having heavy-tailed distribution and jump of means, instead of gradual shifting of density itself for Gaussian models.
Fig. 3. Estimated filtered density $p(a_t \mid y_{1:t})$ by Gaussian model.

Fig. 4. Estimated filtered mean (solid line) and the 5th, 25th, 75th, 95th percentiles (dotted lines) by Gaussian model.

Fig. 5. Kernel smoothed filtered density $p(a_t \mid y_{1:t})$ by Cauchy model.

Fig. 6. Estimated filtered mean (solid line) and the 5th, 25th, 75th, 95th percentiles (dotted lines) by Cauchy model.

B. Tracking of ERD Dynamics

Dataset

The hand movement imagery causes alpha band desynchronization in EEG. The Cauchy models with PF are evaluated on capability to track ERD on motor-imagery data. The evaluation uses Motor-Imagery EEG data (Dataset IIIa) provided by Graz University of Technology (Subset of dataset for BCI Competition III [26]). The dataset consists of 3 subjects. The task was to perform imagery left hand, right hand, foot or tongue movements according to a cue. The experiment consists of several runs with 40 trials each. The first 2 s were quite and at $t = 2$ s an acoustic stimulus indicated beginning of a trial, and a cross (“+”) is displayed for 1 s. Then from $t = 3$ s an arrow to the left, right, up or down was displayed for 1 s, and at the same time subjects imagining a left hand, right hand, tongue or foot movement, respectively, until the cross disappeared at $t = 7$ s. Each of the 4 cues was displayed 10 times within each run in a randomized order [27]. 64-channel EEG data is recorded with sampling frequency 250 Hz. In this paper, ERD of left-hand movement for all subjects are studied. EEG data of labeled trials without artifacts, measured at position C3 and C4 are used. Average EEGs are obtained from these trials subsequent use in ERD analysis.

Model Order Selection

Selection of appropriate orders of AR and ARMA models is important for parametric spectral analysis. Too high a model order causes overfitting and produces spurious spectral peaks while too low an order gives over-smoothed spectrum. The Akaike information criterion (AIC) and Bayesian information criterion (BIC) are used to select the optimal model order.

\begin{align}
\text{AIC} &= -2\ell_T(\hat{\theta}) + 2k \quad (15) \\
\text{BIC} &= -2\ell_T(\hat{\theta}) + \ln(T)k \quad (16)
\end{align}

where $\ell_T(\hat{\theta})$ is log-likelihood of model given an observations of length $T$, $\hat{\theta}$ is ML estimates of model parameters, and $k$ is the number of estimated parameters. The optimal model order is the one that minimizes the AIC and BIC values. For TVAR state-space models, maximizing the analytical likelihood computed by KF for Gaussian case and Monte Carlo approximated likelihood for non-Gaussian case are computationally expensive. The joint state-parameter estimates by PF are only optimal in minimum mean-squared error (MMSE) sense. Thus, the ML $\ell_T(\hat{\theta})$ required for AIC and BIC are not straightforwardly available. Motivated by [28], this paper assumes the non-stationary EEG process to be locally stationary over short-time segment (typically 1 s) with stationary AR and ARMA models fitted to each segment. The model parameters are estimated by ML method and AIC and BIC are evaluated on each segment to select the optimal model order. EEG segments of 1 s period before and after the event for all subjects are evaluated for this study. The average EEGs at position C4 are used. Table I. shows means and standard deviations of optimal AR and ARMA model orders using AIC and BIC criterions, averaged over all 1 s segments for all subjects. AIC tends to overfitting while BIC tends to underestimate. Compromise between AIC and BIC suggests AR(6) and ARMA(4,2). However, the observed spectrum by ARMA(4,2) are over-smoothed and higher AR order of $p = 6$ is selected to capture more accurately the spectral details. Fig. 7 shows AIC and BIC values as a function of order $p$ for...
is also estimated and included.

After event

The state noise \( t \) according to different state noises, i.e. zero mean Cauchy and TVMA coefficients are assumed to evolve randomly. Variance gives smoother spectrum. Hence, the TVAR and TVARMA models in subsequent analysis.

Additional experiments showed that increase of the Gaussian noise is used instead to obtain smoother spectrum. Has been confirmed in additional experiments. For this reason, coefficients increase the variances of driving noises, and thus noisy spectral peaks. This characteristic to its generated large number of small value samples of MA coefficients.

Results and Discussion

Comparison of time-varying spectral estimation for ERD is performed, between Gaussian TVAR(6) model, the proposed non-Gaussian Cauchy TVAR(6) and TVARMA(6,2) models. For the TVAR models, state estimation of the Gaussian and Cauchy models are solved by KF and PF respectively. Observation noise variance \( \sigma^2 = 0.1 \) and initial state distribution \( x_0 \sim N(0, I) \) are set throughout. For Gaussian model, the variance of Gaussian state noise is fixed \( \sigma^2 = 5 \times 10^{-7} \) based on ML criterion. For Cauchy TVAR models, the state and parameter are estimated simultaneously using PF with kernel density parameter estimation with shrinkage modification where \( \delta = 0.99 \) set throughout. To ensure positivity, \( q \) is parameterized by its logarithm as \( \ln q \). Thus the augmented state vector is \( z_t = [x_t^T, \ln q_{ct}]^T \). We assume the initial distribution \( q_{ct} \sim U[0.5 \times 10^{-4}, 2 \times 10^{-4}] \). Here, \( U[a,b] \) denotes uniform distribution over interval \([a,b]\).

For Cauchy TVARMA model, the modeling of evolution of MA coefficients which scale the variances of Gaussian noises in (3) is further discussed here. Larger values of MA coefficients increase the variances of driving noises, and thus smooth the spectrum, on the other hand small values produce peaks. This renders the use of Cauchy noise inappropriate due to its generated large number of small value samples of MA coefficients and hence noisy spectral peaks. This characteristic has been confirmed in additional experiments. For this reason, Gaussian noise is used instead to obtain smoother spectrum. Additional experiments showed that increase of the Gaussian variance gives smoother spectrum. Hence, the TVAR and TVMA coefficients are assumed to evolve randomly according to different state noises, i.e. zero mean Cauchy and Gaussian processes respectively. Thus, by denoting \( w_{l,p} = [w_{l,1}, \ldots, w_{l,p}]^T \) and \( w_{l,q} = [w_{l,1}, \ldots, w_{l,q}]^T \) state noises for AR and MA parts respectively, noise vector in (3) is re-written as \( w_t = [w_{l,p} w_{l,q}]^T \) where \( w_{l,p} \sim C(0, q_{l} I) \), \( w_{l,q} \sim N(0, \sigma^2_{l} I) \). The state noise \( w_t \) is still non-Gaussian and renders estimation performed by PF. \( \sigma^2_{l} \) is also estimated and included in \( z_t = [x_t^T, \ln q_{ct}, \ln \sigma^2_{l}]^T \). \( q_{ct} \sim U[0.5 \times 10^{-4}, 2 \times 10^{-4}] \) and \( \sigma^2_{l} \sim U[5 \times 10^{-4}, 5 \times 10^{-1}] \) is set. Number of particles used is

<table>
<thead>
<tr>
<th>Model</th>
<th>Selection by AIC</th>
<th>Selection by BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Order (p)</td>
<td>AIC value</td>
</tr>
<tr>
<td>AR(p)</td>
<td>7.33±1.75</td>
<td>-3.80±241.55</td>
</tr>
<tr>
<td>ARMA(p, 1)</td>
<td>3.17±1.60</td>
<td>-4.25±238.47</td>
</tr>
<tr>
<td>ARMA(p, 2)</td>
<td>4.33±1.51</td>
<td>-12.92±239.93</td>
</tr>
</tbody>
</table>

AR(p) evaluated on 1 s EEG segments before (left) and after (right) the event at position C4 for subject 3. The selected AR order \( p = 6 \) is indicated with vertical lines. These selected orders for the stationary case are used as approximations for TVAR and TVARMA models in subsequent analysis.

![Fig. 7. The AIC and BIC values as a function of AR order \( p \) for AR(p) on 1 s EEG segments before (left) and after (right) the event at position C4 for subject 3. The selected AR order \( p = 6 \) is indicated with vertical lines.](image)

\( N = 40000 \) with re-sampling threshold \( N_r = N/3 \). Number of particles required depends on dimensionality of state distributions and its posterior uncertainty [29]. Large \( N \) used in this study is due to high dimensionality with additional model parameters and high posterior uncertainty with use of Cauchy noise. The model parameters used are determined to give sufficiently accurate spectral estimates based on of visual inspection, guided by ML criterion. These parameters are used for subsequent experiments unless other specified.

Fig. 8 shows the comparison results of time-varying power spectrums for different methods on average EEG of subject 1 at position C4 for left-hand motor-imagery. The high resolution non-parametric quadratic time-frequency distributions (TFDs) family is used as baseline for comparison. Among the TFDs, we choose the modified B-distribution (MBD) which has been shown to outperform the well-known Wigner-Ville distribution (WVD) [30]. The power spectrums of the parametric methods are presented in decibel (dB) scale. The results for Cauchy AR and ARMA models are obtained by averages over 10 independent runs.

From Fig. 8, spectral estimates for all methods exhibit good time-frequency resolution and show ERD. The decrease in activity after 2 s and after the cue presented at 3 s, is reflected by decreased power at alpha frequency band (8-15 Hz). The three parametric methods especially both of the proposed Cauchy models are shown to outperform the non-parametric modified B-distribution in term of time-resolution. However, the modified B-distribution provides better frequency resolution (narrower frequency band). The time localization of spectral components is more critical for ERD estimation which is a time-varying process. Fig. 8 also shows that the Cauchy models result in better resolution in time-frequency domain than Gaussian model. The Cauchy models are able to localize precisely highly non-stationary spectral details which cannot be detected reliably by Gaussian models, e.g. the short-term components at 3.6 s. This better detection of rapidly varying spectral components will provide more accurate tracking of ERD. It can be seen that both Cauchy models exhibit faster tracking of ERD pattern, compared to Gaussian model which suffers a tracking lag. The Gaussian model shows gradual decrease of power at alpha band, started at 3 s.
On the other hand, Cauchy models results in sharp decrease of alpha rhythm power until almost complete and abrupt blocking of alpha power at 3.5 s. The better representation of highly non-stationary spectrum and hence more reliable ERD estimation by proposed Cauchy models is due to its ability to capture abrupt and smooth changes of the underlying AR parameters as demonstrated in previous simulation. By controlling MA noise variance $\sigma_g^2$, the incorporation of MA coefficients in Cauchy ARMA model introduces spectral zeros and is shown to improve significantly spectral representations of Cauchy AR models by smoothing the spurious peaks of Cauchy AR spectrum, e.g. spectral component at 2.7 s.

We calculate ERD estimation to further evaluate the detection ability, motivated by analysis and discussions in [6]. Fig. 9(a) shows the same average EEG at position C4 for left-hand motor imagery, its spectrum obtained by Cauchy TVARMA model, corresponding band power and ERD estimates. The frequency band for calculating the band power is chosen as upper alpha band 10 – 15 Hz. The reference power is obtained by averaging band power of first 2 s. The ERD estimates are smoothed by using moving average with 20-sample window. From the bottom of Fig. 9(a), the ERD is clearly indicated by power decrease relative to the reference period after the cue is presented at 3 s. The estimated band power and ERD exhibits patterns in accordance of activity changes reflected by the spectrum. The fast response by Cauchy model to capture ERD is indicated by sharp decrease of ERD estimates.

Fig. 10 shows ERD estimates by Cauchy TVARMA model using PF and Gaussian AR model using KF for average EEGs of subject 1 at position C4 and C3 for left-hand motor imagery. For the left-hand motor-imagery, it is clearly seen that the estimates for position C4 shows ERD, but those for C3 do not. The results clearly reconfirm the better detection of abrupt changes in ERD by Cauchy model, reflected by steeper decrease of ERD estimates.

Fig. 11 shows the ERD estimations by Cauchy TVARMA model, Cauchy TVAR model, and Gaussian TVAR model for three different subjects on left-hand motor imagery EEG at position C4. The averaged baseline of estimated sequence $q_{ct}$ over all subjects and all Monte Carlo runs for Cauchy TVAR model is $4.15 \times 10^{-4}$ while corresponding estimates of $q_{ct}$ and $\sigma_g^2$ for Cauchy TVARMA model are $4.18 \times 10^{-4}$ and $1.46 \times 10^{-2}$ respectively. Due to large inter-individual differences, the alpha frequency bands need to be adjusted individually [31]. In this paper, the subject-specific frequency bands are determined empirically based on visual judgment. The obtained alpha frequency bands are 10-15 Hz, 9-15 Hz, and 10-16 Hz for subject 1, 2, and 3 respectively. From Fig. 11, estimations by all methods are able to show band power.
decrease started at 3 s indicating ERD for all subjects. Each analyzed subject shows different characteristics of ERD pattern. Most rapid decrease of power is observed in subject 1, followed by subject 2 and 3. Results for all subjects show consistently the superiority of both proposed Cauchy models in faster tracking of ERD over Gaussian model. The drawback of tracking lag for Gaussian model is evident by much slower power decreases. Among the Cauchy models, TVARMA of tracking lag for Gaussian model is evident by much slower in faster tracking of ERD over Gaussian model. The drawback consistently the superiority of both proposed Cauchy models followed by subject 2 and 3. Results for all subjects show different pattern. Most rapid decrease of power is observed in subject 1, analyzed subject shows different characteristics of ERD

C. Model Evaluation

Objective evaluation for the goodness of fit of competing models is performed using log-likelihood and diagnostic checking. The log-likelihood can be approximated by

$$
\ell_T(\theta) = \log \hat{p}(y_{1:T} | \theta) = \sum_{i=1}^{T} \log \sum_{i=1}^{N} \hat{w}_i^{(i)}
$$

where \( p(y_{1:T} | \theta) \) is marginal likelihood of \( y_{1:T} \) and \( \hat{w}_i^{(i)} \) is un-normalized weights generated by PF. Higher likelihood indicates better model. For diagnostics, if the fitted model is good, sequence \( \{u_k\}_{k=1}^{T} \) with \( u_k = p(Y_k \leq y_k | y_{1:k-1}) \) should be i.d.d. \( u_k \sim U[0,1] \), which holds true for any time series models. \( p(Y_k \leq y_k | y_{1:k-1}) \) can be obtained by counting number of one-step-ahead predictions less than actual observation \( y_k \). Transforming \( u_k \) to normal residuals \( \hat{v}_k = \Psi^{-1}(u_k) \) with \( \Psi \) denotes Gaussian cdf, gives i.d.d. \( \hat{v}_k \sim N(0,1) \). Bowman-Shenton (BS) test and Ljung-Box (LB) test are used to check the hypothesis of normality and randomness of \( \hat{v}_k \) respectively. Details refer [32]. The evaluation results for Gaussian model and Cauchy models fitting the EEG data of each subject at C4 for left-hand motor imagery of Fig. 11 are given in Table II. The results presented are means and variances obtained for 10 Monte Carlo runs. The initial 1.5 s periods of “burn in” are discarded from evaluation. The initial distribution for Gaussian model is \( \delta \) function concentrated on \( x_0 = 5 \times 10^{-7} \). LB statistics of 5 lags are used. The respective 5% critical values for BS and LB statistics are 5.99 and 11.07. The BS tests suggest, for all subjects, the residuals from both Cauchy models are indeed normally distributed with better distributional behavior of TVARMA model, while not for Gaussian model. However, the LB tests indicate significant autocorrelations present in residuals for all models, with significantly less correlation from the Cauchy models than Gaussian model. These results of correlated residuals are consistent with another study using Gaussian TVAR in modeling speech signal [32], which may due to long term correlations in the signal. The fitted likelihoods of Cauchy models are significantly better than the Gaussian model, with Cauchy TVAR slightly better than Cauchy TVARMA model. The results suggest that Cauchy TVARMA and TVAR models are better fitting model to the data than Gaussian models. This also implies that Cauchy models give more accurate estimation of the underlying states of AR and MA coefficients which generate the observation data, and hence explains their higher spectral resolution and superior ERD tracking performance. The advantage of more accurate modeling by the proposed Cauchy models motivates their application beyond spectral estimation to classification problems where better accuracy may be expected.

IV. CONCLUSION

This paper proposes non-Gaussian models with particle filtering for spectral estimation of non-stationary EEG for ERD, and evaluated the methods on a simulated series and real
EEG data. Previous studies on spectral estimation for ERD/ERS used KF which relies on assumption of linearity and Gaussianity of TVAR models. This study formulates the TVAR models into non-Gaussian state-space form where the evolution of TVAR coefficients in state equation is modeled by heavy-tailed noise, with coefficient estimation by PFs. The results show the advantages of the proposed models over Gaussian models in ERD estimation in term of higher time-frequency resolution and better tracking performance. This is due to better detection of the abrupt and smooth changes of underlying AR parameters. This is observed in simulation results that by having tails at one-side and jump of means, the non-Gaussian Cauchy models are more capable to follow the abrupt changes, compared with Gaussian model by gradual shifting of densities. Further support comes from better fitting of Cauchy models to the data evaluated using log-likelihood and statistical diagnostics. The Gaussian model under-fits, gives over-smoothed spectrum and suffers ERD tracking lags. The Cauchy AR model exhibits faster tracking ability, however, may be too fit and produces occasional spurious spikes. The further proposed TVARMA state-space model achieves a balance between reasonable good tracking performance and better spectral representations.

**ACKNOWLEDGMENT**

The authors would like to thank the reviewers for valuable comments and suggestions.

**REFERENCES**


