BAYESIAN DETECTION OF SINGLE-TRIAL EVENT-RELATED POTENTIALS

Maria Rosario Mestre, Simon J. Godsill, William J. Fitzgerald

University of Cambridge, Signal Processing Laboratory, Cambridge, UK
{mrm46, sjg30, wjf1000}@cam.ac.uk

ABSTRACT

The goal of this paper is to build a detector of event-related potentials (ERP) in single-trial EEG data. This problem can be reformulated as a parameter estimation problem, where the parameter of interest is the time of occurrence of the ERP. This type of detector has clinical applications (study of schizophrenia, fatigue), or applications in brain-computer-interfaces. However, the poor signal-to-noise ratio (SNR) and lack of understanding of the noise generating process make this a challenging task. In this paper, we take a Bayesian approach, samples are drawn from the posterior of the parameter of interest using Markov chain Monte Carlo (MCMC). Different noise covariances from Gaussian processes are tested. We show that it is possible to pick up the ERP signal in spite of the poor SNR with an appropriate choice of noise covariance structure.

Index Terms— EEG, event-related potentials (ERP), MCMC, Gaussian process.

1. INTRODUCTION

Electrical activity from the brain can be measured with electroencephalography (EEG). This type of recordings help unveil recurrent patterns, such as event-related potentials (ERPs). An ERP is a type of signal that arises before or after a specific stimulus, usually consisting of a number of positive or negative peaks, also called components. In order to study them, the typical procedure is to present a series of stimuli to a participant while recording their EEG. The whole recording is then cut into smaller intervals containing a single stimulus. Each of these intervals is called a trial. Traditionally, the ERP signal was then recovered by taking the ensemble average over all aligned trials. However, it is commonly accepted the ERP waveform has inter-trial and inter-subject variability, and thus information is lost by taking the average [1].

Single-trial analysis aims at estimating the ERP in a single trial. There is an extensive literature on single-trial analysis, refer to [2] for a review. This type of analysis has applications in the study of certain conditions, such as schizophrenia [3] or fatigue [2]. The detection of single-trial ERPs can also be used in brain-computer-interfaces [4]. Estimating an ERP at single-trial level is a challenging task, primarily due to the poor signal-to-noise ratio (SNR). Secondly, there is little understanding of the noise generating process. Many studies use i.i.d. noise (e.g., [5], [6]), which fails to capture the smooth nature of EEG data. A better understanding of the probabilistic properties of noise could lead to better algorithms. Added to the poor SNR and lack of good models, there is also a lack of ground truth. The typical ERP shape is difficult to identify in most single trials by visual inspection. However, there is sometimes some side information that can be used to validate a model. In this paper, we work with ERPs from the motor cortex associated to the movement of a hand, and it is a known fact that this type of ERP will closely precede the movement.

Our goal is to find an ERP in a sequence of EEG data. Equivalently, if we call τ the time of ocurrence of the ERP in a single trial, our goal is to estimate τ. We start by building a parametric model of the ERP, which in our case has two components: a small positive peak followed by a large negative one. Since we are only interested in τ, we marginalise out the other parameters by using a Bayesian framework. We draw samples from the posterior distribution of τ using MCMC in order to later compute its maximum-a-posteriori (MAP) estimate. Different noise covariance structures are tested, and it is shown how it affects the estimates significantly. We validate our findings using visual validation and hand-labelling.

2. RELATED WORK

Our work presented here is novel from a methodology and model point of view. Many works use parametric models to estimate the amplitudes and latencies of the ERP components. However, most works assume their trials are aligned and have a reduced search space for τ (e.g. [2], [4], [7]). This can be too restrictive an assumption if the goal is to build a detector or for validation purposes. In [4] they allow for a latency variation of 200 ms, while in [5] they allow for latency variation of 60 ms and 120 ms for two components respectively. In [7], they assume τ has small variation around its average value. These methods can only work on aligned trials where the latency variability is very small. Assuming the latency variability to be small is not necessarily appropriate for all types of ERP. Furthermore, by reducing the search space of τ, there is a risk of overfitting, i.e. fitting the model to spurious ongoing EEG waveforms. In our case, our seach space is 740 ms and we do not assume our trials are aligned. However, we assume the ERP is present within the trial. Our detector could have a wealth of applications, such as doing ERP parameter estimation under high latency variation, estimating latency variation [8, 9, 10], performing latency correction in a BCI [11] and offering numerical validation techniques.

In [5], they also work within a Bayesian framework but their model differs from ours in many aspects. Since they compute point estimates using MAP estimation with uniform priors over all parameters, their method is not different from a maximum-likelihood method. In this work, we use non-uniform prior distributions over some parameters and we compute the MAP estimates using MCMC. Additionally, our model is hierarchical which makes it more robust to the choice of the prior parameters. The method in [5] constrains the mean of all latencies to be zero, and can therefore only work on aligned trials, which is an assumption we do not make. Furthermore, they use i.i.d. noise, whereas we explore different types of correlated noise structures. The use of Gaussian process kernels to model the noise is a relatively unexplored technique for single-trial ERP analysis. This work extends previous work in [12], where the

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

noise parameters were found by maximum-likelihood and the signal parameters had discrete empirical priors. In this paper, all parameters have continuous priors, and there is an additional hierarchical level that was not present in [12].

3. THE PARAMETRIC MODEL

We use bold font for vectors, upper case for matrices and lower case for scalars. We use the notation τ_{-i} to denote the set of variables $(\tau_1, \ldots, \tau_{i-1}, \tau_{i+1}, \ldots, \tau_N)$. Similarly, we denote the set of values $(\tau_1, \ldots, \tau_N)$ by $\tau_{1:N}$. Our data matrix is made out of N single trials of dimension T, i.e. $Y = [\mathbf{y}_1^T, \ldots, \mathbf{y}_N^T]$. The i-th single trial is given by

$$\mathbf{y}_i = \boldsymbol{\mu}_i + \mathbf{n}_i, \quad \mathbf{n}_i \sim \mathcal{N}(\mathbf{0}, K(\theta))$$

where $i \in [1 \ldots N]$, $\boldsymbol{\mu}_i$ is the mean of the trial and $\mathbf{n}_i$ is the background noise. We model the background noise as being Gaussian distributed with covariance matrix $K(\theta)$, which we will describe later. Single trials are independent of each other so that $p(Y) = \prod_i p(\mathbf{y}_i)$. The mean of the trial is given by

$$\boldsymbol{\mu}_i = F_i \mathbf{a}_i,$$

where $\mathbf{a}_i = [a_{i,1}\, a_{i,2}]^T$ is a 2-by-1 vector and F_i is a T-by-2 matrix such that $F_i = [\mathbf{f}_{i,1}\, \mathbf{f}_{i,2}]$. The T-by-1 vectors $\mathbf{f}_{i,1}$ and $\mathbf{f}_{i,2}$ are parameterised by τ_i and Δ_i

$$(\mathbf{f}_{i,1})_j = \exp\left(-\frac{(j-(\tau_i+\Delta_i))^2}{2c^2}\right)$$
$$(\mathbf{f}_{i,2})_j = -\exp\left(-\frac{(j-\tau_i)^2}{2c^2}\right),$$

where $j \in [1 \ldots T]$ and the values $(\mathbf{f}_{i,1})_j$ and $(\mathbf{f}_{i,2})_j$ are the j-th entries of vectors $\mathbf{f}_{i,1}$ and $\mathbf{f}_{i,2}$ respectively. In other words, the mean of the trial is the sum of two squared exponential functions represented by $\mathbf{f}_{i,1}$ and $\mathbf{f}_{i,2}$ with amplitudes $a_{i,1}$ and $a_{i,2}$ respectively. Note that the amplitude parameters are positive. This parametric model is commonly used for ERPs (i.e. [6]). An example of a single trial with the fitted model is shown in Figure 2.

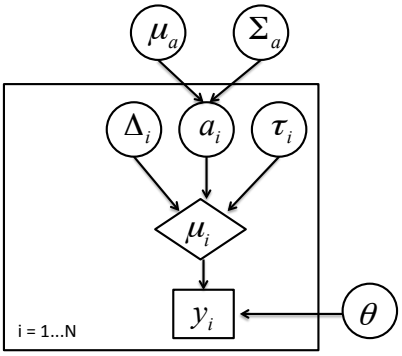

Fig. 1. Hierarchical Bayesian model for the single-trial ERP. Random variables are shown by circles, deterministic variables by diamonds and observations by rectangles.

The goal of this paper is to find τ_i for each trial i. We work in a fully Bayesian framework where all the parameters in the model have prior distributions. We can reformulate our goal as finding the maximum-a-posteriori estimate of τ_i, i.e. $\tau_i^{MAP} = \text{argmax}_{\tau_i}\, p(\tau_i|Y)$. When setting the prior distributions over the model parameters, we assume that we do not have any prior knowledge or information about τ_i. These are the prior distributions we have selected

$$\begin{aligned}
\tau_i &\sim U[\tau_{min}, \tau_{max}] \\
\Delta_i &\sim U[\Delta_{min}, \Delta_{max}] \\
\mathbf{a}_i &\sim \mathcal{N}(\boldsymbol{\mu}_a, \Sigma_a) \\
\boldsymbol{\mu}_a &\sim \mathcal{N}(\boldsymbol{\mu}_0, \Sigma_a/\kappa_0) \\
\Sigma_a &\sim \mathcal{IW}(\Sigma_0, \nu_0),
\end{aligned}$$

where $\mathcal{N}$ denotes the multivariate normal distribution, U denotes the uniform distribution and $\mathcal{IW}$ the inverse Wishart distribution. We chose the multivariate normal distribution for $\mathbf{a}_i$ after looking at the distribution of amplitudes on some labelled trials. We used prior conjugates for $\boldsymbol{\mu}_a$ and Σ_a to make some computations tractable. The hierarchical Bayesian model structure is shown in Figure 1. The first-level hyperparameters τ_{min}, τ_{max}, Δ_{min}, Δ_{max} are fixed so as to have broad distributions for τ_i and Δ_i and encompass all plausible values. The second-level hyperparameters $\boldsymbol{\mu}_0$, κ_0, Σ_0, ν_0 are also fixed after using visual inspection and testing different values. The noise parameters are denoted by θ, they are discussed in the following section.

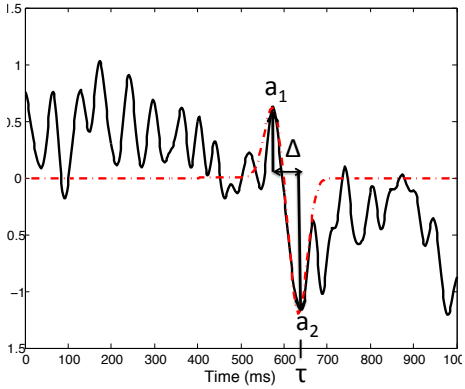

Fig. 2. Example of a single trial with fitted template.

3.1. The noise structure

In this paper, we test different parametric forms for the covariance $K(\theta)$. Our model is a Gaussian process with stationary covariance matrix. Thus, we can build the covariance matrix using a covariance function defined on the separation between two points $r = |i - j|$, i.e. $K_{ij} = k(i, j) = k(r)$ [13]. These are the covariance functions we tested on our data:

- i.i.d. noise: $k_I(r) = \sigma_f^2 \delta_r$, where δ_r is Kronecker's delta,
- Matérn covariances: $k_{\nu=\frac{1}{2}}$, $k_{\nu=\frac{3}{2}}$ and $k_{\nu=\frac{5}{2}}$,
- squared exponential: $k_{SE}(r) = \sigma_f^2 \exp\left(-\frac{r^2}{l^2}\right)$.

The definition of the Matérn covariances can be found in [12]. It is usual to choose ν to be half-integers as it gives a simple form to the covariance. When $\nu \to \infty$, the Matérn covariance becomes the

squared exponential, and so the higher ν, the smoother the noise. Each of these covariance functions has different sets of parameters: $\theta = \sigma_f^2$ for the i.i.d. covariance and $\boldsymbol{\theta} = [\sigma_f^2, l]$ for the others. For both parameters, we choose non-informative priors: the scale parameter l has an improper uniform prior, and the variance has Jeffrey's prior, i.e. $p(\sigma_f^2) \propto \frac{1}{\sigma_f^2}\mathbb{1}_{[0,\infty]}(\sigma_f^2)$.

3.2. The Gibbs sampler

The posterior distribution $p(\tau_i|Y)$ cannot be computed analytically; a Gibbs sampler is used instead to sample from it. We start by sampling from the joint posterior distribution of all parameters $p(\tau_{1:N}, \mathbf{a}_{1:N}, \Delta_{1:N}, \boldsymbol{\mu}_a, \Sigma_a, \theta|Y)$, the marginalisation can be done by keeping only the samples from τ_i. The first step of the Gibbs sampler is to sample from the conditional posterior

$$p(\mathbf{a}_i, \Delta_i, \tau_i|Y, \boldsymbol{\mu}_a, \Sigma_a, \mathbf{a}_{-i}, \Delta_{-i}, \tau_{-i}, \theta) = p(\mathbf{a}_i|\Delta_i, \tau_i, \mathbf{y}_i, \boldsymbol{\mu}_a, \Sigma_a, \theta)p(\Delta_i, \tau_i|\mathbf{y}_i, \boldsymbol{\mu}_a, \Sigma_a, \theta).$$

We can sample $\mathbf{a}_i$ directly from its conditional distribution, but not Δ_i and τ_i. We thus introduce a Metropolis-Hastings (MH) step where we propose new values $(a_i^*, \tau_i^*, \Delta_i^*)$ from a proposal distribution $q(\Delta_i^*, \tau_i^*, \mathbf{a}_i^*|\Delta_i, \tau_i, \mathbf{a}_i) = q(\Delta_i^*) q(\tau_i^*|\tau_i)p(\mathbf{a}_i|\Delta_i, \tau_i, \mathbf{y}_i, \boldsymbol{\mu}_a, \Sigma_a, \theta)$ and compute the acceptance ratio:

$$\frac{p(\mathbf{a}_i^*, \Delta_i^*, \tau_i^*|\mathbf{y}_i, \boldsymbol{\mu}_a, \Sigma_a, \theta)q(\mathbf{a}_i, \Delta_i, \tau_i|\mathbf{a}_i^*, \Delta_i^*, \tau_i^*)}{p(\mathbf{a}_i, \Delta_i, \tau_i|\mathbf{y}_i, \boldsymbol{\mu}_a, \Sigma_a, \theta)q(\mathbf{a}_i^*, \Delta_i^*, \tau_i^*|\mathbf{a}_i, \Delta_i, \tau_i)} = \frac{p(\mathbf{y}_i|\Delta_i^*, \tau_i^*, \boldsymbol{\mu}_a, \Sigma_a, \theta)q(\tau_i, \Delta_i|\tau_i^*, \Delta_i^*)}{p(\mathbf{y}_i|\Delta_i, \tau_i, \boldsymbol{\mu}_a, \Sigma_a, \theta)q(\tau_i^*, \Delta_i^*|\tau_i, \Delta_i)}.$$

To compute this acceptance ratio, it is necessary to solve the integral $\int p(\mathbf{y}_i, \mathbf{a}_i|\Delta_i, \tau_i, \boldsymbol{\mu}_a, \Sigma_a, \theta)\, d\mathbf{a}_i$, which can be done analytically. We propose values for Δ_i from its prior uniform distribution. For the parameter τ_i, we propose from a mixture of Gaussians centered at the N_i local minima $\{\mu_1, \ldots, \mu_{N_i}\}$ of the i-th trial, i.e. $q(\tau_i^*|\tau_i) = \alpha q_0(\tau_i^*) + (1-\alpha)q_1(\tau_i^*|\tau_i)$, where $\alpha \in [0, 1]$ and

$$q_0(\tau_i^*) = \frac{1}{N_i}\sum_{j=1}^{N_i} g(\tau_i^*|\mu_j, \sigma_0^2)$$
$$q_1(\tau_i^*|\tau_i) = g(\tau_i^*|\tau_i, \sigma_1^2).$$

In the above expression, $g(x|\mu, \sigma^2)$ is a Gaussian distribution with mean μ and variance σ^2 evaluated at x. In others words, we sample from a mixture of Gaussians centered at the local minima $\alpha\%$ of the time, the rest of the time we sample from a Gaussian distribution centered at the previous sampled value, i.e. random walk component. Figure 3 illustrates the proposal distribution $q(\tau_1^*|\tau_1 = 278)$. The proposal distribution is very peaked around the previous sampled value. In this step, the Gibbs sampler iterates through all the parameters by increasing i from 1 to N.

The second step of the Gibbs sampler is to sample from $p(\boldsymbol{\mu}_a, \Sigma_a|Y, \mathbf{a}_{1:N}, \Delta_{1:N}, \tau_{1:N}, \theta) = p(\boldsymbol{\mu}_a, \Sigma_a|\mathbf{a}_{1:N})$. Since we selected conjugate priors, this distribution is a multivariate normal-inverse Wishart with known parameters. The acceptance ratio is 1 for this step.

Finally, the last step is to sample from the posterior of the noise parameters $p(\theta|Y, \mathbf{a}_{1:N}, \Delta_{1:N}, \tau_{1:N}, \boldsymbol{\mu}_a, \Sigma_a)$. We cannot sample directly from this distribution, so we use a MH-step again where we propose values from $q(\theta^*|\theta)$ and compute an acceptance ratio for the move. In this case, we choose the same proposal as the prior distribution. The acceptance ratio is

$$\frac{p(Y|\theta^*, \mathbf{a}_{1:N}, \tau_{1:N}, \Delta_{1:N})}{p(Y|\theta, \mathbf{a}_{1:N}, \tau_{1:N}, \Delta_{1:N})}.$$

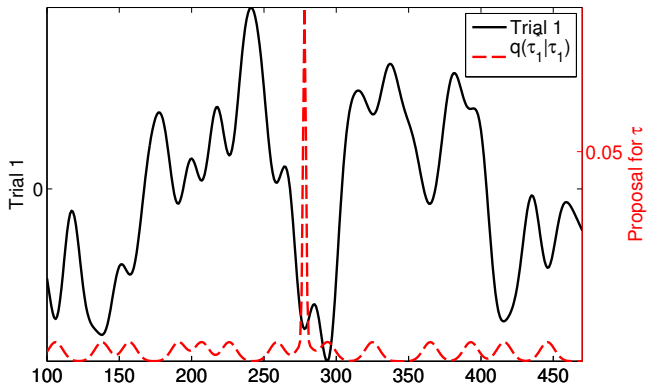

Fig. 3. Proposal distribution for trial 1 and current estimate of $\tau_1 = 278$.

4. THE DATA

The data set was acquired by an Electrical Geodesics system with a 129-channel Hydro-Cell Net at the department of Experiment Psychology in Cambridge, same dataset as [14]. In the experiment, the participants were asked a series of binary questions, and they needed to press a button with either their right or their left hand in order to answer. The time of the button press is the response time, we denote it t_R. This motor action is closely preceded by a movement-related ERP generated in the motor cortex, also called 'lateralized readiness potential', which is defined as the difference between two channels. In our case, we worked on the difference between two motor cortex channels C3 and C4. The sampling frequency was 500 Hz. The preprocessing steps consisted of a bandpass filter with passband [0.01 Hz - 30 Hz] and a notch filter to remove the mains effect. Refer to [14] for details about the experiment or the ERP. We had 3 participants with a total of 142 trials. Every trial was 500 samples long (1 second). In order to allow for the signal to be fully contained inside the search space, we set $\tau_{min} = 100$ and $\tau_{max} = 470$. Thus, our search space was 740 ms.

As mentioned in the introduction, there is no ground truth for this type of data. However, there is some side information that can be used to validate the results. In this case, it is a known fact that the motor ERP occurs soon before the motor response time. We aligned our trials with respect to the response time that we fixed at $t_R = 370$. We thus expected to have a distribution of τ to be centered at approx. $\delta = 100$ ms before the response time (see Supplementary Figure 3 in [14]), and therefore centered at sample 320. We took two approaches to validate our results. Firstly, we used this a-priori knowledge to compare the distribution of $\tau_{1:N}^{MAP}$ using different covariances, and using trials containing only noise. The other validation approach was to hand-label the trials, i.e. $\tau_{1:N}^{l}$. For this, we used the prior knowledge that $t_R - \tau < \delta$, and looked for an identifiable peak. The second validation technique is less reliable than the first one, but it allows for the computation of scores for each model: root-mean-square error (RMSE) and detection accuracy. We deemed we had correctly detected τ_i if $|\tau_i - \tau_i^l| < 5$.

5. THE RESULTS

A well-known issue with Gaussian process covariances is the numerical instability for large T. In our case $T = 500$, we found the matrix to be badly conditioned. A way to circumvent this is to add some i.i.d. noise in the diagonal of the matrix. The final covariance is: $K = \sigma_n^2 I + K_{GP}$. There are some heuristics to set the variance σ_n^2, but in our case the detection was only influenced by the ratio σ_f^2/σ_n^2. We wanted to keep σ_n^2 low enough to take advantage of the correlation structure of the GP matrix as much as possible. We thus set $\sigma_n^2 = 0.01$, which was small given that the signal was usually in the range $[-1, 1]$. We found that a sample size of 10,000 with no thinning and a burn-in period of 100 gave the desired performance. For the k_I and $k_{\nu=1/2}$ covariance matrices, a burn-in period of 500 was used since convergence was slower. We used the following parameter values for the prior and proposal distributions: $\boldsymbol{\mu}_0 = [0.5\ 0.8]^T, [\Delta_{min}\ \Delta_{max}] = [8\ 45], \Sigma_0 = [\mathbf{m}_1\ \mathbf{m}_2], \mathbf{m}_1 = [0.09\ 0.14]^T, \mathbf{m}_2 = [0.14\ 0.21]^T, c = 10, \kappa_0 = 10, \nu_0 = 5, \sigma_0 = 5, \sigma_1 = 1, \alpha = 0.8$. The parameters $\boldsymbol{\mu}_0$, Δ_{min}, Δ_{max}, Σ_0 and c were selected by visual inspection of the signals, the remaining parameters of the priors were selected in a way to allow for flexibility, and the proposal parameters were selected heuristically to improve convergence properties of the chain. We show the posterior $p(\boldsymbol{\mu}_a|Y)$ using the k_{SE} model in Figure 4. This posterior holds information about the single-trial amplitudes of both components. We do not show the posterior of the other parameters due to space constraints.

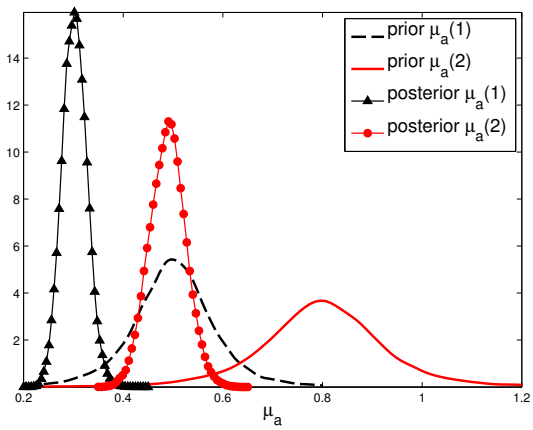

Fig. 4. Kernel density estimation of the prior and posterior distributions of $\boldsymbol{\mu}_a$ using the k_{SE} kernel.

We used the hand-labels $\tau_{1:N}^l$ described earlier to compute a RMSE in samples and percentage of correct detections by comparing $\tau_{1:N}^{MAP}$ to $\tau_{1:N}^l$ for the different covariance matrices (see Table 1). The covariances with smooth noise properties gave the best performance. The i.i.d. covariance did not give the worst performance, it was better than the Matérn covariances with low parameter ν. We also compared the performance of each covariance to a random guess for each τ_i uniformly distributed among the local minima of trial i between 100 and 470. The performance of the k_{SE} model was significantly better than the other models, its detection score was almost 9x better than the random guess. This covariance function yielded a posterior of $\boldsymbol{\mu}_a$ with maxima that were much lower than the maxima of the prior distributions, as shown in Figure 4. We also tried a proposal for τ_i which was a uniform distribution over the local minima in $\mathbf{y}_i$, this proposal was faster and achieved almost the same performance as the one reported here.

Model	RMSE	Correct detections (%)
Random guess	115.60	7.05
$k_{\nu=1/2}$	108.65	31.69
$k_{\nu=3/2}$	98.10	40.85
k_I	88.62	44.37
$k_{\nu=5/2}$	89.10	50.00
k_{SE}	74.79	61.97

Table 1. Comparison of noise models.

Looking at the distribution of the estimates $\tau_{1:N}^{MAP}$ is a form of visual validation. Figure 5 shows the distributions of $\tau_{1:N}^{MAP}$ with the k_{SE} and $k_{\nu=1/2}$ models, as well as the distribution of $\tau_{1:N}^{MAP}$ using k_{SE} in a set of trials that do not have the ERP and are therefore considered noise. We see that the distribution of $\tau_{1:N}^{MAP}$ obtained from the k_{SE} model has a significant peak centered at 320, as expected. In the other cases, the distribution looks more uniform, making it difficult to infer anything from it. This confirms that we are not fitting the parametric ERP model to spurious EEG noise, and that the k_{SE} model picks up a signal indeed.

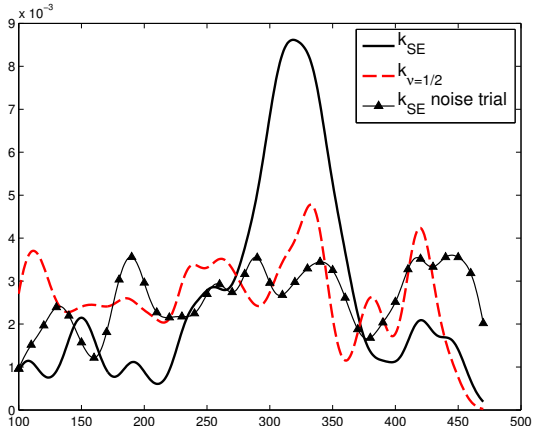

Fig. 5. Kernel-smoothed distributions of $\tau_{1:N}^{MAP}$ with k_{SE}, $k_{\nu=1/2}$ and with k_{SE} on trials containing only noise.

6. CONCLUSION

We have shown how the ERP can be successfully detected in single-trial EEG data. By using a hierarchical Bayesian model, we could capture the inter-trial variability of the ERP signal, while incorporating prior knowledge we might have about the signal such as its mean amplitude. We assumed no prior knowledge on the time of occurrence of the ERP. The estimates were validated by comparing them to hand-labels, and by inspecting their distribution visually. The hand-labels supported the fact that the k_{SE} covariance structure is the most appropriate for the detection of the signal. The distribution of the estimates obtained with the k_{SE} covariance had a clear peak at approx. 100 ms before the motor action, which is an expected property of the ERP. This work shows that a parametric model can be used to find the ERP signal in a single trial, but that careful thought needs to be put into the modelling of the noise. We could extend the work presented here by introducing new smooth covariances (autoregressive, periodic covariance function), and by using multiple channels.

7. REFERENCES

[1] A. Mouraux and G.D. Iannetti, "Across-trial averaging of event-related EEG responses and beyond," *Magnetic resonance imaging*, vol. 26, no. 7, pp. 1041–1054, 2008.

[2] D. Jarchi, S. Sanei, J.C. Principe, and B. Makkiabadi, "A new spatiotemporal filtering method for single-trial estimation of correlated ERP subcomponents," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 1, pp. 132–143, 2011.

[3] Loukianos Spyrou and Saeid Sanei, "Source localization of event-related potentials incorporating spatial notch filters," *Biomedical Engineering, IEEE Transactions on*, vol. 55, no. 9, pp. 2232–2239, 2008.

[4] David E Thompson, Seth Warschausky, and Jane E Huggins, "Classifier-based latency estimation: a novel way to estimate and predict BCI accuracy," *Journal of neural engineering*, vol. 10, no. 1, pp. 016006, 2013.

[5] W. Truccolo, K.H. Knuth, A. Shah, S.L. Bressler, C.E. Schroeder, and M. Ding, "Estimation of single-trial multicomponent ERPs: Differentially variable component analysis (dVCA)," *Biological cybernetics*, vol. 89, no. 6, pp. 426–438, 2003.

[6] H.R. Mohseni, F. Ghaderi, E.L. Wilding, and S. Sanei, "Variational Bayes for spatiotemporal identification of event-related potential subcomponents," *Biomedical Engineering, IEEE Transactions on*, vol. 57, no. 10, pp. 2413–2428, 2010.

[7] Luzhou Xu, Petre Stoica, Jian Li, Steven L Bressler, Xianzhi Shao, and Mingzhou Ding, "ASEO: A method for the simultaneous estimation of single-trial event-related potentials and ongoing brain activities," *Biomedical Engineering, IEEE Transactions on*, vol. 56, no. 1, pp. 111–121, 2009.

[8] P.A. Karjalainen, *Regularization and Bayesian Methods for Evoked Potential Estimation*, Ph.D. thesis, Kuopio University, 1997.

[9] E Callaway, R Halliday, H Naylor, and D Thouvenin, "The latency of the average is not the average of the latencies," *Psychophysiology*, vol. 21, pp. 571, 1984.

[10] Fren TY Smulders, JL Kenemans, and A Kok, "A comparison of different methods for estimating single-trial p300 latencies," *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, vol. 92, no. 2, pp. 107–114, 1994.

[11] Lawrence Ashley Farwell and Emanuel Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.

[12] Maria Rosario Mestre and William J Fitzgerald, "Comparison of Gaussian process models for single-trial event-related potentials," in *Statistical Signal Processing Workshop (SSP), 2012 IEEE*. IEEE, 2012, pp. 428–431.

[13] C.E. Rasmussen and C.K.I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, 2006.

[14] D. Szücs, F. Soltész, and S. White, "Motor conflict in Stroop tasks: direct evidence from single-trial electro-myography and electro-encephalography," *NeuroImage*, vol. 47, no. 4, pp. 1960–1973, 2009.