

Depth of Anaesthesia Assessment with Generative Polyspectral Models

I. Rezek, S. J. Roberts, E. Siva*
Dept. of Engineering Science,
University of Oxford, U.K.
{irezek,sjrob,u97es}@robots.ox.ac.uk

R. Conradt
Dept. of Public Health and Primary Care,
University of Oxford, U.K.
conradtr@web.de

Abstract

The application of anaesthetic agents is known to have significant effects on the EEG waveforms. Information extraction now routinely goes beyond second order statistics analysis, as obtained via power spectral methods, and uses higher order spectral methods. In this paper we present a model which generalises the autoregressive class of polyspectral models by having a semi-parametric description of the residual probability density. We estimate the model in the Variational Bayesian framework and extract higher order spectral features. Testing their importance for depth of anaesthesia classification is done on three different EEG data sets collected under exposure to different agents. The results show that significant improvements can be made over standard methods of estimating higher order spectra. The results also indicate that in two out of three anaesthetic agents, better classification can be achieved with higher order spectral features.

1 Introduction

Hypnosis, besides analgesia and areflexia, forms an integral part of anaesthesia. To assess it, researchers have increasingly suggested analysing the electroencephalogram (EEG) using various statistical methods, such as spectral entropy and bispectral index. The bispectral index (aka BIS) is computed by what is probably the oldest commercially available monitor and has been used in over 200 studies on a range of agents. The chief feature of the BIS index is the bispectrum. It is capable of quantifying interactions in the EEG which standard spectral estimation cannot, since it extracts higher order statistics (up to 3rd order). There are several known shortcomings of bispectrum estimation procedures [11]. For one, the estimator requires extensively averaged estimates which substantially slows down the estimation procedure. It also reduces robustness as stationar-

*now at BAE Systems, Advanced Technology Centre, Filton, Bristol, U.K.

ity assumptions are imposed on very long segments of data. These characteristics of the estimator might be the cause for recent criticism the BIS index has received, such as not being responsive to some anaesthetic agents [1, 6], and not robust across patients [3] or time [7].

The question this paper addresses is whether EEG characterisation can be improved upon while maintaining the route taken by higher order statistical procedures. To achieve this we describe a model that is capable of encoding higher order moments and from which higher order spectra can be computed. The inspiration for this model comes from the robust estimation literature. While robust estimation models are able to capture higher order moments (to account for arbitrary noise characteristics) this information is not used in the subsequent analysis. In contrast we make specific use of the moments for polyspectral estimation. To our knowledge no connection has been made between these models and higher order spectra and this link is the main contribution of this paper. Furthermore, our estimation procedure uses a recently developed paradigm (known as Variational Bayesian Learning) which leads to fast estimation algorithms.

We begin with a brief review of traditional higher order spectral estimation using parametric models, with special focus on autoregressive (AR) models. We then formulate a generative AR model which is capable of capturing all higher order spectral information and also can be estimated in a robust and fast fashion. We finally compare the traditional estimators with our estimators, both on real and synthetic data.

2 Brief Review of Higher-order Spectra

The polyspectrum (also known as n^{th} order spectrum) of a process, say y_t , is defined as the Fourier transform of the process' n^{th} moment or cumulant [9]. In particular, the *bispectrum* (or 3rd-order polyspectrum) is the transform of the third-order moment, $R_y(\mu, \nu)$, $B_y(u, v) = \iint_{-\infty}^{\infty} R_y(\mu, \nu) e^{-j(u\mu + v\nu)} d\mu d\nu$ where $R_y(\mu, \nu) = E_t\{y_t y_{t+\mu} y_{t+\nu}\}$. The

trispectrum is defined similarly as the transform $T_y(u, v, w) = \iint\int_{-\infty}^{\infty} R_y(\mu, \nu, \omega) e^{-j(u\mu + v\nu + w\omega)} d\mu d\nu d\omega$ with $R_y(\mu, \nu, \omega) = E_t\{y_t y_{t+\mu} y_{t+\nu} y_{t+\omega}\}$ being the fourth-order moment. For system identification the input process, say e_t , is typically unknown and assumed to be a strictly stationary (non-Gaussian) random noise sequence with finite n^{th} -order moment. If, furthermore, the observed process y_t is a (parametric) autoregressive process of order p , $y(t) + \sum_p y(t-p)a_p = e_t$, then the observed process's polyspectrum is directly proportional to the system's own polyspectrum. For example, for a system with transfer function $h(t)$, the bi- and trispectrum relations are

$$\begin{aligned} B_y(u, v) &= Q_B H(u)H(v)H^*(u+v) \\ T_y(u, v, w) &= Q_T H(u)H(v)H(w)H^*(u+v+w) \end{aligned} \quad (1)$$

where $H(f) = \int h(t)e^{-j(ft)} dt$, $Q_B = B_e(0, 0)$ and $Q_T = T_e(0, 0, 0)$.

Estimation typically involves constructing a set of linear equations in which a Toeplitz matrix \mathcal{R} of n^{th} -order moments is inverted to solve for the autoregressive (AR) coefficients [11]. The AR coefficients thus describe the system properties with regard to one particular moment order, i.e. bispectrum and trispectrum lead to different sets of AR coefficients.

There are two major drawbacks of, for instance, bispectral estimators as they are most commonly used. The first is its large sample assumption required for a stable estimation of the third order moments from data. The second is the point estimate of the model parameters. As to the first assumption, it should be noted that triple products, as they arise in third-order moment estimation, are naturally more sensitive to outliers - simply because estimation involves raising to a higher power. Only by using more data can the impact of this assumption be minimised. The second limitation requires additional model selection criteria to be applied to repeated parameter estimates of varying model orders p . This is due to the fact that traditional estimation does not naturally provide bounds on the parameter estimates. To obtain these, additional computations are needed which are time costly. In addition, increased model sizes compromises estimating algorithm stability. A principled Bayesian approach would largely avoid such problems and return distributions over AR model coefficients as well as model scores.

Ideally, therefore, we would like a Bayesian polyspectral estimator. However, parametric and therefore fast estimators are only available for normal spectra, that is only standard AR model estimation is done in a Bayesian framework. To retain fast estimators yet be able to compute higher order spectra, we represent the input noise process e_t of the AR model by a mixture of K Gaussian densities each weighted by π_k and with mean and precision parameters $\theta_k = \{\mu_k, \beta_k\}$, $p(e_t) = \sum_{k=1}^K \pi_k \mathcal{N}(e_t; \theta_k)$. Since mix-

ture densities can model any smooth density with arbitrary accuracy, they can also be used to compute the underlying density's higher order moments analytically (see [13] for formulae). Thus, we can compute all higher order spectra, by extension of (1), using

$$P_y(f_1 \cdots f_n) = Q_{P_n} H(f_1) \cdots H(f_{n-1}) H^*(\sum_n f_n) \quad (2)$$

where $Q_{P_n} = P_e(0, \cdots, 0)$ is essentially the n^{th} moment of e_t [9].

The interpretation above has a strong correspondence to robust estimation [2, 4, 14, 15] which reduces the model's sensitivity to outliers via use of heavy tailed distributions, L1 norms or mixture models. We are unaware, however, about any work pointing to the model's relationship to polyspectral models. Also, our work uses fast variational estimation, rather than MCMC sampling approaches commonly used in the relevant literature. We also extend our work [10] by augmenting the model with a hidden Markov chain for the mixture components. This is required to test the validity of the mixture model assumption - which, for (2) to hold, must reject the Markov assumption.

Simulated Example We compare bispectral estimates obtained by the generalised AR model with those obtained using standard methods [11]. For a fair comparison, we simulated data as described in [11]. To recap, data is generated using a 4-th order autoregressive model with AR coefficients set to $a = [0.1 \ 0.2238 \ 0.0844 \ 0.02994]$. The bispectrum based on the AR coefficients can be obtained according to $B(w_1, w_2) = \beta H(w_1)H(w_2)H^*(w_1 + w_2)$ where w is the angular frequency and $H(w) = 1/(1 + \sum_{p=1}^4 a_p \exp\{-jwp\})$ for $|w| \leq \pi$. Figure 1 shows the estimated bispectrum calculated from the coefficients estimated by the variational generalised AR method. Only 128 samples were needed to estimate the AR coefficients within 2% of the true values. That is, in comparison with [11] a 12-fold saving in the amount of data required to estimate a bispectrum. Since in our approach only Gaussian densities are used for the error term, estimation of only first and second moments is required and any higher moment can be computed by simply plugging in the first 2 moments in a moment integral (see below). This adds significantly to the stability and reduces the amount of data needed for estimation. In turn, our algorithm is faster.

3 Generalised AR model

As mentioned earlier, we aim to explain the data in terms of a linear autoregressive (AR) model and temporally correlated residuals which follow an arbitrary distribution. In this section, we describe in more detail the model and the estimation procedure.

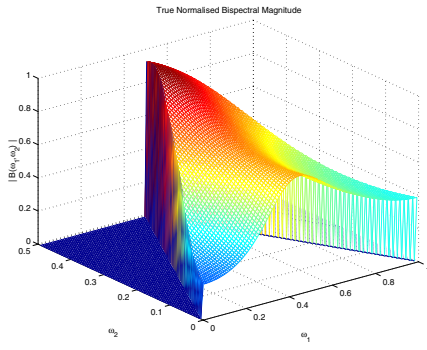


Figure 1. Magnitude of bispectrum estimated using simulated data of a 4th order autoregressive model of Example 1 in [11].

Following the definition of AR models, at each time instant, t the data sample, y_t , is a weighted combination of past samples, $y_{t-1}, y_{t-2}, \dots, y_{t-p}$ with residual error e_t $y_t - x_t a = e_t$. In our notation, y_t is the observation at time instant t , $a \in \mathbb{R}^p$ is the vector of p autoregressive coefficients, $x_t = [y_{t-1}, y_{t-2}, \dots, y_{t-p}] \in \mathbb{R}^p$ is the vector containing p past observations and e_t is the autoregressive driving noise. In order to keep the residual probability distribution as flexible as possible we use a mixture model approach and model them as arising from one of K 1-dimensional Gaussian components, each parameterised by a mean $x_t a + \mu_k$ and precision β_k . An indicator variable s_t is used to select one of the K components at time t , $p(e_t | s_t, \mu, \beta, a) \propto \prod_{k=1}^K \mathcal{N}(e_t | x_t a + \mu_k, \beta_k)^{\delta(s_t=k)}$ where the means and variances of the components are collated in the parameter vectors $\beta = \{\beta_1, \dots, \beta_K\}$ and $\mu = \{\mu_1, \dots, \mu_K\}$. We model the temporal dependence between the residuals by imposing a first order Markov dependence¹ on the component indicator variables s_t . The probability of drawing component k at time t is thus dependent upon the component l drawn at time $t-1$. Since the component labels form a discrete set, the conditional probability is a $K \times K$ dimensional Multinomial (\mathcal{Mn}) distribution with parameters π_{lk} , $p(s_t | s_{t-1}, \pi) = \prod_{l=1}^K \mathcal{Mn}_K(s_t | \pi_l) \propto \prod_{k,l=1}^K \pi_{lk}^{\delta(s_t=k, s_{t-1}=l)}$ with $\pi = \{\pi_{lk}\}, \forall k, l = 1, \dots, K$. The probability of the first component indicator, s_0 , is parameterised by a K -dimensional Multinomial distribution with parameters π_{0k} , $p(s_0 | \pi_0) = \mathcal{Mn}_K(s_0 | \pi_0) \propto \prod_{k=1}^K \pi_{0k}^{\delta(s_0=k)}$. For data set $Y = \{y_1, \dots, y_T\}$, the full

¹While the polyspectral model does precludes this Markov property, we deliberately incorporate in order to test, though model scoring, the correctness of our polyspectral model.

log-likelihood is given as

$$\begin{aligned} \log p(\theta, S) &= \sum_{k=1}^K \delta(s_0 = k) \log \pi_{0k} + \\ &\sum_{t=1}^T \sum_{k=1}^K \sum_{l=1}^K \delta(s_t = k, s_{t-1} = l) \log \pi_{lk} + \\ &\sum_{t=1}^T \sum_{k=1}^K \delta(s_t = k) \left[\log \left(\frac{\beta_k}{2\pi} \right)^{-\frac{1}{2}} - \frac{\beta_k}{2} (y_t - \mu_k - x_t a)^2 \right] \end{aligned} \quad (3)$$

where we have introduced $S = \{s_0, \dots, s_T\}$, the set of all component indicator variables, and the parameter vector θ consisting of all other model parameters.

Model estimation is done using the variational Bayesian learning paradigm [10]. The aim of variational Bayesian learning (see [5] for an excellent tutorial) is to perform an approximate Bayesian integration by minimising a Kullback-Leibler (KL) divergence between the exact posterior distribution and a simpler form of it. For Bayesian inference, all parameters θ are associated with conjugate prior distributions, $p(\theta)$, and posterior distributions, $q(\theta)$. Computation of the posterior distribution parameters is then very similar to that described in [10]². The variational approach iterates until the parameters of the posterior distribution have converged to fixed points. The obtained KL-divergence assumes predetermined model order, i.e. number of components for residual density and autoregressive model coefficients is fixed. To perform optimal model order estimation the estimation is repeated for various model orders and subsequently the lowest divergence among repeats used to decide on the the optimal model order.

Higher order AR Spectrum The classification of the data is based on a number of features extracted from our model. The premise of this work is that there is valuable information in the EEG not readily captured by standard spectral techniques and that higher order spectral estimation is required. Using the mean of the autoregressive coefficients posterior distribution, \hat{a} , we can readily extract the higher order spectrum using equation (2). To reduce the dimensionality of the higher order spectral measures, we have opted for the total spectral power in the δ -range ($0.5Hz - 4Hz$) as an indicator of speed differences in the EEG. We compute the total higher order spectral power in this band based on spectral evaluation at 1000 discrete and evenly spaced frequencies.

²The main difference between the methods here and in [10] is the calculation of the posterior distribution $p(S)$ needs to be performed using the Baum-Welch equations with a likelihood term corresponding to $E\{\log p(\theta, S)\}_{q(\theta)}$.

4 Anaesthesia Data and Recording Procedure

We applied the model to three data sets. The first EEG data was recorded during anaesthesia with a combination of Remifentanyl/Propofol. Eight patients agreed to take part in the study, which was performed prior to patients undergoing abdominal surgery. EEG was recorded from 2 electrodes placed at either side of the outer malar bone (At_1 and At_2) with reference and ground electrodes placed at Fpz and Fp_1 , respectively. The recording lengths were on average 8 minutes long (range between 5 and 13 minutes). The second data set consisted of EEG recordings of nine patients, from the forehead to the left mastoid (with the right mastoid as common) during Desflurane induced anaesthesia. The EEG was recorded for 10 minutes while the desflurane concentration was near constant (monitored by the gas analyser). However, only the final 5 minutes of each recording period were used for experimental analysis. Finally, the third data set was recorded under the influence of propofol. From six patients the EEG was again recorded from the forehead to the left mastoid (with the right mastoid as common). The full 11 minute long EEG recording periods were used for the analysis during which agent concentration was near constant. For each state, the total available training sequence was approx 2h long (equivalent to $1.5 \cdot 10^6$ samples).

5 Results

A series of experiments was performed to examine the effect of additional residual information on the performance of a classifier using the Netlab Software [8]. For classification we used a multi-layer perceptron (MLP) with one output unit and 10 hidden units. The MLP was trained using the evidence framework [8] to avoid over-fitting. All results presented below are obtained from a patient-fold cross-validation. By this we mean, leaving out one patient for testing and then cycling through all patients. The justification is this by predicting across patients were obtain a better understanding of the true "across-patient" informativeness of our features. After training the MLP classifier, we computed a confusion matrix and classification rate. The classification rates and confusion matrices for the different experiments presented here are the sample averages of the confusion matrices and classification rates resulting from each fold within one experiment. The classification rate's standard deviation reported in the results is also a sample estimate.

Testing the Mixture Model Assumption The validity of the polyspectral model of system identification assumes an i.i.d. noise sequence e_t . The true model for e_t must therefore be a mixture model and no long-term time dependencies must exist. We tested, using model scores, whether a

mixture model is favoured over the 1st order Markov assumption. To do this, each second of EEG data was used to train a set of generalised autoregressive models, with different autoregressive model orders and different number of mixture kernels. The optimal number of autoregressive coefficients and kernels which resulted in the minimal KL divergence were identified.

All data sets support mixture models with minimum 2 mixture components, figure 5 for this application. The tests showed that neither remifentanyl/propofol nor propofol favoured the hidden Markov chain coupling of the residual noise components. Only the desflurane data favoured a Markov chain, but not strongly. This suggest that while the residual model is needed, it does not exhibit temporal dependencies and thus supports the parametric polyspectral model assumptions.

Figure 5 shows the distribution of the optimal choices for each of the three data sets. There are clear differences in the way anaesthetic state and agent affect the model estimation. The effect of remifentanyl/propofol on the EEG is only noticeable on the AR part of the model. Whilst the residual density requires 2 components in both the anaesthetised and wake state (shown in black and white in figure 5, respectively) the number of autoregressive coefficients is different. Wake requires a higher number of AR coefficients to fit the EEG. Agent desflurane's effect seems to be the opposite of that of remifentanyl/propofol. In the desflurane case, the number of AR coefficients required by the model hardly changes between sedated and wake state, whilst the residuals require an extra mixture component in the wake state. Propofol falls in-between the two extreme effects in the remifentanyl/propofol and desflurane data. The number of AR coefficients as well as the number of residual mixture components are affected by the propofol agent, which is reduced and increased, respectively, in the anaesthetised state.

Comparison of Bi- and Trispectral estimators The results using simulated data in section 2 suggested that our model improved considerably on standard methods for higher order spectral estimation. Our evaluation in the anaesthesia data further supports this finding. Table 1 below shows the classification results obtained by using either the bispectrum or trispectrum for anaesthetic state classification on all our data. The table lists, for each agent, the classification rate as well as the empirical standard deviation of the classifiers, computed across subjects.

For desflurane and remifentanyl/propofol variational performs consistently better in terms of classification rate, which is 15% higher than the classical approach. Also classification robustness is improved by at least 1/3. This statement holds for anaesthetic agents which require higher order spectral estimation. For example, Propofol is an agent

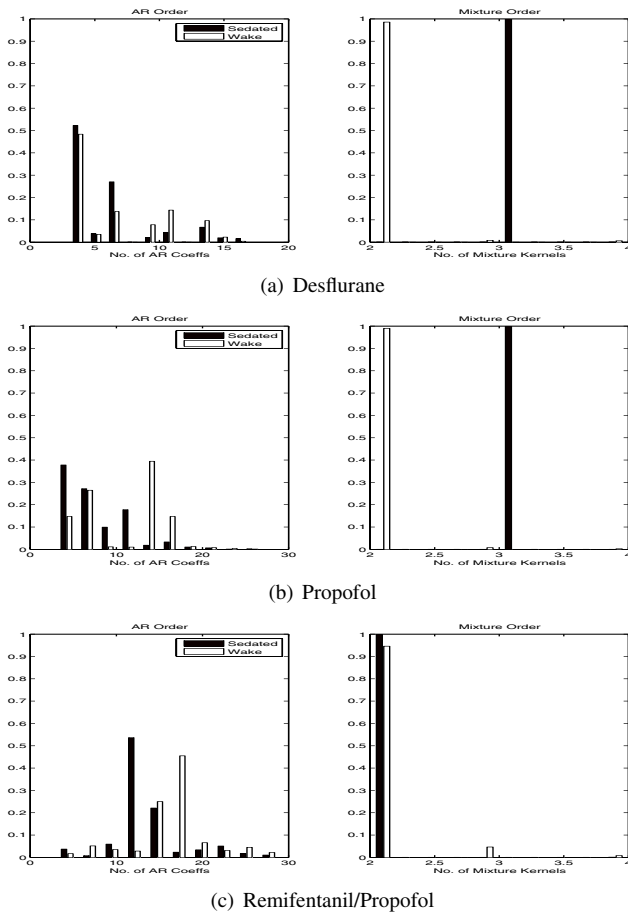


Figure 2. Empirical distribution of optimal number of AR coefficients (left histogram) and residual mixture components (left histogram) for three data sets, with wake state marked in white and sedated in black.

which, based on our observations, does not require sophisticated higher order spectral modelling. In fact classification based on standard spectrum analysis performs equally well (not shown here).

Depth of Anaesthesia Classification What the results of the previous section have shown is that higher order spectra are informative at least in two anaesthetic agents out of three, in those data sets that we have studied. What remains to be seen is which spectral features are most discriminative. In order to decide which spectral are the most discriminative we have investigated the classification performances on each spectral feature individually and also combinations of features. The results based on individual features are shown in table (1). Tables (2)-(4) below show the classification results using various combinations of spectral features. Table (2) shows the classification rates using the normal spec-

Table 1: Average (Std. Deviation) of classification rates for different agents and higher order spectral features

| | Propofol | Desflurane | Remifentanil/ Propofol |
|-------------|--------------|---------------|---------------------------|
| Variational | | | |
| Spectrum | 98.93 (0.97) | 91.45 (7.44) | 82.40 (6.95) |
| BiSpec | 98.56 (1.51) | 92.60 (7.86) | 86.03 (6.48) |
| TriSpec | 98.69 (1.49) | 93.62 (5.85) | 86.37 (6.47) |
| Traditional | | | |
| Spectrum | 99.00 (0.94) | 81.04 (10.31) | 70.41 (8.65) |
| BiSpec | 98.87(0.96) | 80.95 (10.60) | 70.91 (9.41) |
| TriSpec | 98.93 (0.87) | 80.69 (10.57) | 69.48 (8.41) |

trum combined with the bispectrum, table (3) using the normal spectrum combined with the trispectrum and table (4) using normal spectrum combined with bi- and trispectrum. In addition to the actual classification rates and accuracies, the tables (2)-(4) also depict the confusion matrices. As mentioned earlier, good propofol classification can be performed using normal spectra. Hence reference to propofol is omitted from the tables below.

For desflurane we achieve the best classification using, both the standard spectrum as well as the bispectrum. The actual classification rate may only be 2% lower, but the accuracy is doubled. These two effects make the difference between this and all other classifiers statistically significant for desflurane (McNemar [12], $p < 10^3$). As for remifentanil/propofol, any single higher order spectral feature is sufficient for the classification task, with the best classification rate being around 86%. The advantage of using of the higher order spectral features is evident in the classification of the anaesthetic state (bottom row of confusion matrix). Wake state classification is hardly affected by higher order features.

6 Discussion and Conclusions

In this work we have posed the question whether useful information, for classification purposes, is contained in the higher order statistics of anaesthetised EEG and whether more robust estimation methods can be found to extract them. The question arose naturally considering some of the criticism higher order methods have attracted from the medical community. Our results support the use of monitors, such as BIS with its main component being the bispectrum [11]. However, our result also imply that criticism of BIS is most likely due to the lack of robust bispectral estimation.

Our approach is, in essence, the reverse of that (de facto standard) of [11]. In [11] the derivation is algorithmic and focuses on the higher order moments and their estimation. Only subsequently are models postulated for these algo-

Table 2: Classification Performance for Normal & Bispectrum Combination

| | Desflurane | Remifentanil/ Propofol |
|-------------------------|--|--|
| Conf. Matrix | $\begin{bmatrix} 0.94 & 0.06 \\ 0.05 & 0.95 \end{bmatrix}$ | $\begin{bmatrix} 0.78 & 0.22 \\ 0.06 & 0.94 \end{bmatrix}$ |
| Class. Rate (Std. Dev.) | 94.20 (4.82) | 84.59 (8.70) |

Table 3: Classification Performance for Normal & Trispectrum Combination

| | Desflurane | Remifentanil/ Propofol |
|-------------------------|--|--|
| Conf. Matrix | $\begin{bmatrix} 0.95 & 0.05 \\ 0.07 & 0.93 \end{bmatrix}$ | $\begin{bmatrix} 0.79 & 0.21 \\ 0.04 & 0.96 \end{bmatrix}$ |
| Class. Rate (Std. Dev.) | 93.78 (5.71) | 86.37 (7.60) |

Table 4: Classification Performance for Normal & Bi- & Trispectrum Combination

| | Desflurane | Remifentanil/ Propofol |
|-------------------------|--|--|
| Conf. Matrix | $\begin{bmatrix} 0.94 & 0.06 \\ 0.07 & 0.93 \end{bmatrix}$ | $\begin{bmatrix} 0.79 & 0.21 \\ 0.05 & 0.95 \end{bmatrix}$ |
| Class. Rate (Std. Dev.) | 93.21 (6.10) | 85.83 (8.89) |

gorithms. In contrast, our model is a generative model, that is we begin with the assumption of the data generating process and then we derive the estimation procedure. In our approach, the noise is represented by a fully parameterised probability distribution. This gives us the flexibility to extract all moments analytically. However, we never need to estimate any moments other than the mean and variance. Estimation of all model parameters was performed in the variational Bayesian framework [5], which has the advantage of providing a model score, leading to analytic solutions (set of coupled equations that were iterated until convergence) and faster learning algorithms (compared to MCMC sampling).

Our results indicate that extraction of higher order features is more robust, faster and more accurate using the generative model and the variational estimation procedure. This was demonstrated on synthetic as well as real EEG data. Our analysis suggests that Propofol has the strongest effect on the EEG signal and therefore no sophisticated estimation is required. However, Desflurane and Remifentanil/Propofol require higher order spectral features to char-

acterise them accurately.

Acknowledgement

The authors would like to thank W. Penny for contributing his software and S. Reece for extremely helpful suggestions. Also thanks to Chris Jordan from the Academic Department of Anaesthesia based at Northwick Park Hospital in Harrow, London for contributing the propofol and desflurane data set, and O. Ortolani from the Dipartimento Area Critica Medico Chirurgica, Sez. di Anestesiologia e Rianimazione, Università di Firenze, Italy, for contributing the remifentanil/propofol data. This work was funded by EPSRC Grant GR/N03471.

References

- [1] G. Barr, J. Jakobson, A. Owall, and R. Anderson. Nitrous oxide does not alter bispectral index: study with nitrous oxide as a sole agent and as a adjunct to i.v. anesthesia. *Br J. Anaesth.* 82:8227–8230, 1999.
- [2] S. Godsill and P. Rayner. Statistical reconstruction and analysis of autoregressive signals in impulsive noise. *IEEE Transactions on Speech & Audio Processing*, 6(4):352–372, 1998.
- [3] J. D. Hall and G. G. Lockwood. Bispectral index: comparison of two montages. *British Journal of Anaesthesia*, 80(3):342–344, 1998.
- [4] P. Huber. *Robust Statistics*. J. Wiley, 2003.
- [5] T. Jaakkola. Tutorial on Variational Approximation Methods. In M. Opper and D. Saad, editors, *Advanced Mean Field Methods: Theory and Practice*. MIT Press, 2000.
- [6] J. Johansen and P. Sebel. Development and clinical application of electroencephalographic bispectrum monitoring. *Anesthesiology*, 93:1336–1344, 2000.
- [7] K. Kuizenga, J. Wierda1, and C. Kalkman. Biphasic EEG changes in relation to loss of consciousness during induction with thiopental, propofol, etomidate, midazolam or sevoflurane. *British Journal of Anaesthesia*, 86(3):354–360, 2003.
- [8] I. Nabney. *Netlab: Algorithms for Pattern Recognition*. Springer Verlag, 2002.
- [9] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Singapore, 1991.
- [10] W. Penny and S. Roberts. Variational bayes for generalised autoregressive models. *IEEE Transactions on Signal Processing*, 50(9):2245–2257, 2002.
- [11] M. Raghuvver and C. Nikias. Bispectrum Estimation: A Parametric Approach. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(4):1213–1230, 1985.
- [12] B. Ripley. *Pattern Recognition and Neural Networks*. Cambridge University Press, 2000.
- [13] A. Stuart and J. Ord. *Distribution Theory*, volume 1 of *Kendall’s Advanced Theory of Statistics*. Edward Arnold, 1994.
- [14] R. Tibshirani. Regression selection and shrinkage via the lasso. *J Royal Stat Soc B*, 1(267-288), 1996.
- [15] M. West. Outlier models and prior distributions in bayesian linear regression. *J. R. Stat. Soc. B*, 46(3):431–439, 1984.