

# Depth of Anaesthesia Assessment with Generative Polyspectral Models

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

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

April 4, 2006

## Abstract

The application of anaesthetic agents is known to have significant effects on the EEG waveform. Information extraction now routinely goes beyond second order spectral 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.

**Keywords** anaesthesia, bispectrum, trispectrum, mixture model, variational Bayes.

# 1 Introduction

Assessment of anesthesia now routinely includes some analysis of the electroencephalogram. The CFAM monitor [1] is one of the earliest monitors and analysed EEG spectrum and amplitude. Recent devices use methods, such as spectral entropy [2–4], evoked potentials [5], and the BIS index.

The BIS monitor [6, 7] is probably the most popular commercially available monitor. Its BIS index is a single figure derived from a set of time domain and frequency domain measures [8]. Among the time domain features are, for example, the burst suppression ratio (BSR measures the time fraction, per epoch, of EEG burst suppression) and QUAZI suppression (a baseline-drift insensitive detection of burst suppression). Among the spectral domain based features are the relative beta ratio (the difference between the log total power within the 30-47 Hz and 11-20 Hz bands ) and the SynchFastSlow value (the difference between log total bispectral power in the 40-47 Hz band and 1-47 Hz band). The features are then weighted and fused to give the overall single BIS index, ranging from 0-100.

The BIS index offers considerable advantages, most notably its non-invasive nature and extensive clinical validation [9]. Interestingly, the BIS index has recently received some critical press, such as being redundant [10], not responsive to some anaesthetic agents [11–13], and not robust across patients [14] or time [15]. In fact, alternatives to BIS are being contemplated [16, 17]. It is difficult to pinpoint the main cause for this criticism, considering that BIS is a composite measure that weights features differently towards the overall index depending on anaesthetic depth<sup>1</sup>. However, this paper starts with the premise that, if improvements are to be made, there exist for most features, forming the BIS index, robust estimation methods. For example, standard frequency spectra can be estimated using Bayesian approaches [18] or parametric methods [19], as can fusion methods [20]. The exception to this list is the bispectrum. Hence, the question this paper addresses, in more detail than its already published counterpart [21], is whether EEG characterisation can be improved upon while maintaining the route taken by higher order statistical procedures.

The bispectrum is capable of quantifying interactions in the electroencephalogram (EEG) which standard spectral estimation cannot and for which it extracts higher order statistics (the 3<sup>rd</sup> order, to be precise; trispectra require 4<sup>th</sup> order a.s.o.). However, there are several known shortcomings

---

<sup>1</sup>We use the rather vaguely defined term “depth of anaesthesia” to mean the level of the hypnotic component of anaesthesia throughout this paper.

of the bispectrum estimation procedure, as described in [22, 23]. For one, the estimator requires extensively averaged estimates at several levels of the estimation procedure. This slows down the estimation substantially. It also reduces its robustness as stationarity assumptions are imposed on a larger segment of the data. To achieve a better estimation 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.

## 2 Understanding the Problem Domain

### 2.1 Brief Review of Higher-order Spectra

The polyspectrum (also known as  $n^{\text{th}}$ -order spectrum) of a process, say  $y_t$ , is defined as the Fourier transform of the process'  $n^{\text{th}}$  moment or cumulant [24]. In particular, the *bispectrum* (or  $3^{\text{rd}}$ -order polyspectrum) is the transform of the third-order moment,  $R_y(\mu, \nu)$ ,  $B_y(u, v) = \int \int_{-\infty}^{\infty} R_y(\mu, \nu) e^{-j(u\mu + v\nu)} d\mu d\nu$  where  $R_y(\mu, \nu) = E_t\{y_t y_{t+\mu} y_{t+\nu}\}$ . The *trispectrum* is defined similarly as the transform  $T_y(u, v, w) = \int \int \int_{-\infty}^{\infty} R_y(\mu, \nu, \omega) e^{-j(u\mu + v\nu + w\omega)} d\mu d\nu d\omega$  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  $n^{\text{th}}$ -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  $n^{\text{th}}$ -order moments is inverted to solve for the

autoregressive (AR) coefficients [23]. 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 the 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 samples to a higher power. Only by using more data can the impact of this assumption be minimised. In the case of the Gamma distribution, for example, the number of samples required to estimate normalised moments increases exponentially with the order of the moment [25]. 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 compromise 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  $\pi_k$  and with mean and precision (i.e. inverse variance) parameters  $\theta_k = \{\mu_k, \beta_k\}$ ,  $p(e_t) = \sum_{k=1}^K \pi_k \mathcal{N}(e_t; \theta_k)$ . Since mixture 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 [26] for formulae). Thus, we can compute all higher order spectra, by extension of (1), using

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

where  $Q_{P_n} = P_e(0, \cdots, 0)$  is essentially the  $n^{\text{th}}$  moment of  $e_t$  [24].

The interpretation above has a strong correspondence to robust estimation [27–30] 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. This makes our algorithm faster but prone to entrapment in local minima. We also extend our work [31] 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 [23]. For a fair comparison, we simulated data as described in [23]. 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 which is identical to the one obtained using standard estimation methods [23]. However, Only 128 samples were needed to estimate the AR coefficients within 2% of the true values. That is, in comparison with [23] 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.

## 2.2 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.

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  $e_t = y_t - x_t a$ . In our notation,  $y_t$  is the observation at time instant  $t$ ,  $a \in \mathbb{R}^{p \times 1}$  is the (column) vector of  $p$  autoregressive coefficients,  $x_t = [y_{t-1}, y_{t-2} \dots, y_{t-p}] \in \mathbb{R}^{1 \times p}$  is the (row) 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,  $\beta_k$ . 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  $\mu = \{\mu_1, \dots, \mu_K\}$ . The indicator variable,  $s_t$ , is simply a vector of zeros and ones. If the vector's  $k^{\text{th}}$  component is set to one,  $\delta(s_t = k) = 1$ , otherwise it  $\delta(s_t = k) = 0$  if not. Thus, the role of  $\delta(s_t = k)$  is to select only the  $k^{\text{th}}$  component of the product of Normal distributions. We model the temporal dependence between the residuals by imposing a first order Markov dependence<sup>2</sup> on the component indicator variables  $s_t$ . The probability of selecting component  $k$  at time  $t$  is thus dependent upon the component  $l$  selected 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,  $p(s_t|s_{t-1}, \pi) = \prod_{l=1}^K \mathcal{Mn}_K(s_t|\pi_l) \propto \prod_{k,l=1}^K \pi_{l_k}^{\delta(s_t=k, s_{t-1}=l)}$  with parameters  $\pi = \{\pi_{l_k}\}, \forall k, l = 1, \dots, K$ . The probability of the first component indicator,  $s_0$ , is parameterised by a  $K$ -dimensional Multinomial distribution,  $p(s_0|\pi_0) = \mathcal{Mn}_K(s_0|\pi_0) \propto \prod_{k=1}^K \pi_{0_k}^{\delta(s_0=k)}$  with parameters  $\pi_{0_k}, \forall k = 1, \dots, K$ . The full log-likelihood for a data set  $Y = \{y_1, \dots, y_T\}$  is given as

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

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

$$\{a, \mu_1, \dots, \mu_K, \beta_1, \dots, \beta_K, \pi_{0_1}, \dots, \pi_{0_K}, \pi_{1_1}, \dots, \pi_{1_K}, \dots, \pi_{K_1}, \dots, \pi_{K_K}\}$$

Model estimation is done using the variational Bayesian learning paradigm [31]. The aim of variational Bayesian learning (see [32] 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  $\theta$  are associated with conjugate prior distributions,  $p(\theta)$ , and posterior distributions,  $q(\theta)$ . Computation of the posterior distribution parameters is then

---

<sup>2</sup>While the polyspectral model does preclude this Markov property, we deliberately incorporate it in order to test, through model scoring, the correctness of our polyspectral model.

very similar to that described in [31]<sup>3</sup>. Briefly, by minimising the KL divergence between the exact and approximate distributions we obtain the individual update equations for each of the model parameters  $\theta$  and the set of all component labels  $S$ . The update equations are then iterated in the sequence

1. run one pass forward/backward algorithm, using previously obtained parameter values for the distribution of model parameters, to update the component labels.
2. Using the component labels, update the distribution parameters the initial state and the state transition probabilities.
3. Update the distribution parameters of the mean of the mixture model components.
4. Update the distribution parameters of the precisions (inverse variances) of the mixture model components.
5. Update the autoregressive coefficient distribution mean.
6. Update the autoregressive coefficient distribution precisions (inverse variance).

The variational approach iterates until the parameters of the posterior distribution have converged to fixed points. The obtained KL-divergence assumes a predetermined model order, i.e. number of components for residual density and autoregressive model coefficients is fixed. To perform optimal model order estimation model training is repeated for various model orders and subsequently the lowest divergence among the 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 posterior distribution mean,  $\hat{a}$ , of the autoregressive coefficients 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 higher order spectral power as an indicator of speed differences in the EEG. The frequency ranges, within which the total power was computed, depended on the anaesthetic agent and was based on

---

<sup>3</sup>The main difference between the methods here and in [31] is the calculation of the posterior distribution  $p(S)$  needs to be performed using the Baum-Welch equations with a likelihood term corresponding to  $E\{\log p(\theta, S)\}_{q(\theta)}$ . Precise formulae for the update equations can be found on [www.robots.ox.ac.uk/~parg](http://www.robots.ox.ac.uk/~parg) or requested from the I. Rezek.

1000 discrete and evenly spaced spectral evaluation at within the frequency range.

### 3 Understanding the Data

Electrical brain wave recording, measured by surface electrodes attached to the scalp, is known as the Electroencephalogram (EEG). EEG signals are recorded at various locations on the scalp. Regular potential changes are expressed in terms of frequencies [33]: delta range ( $0.5 - 4Hz$ ), theta range ( $4 - 8Hz$ ), alpha range ( $8 - 13Hz$ ) and beta range ( $13 - 24Hz$ ). Various pathological and cognitive conditions can characteristically induce changes in the EEG, such as frequency, amplitude, slope and waveform location.

In anaesthesia monitoring, the recorded EEG depends on the type of anaesthetic agent [34]. For instance, intravenous agents, such as Propofol, exhibit a non-reactive and slowing EEG which is observed especially in the anterior regions of the scalp. As an example, figure 2 shows a  $20sec$  section a raw EEG signal under unsedated and sedated conditions. The most striking distinguishing feature between the two conditions is the slowness and elevated amplitude of the sedated EEG.

#### 3.1 Anaesthesia Data and Recording Procedure

We applied the model to a set of three EEG data sets. Two sets were recorded during anaesthesia induced either by Propofol and Desflurane. A third recording set consisted of Propofol with Remifentanil administered as an additional analgesic. The total recording length available for training was, for each state, approximately 2 hours long or, equivalently,  $1.5 \cdot 10^6$  samples. The details of the recording procedures are described below.

##### 3.1.1 Remifentanil/Propofol

The following set of experiment data were obtained from the Dipartimento Area Critica Medico Chirurgica, Sez. di Anestesiologia e Rianimazione, of the University of Florence and were used in previous studies [17]. Eight patients, who were to undergo abdominal surgery, agreed to take part in the study. Patients were of either sex, aged between 20 and 55 years, weighed 60kg to 75kg, were of normal size and categorised as ASA I. Mechanically ventilated and pre-medicated with  $0.01mg/kg$  atropine and 700ml ringer lactate, anaesthesia consisted of i.v. administration of Remifentanil ( $15 \mu g/kg/h$ ) and Propofol, ( $1.5mg/kg$  as a bolus) followed by an infusion of

Propofol 10 mg/kg/h, Pancuronium .1 mg/kg. Remifentanil and Propofol were constantly monitored and varied according to concurrently monitored BIS values, anaesthetist's judgement, and OAA/S score. No patients required more than 10% from the initial infusion rates adjustment.

EEG was recorded from 2 frontal electrodes place 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 original recordings, which had an average length of 25 minutes, were contaminated by movement artifacts. Corrupted sections lead to failures in the model initialisation and had to be excluded from the analysis. The resulting recording lengths were on average 8 minutes long (with the shortest duration being 5 minutes and the longest 13 minutes).

### 3.1.2 Desflurane and Propofol

The following two sets of experiment data were obtained from the Academic Department of Anaesthesia based at Northwick Park Hospital in Harrow, London and were used in previous studies [2]. Fifteen patients, aged between 31 and 61, formally agreed to take part in the study. The EEG signal was recorded from the forehead to the left mastoid (with the right mastoid as common) using a 82 dB preamplifier (with a bandwidth of .5 – 400Hz, a 1st order high-pass and 3rd order low-pass Butterworth pre-filter) and digitised to a 12bit accuracy at 1 kHz sampling rate. Post digitising, the signal was down-sampled to 250Hz after low-pass removal of frequencies above 100Hz cut-off using an finite impulse response filter (47th order).

**Group A: Desflurane** Eleven patients, pre-medicated with 10mg morphine and .04mg atropine, underwent anaesthetic sedation through inhalation of Desflurane. Muscle relaxant Vecuronium was also administered by infusion to maintain good muscular relaxation throughout. The sedation procedure consisted of 3 consecutive applications of various Desflurane dosages (1.5, 3, and 6 %). The sequence of concentrations was selected randomly order in order to limit any carryover effects from one concentration to another. The end-expiratory concentration of Desflurane,  $N_2O$  and  $CO_2$  was measured (using a calibrated Datex Ultima gas analyser) to determine whether the patient was in a low, medium or deep state of 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. For classification, only the smallest and strongest concentrations were used and labelled

wake and anaesthetised, respectively.

**Group B: Propofol** Six patients, pre-medicated with 10mg of morphine and .04mg of atropine and followed by induction of anaesthesia with thiopentone (2-4mg/kg), underwent anaesthetic sedation through intravenous infusion of Propofol. Seven to ten minutes after induction with thiopentone, Propofol was infused in five equal 10 minute steps, starting with concentrations of 40 mg/kg/min and ending with 200 mg/kg/min. The Propofol blood concentration was measured by taking venous blood measurements the arm contralateral to the infused arm. The Propofol blood concentration as used as the indicator of patient level of anaesthesia (levels low, three distinct medium and high). These labels were also the classification labels given to the EEG which was recorded concurrent to the administration and during which agent concentration was near constant. For the subsequent classification analysis the concentrations used were 40 mg/kg/min and 200 mg/kg/min, which were labelled wake and deeply anaesthetised, respectively.

## 4 Preparation of the Data

We used 2.5sec long non-overlapping data segments to extract the higher order spectral features. As we are only interested in the distribution moments of order  $> 2$ , each data segment was first standardised to zero mean and unit variance. It would be impossible to distinguish higher order information content in the spectrum from second order information, if normalisation is committed at this stage. Subsequently, the normalised signal segments were used to train models with model orders ranging from 1 – 30 for the AR coefficients and 1 – 4 for the number of residual components. The AR coefficients were subsequently used to compute the higher order spectra as well as the total power in agent-dependent frequency bands. For Propofol, we empirically found the  $\delta$  band (.25Hz – 4Hz) to be most discriminative. For Desflurane the .25Hz – 8Hz band and Propofol/Remifentanil the 4Hz – 8Hz frequency band.

**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 1<sup>st</sup> 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, see figure 3. The tests showed that neither Remifentanil/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 suggests that while the residual model is needed, it does not exhibit temporal dependencies and thus supports the parametric polyspectral model assumptions.

Figure 3 shows the distribution of the optimal model order 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 Remifentanil/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 3, 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 Remifentanil/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 Remifentanil/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.

**Spectrum Calculation** For each segment the optimal model order was computed and the model parameters used for subsequent higher order spectral estimation. The entire spectral range (0 - Nyquist frequency) was evaluated at 1024 frequency values and the total spectrum, bispectrum, trispectrum amplitude in the frequency ranges  $.5 - 4 Hz$  computed from the AR coefficients using equation (2). The feature was then log-transformed prior to any classification experiments.

The same standardised data segments were also used to extract the spectrum, bispectrum and trispectrum using the traditional methods [23] implemented in the Matlab Higher Order Spectral Toolbox [35]. The resulting AR coefficients were evaluated identically to those estimated by the variational

estimator, that is computing the cumulative total poly-spectral magnitude in the frequency range of  $.25 - 4 Hz$  and then log-transformed for later use in the classification experiments.

## 5 Data Mining

Having extracted polyspectral features using our model we can return to the two main questions we wanted to address with our model. The first related to the robustness of feature estimation method and the second to the informativeness with regard to classifying depth of anaesthesia. To quantify the informativeness of the features, we followed earlier examples [2, 17, 36, 37] and performed two sets of classification experiments using artificial neural networks. One set of experiments compared traditional and variational spectral features in their ability to discriminate 2 levels of depth of anaesthesia: wake and deeply anaesthetised<sup>4</sup>. A second set of experiments compared depth of anaesthesia classification performances of various combinations of variational spectral features. Using the Netlab Software [38], for classification we considered a single-layer perceptron, i.e. generalised linear model (GLM) with logistic output, and a multi-layer perceptron (MLP), each with one output unit and the MLP with initially 10 hidden units. Both were trained using the evidence framework [39]. A comparison of the classification performance by the MLP and the GLM showed hardly any difference, suggesting that the decision boundary was sufficiently simple to be captured by the simpler GLM.

All results presented below are obtained from a patient-fold cross-validation. By this is meant, leaving out one patient for testing and cycling through all patients. The justification is that by predicting across patients we obtain a better understanding of the true "across-patient" informativeness of our features. After training the GLM classifier we computed the area under the receiver operating curve for each validation fold. The area under the receiver operating curves (AUC) presented here are the sample averages and one standard deviation confidence bounds of the individual fold's logit-transformed AUC [40].

---

<sup>4</sup>The reason for the choice of only 2 levels is pragmatic one as one data set we had available was recorded only under these two conditions.

## 5.1 Comparison of Standard and Variational Polyspectral Features

We compared the classification performances of the variational polyspectral features with those estimated using traditional methods. The average AUCs together with their  $\pm 1$  standard deviation confidence bounds are shown in table 1 for each of the anaesthetic agents and classified higher order spectral features. What the results suggest is that variational higher order spectral estimation generally results in substantially improved classification. The improvements are least in the Propofol data-set. However, there does not seem to be any need to resort to higher order spectral estimation when classifying Propofol induced anaesthesia. Propofol’s effect on the EEG is readily captured by standard spectral estimators.

We used the AUC values, obtained from every one of the patient cross-validation experiments, to quantitatively compare the improvements between the spectral features. Using a Kolmogorov-Smirnov test we examined the hypothesis that the distribution of AUC values obtained using the standard feature classifier is smaller than that of the variational classifier, and vice versa. Only the former test resulted in hypotheses that could not be rejected at the 95% level for the Desflurane and Remifenatil/Propofol data-sets. The tests, with probability of the KS statistics shown in table 2, confirm the significance of the improvements seen in table 1. Variational higher order spectral features substantially improve classification performance in Desflurane and Remifenatil/Propofol. The differences in performances between the Propofol and the Remifenatil/Propofol data-set we believe to be due to operating conditions: Propofol is collected as part of a dedicated study while the Remifenatil/Propofol data was collected prior to surgical procedures.

## 5.2 Comparison of Combinations of Variational Polyspectral Features

We’ve established that improved robustness of variational model greatly enhances the ability of polyspectral features to classify anaesthetic state. However, as the results for Propofol already seemed to indicate, higher order spectral estimation may not always be required to achieve a good description of the EEG. Thus, a naturally arising question is whether higher order spectral features carry information about the EEG that is complimentary to standard spectra. Equally we ask whether, in general, spectra of certain orders carry information complementary to those of other order.

	Propofol	Desflurane	Remifentanil/ Propofol
Variational Estimation			
Spectrum	0.99 (0.97,1.00)	0.95 (0.80,1.00)	0.94 (0.89,0.99)
Bispectrum	0.99 (0.98,1.00)	0.97 (0.94,0.98)	0.98 (0.95,1.00)
Trispectrum	0.99 (0.98,1.00)	0.96 (0.88,1.00)	0.98 (0.95,1.00)
Traditional Estimation			
Spectrum	0.99 (0.98,1.00)	0.74 (0.59,0.88)	0.92 (0.84,0.97)
Bispectrum	0.91 (0.84,0.96)	0.58 (0.47,0.69)	0.82 (0.73,0.92)
Trispectrum	0.93 (0.91,0.96)	0.59 (0.47,0.71)	0.77 (0.65,0.88)

Table 1: Differences, in classification performance, between traditionally estimated polyspectra and those estimated using variational model training. Shown are the averages and the confidence bounds of areas under receiver operating curve for different agents and higher order spectral features

We have used a McNemar [41] test to investigate differences between classifiers trained on second and higher order spectral features extracted from the same data. Obviously, improved classification implies sample-by-sample classification differences and so there is no need to compare again traditionally estimated features with variational ones. The reverse however, is not true and thus requires tests beyond those of the previous section. For each fold, which we apply to combinations of variationally estimated features, we estimated the McNemar statistic and compared the statistic of any two classifiers using a 2-sided T-test.

The average AUCs together with their  $\pm 1$  standard deviation confidence bounds are shown in table 3, for each of the anaesthetic agents and every order combination of variational spectral features. In terms of average classification performance the improvements are marginal and certainly not statistically significant for Propofol. However, the results using our Remifentanil/Propofol and Desflurane data sets show, that higher order spectral features can improve matters by, primarily, helping to identify the anaesthetised state. In the case of Remifentanil/Propofol, there appears to be a substantial reduction in the lower confidence bound when classification is performed using the bispectrum as the informative feature. Using the McNemar test described earlier the null hypothesis, that the classifier make identical errors, could be rejected (see table 4 which shows the probability of the McNemar test statistic for all possible test pairings). The results in table suggest, that the reduction of the confidence bands, seen in table 3, be-

	Variational		
	Spectrum	Bispectrum	Trispectrum
Std. Spectrum			
Propofol	×	×	×
Desflurane	$< 10^{-2}$	$< 10^{-3}$	$< 10^{-3}$
Remifentanil	×	$< 10^{-1}$	$< 10^{-2}$
Std. Bispectrum			
Propofol	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$
Desflurane	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$
Remifentanil	$< 10^{-1}$	$< 10^{-2}$	$< 10^{-3}$
Std. Trispectrum			
Propofol	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
Desflurane	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$
Remifentanil	$< 10^{-2}$	$< 10^{-4}$	$< 10^{-4}$

Table 2: Kolmogorov-Smirnov testing for significant differences, in the classification performance, between classifiers trained with standard and those trained with new polyspectral features. Shown are the probabilities of the KS test statistic, if not rejected at  $\alpha = 0.05$ . A  $\times$  indicates a rejected test. The 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> sub-rows refer to the Propofol, Desflurane and Remifentanil/Propofol data sets, respectively.

comes significant when using higher order spectral features. The results also indicate that either higher order spectral feature improve the classification performance to an almost equal degree. The situation is slightly different in the case of Desflurane (and harder to interpret). Again, higher order spectral feature improve the classification performance and the reduction passes the significance test (see table 4). However, the bispectrum and trispectrum are not quite complementary, as was the case for the Remifentanil/Propofol data. Both, bispectrum and trispectrum lift performance but the best performance classifier is obtained when they are used in conjunction.

## 6 Evaluation of Discovered Knowledge

The improved accuracy of the polyspectral features, derived from the generalised AR model, is most evident when comparing the new features with the traditionally estimated higher order spectral features. The improvements can be observed, provided higher order spectral description is actually needed to quantify anaesthetic depth for a particular agent.

	Propofol	Desflurane	Remifentanil/ Propofol
Unispectrum	0.99 (0.97,1.00)	0.95 (0.80,1.00)	0.94 (0.89,0.99)
Bispectrum	0.99 (0.98,1.00)	0.97 (0.94,0.98)	0.98 (0.95,1.00)
Trispectrum	0.99 (0.98,1.00)	0.96 (0.88,1.00)	0.98 (0.95,1.00)
Uni- & Bispec.	0.98 (0.96,1.00)	0.97 (0.87,1.00)	0.99 (0.96,1.00)
Uni- & Trispec.	0.98 (0.90,1.00)	0.96 (0.89,0.99)	0.99 (0.99,1.00)
Bi- & Trispec.	0.98 (0.92,1.00)	0.95 (0.86,0.99)	0.98 (0.95,1.00)
Uni- & Bi- & Trispec.	0.97 (0.85,1.00)	0.98 (0.96,0.99)	0.99 (0.97,1.00)

Table 3: Comparing the informativeness of the new polyspectral features and all their combinations. The averages and confidence bounds of the area under receiver operating curve for different agents and combinations of Variational higher order spectral features. (Note Unispectrum is the standard  $2^{nd}$  order spectrum).

In our study, Propofol separately has a very strong effect on the EEG which can be captured by simple spectral estimation alone: Statistical test for significant improvements all fail to reject the null hypothesis that classification rates are drawn from the same distribution. Unlike the Propofol data set, Remifentanil/Propofol was recorded under more realistic conditions (the patients are mechanically ventilated and EEG is recorded instantaneously with some un-stationary effects). For Remifentanil/Propofol variational spectral features show a statistically significant improvement at the 95% level compared to standard higher order spectral features. The effect on the EEG of the anaesthetic agent Desflurane differs from that of Propofol [34, 42] and, in our data set, the characteristic changes in EEG are harder to extract. It is in the case of Desflurane that the difference in performance between standard and variational higher order spectral estimation is most prominent: The hypothesis that classification performances of both feature extraction methods are drawn from the same distribution is more strongly reject for Desflurane than for any of the other agents.

## 7 Conclusion

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 the known short-

comings of bi- and trispectral estimation and their role in forming the BIS index, which has recently attracted some criticism from the medical community [16, 17]. Being more rather more like an expert system, our results cannot directly help finding the cause for the criticism. However, our results demonstrate that, when compared to traditional estimation methods [23] at least in the case of polyspectral estimation a more robust estimation can be achieved. This, in turn, could help make the BIS index a more robust measure.

Our approach is, in essence, the reverse (and de facto standard) of [23]. In [23] the derivation is algorithmic and focuses on the higher order moments and their estimation. Only subsequently are models postulated for these algorithms. 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 preformed in the variational Bayesian framework [32], 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 characterise them accurately. To classify only two anaesthetic states was practical, rather than scientific, choice. Consequently, we cannot make any statements about the ability of higher order spectral features to discriminate any intermediate anaesthetic states. It is worth remembering, however, that the BIS index algorithm gives more weight to polyspectral features at higher anaesthetic concentrations and, so, some conclusion can still be drawn.

On 3 different agent or anaesthetic medications our results showed that, while there is information present in the higher order spectra, it is not always needed for classifying EEG as sedated or wake. As our reviewers pointed out, the benefit of including other frequency bands in the analysis should not be ignored. There is also the very important issue of describing intermediate states. Our currently available data is insufficient to properly address this

issue and more data needs to be gathered and experiments performed. It is, therefore, only explorative and any conclusions need to be made with care. The message from this data is evidently that, by optimising the model, including the AR part, and choosing the appropriate spectral band, there are some improvements to be made when using higher order spectral information for classification of anaesthetic depth.

### **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 Remifentanyl/Propofol data. Particular thanks also to the reviewers for their exciting and excellent comments. This work was funded by EPSRC Grant GR/N03471.

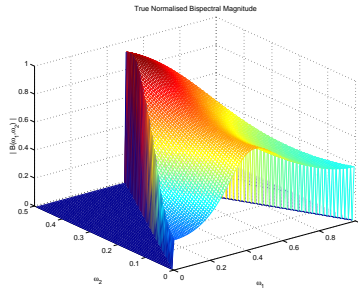


Figure 1: Magnitude of bispectrum estimated using simulated data of a 4th order autoregressive model of Example 1 in [23].

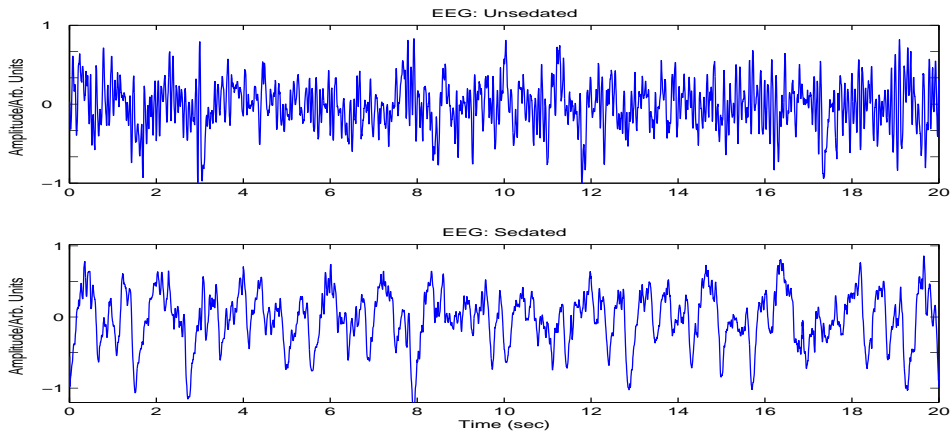


Figure 2: Typical raw EEG recording taken under wake conditions (top trace) and sedated conditions using Desflurane as the anaesthetic agent.

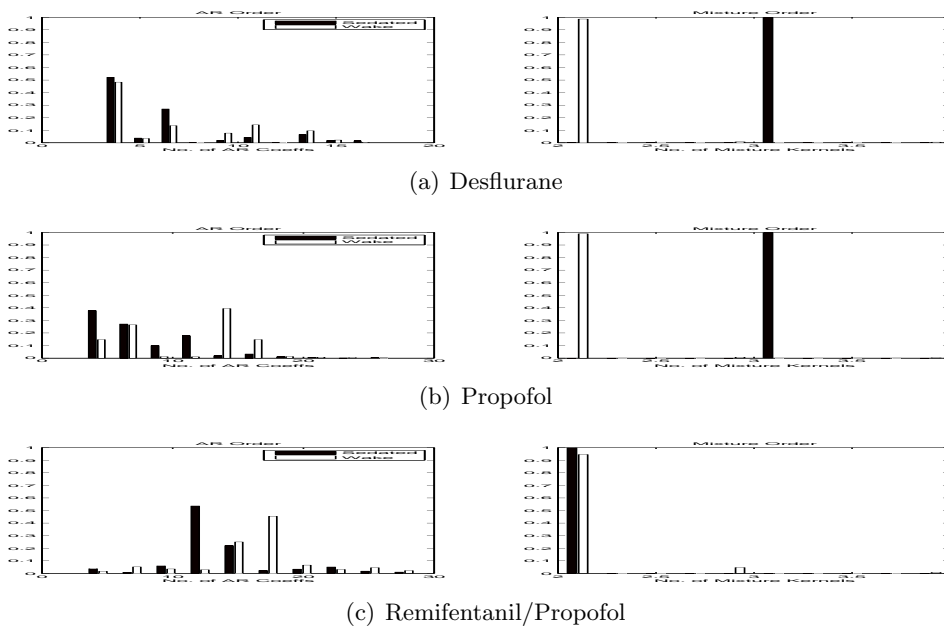


Figure 3: Empirical distribution of optimal number of AR coefficients (left histogram) and residual mixture components (right histogram) for three data sets, with wake state marked in white and sedated in black.

	Bispectrum	Trispectrum	Uni- & Bispectrum	Uni- & Trispectrum	Bi- & Trispectrum	Uni- & Bi- & Trispectrum
Unispectrum Remifentanil - Desflurane -	$< 10^{-2}$ $\times$	$< 10^{-3}$ $< 10^{-1}$	$< 10^{-2}$ $< 10^{-1}$	$< 10^{-3}$ $< 10^{-2}$	$< 10^{-3}$ $< 10^{-1}$	$< 10^{-3}$ $< 10^{-1}$
Bispectrum Remifentanil Desflurane	- -	$\times$ $\times$	$\times$ $< 10^{-1}$	$< 10^{-1}$ $< 10^{-1}$	$< 10^{-1}$ $< 10^{-2}$	$< 10^{-1}$ $< 10^{-1}$
Trispectrum Remifentanil Desflurane		- -	$\times$ $< 10^{-1}$	$< 10^{-1}$ $\times$	$\times$ $< 10^{-1}$	$< 10^{-2}$ $< 10^{-1}$
Uni- & Bispectrum Remifentanil Desflurane			- -	$< 10^{-1}$ $< 10^{-1}$	$< 10^{-1}$ $< 10^{-1}$	$\times$ $< 10^{-1}$
Uni & Trispectrum Remifentanil Desflurane				- -	$< 10^{-2}$ $< 10^{-3}$	$\times$ $< 10^{-1}$
Bi- & Trispectrum Remifentanil Desflurane					- -	$< 10^{-2}$ $< 10^{-1}$

Table 4: Testing for significant differences between classifiers trained using various combinations of variational polyspectral features. Shown are the probabilities of the McNemar test statistic, if not rejected at  $\alpha = 0.05$  and a  $\times$  indicates a rejected test. The 1<sup>st</sup> and 2<sup>nd</sup> entries refer to the Remifentanil/Propofol and Desflurane data sets, respectively.

## References

- [1] D.E. Maynard and J.L. Jenkinson. The cerebral function analysing monitor. Initial clinical experience, application and further development. *Anaesthesia*, 39:678–690, 1984.
- [2] M. Krkic, S.J. Roberts, I. Rezek, and C. Jordan. EEG-Based Assessment of Anaesthetic-Depth Using Neural Networks. *IEE Proceedings on AI Methods in Biosignal Analysis*, pages 10/1–10/6, Apr 1996.
- [3] A.L. Vanluchene, H. Verecke, O. Thas, E.P. Mortier, S.L. Shaferand, and M.M. Struys. Spectral entropy as an electroencephalographic measure of anesthetic drug effect: a comparison with bispectral index and processed midlatency auditory evoked response. *Anesthesiology*, 101(1):34–42, 2004.
- [4] H. Viertiö-Oja, V. Maja, M Särkelä, P Talja, N. Tenkanen, H. Tolvanen-Laakso, M. Paloheimo, A. Vakkuri, A. Yli-Hankala, and P. Meriläinen. Description of the Entropy algorithm as applied in the Datex-Ohmeda S/5 Entropy Module. *Acta Anaesthesiol Scandinavica*, 48(2):154–61, 2004.
- [5] C. Thornton and R.M. Sharpe. Evoked responses in anaesthesia. *British Journal of Anaesthesia*, 81:771–781, 1998.
- [6] M. Holt, M.A. Tooley, F.C. Forrest, C. Prys-Roberts, and L. Tarassenko. Use of parametric modelling and statistical pattern recognition in detection of awareness during general anaesthesia. *IEE Proceedings - Science, Measurement and Technology*, 145(6):307–316, 1998.
- [7] P.S. Myles, K. Leslie, J. McNeil, A. Forbes, and M.T.V. Chan. Bispectral index monitoring to prevent awareness during anaesthesia: the B-Aware randomised controlled trial. *The Lancet*, 363(9423), 2004.
- [8] I.J. Rampil, J.S. Kim, R. Lenhardt, C. Negishi, and D.I. Sessler. Bispectral EEG index during nitrous oxide administration. *Anesthesiology*, pages 671–677, 1998.
- [9] R. Myles and D. Williams. Patient satisfaction after anaesthesia and surgery: results of a prospective survey of 10,811 patients. *British Journal of Anaesthesia*, 84, 2000.

- [10] A. Miller, J.W. Sleight, J. Barnard, and D.A. Steyn-Ross. Does bispectral analysis add anything but complexity? BIS sub-components may be superior to BIS for detection of awareness. *British Journal of Anaesthesia*, 93(4):596–597, 2004.
- [11] G. Barr, J.G. Jakobson, A. Owall, and R.E. 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.
- [12] J.W. Johansen and P.S. Sebel. Development and clinical application of electroencephalographic bispectrum monitoring. *Anesthesiology*, 93:1336–1344, 2000.
- [13] N. Morioka, M. Ozaki, T. Matsukowa, D.I. Sessler, K. Atarashi, and H. Suzuki. Ketamine causes a paradoxial increase in the bispectral index. *Anesthesiology*, 87:A502, 1997.
- [14] J. D. Hall and G. G. Lockwood. Bispectral index: comparison of two montages. *British Journal of Anaesthesia*, 80(3):342–344, 1998.
- [15] K. Kuizenga, J.M.K.H. Wierda1, and C.J. 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.
- [16] D.K. Tempe and L. Satyanarayana. Is there any alternative to the Bispectral Index Monitor? *British Journal of Anaesthesia*, 29(1):1–3, 2004. Editorial.
- [17] O. Ortolani, A. Conti, A. Di Filippo, C. Adembri, E. Moraldi, A. Evangelisti, M. Maggini, and S.J. Roberts. EEG signal processing in anaesthesia. Use of a neural network technique for monitoring depth of anaesthesia. *British Journal of Anaesthesia*, 88(5):644–8, 2002.
- [18] G.L. Bretthorst. *Bayesian Spectrum Analysis and Parameter Estimation*. New-York: Springer Verlag, 1989.
- [19] S.M. Kay and S.L. Marple. Spectrum Analysis - A Modern Perspective. *Proceedings of the IEEE*, 69(11):1380–1419, Nov 1981.
- [20] A. Webb, editor. *Statistical Pattern Recognition*. John Wiley, 2002.
- [21] I. Rezek, R. Conradt, S.J. Roberts, and E. Siva. Depth of Anaesthesia Assessment with Generative Polyspectral Models. In *International Conference on Machine Learning and Applications*, 2005.

- [22] I.J. Rampil. A primer for EEG signal processing in anaesthesia. *Anesthesiology*, 89:981–1001, 1998.
- [23] M.R. Raghuveer and C.L. Nikias. Bispectrum Estimation: A Parametric Approach. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(4):1213–1230, 1985.
- [24] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Singapore, 1991.
- [25] R.J.A. Tough. The estimators of normalized moments of the gamma distribution. *Journal of Physics A: Mathematical and General*, 27:7883–7892, 1994.
- [26] A. Stuart and J.K. Ord. *Distribution Theory*, volume 1 of *Kendall's Advanced Theory of Statistics*. Edward Arnold, 1994.
- [27] 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.
- [28] M. West. Outlier models and prior distributions in Bayesian linear regression. *J. R. Stat. Soc. B*, 46(3):431–439, 1984.
- [29] P.J. Huber. *Robust Statistics*. J. Wiley, 2003.
- [30] R. Tibshirani. Regression selection and shrinkage via the lasso. *J Royal Stat Soc B*, 1(267-288), 1996.
- [31] W. Penny and S. Roberts. Variational Bayes for Generalised Autoregressive Models. *IEEE Transactions on Signal Processing*, 50(9):2245–2257, 2002.
- [32] T.S. Jaakkola. Tutorial on Variational Approximation Methods. In M. Opper and D. Saad, editors, *Advanced Mean Field Methods: Theory and Practice*. MIT Press, 2000.
- [33] W. Pschyrembel. *Klinisches Wörterbuch*. Walter de Gruyter, 260th edition, 2002.
- [34] H.C. Hemmings and P.M. Hopkins. *Foundations of Anesthesia*. Elsevier, 2005.

- [35] A. Swami, J.M. Mendel, and C.L. Nikias. *Higher-Order Spectral Analysis Toolbox User's Guide, R12 For Use with MATLAB<sup>®</sup>*. United Signals and Systems, Inc, 2000.
- [36] D.A. Linkens and L. Vefghi. Recognition of patient anaesthetic levels: neural network systems, principle analysis, and canonical discriminant variates. *Artificial Intelligence in Medicine*, 11(2):155–173, 1997.
- [37] H.U. Rehman, D.A. Linkens, and A.J. Asbury. Neural networks and nonlinear regression modelling and control of depth of anaesthesia for spontaneously breathing and ventilated patients. *Computer Methods and Programming In Biomedicine*, 40:227–247, 1993.
- [38] I. Nabney. *Netlab: Algorithms for Pattern Recognition*. Springer Verlag, 2002.
- [39] D. MacKay. The Evidence Framework Applied to Classification Networks. *Neural Computation*, 4(5):720–736, 1992. [citeseer.nj.nec.com/mackay92evidence.html](http://citeseer.nj.nec.com/mackay92evidence.html).
- [40] S. A. Macskassy and F. Provost. Confidence Bands for ROC Curves: Methods and an Empirical Study. In *Proceedings of the First Workshop on ROC Analysis in AI (ROCAI) at ECAI, 2004*.
- [41] B. Ripley. *Pattern Recognition and Neural Networks*. Cambridge University Press, 2000.
- [42] R. K. Stoelting and S.C. Hillier. *Handbook of Pharmacology and Physiology in Anesthetic Practice*. Lippincott Williams & Wilkins, 2006.

### Figure Captions

1. Magnitude of bispectrum estimated using simulated data of a 4th order autoregressive model of Example 1 in [23].
2. Graph showing the probabilistic relationships between the model's variables.
3. Typical raw EEG recording taken under wake conditions (top trace) and sedated conditions using Desflurane as the anaesthetic agent.
4. Empirical distribution of optimal number of AR coefficients (left histogram) and residual mixture components (right histogram) for three data sets, with wake state marked in white and sedated in black.