Hidden Markov models for complex stochastic processes: A case study in electrophysiology.

M. White, Nicole, Helen Johnson, Peter Silburn, Judith Rousseau, Kerrie Mengersen

Hidden Markov models for complex stochastic processes: A case study in electrophysiology.

Nicole White[1⋆], Helen Johnson[1], Peter Silburn[2], Judith Rousseau[3] & Kerrie Mengersen[1]

[1] Mathematical Sciences, Queensland University of Technology, Brisbane, Australia

[2] St. Andrew's War Memorial Hospital and Medical Institute, Brisbane, Australia

[3] CEREMADE, Universitè Paris Dauphine, Paris, France

⋆ nm.white@qut.edu.au

1. Introduction

Understanding the complexities of human physiology remains an exciting and challenging field in modern medicine. Of the many research streams in this field, a popular area is the study of action potentials (APs), or electrophysiology. Defined by a rapid rise and fall of electrical potential in an activated cellular membrane, an AP is visibly characterised by a unique waveform shape or trajectory, that is considered as an event separate from background noise. Biologically, APs play a central role in the activation of intracellular processes in the human body, including heart and muscle contraction, the release of insulin from the pancreas, facilitation of communication between neurons in the brain

and motor sensory signals between the brain and muscles and tissues in the body. As such, the study of APs has the potential to gain understanding of these processes and how they are affected under different environmental, genetic and physical conditions.

The understanding of this and other complex phenomena involves statistical analyses of stochastic processes. What is essentially random behaviour over time and/or space, the collection of data on stochastic processes is carried out in a myriad of contemporary research areas, including but not limited to finance, economics, bioinformatics, signal processing and machine learning. Regardless of their origin, analysis of these data centre on the uncovering of patterns or trends amidst what is otherwise perceived as unpredictable behaviour or noise. For example, in speaker diarization (Gales and Young, 2008; Tranter and Reynolds, 2006), an important topic in speech processing, analysis aims to not only recognise voices in an audio recording originating from different speakers, but also transitions between speakers and the number of speakers, at the same time accounting for natural variation in an individual's tone, volume and pitch. Likewise, in bioinformatics, the alignment of multiple, highly variable DNA sequences is central to the identification of potential regions of functionality or structural importance (Holmes and Bruno, 2001; Eddy, 1995).

Of the statistical tools available for analysis, *hidden Markov models* (HMMs) have proven successful in light of their relative simplicity and flexibility in describing a wide variety of stochastic processes (Cappé et al., 2005; MacDonald and Zucchini, 1997). A form of latent variable model, HMMs aim to describe the observed outputs of a stochastic process by a finite alphabet of unobservable, discrete valued "states", where different states are taken to represent different features of the process' behaviour and are inferred from the data. Relating these models to other chapters in this book, a HMM is a type of Bayesian Network but may also be conceptualised as a finite mixture model for time dependent observations. In this chapter, we aim to introduce these models to the reader and to demonstrate their applicability to a case study in electrophysiology. Specifically, we seek to apply a HMM to the identification and sorting of action potentials in extracellular recordings (White, 2011).

2. Case study: Spike identification and sorting of extracellular recordings

In this chapter, we focus on the analysis of action potentials in brain, made possible by the collection of extracellular recordings. Extracellular recordings consist of measurements of electrical potential discharged

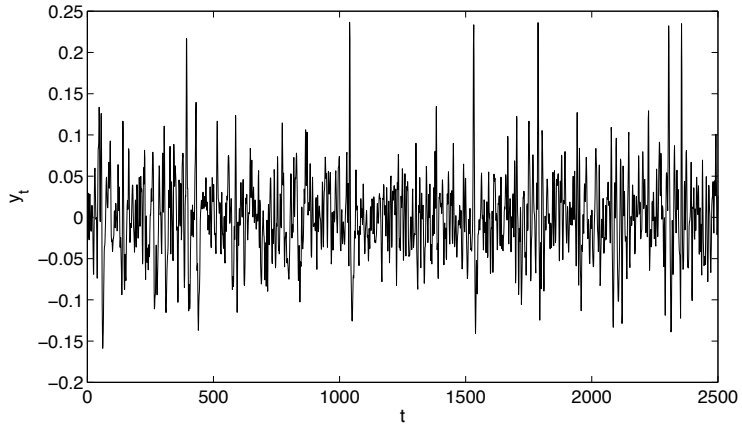

FIGURE 1. Example of an neural extracellular recording, to be analysed in Section 4

by either a single or multiple cells, in this case neurons, over time. An example of an extracellular recording is given in Figure 1.

Applying the definition of an AP, their presence in Figure 1 are visualised as "spikes", with an amplitude notably higher (or lower) than the background noise. That said, it is not clear which of these spikes genuinely correspond to APs and which merely form part of the background noise. For this reason, this area of research has witnessed substantial literature on statistical methods for reliable spike detection, from simple thresholding rules (Freeman, 1971; Thakur et al., 2007; Rizk and Wolf, 2009) to more adaptive alternatives, including nonlinear and wavelet-based detection methods (Yang and Shamma, 1988; Kim and Kim, 2000; Nenadic and Burdick, 2005; Mtetwa and Smith, 2006).

In this chpater, HMMs are considered for the unsupervised detection of APs in extracellular recordings. By assuming the behaviour of an AP as a sequence of unobservable or latent states, HMMs model transitions between these states and therefore provide a means of unsupervised spike detection, without the need to set thresholds. In the next Section, we first consider a single-sequence HMM and how it may be specified for the spike detection problem. This model is then extended to multiple, independent HMMs, in an attempt to model spikes with distinct trajectories in the same recording. This concept was first proposed by Herbst et al. (2008), as a novel solution for the simultaneous identification and assignment of spikes to source cells, the latter commonly referred to as the "spike sorting" problem. In this Chapter, the model proposed by Herbst et al. (2008) is recast into the Bayesian framework, and is a summary of work presented in White (2011).

To explore these ideas, we consider a case study of extracellular recordings taken during Deep Brain Stimulation, a popular treatment for advanced Parkinson's disease. Deep Brain Stimulation is a surgical procedure involving the placement of electrodes in an affected part of the brain, to provide a constant source of electrical stimulation. Best described as a "brain pacemaker", this constant supply of electrical pulses has been consistently shown to alleviate symptoms associated

with Parkinson's disease (Limousin et al., 1998; Kumar et al., 1998; Krack et al., 2003; Kleiner-Fisman et al., 2003). In this Chapter, we model an extracellular recording taken at the Subthalamic Nucleus (STN), a popular surgical target for Deep Brain Stimulation.

3. Models and Methods

3.1. **What is a hidden Markov model?** As described in the Introduction, a HMM is defined by its use of latent or "hidden" states to describe the behaviour of each value in an observed stochastic process, defined as $\mathbf{y} = y_1, y_2, \ldots, y_T$ or in shorthand $\mathbf{y}_{1:T}$. Given this data, the simplest HMM exists when $\mathbf{y}_{1:T}$ is modelled using a single latent sequence, say $\mathbf{z}_{1:T}$, taking a finite number of discrete values, $z_t \in (1, \ldots, G)$. This scenario is shown in Figure 2. The elements of

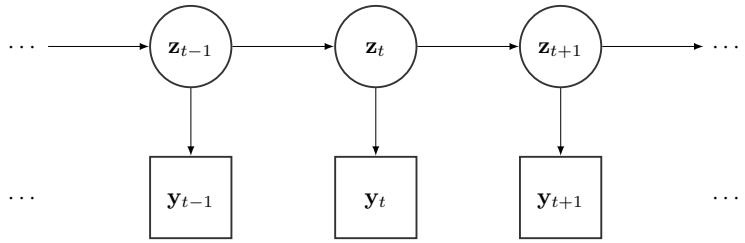

Figure 2. Directed Acyclic Graph of a simple HMM, with dependencies observed in the observed data y are attributed to a latent variable z, modelled by a first order Markov chain.

$\mathbf{z}_{1:T}$ in Figure 2 are connected due to their generation by a first order Markov chain, whereby the distribution of z_t is dependent only on the state inferred at time $t-1$, as opposed to its entire history.

The transition from z_{t-1} to z_t is governed by a time homogeneous transition matrix, $\mathbf{Q}$, such that the distribution of each z_t is given by

$$z_t|z_{t-1}, \mathbf{Q} \sim \mathbf{Q}_{z_{t-1},:} \tag{1}$$

where $\mathbf{Q}_{z_{t-1},:}$ denotes the row of $\mathbf{Q}$ corresponding to state z_{t-1}. The elements of $\mathbf{Q}$ are defined as $q_{ij} = Pr(z_t = j|z_{t-1} = i)$. Note that as a result of introducing $\mathbf{z}_{1:T}$, the time dependence exhibited by $\mathbf{y}_{1:T}$ is now assumed to be completely attributed to the latent sequence.

Accompanying this transition matrix and completing the model, we also specify a distribution for y_t, conditional on z_t. In doing so, changes in the behaviour of the stochastic process are defined by changes in the parameters of the proposed distribution. For example, given Poisson distributed data, a suitable choice for y_t may be

$$y_t|z_t, \mathbf{Q}, \lambda \sim Poisson(\lambda_{z_t}) \tag{2}$$

such that the behaviour of $\mathbf{y}_{1:T}$ is assumed to be driven by changes in the unknown rate, λ, which depends on $\mathbf{z}_{1:T}$.

Taking this basic definition, the remainder of this Section aims to demonstrate how HMMs can be adapted to the problems of spike identification and spike sorting. This discussion begins with the application of the single HMM in Figure 2, where the specification of the transition matrix $\mathbf{Q}$ can be done in such a way to model the dynamics of a single AP. This model is then extended to multiple independent latent sequences, commonly referred to as a *factorial* HMM (fHMM) (Ghahramani and Jordan, 1997).

3.2. **Modelling a single AP: Application of a simple HMM.** Given the model defined in Section 3.1, we first consider modelling a single AP over time, given an extracellular recording consisting of observations $\mathbf{y}_{1:T}$, which are subject to background noise. The latent variable z_t therefore represents the state of behaviour of the AP at time t.

Central to the modelling of this behaviour is the specification of the transition probability matrix $\mathbf{Q}$. To do this, we consider the behaviour of a typical action potential over time. Explicitly, we acknowledge that at any given time, an AP exists exclusively in one of three phases, depicted in Figure 3.

The first of these phases, *resting*, corresponds to times of equilibrium membrane potential and therefore, the AP is indistinguishable

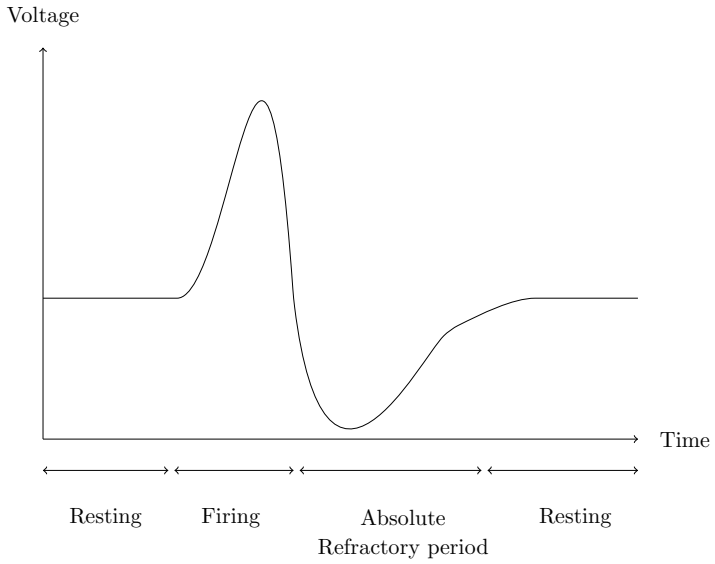

FIGURE 3. Schematic for the behaviour of an AP, where its behaviour can be divided into one of three phases: resting, firing and the refractory period.

from the background noise. When this equilibrium is interrupted and the potential rises above a certain threshold, the AP's trajectory is initiated, a state which we refer to as the *firing* state and refers to the processes of depolarization and repolarization. Upon completion of the firing phase, the AP enters what is known as the *absolute refractory period*, before returning to the resting period.

When transitioning between these phases it is assumed that, from the time of firing until the end of the absolute refractory period, it is biologically impossible for the AP to refire. Given this description, **Q**

is given by,

$$\mathbf{Q} = \begin{bmatrix} q_{11} & 1-q_{11} & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}_{G \times G} \tag{3}$$

where $z_t = 1$ denotes that the AP is in the resting phase at time t. The number of states G is chosen proportional to the sampling frequency of $\mathbf{y}_{1:T}$ and enforces the aforementioned assumption. For example, if an extracellular recording has a sampling rate of 20KHz, $G = 20$ means from the moment the AP fires, with probability $1 - q_{11}$, it is unable to refire for 1ms.

Conditional on z_t, y_t is assumed Normally distributed with unknown mean and common unknown variance, namely

$$y_t | z_t = g, \mathbf{Q}, \mu, \sigma^2 \sim N\left(\mu_g, \sigma^2\right) \qquad g = 1, \ldots, G. \tag{4}$$

This distributional form is chosen for two reasons. Firstly, the common variance term, σ^2, provides an estimate of the background noise. Secondly, by conditioning the unknown mean on z_t, we obtain an estimate of the average voltage across all defined states of the AP's behaviour.

By ordering these means for 1 to G, the predicted average shape or trajectory of the AP is produced, commonly referred to as the spike template.

For all unknown parameters, prior distributions are conjugate and of the form,

$$\sigma^2 \sim IG(\alpha, \beta) \tag{5}$$

$$\mu|\mathbf{Q}, \sigma^2, b \sim N\left(b, \sigma^2\tau^2\right) \tag{6}$$

$$q_{11}|\mathbf{Q} \sim Beta(\gamma, \phi) \tag{7}$$

The inclusion of σ^2 in Equation (6) allow for the derivation of closed-form full conditionals for model estimation, discussed further in Section 3.5.

In summary, by specifying an appropriate transition probability matrix, one is able to not only predict when the AP enters the firing state but also its expected trajectory. This second property may become useful when multiple APs are present, addressed in the next Section, as a possible solution to the spike sorting problem.

3.3. **Multiple neurons: An application of a factorial HMM.** In some cases, it is possible that an extracellular recording contains spikes from more than a single AP. With this comes the additional task of spike sorting, whereby identified spikes are classified by predicted

trajectory, as this is indicative of their origin from different cells. In order to accommodate this general case, the current HMM must be extended to include N latent sequences, one for the behaviour of each distinct AP. This collection of sequences is denoted thorughout this Chapter by $\mathbf{z}_{1:T}^{1:N} = (\mathbf{z}_{1:T}^{1}, \ldots, \mathbf{z}_{1:T}^{N})$.

The introduction of an additional $N - 1$, independent latent sequences results in the specification of a factorial HMM (fHMM) (Ghahramani and Jordan, 1997), depicted in Figure 4.

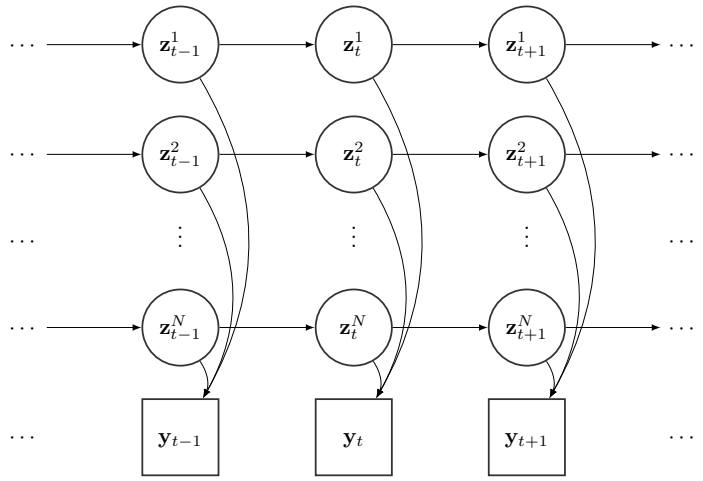

FIGURE 4. Factorial HMM for N independent APs.

The independence amongst latent sequences assumes that the behaviour of APs originating from different neurons is independent. As a consequence, the behaviour of each AP can be expressed by its own

transition probability matrix, $\mathbf{Q}^n$; $n = 1, \ldots, N$. Furthermore, it means that, for all t, the joint transition probability for $t-1 \rightarrow t$ is the product of individual AP transition probabilities,

$$Pr(z_t^{1:N}|z_{t-1}^{1:N}) = \prod_{n=1}^{N} q^n_{z_t^n, z_{t-1}^n}. \tag{8}$$

This assumption is chosen entirely for of computational convenience, with model estimation described in Section 3.4.

We also assume for this model that the observed voltage at any given time is the sum of individual voltages from all APs. This a biologically relevant assumption and allows for the possibility of overlapping APs, namely when more than one AP enters the firing state simultaneously or one AP fires during another AP's refractory or resting phase. Taking this second assumption into account, Equation (4) becomes

$$y_t|\mathbf{y}_{1:t-1}, z_t^{1:N}, \mathbf{Q} \sim N\left(\sum_{n=1}^{N} \mu^n_{z_t^n}, \sigma^2\right) \tag{9}$$

where $\mu^n_{z_t^n}$ is the mean for the n^{th} AP conditional on z_t^n. As a result of this modification to the distribution of y_t, each AP is defined by its own spike template, $\boldsymbol{\mu}^n = (\mu_1^n, \ldots, \mu_G^n)$, and common variance, σ^2, as per the single AP model. This model was first introduced by Herbst et al. (2008) in the maximum likelihood framework.

In this Chapter, the extension of the model into the Bayesian framework requires the extension of priors in Equations (5) to (7) to accommodate the $N > 1$ case. While Equation (5) remains unchanged, this extension involves the hierarchical prior on μ defined by Equation (6) being replaced by an independence prior over n. Also, for each q_{11}^n, we specify a $Beta(\gamma, \phi)$ distribution.

3.4. **Model Estimation and Inference.** Estimation of any HMM, regardless of its structure, consists of two key tasks: the estimation of *(i)* the latent sequence/s and *(ii)* the parameters associated with each latent state. In a Bayesian framework, we can achieve these tasks using a combination of MCMC, in this case Gibbs sampling, and a unique algorithm for latent sequence estimation, henceforth referred to as the forward-backward algorithm. For the sake of brevity, we restrict our attention to the factorial HMM, as this model reduces to a single sequence HMM when $N = 1$.

3.4.1. *The Forward-Backward Algorithm.* First developed by Rabiner (1989) in the classical statistical framework, the forward-backward algorithm aims to produce samples from the joint distribution of the latent sequence or sequences, represented by

$$\mathbf{z}^{1:N} \sim p(\mathbf{z}_{1:T}^{1:N} | \mathbf{y}_{1:T}, \mathbf{Q}^{1:N}, \mathbf{\Theta}) \tag{10}$$

Here, $\boldsymbol{\Theta}$ defines the parameter set $\{\boldsymbol{\mu}, \boldsymbol{\sigma}^2\}$. The generation of samples from Equation (10) is simplified by first re-expressing this joint distribution by the decomposition,

$$
\begin{aligned}
p(\mathbf{z}_{1:T}^{1:N}|\mathbf{y}_{1:T}, \mathbf{Q}^{1:N}, \boldsymbol{\Theta}) &= p(\mathbf{z}_T^{1:N}|\mathbf{y}_{1:T}, \mathbf{Q}^{1:N}, \boldsymbol{\Theta}) \times \ldots \\
&\quad \times p(\mathbf{z}_t^{1:N}|\mathbf{z}_{t+1}^{1:N}, \ldots, \mathbf{z}_T^{1:N}, \mathbf{y}_{1:T}, \mathbf{Q}^{1:N}, \boldsymbol{\Theta}) \times \ldots \\
&\quad \times p(\mathbf{z}_1^{1:N}|\mathbf{z}_2^{1:N}, \ldots, \mathbf{z}_T^{1:N}, \mathbf{y}_{1:T}, \mathbf{Q}^{1:N}, \boldsymbol{\Theta}). \qquad (11)
\end{aligned}
$$

This implies that simulation of $\mathbf{z}^{1:N}$ involves calculation of each probability mass function on the right hand side of Equation (11). This task is further simplified by the knowledge that the dependence exhibited by $\mathbf{z}^{1:N}$ is restricted to first order, as illustrated in both Figures 2 and 4. Chib (1996) showed that each of these terms, through the application of Bayes' rule, is given by,

(12)

$$
p(\mathbf{z}_t^{1:N}|\mathbf{z}_{t+1:T}^{1:N}, \mathbf{y}_{1:T}, \mathbf{Q}^{1:N}, \boldsymbol{\Theta}) = \frac{p(\mathbf{z}_t^{1:N}|\mathbf{y}_{1:t}, \mathbf{Q}^{1:N}, \boldsymbol{\Theta}) p(\mathbf{z}_{t+1}^{1:N}|\mathbf{z}_t^{1:N}, \mathbf{Q}^{1:N})}{\sum_{\mathbf{z}_{t+1}^{1:N}} p(\mathbf{z}_t^{1:N}|\mathbf{y}_{1:t}, \mathbf{Q}^{1:N}, \boldsymbol{\Theta}) p(\mathbf{z}_{t+1}^{1:N}|\mathbf{z}_t^{1:N}, \mathbf{Q}^{1:N})}.
$$

with $p(\mathbf{z}_{t+1}^{1:N}|\mathbf{z}_t^{1:N}, \mathbf{Q}^{1:N})$ corresponding to the joint transition matrix probability defined in Equation (8). The term $p(\mathbf{z}_t^{1:N}|\mathbf{y}_{1:t}, \mathbf{Q}^{1:N}, \boldsymbol{\Theta})$ is computed recursively for each t, a process known as forward filtering. Beginning at $t = 1$, the recursive scheme consists of iterating between

the two steps:

$$p(\mathbf{z}_t^{1:N}|\mathbf{y}_{1:t}, \mathbf{Q}^{1:N}, \mathbf{\Theta}) = \frac{p(\mathbf{z}_t^{1:N}|\mathbf{y}_{1:t-1}, \mathbf{Q}^{1:N}, \mathbf{\Theta})p(y_t|\Theta, z_t^{1:N})}{\sum_{z_t^{1:N}} p(\mathbf{z}_t^{1:N}|\mathbf{y}_{1:t-1}, \mathbf{Q}^{1:N}, \mathbf{\Theta})p(y_t|\Theta, z_t^{1:N})} \tag{13}$$

$$p(\mathbf{z}_t^{1:N}|\mathbf{y}_{1:t-1}, \mathbf{Q}^{1:N}, \mathbf{\Theta}) = \sum_{z_{t-1}^{1:N}} p(\mathbf{z}_t^{1:N}|\mathbf{z}_{t-1}^{1:N}, \mathbf{Q}^{1:N})p(\mathbf{z}_{t-1}^{1:N}|\mathbf{y}_{1:t-1}, \mathbf{Q}^{1:N}, \mathbf{\Theta}) \tag{14}$$

Upon computing Equation (13) for all t, sampling from the joint distribution of $\mathbf{z}^{1:N}$ is achieved via Equation (12). Explicitly, sampling is performed backwards in time and is thus referred to as backward smoothing. This process is outlined in Equation (15).

$$\begin{aligned}
\mathbf{z}_T^{1:N} &\sim p(\mathbf{z}_T^{1:N}|\mathbf{y}_{1:T}, \mathbf{Q}^{1:N}, \mathbf{\Theta}) \\
\mathbf{z}_{T-1}^{1:N} &\sim p(\mathbf{z}_{T-1}^{1:N}|\mathbf{z}_T^{1:N}, \mathbf{y}_{1:T}, \mathbf{Q}^{1:N}, \mathbf{\Theta}) \\
&\vdots \\
\mathbf{z}_1^{1:N} &\sim p(\mathbf{z}_1^{1:N}|\mathbf{z}_2^{1:N}, \ldots, \mathbf{z}_T^{1:N}, \mathbf{y}_{1:T}, \mathbf{Q}^{1:N}, \mathbf{\Theta}).
\end{aligned} \tag{15}$$

The forward backward algorithm has found success in many areas of research, in particular speech processing, also appearing under the guise of the Baum-Welch (Baum et al., 1970) and Viterbi (Forney Jr, 1973) algorithms. For a more comprehensive discussion of this algorithm in the Bayesian framework, see Frühwirth-Schnatter (2006).

3.4.2. *Gibbs Sampler.* The choice of conjugate priors on all unknown parameters means model estimation can be implemented using Gibbs

sampling (Geman and Geman, 1984; Smith and Roberts, 1993; Chib, 1996). For our model, the Gibbs sampler involves sampling from the following full conditionals:

$$\mathbf{z}^{1:N} \mid \mathbf{y}_{1:T}, \mathbf{Q}^{1:N}, \mathbf{\Theta} \tag{16}$$

$$\mathbf{Q}^{1:N} \mid \mathbf{y}_{1:T}, \mathbf{z}^{1:N} \tag{17}$$

$$\mathbf{\Theta} \mid \mathbf{y}_{1:T}, \mathbf{z}^{1:N} \tag{18}$$

Equation (16) is sampled from using the forward-backward algorithm described in the previous Section, given current realisations of $\mathbf{Q}^{1:N}$ and $\mathbf{\Theta}$. Given $\mathbf{z}^{1:N}$, $\mathbf{Q}^{1:N}$ and $\mathbf{\Theta}$, are then updated via their corresponding full conditionals. For $\mathbf{Q}^{1:N}$, this involves updating each q_{11}^n, for $n = 1, \ldots, N$. These updates are in the of Beta distributions,

$$q_{11}^n = Beta(\gamma + m_{11}^n, \phi + 1) \tag{19}$$

where $m_{11}^n = \#(z_t^n = 1 | z_{t-1}^n = 1)$, or the number of transitions where the latent sequence remains in the resting state. For Θ, we note that the posterior can be decomposed to give

$$p(\boldsymbol{\mu}, \boldsymbol{\sigma}^2 | \mathbf{y}_{1:T}, \mathbf{z}^{1:N}) = p(\boldsymbol{\mu} | \boldsymbol{\sigma}^2, \mathbf{y}_{1:T}, \mathbf{z}^{1:N}) \times p(\sigma^2 | \mathbf{y}_{1:T}, \mathbf{z}^{1:N}). \tag{20}$$

The full conditional for $\boldsymbol{\mu}$ follows a Multivariate Normal distribution,

$$
\begin{aligned}
p(\boldsymbol{\mu}|\sigma^2, \mathbf{y}_{1:T}, \mathbf{z}^{1:N}) &\sim MVN\left(\left(\frac{b}{\tau^2} + d(\mathbf{y})\right)^T \Sigma, \sigma^2\Sigma\right) \\
\mu &= \left[\mu_1^1, \mu_2^1, \ldots, \mu_G^1, \mu_1^2, \ldots, \mu_G^N\right]^T \\
d(\mathbf{y}) &= \left[d_{z_t^1=1}(\mathbf{y}), \ldots, d_{z_t^1=G}(\mathbf{y}), d_{z_t^2=1}(\mathbf{y}), \ldots, d_{z_t^N=G}(\mathbf{y})\right]^T \\
d_{z_t^n=g}(\mathbf{y}) &= \sum_{t=1}^{T} \mathbb{I}\left\{z_t^n = g\right\} y_t.
\end{aligned}
\tag{22}
$$

Here, μ has been defined as a row vector, containing all possible combinations or $n = 1, \ldots, N$ and $g = 1, \ldots, G$. This allows for all mean parameters to be sampled in a single, vectorised step and is done so for computational convenience.

The remaining term, σ^2, is updated by an Inverse Gamma distribution, having integrated out $\boldsymbol{\mu}$.

$$
\sigma^2|\mathbf{y}_{1:T}, \mathbf{z}^{1:N} \sim IG\left(\alpha + \frac{T}{2}, \beta + \frac{1}{2}\left(\sum_{t=1}^{T} y_t^2 - d(\mathbf{y})^T \Sigma d(\mathbf{y})\right)\right). \tag{23}
$$

The covariance matrix Σ is a diagonal matrix with entries in the form of counts and ordered identically to $d(\mathbf{y})$. The diagonal of the inverse of Σ is given by

$$
diag(\Sigma^{-1}) = \frac{1}{\tau^2} + N(\mathbf{y}). \tag{24}
$$

with $N(\mathbf{y})$ being a row vector of latent state counts for each action potential, written as

$$N(\mathbf{y}) = [N_{z^1=1}(\mathbf{y}), \ldots, N_{z^N=G}(\mathbf{y})]^T \tag{25}$$

$$N_{z^n=g}(\mathbf{y}) = \sum_{t=1}^{T} \mathbb{I}\{z_t^n = g\}. \tag{26}$$

The primary goal of model inference is to predict the latent sequences $\mathbf{z}^{1:N}$. In the Bayesian framework, posterior inference on each z_t^n given $\mathbf{y}_{1:T}$ is possible through the Monte Carlo estimate of the forward filter (Chib, 1998). For D Gibbs iterations, this probability is given by,

$$\begin{aligned} Pr(z_t^n|\mathbf{y}) &= \int p(z_t^n|y_{1:t-1}, \mathbf{\Theta}, Q^n) p(\Theta, Q^n|\mathbf{y})\, d\mathbf{\Theta}\, dQ^n \\ &= \frac{1}{D}\sum_{d=1}^{D} p(z_t^{n(d)}|y_{1:t-1}, \mathbf{\Theta}^{(d)}, Q^{n(d)}) \end{aligned} \tag{27}$$

which is simply the average over Equation (14), marginalised over all action potentials $n' \neq n$.

4. DATA ANALYSIS AND RESULTS

In this Section, we apply the model defined in Section 3.4 to both a simulated and real dataset, the latter being extracellular recordings collected during Deep Brain Stimulation, a popular treatment for patients diagnosed with advanced Parkinson's disease.

For each dataset, the Gibbs sampler presented in the previous Section was run for 10,000 iterations, discarding the first 5,000 as the burnin period. The results presented are based on the following choice of hyperparameters: $\tau^2 = 10$, $b = 0$, $\gamma = 10$, $\phi = 1$, $\alpha = 2.5$ and $\beta = 0.5s^2$, where s^2 is the sample variance of the observed data.

4.1. **Simulation study.** Data for this study was simulated assuming two distinct APs ($N = 2$) and a sampling frequency of 15kHz, leading to the specification of fifteen latent states for each AP ($G = 15$) or an AP duration of 1ms. This dataset is illustrated in Figure 5.

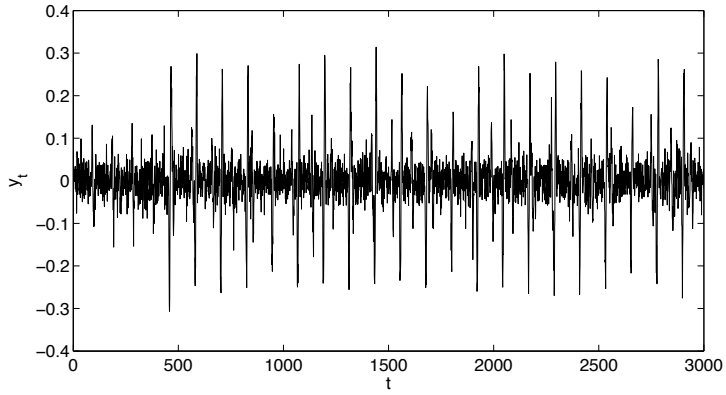

FIGURE 5. Simulated dataset with $N = 2$ and $G = 15$

Spike onset locations were set at every 95^{th} time step for the first AP, beginning at $t = 85$ and at every 122^{nd} time step for the second AP, starting at $t = 453$. This resulted in one full overlap of the two APs (each AP fires simultaneously) and two partial overlaps (one AP

fires during the refractory period of the other AP). From Figure 5, the presence of the second AP is clear, characterised by an amplitude between 0.2 and 0.3, however the locations of the first AP are less clear and the locations of overlaps even less so.

For the Gibbs sampler in this study, two chains were run with different initial templates $(\mu_1, \ldots, \mu_G)$, for each AP. In the first instance, initial templates were set to the true, simulated template for each AP. For the second chain, each μ_g^n was randomly drawn from a $N(0, \tau^2\sigma^2)$ distribution with each μ_1^n then set to zero to correspond to the assumed mean amplitude of the background noise. The comparison of inferences between these two chains did not result in any major discrepancies. Thus, the results that follows are based on MCMC output for the second chain.

The posterior probabilities of firing for isolated or non-overlapping spikes are summarised in Figure 6. Given each latent sequence was sampled backwards recursively, Figure 6 gives posterior probabilities for each AP in the form $Pr(z_t^n = G, z_t^{n'} = 1|\mathbf{y}_{1:T}, \boldsymbol{\mu}, \sigma^2, \mathbf{Q}^{1:N})$.

In Figure 6, there are the high posterior probabilities associated with the locations of each isolated AP. Assuming a probability threshold of 0.5, Figure 7 shows posterior inference performed well with recovering the true locations of each AP. However for the inferred locations of each

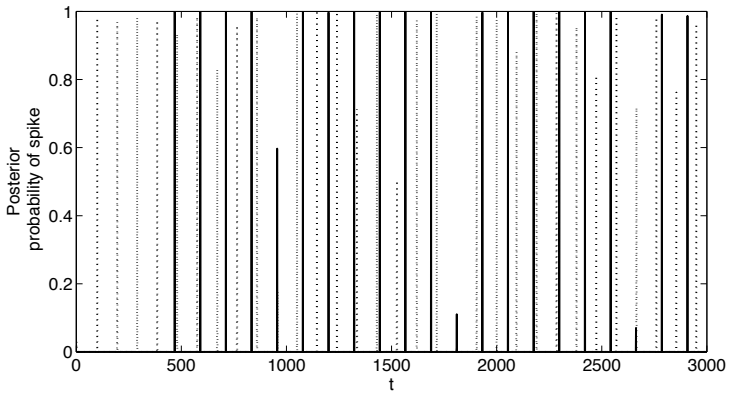

FIGURE 6. Plot of isolated firing probabilities for the simulated dataset, under the assumption of Normally distributed emissions. Neuron 1 = Dotted line, Neuron 2 = Solid line

AP, one sees some omissions when compared to the true sequences, that may be the result of overlaps.

To explore the possibility of overlapping APs being present, Figure 8 illustrates the posterior probabilities of partially overlapping APs over all locations, for each AP individually. In this case, these probabilities are of the form $Pr(z_t^n = G, z_t^{n'} \neq 1 \cap z_t^{n'} \neq G|\mathbf{y}_{1:T}, \boldsymbol{\mu}, \sigma^2, Q^{1:N})$ or, the probabilty that the n^{th} AP fires given the other AP is in the resting period.

Given the first row of plots in Figure 8, two partially overlapping APs have been recovered, one for each AP. For the first AP, we see that it is initiated during the refractory period of the second AP at

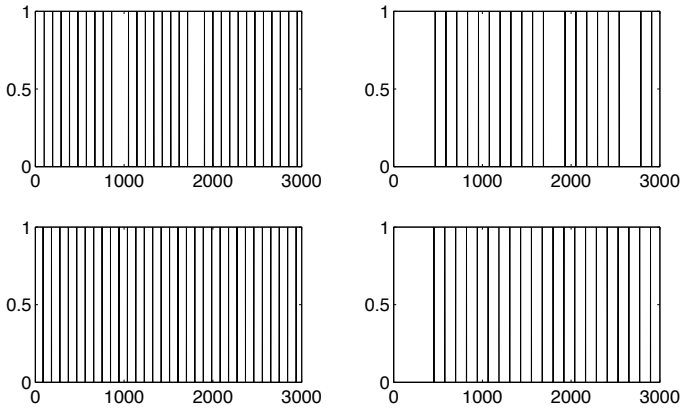

FIGURE 7. Comparison of inferred locations of spike onset (top row) versus true spikes given a posterior probability threshold of 0.5 (bottom row). Comparisons are given for each individual AP: first AP (left column) and second AP (right column). For each plot, the horizontal axis represents time, and the vertical axis represents the indicator variable for the posterior probability exceeding 0.5.

approximately $t = 1000$. Likewise, the second AP fires during the refractory period of the first AP later in the dataset between $t = 2500$ and $t = 3000$. Comparing this result with the omissions in Figure 7, one sees that these partial overlaps for two of the four omissions between the inferred and true latent sequences.

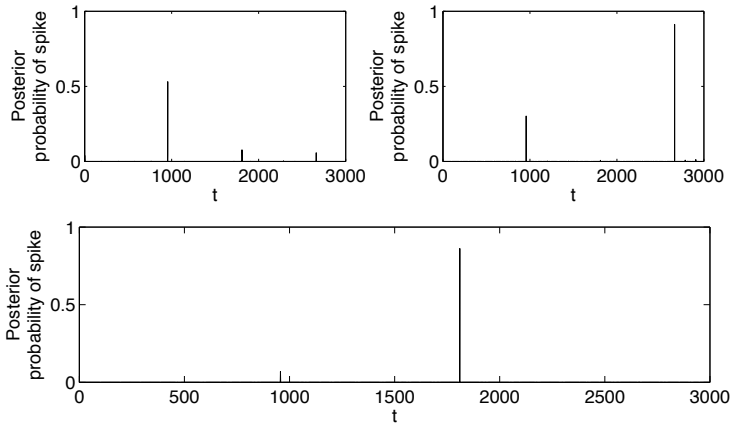

FIGURE 8. Plots of partially overlapping firing probabilities (top row) and simultaneous firing probabilities (bottom row) for the simulated dataset, under the assumption of Normally distributed emissions. Partially overlapping probabilities are given for Neuron 1 (top left) and Neuron 2 (tow right)

For the identification of fully overlapping APs, summarised in the second row of Figure 8 by the posterior probability $Pr(z_t^{1:N} = G|\mathbf{y}_{1:T}, \boldsymbol{\mu}, \sigma^2, \mathbf{Q}^{1:N})$, the final omissions in Figure 7 are recovered, corresponding to the full overlap between $t = 1500$ and $t = 2000$.

Given this favourable performance of this modelling approach, Figure 9 compares the true versus predicted templates for each AP, given by $\mu_1^n, \ldots, \mu_G^n; n = 1, 2$. Given the presence of label switching throughout the course of the Gibbs sampler, predicted templates were constructed

by choosing the estimates of $\boldsymbol{\mu}$ corresponding to the approximate *maximum a posteri* (MAP) estimate. For each Gibbs iteration, the approximate MAP estimate was calculated by multiplying the observed data likelihood with the prior density $p(\mu^{1:N}, \sigma^2)$. Given these estimates for

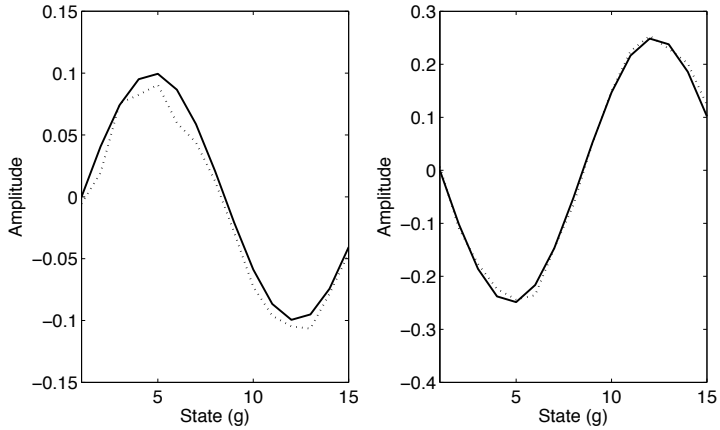

FIGURE 9. Plot of Simulated templates (solid line) superimposed on predicted templates (dashed line), based on the approximate MAP estimate for each μ. The left and right plots correspond to the first and second AP respectively.

each AP, Figure 9 further supports the favourable performance of the model, with minimal discrepancies between true and predicted templates.

4.2. **Case study: Extracellular recordings collected during Deep Brain Stimulation.** Results are now presented on an extracellular

recording taken from the STN. Given the computational burden involved in estimating HMMs, analysis is restricted to $N = 2$ and to a small section of real time recording of approximately 0.1 seconds. The dataset is given in Figure 1. Potential extensions to allow for the feasible computation of larger datasets is left to the Discussion.

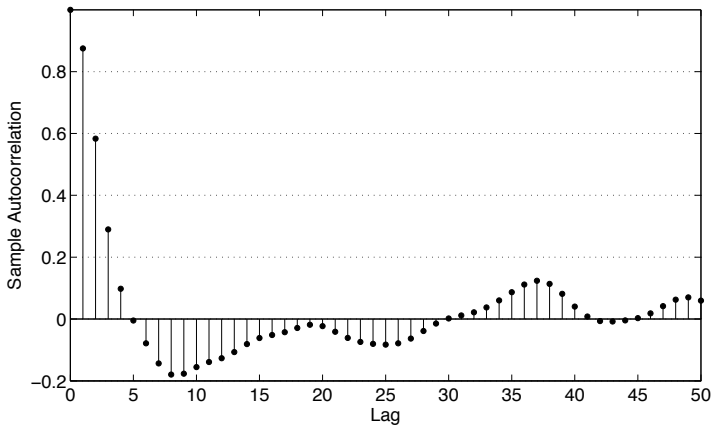

FIGURE 10. Sample autocorrelation function (ACF) up to and including lag fifty for the real dataset given in Figure 1.

To estimate the number of latent states G to model to trajectory of each AP, the sample autocorrelation function (ACF) over the entire dataset was calculated and is summarised in Figure 4.2. G was then approximated by identifying the nearest lag corresponding to the ACF crossing zero for the second time, to take into account the biphasic nature of an AP, as illustrated by the schematic in Figure 3. This

resulted in setting G equal to 30, representing a refractory period of approximately 1.3ms.

Based on the last 5000 iterations of the Gibbs sampler, Figure 11 summarises the total number of spikes detected over the course of the Gibbs sampler.

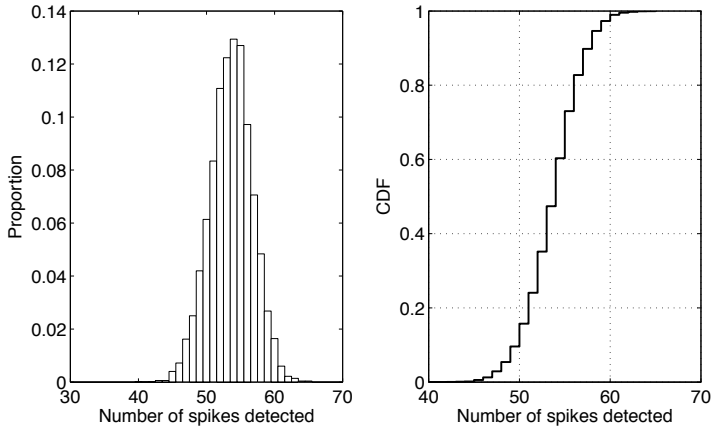

FIGURE 11. Summary of the total number of APs detected by a histogram (left) and empirical CDF (right).

For both neurons, the posterior probability of remaining in the resting state was very high, with posterior expectations both equalling approximately 0.94. This result indicated that, for both modelled APs, spiking events were rare. Reviewing the number of identified spikes in Figure 11, some variation was noted, with an expectation of 54 spikes and a 95% credible interval of between 47 and 60 spikes.

This analysis of real data, however, provided inferences there were not as conclusive as those obtained for the simulation study. In Figure 12, posterior inference for the chosen model for summarised for isolated, partially overlapping and fully overlapping APs. For isolated APs, the majority of nonzero posterior probabilities did not exceed 0.5 and therefore suggested substantial uncertainty about the presence of many APs. This result was somewhat expected, in part due to the higher level of background activity visible in Figure 1. Furthermore, investigation into this result revealed that many posterior probabilities regarding the locations of isolated APs were in fact shared between neighbouring locations. This suggested that the uncertainty in Figure 12 not so much suggested uncertainty in the presence of APs but more so the uncertainty in the exact locations where APs were initiated. A similar statement was made for possible locations of partially overlapping APs, also summarised in Figure 12. That said, in inferring the most likely locations for each AP, it was thought feasible to set the posterior probability threshold lower than 0.5. Finally, in this case, there was no evidence of fully overlapping APs, indicated by the bottom right plot in Figure 12.

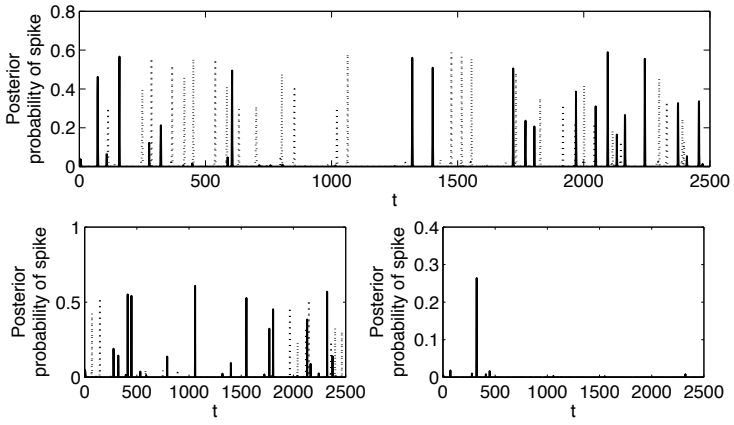

FIGURE 12. Posterior probability plots for isolated APs (top), partially overlapping APs (bottom left) and fully overlapping APs (bottom right).

Concerns about the use of this model also arose following inference on the average templates for each of the two assumed APs. This inference is summarised in Figure 13.

In Figure 13, predicted APs given posterior probabilities of onset greater than 0.4 are given and compared with their predicted template. In the chosen model's favour, many APs assigned to the second neuron (right hand plots) resemble the predicted the predicted average template well, however there appeared to be a small subset of APs with notably higher amplitude. This raised concern that this subset in fact corresponded to a third neuron and were thus incorrectly sorted. Representing a less favourable result, the predicted APs attributed to

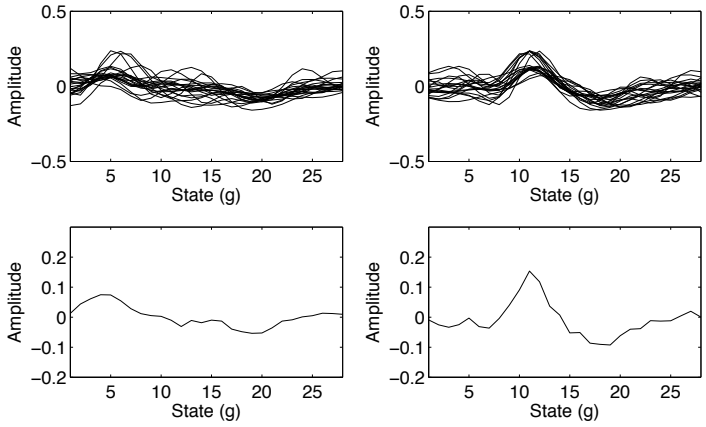

FIGURE 13. Identified APs given a probability threshold of 0.4 for the first (top left) and second AP (top right). The predicted templates based on the MAP estimate are also given for each AP (bottom left and right).

the first neuron (left hand plots) did not appear to consistently follow the predicted average template, for either threshold. This was thought to be the result on two possible reasons. Firstly, given the maximum and minimum amplitudes of the predicted template are closer to the background noise than the other template for the second AP, it was possible that the corresponding spikes were corrupted by noise to a greater degree. Secondly, and of greater concern, was that these spikes were false positives and did not correspond to true APs at all.

5. DISCUSSION

In this Chapter, we have highlighted the use of HMMs for modelling data-rich, complex stochastic processes. Furthermore, taking the simplest definition of a HMM, we have shown its flexibility in adapting to different modelling situations, in this case the extension of a single sequence HMM to a factorial setting, in an attempt to simultaneously identify and sort spikes arising from multiple APs.

In the simulation study, the model provided promising results and was able to correctly identify both APs and locations where they overlapped. However, in the analysis of the real dataset, a number of key concerns arose, presenting opportunities for future research in this area and form the focus of this discussion. Broadly speaking, these concerns could be categorised as either computational or methodological.

An analysis not considered in this Chapter is that of model comparison. This comparison could take place for models with different values of N or, as explored in White (2011), models with different distributional assumptions for y_t and/or different transition matrix specifications. Analyses of this form could be conducted with the use of goodness-of-fit criteria, such as the Deviance Informatin Criterion (Spiegelhalter et al., 2002; Celeux et al., 2006), as is used in White (2011), or by Bayes' factors (Kass and Raftery, 1995).

For many segments of real data analysed but not presented in this Chapter, the specification of G also led to problems, in terms of the number of APs detected and the posterior probabilities of spikes occurring at a given location. Furthermore, for the real data analysis in Section 4.2, a number of different values for G were trialled. Although these additional analyses are not included in this Chapter, it was seen that smaller values of G than the one chosen resulted in many false positives, compared to larger values of G that resulted in the detection of very few or no spikes in the observed data. The most likely reason for sensitivity regarding the choice of G is that, in recordings of multiple cells, the refractory period may vary considerably. It was this reason that motivated the alternative choice of transition matrix however, in both analyses presented, this extension did not appear to provide an improved fit to the data. For the real dataset, evidence for the presence of two different refractory periods could be seen in the sample ACF, with it almost crossing zero for the second time at lag twenty. To implement different refractory periods in the current model would involve the specification of identifiability constraints and would be suitable extension for future work.

To conclude, while implementation of the proposed model in the Bayesian framework was appealing, in terms of the range of probability statements possible, the cost of model computation by MCMC was found to be a prohibitive factor, with computation taking approximately 2hours per 0.1seconds of real recording. That said, less expensive alternative to MCMC would be desirable for future work in this area. Examples of viable alternatives to MCMC include Variational Bayes (McGrory and Titterington, 2009; Ghahramani and Hinton, 2000; McGrory et al., 2011), Approximate Bayesian Computation (Jasra et al., 2010) and Sequential Monte Carlo (Doucet et al., 2000).

References

Baum, L., Petrie, T., Soules, G., Weiss, N., 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. The Annals of Mathematical Statistics 41, 164–171.

Cappé, O., Moulines, E., Rydén, T., 2005. Inference in hidden Markov models. Springer Verlag.

Celeux, G., Forbes, F., Robert, C., Titterington, M., 2006. Deviance information criteria for missing data models. Bayesian Analysis 1 (4), 651–674.

Chib, S., 1996. Calculating posterior distributions and modal estimates in Markov mixture models. Journal of Econometrics 75 (1), 79–97.

Chib, S., 1998. Estimation and comparison of multiple change-point models. Journal of Econometrics 86 (2), 221–241.

Doucet, A., Godsill, S., Andrieu, C., 2000. On sequential Monte Carlo sampling methods for Bayesian filtering. Statistics and computing 10, 197–208.

Eddy, S., 1995. Multiple alignment using hidden Markov models. Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology 3, 114–120.

Forney Jr, G., 1973. The viterbi algorithm. Proceedings of the IEEE 61 (3), 268–278.

Freeman, J., 1971. A simple multi-unit channel spike height discriminator. Journal of Applied Physiology 31, 939–941.

Frühwirth-Schnatter, S., 2006. Finite mixture and Markov switching models. Springer Verlag.

Gales, M., Young, S., 2008. The application of hidden Markov models in speech recognition. Foundations and Trends in Signal Processing 1 (3), 195–304.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayeisan restoration of images. IEEE Transactions on

Pattern Analysis and Machine Intelligence 6, 721–741.

Ghahramani, Z., Hinton, G., 2000. Variational learning for switching state-space models. Neural computation 12 (4), 831–864.

Ghahramani, Z., Jordan, M., 1997. Factorial hidden Markov models. Machine learning 29 (2), 245–273.

Herbst, J., Gammeter, S., Ferrero, D., Hahnloser, R., 2008. Spike sorting with hidden Markov models. Journal of neuroscience methods 174 (1), 126–134.

Holmes, I., Bruno, W., 2001. Evolutionary HMMs: a Bayesian approach to multiple alignment. Bioinformatics 17 (9), 803.

Jasra, A., Singh, S., Martin, J., McCoy, E., 2010. Filtering via approximate Bayesian computation. Statistics and ComputingTo appear.

Kass, R., Raftery, A., 1995. Bayes factors. Journal of the American Statistical Association 90 (430), 773–795.

Kim, K., Kim, S., 2000. Neural spike sorting under nearly 0-dB signal-to-noise ratio using nonlinear energy operator and artificial neural-network classifier. IEEE Transactions on Biomedical Engineering 47, 1406–1411.

Kleiner-Fisman, G., Fisman, D., Sime, E., Saint-Cyr, J., Lozano, A., Lang, A., 2003. Long-term follow up of bilateral deep brain stimulation of the subthalamic nucleus in patients with advanced Parkinson's disease. Journal of neurosurgery 99, 489–495.

Krack, P., Batir, A., Van Blercom, N., Chabardes, S., et al, 2003. Five-year follow-up of bilateral stimulation of the subthalamic nucleus in advanced Parkinson's disease. New England Journal of Medicine 349, 1925–1934.

Kumar, R., Lozano, A., Kim, Y., Hutchison, W., Sime, E., Halket, E., Lang, A., 1998. Double-blind evaluation of Subthalamic Nucleus Deep Brain Stimulation in advanced Parkinson's disease. Neurology 51, 850–855.

Limousin, P., Krack, P., Pollak, P., Benazzouz, A., Ardovin, C., Hoffman, D., Benabid, A., 1998. Electrical stimulation of the Subthalamic Nucleus in advanced Parkinson's disease. New England Journal of Medicine 339, 1105–1111.

MacDonald, I. L., Zucchini, W., 1997. Hidden Markov and Other Models for Discrete-Valued Time Series. Chapman and Hall: New York.

McGrory, C., Titterington, D., 2009. Variational Bayesian analysis for hidden Markov models. Australian & New Zealand Journal of Statistics 51 (2), 227–244.

McGrory, C., White, N., Mengersen, K., Pettitt, A., 2011. A Variational Bayes approach to fitting Hidden Markov Models in Parkinsons Disease Research, submitted to Biometrics, October 2011.

Mtetwa, N., Smith, L., 2006. Smoothing and thresholding in neuronal spike detection. Neurocomputing 69, 1366–1370.

Nenadic, Z., Burdick, J., 2005. Spike detection using the continuous wavelet transform. IEEE Transactions on Biomedical Engineering 52, 74–87.

Rabiner, L., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77 (2), 257–286.

Rizk, M., Wolf, P., 2009. Optimizing the automatic selection of spike detection thresholds using a multiple of the noise level. Med Biol End Computation 47, 955–966.

Smith, A., Roberts, G., 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. Journal of the Royal Statistical Society. Series B (Methodological) 55 (1), 3–23.

Spiegelhalter, D., Best, N., Carlin, B., Van der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 64 (4), 583–639.

Thakur, P., Lu, H., Hsiao, S., Johnson, K., 2007. Automated optimal detection and classification of neural action potentials in extracellular recordings. Journal of Neuroscience Methods 162, 364–376.

Tranter, S., Reynolds, D., 2006. An overview of automatic speaker diarization systems. Audio, Speech, and Language Processing, IEEE Transactions on 14 (5), 1557–1565.

White, N., 2011. Bayesian mixtures for modelling complex medical data: A case study in Parkinson's disease. Ph.D. thesis, Queensland University of Technology.

Yang, X., Shamma, S., 1988. A totally automated system for the detection and classification of neural spikes. IEEE Transactions on Biomedical Engineering 35, 806–816.