AN OVERVIEW OF MODELING TECHNIQUES FOR HYBRID BRAIN DATA

Apratim Guha[1,2] and Atanu Biswas[3]

[1]*National University of Singapore,* [2]*University of Birmingham*
and [3]*Indian Statistical Institute*

Abstract: Constructing models for neuroscience data is a challenging task, more so when the data sets are of hybrid nature, and there exists very little work. The models have to be physiologically meaningful, as well as statistically justifiable. Here we introduce various techniques for fitting a model to bivariate hybrid time series data from the field of neuroscience. As an example, we use a data set on the local field potentials (which is a continuous time series) and nerve cell firings (which is a point process) of anesthetized mice. We extend various available methodologies for modeling nerve cell spike trains to the hybrid set-up, and present a model that has not been previously explored in neuroscience literature. We illustrate the fit of the data set by some Markov chain-based models, some models with crossed dependence including an Inhomogeneous Markov interval type (IMI) model, some ARMA-type models, and also some state space models. We compare the proposals with two existing models. We aim to provide an overview of various possible modeling strategies, and to provide a comparison of the fit and estimation of different models in terms of various standard model selection criteria like AIC and BIC. A detailed simulation study is performed to assess the performance of different models.

Key words and phrases: AIC, BIC, bivariate time series, continuous time series, IMI, jump Markov linear Gaussian system, hybrid process, Markov model, maximum log-likelihood, point process.

1. Introduction

A multivariate process, where one component is a continuous time series and the other is a point process (or a discrete valued time series) is referred to as a hybrid process. Although not many detailed studies of hybrid models exist, examples of hybrid situations in neuroscience are numerous. One of the examples that Willie (1979) discusses is the relation between transmembrane current (a time series) and the firing of voltage impulses (a point process) by a nerve cell. In a more recent work, Brown, Frank, Tang, Quirk and Wilson (1998) modeled spike trains jointly with the location of a rat, treating the location of the rat as a bivariate Gaussian ranodm vector. See Valentine and Eggermont (2001) and

Noreña and Eggermont (2002) for some other examples in the neuroscience field that deal simultaneously with spike trains and local field potentials.

In this paper, we aim to study various models that capture the inter-relationship between spike trains emanating from neurons in the temporal cortex of mice, which is a part of the auditory cortex, and local field potentials in the vicinity. The need is for a model that justifiably combines model for a spike train with for the continuous local field potential values. Toward that, we extend some of the existing methods for modeling spike train data to hybrid models. The aim of this work is not to suggest one "best" method, but to provide an overview of the tools available, and to suggest directions to possible reasonable models.

There are numerous competing methods available for modeling univariate data, both continuous and discrete, in neuroscience literature. We now provide a brief overview in this direction. The most basic approach to modeling local field potential data recorded over time, denoted $\{X(t)\}$ henceforth, is to employ the linear model

$$X(t) = \mu + \int_l^u a(s)X(t-s)ds + \epsilon(t),$$

where μ is some constant and $a(\cdot)$ is a suitable filtration, (l, u) represent the past on which the present process is dependent, and $\epsilon(\cdot)$ is some error process. When the process is observed over discrete time, the above model reduces to

$$X(t) = \mu + \sum_{s=l}^u a(s)X(t-s) + \epsilon(t). \tag{1.1}$$

This model is known as variable signal plus ongoing activity (VSPOA) model. References are numerous, see Chen et al. (2006) and the references therein. Other important approaches include employing wavelet techniques, for which some recent references are Roux et al. (2007) and Vialatte et al. (2007). For a neural network model, see Quenet and Horn (2003).

A classical approach to modeling a spike train is to employ a threshold on a simultaneously observed action potential process, if present (Brillinger and Segundo (1979) and Brillinger (1988)). As a variation of this model, the action potential of the nerve cell is modeled as a function of the firing times and then a threshold is used (Brillinger (1992) and Brillinger, Bryant and Segundo (1976)).

The spike train data are more usually modeled as counting processes, although some recent works explore wavelet techniques (Laubach (2004)). The simplest case of a counting process is when the "rate", or the conditional intensity, of the process depends only on time t. Representing the count of spikes at time t by $N(t)$, the above is equivalent to saying that $\lambda(t) = \lim_{\delta \to 0} E(N(t+\delta) - N(t)|$history till time t), the conditional intensity function of the process,

depends only on t, see Ventura et al. (2001). The peristimulus time histogram (PSTH) is often used as an estimate of the time-varying firing rate (Kass and Ventura (2001)). When large numbers of neurons are considered, the spike times for the collection of neurons can be modeled as a Poisson process. Sometimes, as an alternative, nonhomogeneous Poisson models are employed for interspike intervals (ISI), see Barbieri et al. (2001) and Kass and Ventura (2001). However, when neuron firing times individual trials are of interest, the joint distribution of the spikes cannot be obtained from the Poisson models. Such models are usually inadequate for these situations because they do not account for the refractory period, a brief period during which the neurons can not fire after a spike-burst. This requires a more complex model than the memoryless Poisson. Kiang et al. (1965) modeled the spike train process as a renewal process to account for the dependence on past. A simple Markovian model to achieve this has

$$\lambda(t) = f(t, t - w_N(t)), \qquad (1.2)$$

where $\lambda(t)$ is the conditional intensity function, f is a nonnegative function, $w_j(t)$ is the time of the time of the jth spike burst before time t, and N is the number of spikes till time t. See Chan and Loh (2007) for examples of various methods of estimating f. An important subclass of the above model has the multiplicative form (Johnson and Swami (1983), Berry, Warland and Meister (1997) and Kass and Ventura (2001))

$$\lambda(t) = s(t)r(t - w_N(t)), \qquad (1.3)$$

where $s(t)$ is the firing function and $r(\cdot)$ is a recovery function. One could take $r(t - w_N(t))$ to be zero for values of $t - w_N(t)$ less than some value, assuming the existence of a constant refractory period (Johnson and Swami (1983) and Berry, Warland and Meister (1997)). More complicated models include a weighted fit (Berry, Warland and Meister (1997)) and a spline fit (Kass and Ventura (2001)). In this article we limit our discussion to the constant refractory period, but in most cases extensions are immediate, at least theoretically.

Sometimes the neural spike trains are assumed to be governed by a stimulus. The stimulus could be explicit (Wilson and McNaughton (1993) and Berry, Warland and Meister (1997)) or implicit (Riehle et al. (1997) and McEchron, Weible and Disterhoft (2001)). For explicit stimuli, a parametric non-homogeneous Poisson model fit has been used by Brown, Frank, Tang, Quirk and Wilson (1998). For implicit stimuli, one could fit a state space model, as in Smith and Brown (2003). The state space models for spike bursts usually assume that the conditional intensity function depends on some hidden latent process, that can be continuous (Smith et al. (2003)) or discrete (Schneidman et al.

(2006)). One possible alternative could be to employ the marked point process methodology. This method is hugely popular in the field of seismology for earthquake modeling (Vere-Jones (1995) and Schoenberg (2004)), but applications to the field of neuroscience are not available to our knowledge. Other popular methods include a mixture of Poisson (MP) model (Wiener and Richmond (2003) and Nakahara, Amari and Richmond (2006)). However, the MP model does not work well for spike train generation without additional assumptions. For a recent and more comprehensive review of methods in modeling firing time data, see Kass, Ventura and Brown (2005).

Most of the existing works on modeling hybrid data assume some type of causal structure, some of which are found in Sections 3 and 4. A notable exception is Jørgensen et al. (1996); it explores the application of state-space modeling techniques to non-stationary hybrid series. Other notable references for hybrid models without conditional structures include Andrieu, Davy and Doucet (2003) and Davis and Ensor (2007); the former uses a particle filtering method to perform optimal estimation in jump Markov systems, and the later utilizes a logistic smooth transition regression model.

Unlike the relationship of spike trains and action potentials, where the former depends on the later ((Brillinger (1992) and Brillinger, Bryant and Segundo (1976)), if there does not exist any scientific justification of any causal relationship between spike trains and local field potentials, and hence any model with a causal structure, whereas simpler to fit, might not explain the data well. In such a situation, a state-space model would be expected to perform much better than fitted models like the two above. However, as one aim of this paper is to give an overview of various modeling techniques for hybrid processes, we also explore some causal models with the hope that such models may serve as a useful reference for future hybrid modeling exercises.

We fit the following models to data: (a) some Markov models for spikes with LFPs conditionally normal; (b) some inhomogeneous Markov interval (IMI) models for the spike train with an autoregressive (AR) model for the LFP; (c) some mixture of a discrete AR and a continuous process; and (d) state space models, fitted using hidden Markov model fitting techniques. Comparison of the proposed models along with two existing models is based on maximum log-likelihood, AIC and BIC.

The rest of the paper is organized as follows. In Section 2 we provide some pertinent background material about the data. Section 3 discusses some Markov models for such hybrid processes. Some models with crossed dependence, including some IMI-type models and some mixture ARMA-type models, are discussed in Section 4. Section 5 provides some hidden Markov model for hybrid data. Section 6 discusses the results of some simulations and compares different models in terms of AIC and BIC and Section 7 concludes.

2. Data Description

The area of the auditory cortex in a mouse was demonstrated by Stiebler et al. (1997) based on neuronal response characteristics. They concluded that the auditory cortex of the mouse is situated in the caudal (toward the back of the head) half of the parietal cortex area of the mouse brain. The local field potential (LFP) is supposedly related to the so-called 'synaptic activity' of neurons, i.e., the compound electric field generated by the post-synaptic potentials. It is generally assumed that there exists a threshold of the 'action potential' of these synaptic activities, where the spikes are related exclusively to supra-threshold activity, whereas sub-threshold synaptic potentials provide a contribution to LFP. So, one can think of the LFP as the total spiking activity of all units (neurons) within a certain region. For a more detailed description, see Kandel, Schwartz and Jessell (2000).

The data set we analyze consists of local field potential recordings (sampled continuous time series) and nerve cell firings (point processes) recorded by four electrodes located in the temporal cortex of anesthetized mice. Some recording details for the experiments are provided in Villa et al. (1998). The original data set consists of spike trains and local field potentials (LFP), collected simultaneously by electrode placed in the auditory cortex of an anesthetized mouse. The experiment is a succession of 18 sub-experiments, each of which lasts for approximately 100 seconds, and which are separated by around 20-30 seconds. The first nine of these are associated with spontaneous firing of the auditory neurons, whereas the last nine correspond to the application of an auditory stimuli. We consider only the spontaneous recordings here. The sampling rate of the LFP is 500 samples/sec (Hz). The firing times are recorded to the nearest millisecond (ms).

3. Fits with Markov Models for Spikes

Basic models for hybrid time series are constructed assuming causal relationships, i.e., assuming that one of the two series drives the others. In this section, we explore two types of such models. A usual strategy divides the observation period into small intervals of equal width so that there is at most one spike per interval (Smith and Brown (2003) and Kass and Ventura (2001)). For this purpose, and also to utilize the LFP observations fully, we observe the increment of the spike process every 2 ms, which we denote as $Y(t) = N(t) - N(t - 1)$. Clearly, $Y(t)$ is a binary series.

We begin with a causal structure where the 0-1 valued spike series, denoted by $\{Y(t)\}$, follows a Markov chain and the continuous valued LFP series, which we denote by $\{X(t)\}$, has a distribution dependent on the present and past values
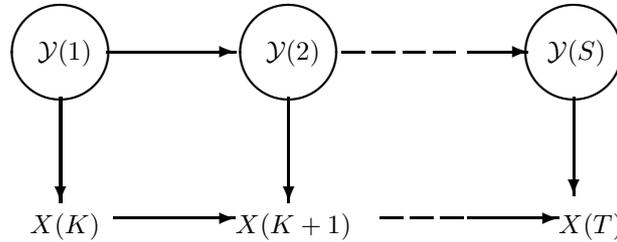
Figure 1. The Markov model, where each $\mathcal{Y}(j) = (Y(j), Y(j+1), \cdots, Y(j+K-1))$ is a vector of size K and discrete; the $\mathcal{Y}$'s form a stationary Markov chain and the X's, the continuous variables represented by (3.2), are conditionally normal given the $\mathcal{Y}$'s.

of itself and also of Y. One example of such a model is the jump Markov linear system (JMLS), where the parameters evolve with time according to a finite state Markov chain, see Krishnamurthy and Evans (1998) and Doucet and Andrieu (2001). One popular approach is to assume the system to be Gaussian and the process to be dependent on its immediate past through a linear system equation

$$X(t) = a_{Y(t)}X(t-1) + b_{Y(t)}\epsilon(t), \tag{3.1}$$

where the $Y(t)$'s form a finite state Markov chain, and $\epsilon(t) \sim N(0,1)$. This system is known as the jump Markov linear Gaussian system (JMLGS) of order 1. It is often of practical interest to relax the assumptions of linearity and Gaussian distribution, see Andrieu, Davy and Doucet (2003). In the following section, we discuss some simple extensions of the model (3.1) to neuroscience data.

3.1. A Markov model

Assume that the sets of 0's and 1's occur in blocks, say blocks of size K for some positive integer K, and the binary process has a lag K dependence structure, i.e., the distribution of the 0-1 valued variable at a time instant is assumed to be dependent on the values of previous K time instants, subject to some restrictions due to the presence of the refractory period. The continuous variable at time t is conditionally normal given the values of the 0-1 valued variable at the last K time instants. The model is described by a generalization of (1.1) (and also of (3.1)) as

$$X(t) = \mu_j + \sum_l^u a_j(s)X(t-s) + \sigma_j\epsilon(t), \tag{3.2}$$

for suitable choices of l and u, where j is a vector of the values of Y for the previous K time instants, and $\epsilon(t) \sim N(0,1)$. We adopt the convention that

$a_j(0) = 0$, so that for $K = 1$, $l = u = 0$, we get the usual Markov chain model for the discrete series, with continuous variables being conditionally normal given the value of the discrete variable at a time instant; and for $K = 1$, $l = 0$ and $u = 1$, we get the JMLGS described in (3.1). Also note that when μ_j, $\alpha_j(\cdot)$ and σ_j^2 are independent of j, the two series are independent.

Figure 1 describes the model. Write $\mathcal{Y}(j) = (Y(j), Y(j+1), \ldots, Y(j+K-1))$ for $j = 1, 2, \ldots, S = T - K + 1$, and assume that the $\mathcal{Y}(j)$'s form a Markov chain. This Markov chain has only finitely many possible different states. The maximum possible number is 2^K, but the actual number may be less. Some of the states of the Markov chain would be ruled out due to the presence of the refractory period; e.g., if $K = 2$ and the refractory period is two ms (assumed so that the minimum gap between two consecutive spikes is 4 ms), then the possible states are $00, 01, 10$ and, correspondingly, there are three different states. Also note that certain transition probabilities, for example $p(01 \to 01)$ would have to be 1 in this model. Hence $P_{10} = P_{01}$, where the P_j's describe the intial distribution of the states. These quantities are estimated by the estimated stationary distribution of the states. Let us denote the number of states by L. The $X(j + k - 1)$'s are conditionally normal given $\mathcal{Y}(j)$, while the first $(K - 1)$ X values are ignored in this fit.

There could be various added complexities: $X(\cdot)$ and $Y(\cdot)$ could be multivariate; the dependence structure of Y could be more complex, for example a popular choice would be to choose a variable K, say $K(t) = t - w_N(t)$, where $w_N(t)$ is as described in Section 1; or the refractory period could be random. We now briefly discuss a model with multivariate $X(\cdot)$ and $Y(\cdot)$, and leave the other situations to the interested reader.

3.2. A Multivariate Markov model

We now describe an alternative Markov model where the states are assumed multivariate, with spikes over K consecutive time instants forming a Markov chain: $\mathbf{Y}(j) = (Y((j-1)K+1), Y((j-1)K+2), \ldots, Y(jK))$ for $j = 1, 2, \ldots, M = [T/K]$, and $\mathbf{X}(j) = (X((j-1)K+1), X((j-1)K+2), \ldots, X(jK))$'s, for $j = 1, 2, \ldots, M$, being conditionally normally distributed given the spike states, which corresponds to $u = 0$ in (3.2). See Figure 2 for a pictorial representation of the model.

The likelihood can be written as

$$P(\mathbf{X}, \mathbf{Y}) = P(\mathbf{Y}(1)) \prod_{t=2}^{M} P(\mathbf{Y}(t)|\mathbf{Y}(t-1)) \prod_{t=1}^{M} P(\mathbf{X}(t)|\mathbf{Y}(t)),$$
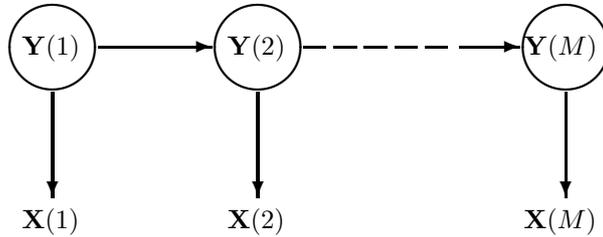
Figure 2. The Second Type of Conditional Model. $\mathbf{Y} = (Y((j-1)K+1), Y((j-1)K+2), \ldots, Y_{(jK)})$ form a Markov chain and $\mathbf{X}(j) = (X((j-1)K+1), X((j-1)K+2), \ldots, X(jK))$ are conditionally normal.

and the maximum likelihood estimators can be obtained in the standard way.

3.3. Discussion of Results for the Simple Markov Model

Here we present the fitted models for $K = 2$ and 3 corresponding to $l = 0$, and $u = 0, 1, 2$. Fits for higher values of K are also investigated but are not found to improve the model by much, and hence not presented. The other values of l and u could also be tried if a scientific justification could be established, but here we discuss only some of the simplest cases. Table 1 summarizes the goodness of fit of these models in terms of maximum log likelihood, AIC, and BIC. The best fits are denoted in bold. The fitted parameters for the models are provided in Tables 2 and 3. In Table 2, $\mathcal{Y}_K(t) = (Y(t), Y(t+1), \ldots, Y(t+K-1))$ for a fixed K, so that for $K = 2$, $\mathcal{Y}_K$ takes values $(0,0), (0,1)$ and $(1,0)$. The notation is simplified by writing, for example, "00" for $(0,0)$, and dropping the subscript "K" when there is no scope of confusion. Table 3 provides the fit statistics. Note that the transition probabilities from states 01, 001, 101 are not reported because they lead to the state 0 with probability 1. The parameter estimates corresponding to the state 101 were not obtained because there was only one instance of it. It may be seen from the fit that immediately after a spike firing, the LFP values tend to be higher on the average for states 10 and 010 than for the rest, which points to some causal relationship. We also observe that, on the basis of BIC, there is little to choose between $K = 2$ and $K = 3$, especially for the case $u = 2$. This indicates that the LFP values are perhaps not too affected by the spike firings except for the immediate past.

When the LFP and the spike trains are independent, under the above model, one has

$$X(t) = \mu + \sum_{l}^{u} a(s)X(t-s) + \epsilon(t),$$

which is similar to (3.2) except for the lack of dependence on spike-states. For comparison, goodness of fit statistics for $K = 3$ and $u = 2$, the best model under

Table 1. Comparison of various goodness of fit statistics for the fitted Markov models. The best fits are shown in bold. K denotes the degree of dependence of the spike train on the past.

Markov model for spikes with univariate conditional LFPs						
	-2 Likelihd	AIC	BIC	-2 Likelihd	AIC	BIC
	$K = 2$			$K = 3$		
$u = 0$	587036.0	587062.0	587176.5	586998.6	587036	587203.5
$u = 1$	562337.9	562365.9	562489.3	562277.3	562317.3	562493.6
$u = 2$	557479.2	557509.2	557641.4	**557413.6**	**557467.6**	**557705.6**
When the two series are independent						
$u = 2$	557563.5	557579.5	557650	557547.7	557569.7	557666.6
Markov model for spikes with bivariate conditional LFPs						
	-2 Likelihd		AIC		BIC	
Full model	574269.4		574321.4		574550.6	
Same Var., varying mean	574320.6		574356.6		574515.3	
Ind. case	574329.2		574357.2		574480.6	

independence is provided in Table 1. The dependence model does perform better, although the difference is not very significant, especially in terms of BIC. Hence, this model perhaps does not improve much from the independence model.

3.4. Discussion of results for the multivariate Markov model

We present only the $K = 2$ case for simplicity, the models for higher values of K would be similar. We consider three different settings for the continuous process: (a) having varying mean and variance depending on the states of the spike train; (b) having the mean dependent but the variance independent of the spikes; and (c) fully independent of the spikes. The fit of the model parameters are presented in Table 4, and the goodness of fit statistics are given in Table 1. This model performs better than the $u = 0$ case for $K = 2$, as well as $K = 3$ for the univariate model, but is worse than the rest.

4. Models with Crossed Dependence

In the previous section, we fitted a conditional model which assumes that the spike train drives the LFP data. Models fitted assuming the opposite conditional direction are also common in neuroscience literature, where the neural spike trains are assumed to be governed by a stimulus. The stimulus could be explicit (Wilson and McNaughton (1993) and Berry, Warland and Meister (1997)) or implicit (Riehle et al. (1997) and McEchron, Weible and Disterhoft (2001)).

Table 2. Fitted parameters for various values of u for the Markov model fitted in Section 3 for $K = 2$.

Fitted parameters				
Stationary Distribution: $P_{00} = 0.992$, $P_{01} = 0.004$, $P_{10} = 0.004$				
$\mathcal{Y}(t-1)$	Transition Probability $P(Y(t)\mid \mathcal{Y}(t-1))$			
	$Y(t) = 0$		$Y(t) = 1$	
00	0.996		0.004	
10	0.995		0.005	
LFP model Parameters	Constant Terms	Lag 1 Coefficients	Lag 2 Coefficients	Error Variances
$u = 0$	$\mu_{00} = -9.92,$ $\mu_{10} = -11.24,$ $\mu_{01} = -24.46$	-	-	$\sigma^2_{00} = 7457.2,$ $\sigma^2_{10} = 12139.6,$ $\sigma^2_{01} = 10279.9$
$u = 1$	$\mu_{00} = -3.74,$ $\mu_{10} = 3.17,$ $\mu_{01} = -10.61$	$a_{00}(1) = 0.63,$ $a_{10}(1) = 0.59,$ $a_{01}(1) = 0.50$	-	$\sigma^2_{00} = 4529.3,$ $\sigma^2_{10} = 8615.6,$ $\sigma^2_{01} = 7828.7$
$u = 2$	$\mu_{00} = -4.88,$ $\mu_{10} = -1.57,$ $\mu_{01} = -9.44$	$a_{00}(1) = 0.82,$ $a_{10}(1) = 0.73,$ $a_{01}(1) = 0.74$	$a_{00}(2) = -0.30,$ $a_{10}(2) = -0.30,$ $a_{01}(2) = -0.37$	$\sigma^2_{00} = 4110.1,$ $\sigma^2_{10} = 7978.1,$ $\sigma^2_{01} = 6740.4$

For explicit stimuli, a parametric non-homogeneous Poisson model fit has been used by Brown, Frank, Tang, Quirk and Wilson (1998). For implicit stimuli, one could fit a state space model, as in Smith and Brown (2003).

In this section, we study two conditional models where the LFP is considered as an stimulus for the spike train. The first model is a modification of Smith and Brown (2003); the second model is a mixture of a discrete ARMA type process and an AR model.

4.1. Inhomogeneous Markov Interval Models

A drawback of the model fitted by Smith and Brown (2003) is that it does not consider the refractory period of the neural process. Here, we apply a modification as in Johnson and Swami (1983), where the conditional intensity process is defined by (1.2). Hence, the conditional intensity function of this process depends only on the last spike. For this Markov property, the processes are sometimes referred to as inhomogeneous Markov interval (IMI) process (Kass and Ventura (2001)). Note that in (1.2), when the conditional intensity does not depend on the second argument, the process becomes an inhomogeneous Poisson process. An important subclass of the above model is the multiplicative form, defined previously by (1.3). Smith and Brown (2003) suggest the firing function

$$s(t) = \exp(\alpha_0 + \alpha_1 X(t)), \tag{4.1}$$

Table 3. Fitted parameters for various values of u for the Markov model fitted in section 3 for $K = 3$.

Fitted parameters				
Stationary Distribution: $P_{000} = 0.9889, P_{100} = 0.0037, P_{010} = 0.0037, P_{001} = 0.0037, P_{101} = 0.00002$				
$\mathcal{Y}(t-1)$	Transition Probability $P(Y(t)|\mathcal{Y}(t-1))$			
	$Y(t) = 0$		$Y(t) = 1$	
000	0.996		0.004	
100	0.995		0.005	
010	0.995		0.005	
LFP model Parameters	Constant Terms	Lag 1 Coefficients	Lag 2 Coefficients	Error Variances
$u = 0$	$\mu_{000} = -10.02,$ $\mu_{100} = 17.78,$ $\mu_{010} = -11.24,$ $\mu_{001} = -25.50$	-	-	$\sigma^2_{000} = 7446.0,$ $\sigma^2_{100} = 9765.4,$ $\sigma^2_{010} = 12139.4,$ $\sigma^2_{001} = 10138.9$
$u = 1$	$\mu_{000} = -3.85,$ $\mu_{100} = 23.51,$ $\mu_{010} = 3.17,$ $\mu_{001} = -11.63$	$a_{000}(1) = 0.63,$ $a_{100}(1) = 0.52,$ $a_{010}(1) = 0.59,$ $a_{001}(1) = 0.50$	-	$\sigma^2_{000} = 4519.9,$ $\sigma^2_{100} = 6448.1,$ $\sigma^2_{010} = 8613.7,$ $\sigma^2_{001} = 7630.0$
$u = 2$	$\mu_{000} = -4.97,$ $\mu_{100} = 19.04,$ $\mu_{010} = -1.57,$ $\mu_{001} = -10.36$	$a_{000}(1) = 0.82,$ $a_{100}(1) = 0.64,$ $a_{010}(1) = 0.73,$ $a_{001}(1) = 0.74$	$a_{000}(2) = -0.31,$ $a_{100}(2) = -0.24,$ $a_{010}(2) = -0.30,$ $a_{001}(2) = -0.37$	$\sigma^2_{000} = 4100.0,$ $\sigma^2_{100} = 6056.0,$ $\sigma^2_{010} = 7978.1,$ $\sigma^2_{001} = 6593.1$

using the same notation as in Section 3. As we deal with discrete time, $s(t)$ could be interpreted as the firing probability at time t; hence, to ensure that $0 \leq s(t) \leq 1$, we use the logit model

$$\log\left(\frac{s(t)}{1 - s(t)}\right) = \alpha_0 + \alpha_1 X(t) \tag{4.2}$$

instead of (4.1). Note that if the two series are independent, the firing function is a constant.

Now, as in Johnson and Swami (1983) and Berry, Warland and Meister (1997), assume that the recovery function is

$$r(t) = \begin{cases} 1 \text{ if } 0 \leq t \leq 2 \text{ ms}, \\ 0 \text{ otherwise}. \end{cases}$$

In the present set-up, this translates to $r(t) = 0$ for $t \geq 1$. More complex firing and recovery functions could also be accommodated easily, for example Kass and Ventura (2001) employ a spline model.

APRATIM GUHA AND ATANU BISWAS

Table 4. The fit of the multivariate Markov model for various situations.

Fitted parameters			
Stationary Distribution:$P_{00} = 0.9926, P_{01} = 0.0018, P_{10} = 0.0056$			
$\mathbf{Y}(t-1)$	$\mathbf{Y}(t) = 00$	$\mathbf{Y}(t) = 01$	$\mathbf{Y}(t) = 10$
	Transition Probability $P(\mathbf{Y}(t)|\mathbf{Y}(t-1))$		
00	0.9926	0.0018	0.0056
10	0.9928	0.0072	0
01	1	0	0
Both mean and variance of the LFP vary over different spike states			
Mean	$\begin{pmatrix} -9.60 \\ -10.11 \end{pmatrix}$	$\begin{pmatrix} -28.50 \\ -33.26 \end{pmatrix}$	$\begin{pmatrix} -26.01 \\ -23.12 \end{pmatrix}$
Variance	$\begin{pmatrix} 7468.94 & 4732.60 \\ 4732.60 & 7453.96 \end{pmatrix}$	$\begin{pmatrix} 10301.24 & 5963.09 \\ 5963.09 & 11150.60 \end{pmatrix}$	$\begin{pmatrix} 10671.39 & 5236.21 \\ 5236.21 & 10346.82 \end{pmatrix}$
Mean of LFP varies but variance remains same			
Mean	Same as above process		
Variance	$\begin{pmatrix} 7493.19 & 4739.14 \\ 4739.14 & 7477.85 \end{pmatrix}$		
Spikes and LFPs are independent			
Mean	$\begin{pmatrix} -9.73 \\ -10.22 \end{pmatrix}$		
Variance	Same as above process		

Finally, we assume an autoregressive model for the LFP process:

$$X(t) = \mu + \sum_{l}^{u} a(s)X(t-s) + \sum_{l}^{u} b(s)Y(t-s) + \epsilon_t,$$

where we include a spike train term as in Section 3. It was observed that the spike values from the immediate past have some effect on the LFP values. As in Smith and Brown (2003), we consider only the special case

$$X(t) = \mu + a\,X(t-1) + b\,Y(t-1) + \epsilon_t. \tag{4.3}$$

The basic structure of the model is given in Figure 3. Note that in case the two series are independent, the firing function is constant and $b = 0$ in equation (4.3).

The fit of the IMI model is straightforward, with parameter estimates

$$(\hat{\alpha}_0, \hat{\alpha}_1) = (-5.69(0.006), -0.004(7.44 \times 10^{-7}))$$
$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = (-3.77(2.05 \times 10^{-5}), 7.82(0.005), 0.63(2.69 \times 10^{-9})),$$

where the values in the parenthesis are the corresponding squared standard errors. The estimated value of σ^2 is 4,558.85. The loglikelihood, AIC and BIC values
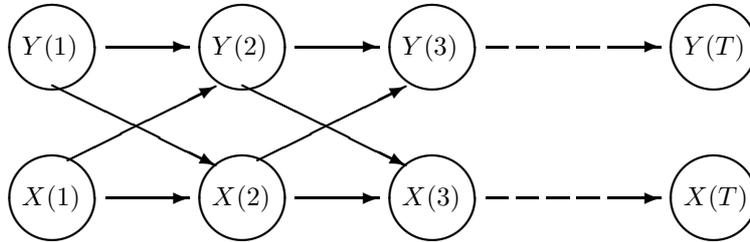
Figure 3. IMI: conditional intensity of $Y(t)$'s depend on $X(t)$ and $X(t-1)$; $X(t)$'s are normal, depending linearly on $X(t-1)$ and $Y(t-1)$'s.

are presented in Table 5, along with the independence case. Note that here the independent model performs much worse, which is an indication that this simple model captures the dependence structure of the data well.

We also considered a more general version of the IMI model, where the firing function was modified from (4.1) to

$$s(t) = \exp(\alpha_0 + \alpha_1 X(t) + \alpha_2 Y(t-1)),$$

and the LFP's were again modeled by (4.3). However, after fitting the model it was observed that the estimate of α_2 had a large p-value, and there was no significant change in the likelihood. Hence, we do not pursue this model.

Table 5. Goodness of fit for the IMI and ARMA models.

	$-2\times$ Likelihood	AIC	BIC
IMI model with dependence	562427.2	562439.2	562492.1
ARMA(1,1)-type model	567627.4	567641.4	567703.1
ARMA(2,1)-type model	567627.2	567643.2	567713.7
ARMA(1,3)-type model	567567.6	567589.6	567686.6
The independent case with AR(1) LFP	628682.7	628690.7	628725.9

4.2. A mixture ARMA-type model

Here we propose a new model for the time series of hybrid data. Note that data at the time point t is to be $(Y(t), X(t))$, $t = 1, 2, \ldots, T$, where $Y(t)$ is binary and $X(t)$ is continuous. In this model $Y(t)$ depends on $(Y(t-1), X(t-1))$ and $X(t)$ depends on $(Y(t), Y(t-1), X(t-1))$. See Figure 4 for an illustration.

Suppose $E(Y(t)) = \mu$, and $X(t)$ has a distribution F such that $E(X(t)|Y(t)) = \mu_0 + (\mu_1 - \mu_0)Y(t)$ and $\text{Var}(X(t)|Y(t)) = \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)Y(t)$. Then $E(X(t)|Y(t) = 0) = \mu_0$, $\text{Var}(X(t)|Y(t) = 0) = \sigma_0^2$, $E(X(t)|Y(t) = 1) = \mu_1$, and $\text{Var}(X(t)|Y(t)$
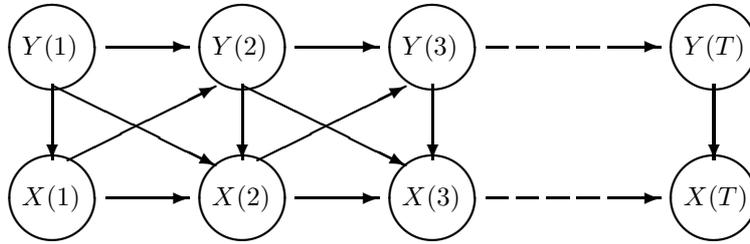
Figure 4. ARMA type model: $Y(t)$'s have a discrete ARMA-type structure and depend on $X(t-1)$ and $Y(t-1)$; $X(t)$ is normal, and depends on $X(t-1)$, $Y(t-1)$ and $Y(t)$.

$= 1) = \sigma_1^2$. Keeping this in mind, we define the mixture model for $Y(t)$ as

$$P(Y(t+1) = 1 | Y(t), X(t)) = \left(\frac{\mu}{1-\mu} + \phi[\psi(X(t)) - \psi_0] \right)(1 - Y(t)), \quad (4.4)$$

where ψ is a function which transforms $X(t)$ (which is continuous, possibly with domain $\mathcal{R}$, the real line) onto $[0,1]$. So ψ can be, for example, the cumulative distribution function (cdf) of some random variable. An obvious choice of ψ is Φ, the cdf of a standard normal random variable. Here

$$\psi_0 = E[\psi(X(t)) | Y(t)] = E[\psi(X(t)) | Y(t) = 1]Y(t) + E[\psi(X(t)) | Y(t) = 0](1 - Y(t))$$

is a function of $Y(t)$.

The AR(1)-type model for $X(t)$ (given $X(t-1), Y(t), Y(t-1)$) is

$$\{X(t) - (\mu_0 + (\mu_1 - \mu_0)Y(t))\} = \rho \{X(t-1) - (\mu_0 + (\mu_1 - \mu_0)Y(t-1))\} + \epsilon(t) \tag{4.5}$$

where, given $Y(t)$, $\epsilon(t)$ has the same distribution F with mean 0 and variance $\sigma_0^2 + (\sigma_1^2 - \sigma_0^2)Y(t)$.

Under this model, we have $E(Y(t+1) | Y(t)) = (\mu/(1-\mu))(1 - Y(t))$, and $E(Y(t)Y(t+1)) = 0$, and $(Y(t), Y(t+1)) \neq (1, 1)$ for all t. Note that, under the model (4.4) and (4.5), the maximum likelihood estimates are very difficult to obtain. Hence we propose some moment estimates of the parameters, which are consistent. We obtain the log-likelihood by plugging these estimates into the log-likelihood. We assume F to be the normal cdf to find the log-likelihood. To find the moment estimates, we proceed as follows. Routine derivation gives, for $s \geq 1$,

$$E(Y(t+s) | Y(t)) = E[c(1 - E(Y(t+s-1) | Y(t)))] = \cdots = c(1-c)^{s-1} + (-1)^s c^s Y(t),$$

for $c = \mu/(1 - \mu)$. Further,

$$\text{Cov}\,(X(t), X(t + s))$$
$$= \text{Cov}\,(\mu_0 + (\mu_1 - \mu_0)Y(t), \mu_0 + (\mu_1 - \mu_0)Y(t + s)) + E[\rho^s(\sigma_0^2 + (\sigma_1^2 - \sigma_0^2)Y(t))]$$
$$= (\mu_1 - \mu_0)^2\mu\left[c(1 - c)^{s-1} + (-1)^s c^s - \mu\right] + \rho^s(\sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\mu), \qquad (4.6)$$

$$\text{Var}\,(X(t)) = (\mu_1 - \mu_0)^2\mu(1 - \mu) + \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\mu. \qquad (4.7)$$

From (4.6) and (4.7) we get the autocorrelation function of $\{X(t)\}$ as

$$\rho_{t,t+1}^X = \frac{-(\mu_1 - \mu_0)^2\mu^2 + \rho(\sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\mu)}{(\mu_1 - \mu_0)^2\mu(1 - \mu) + \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\mu}.$$

The parameter ϕ appears in the cross-correlation like $\rho_{t,t+1}^{XY}$. Writing $\psi_0^* = E(\psi_0)$ and $\mu_0 + (\mu_1 - \mu_0)\mu = \mu^*$, we have

$$E(X(t)Y(t + 1)) = E[X(t)\,E(Y(t + 1)|X(t), Y(t))]$$
$$= \mu\mu^* + \phi\left[E(X(t)\psi(X(t))) - \psi_0^*\mu_*\right],$$

and we obtain

$$\rho_{t,t+1}^{XY} = \frac{\phi\left[E(X(t)\psi(X(t))) - \psi_0^*\mu^*\right]}{\sqrt{\mu(1 - \mu)}\sqrt{(\mu_1 - \mu_0)^2\mu(1 - \mu) + \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\mu}}.$$

The parameters can be estimated in the following way: (i) Obtain $\widehat{\mu}$ from the proportion of $Y(t) = 1$; (ii) obtain $\widehat{\mu}_1, \widehat{\sigma}_1^2$ from the $X(t)$'s for which the corresponding $Y(t) = 1$; (iii) get the estimates $\widehat{\mu}_0, \widehat{\sigma}_0^2$ from the $X(t)$'s for which the corresponding $Y(t) = 0$; (iv) obtain $\widehat{\rho}$ from the sample value of $\rho_{t,t+1}^X$ plugging in the estimates of μ, μ_1, σ_1^2, μ_0 and σ_0^2, obtained above; (v) obtain $\widehat{\phi}$ from the sample value of $\rho_{t,t+1}^{XY}$ using the estimates obtained above.

From the data we obtain $\widehat{\mu} = 0.0037$, $\widehat{\mu}_1 = -24.4620$, $\widehat{\sigma}_1^2 = 101.3764^2$, $\widehat{\mu}_0 = -9.92$, $\widehat{\sigma}_0^2 = 86.46^2$, $\widehat{\rho} = 0.6251$, and $\widehat{\phi} \simeq 0$. Note that $\widehat{\phi} \simeq 0$ implies that the Y_t's do not depend on the X_{t-1}'s, so far the given model is concerned. In the independence case, the model reduces to the independent IMI model discussed in the previous section.

4.3. Higher order ARMA-type models

The above model can be further extended by modeling $X(t)$ and $Y(t)$ using $(X(t - i), Y(t - i))$, $i = 1, 2, \ldots$. For example, we can write

$$P(Y(t) = 1|Y(t - 1), X(t - 1), Y(t - 2), X(t - 2), \ldots)$$
$$= \left(c_p\left\{1 + \sum_{i=2}^p (Y(t-i) - \mu)\right\} + \phi[\psi(X(t-1), X(t-2), \ldots) - \psi_0]\right)(1 - Y(t-1)), \quad (4.8)$$

$$\left\{ X(t) - \mu_0 - \sum_{i=1}^{q} \mu_i Y(t-i) \right\} = \rho \left\{ X(t-1) - \mu_0 - \sum_{i=1}^{q} \mu_i Y(t-1-i) \right\} + \epsilon(t),$$
$$(4.9)$$

where c_p is so chosen that $E(Y(t)) = \mu$. The model (4.8) can be looked as an AR(p)-type mixture model for Y, and the model (4.9) can be looked as an MA(q)-type model for X. Their suitable combination might give ARMA-type models for Y and X.

Note that the model discussed in Subsection 4.2 is ARMA(1,1)-type. We also tried ARMA(2,1)-type model and ARMA(1,3)-type model with our data, and the results are given in Table 5.

4.4. A modified ARMA model

The ARMA model does not seem to fit the data very well compared to some other models, as is evident from Table 10. However, a very slight modification to the ARMA models, to make the correlation parameter also dependent on the value of the spike process, improves results significantly. In that case, the continuous variable is defined as

$$\{ X(t) - (\mu_0 + (\mu_1 - \mu_0)Y(t)) \}$$
$$= (\rho_0 + (\rho_1 - \rho_0)Y(t)) \{ X(t-1) - (\mu_0 + (\mu_1 - \mu_0)Y(t-1)) \} + \epsilon(t) \quad (4.10)$$

where, given $Y(t)$, $\epsilon(t)$ has the same distribution F with mean 0 and variance $\sigma_0^2 + (\sigma_1^2 - \sigma_0^2)Y(t)$. The parameters could be fitted by a version of the E-M algorithm. Alternatively, one can consider combining the discrete ARMA model for $Y(t)$ with the following JMLGS type-equation:

$$X(t) = (a_0 + (a_1 - a_0)Y(t)) + (b_0 + (b_1 - b_0)Y(t))X(t-1) + \epsilon(t), \quad (4.11)$$

where $\epsilon(t)$ is distributed as before. It may be noted that (refar1-xxx) is only a very slight modification of (4.10), as it can be re-written as

$$\{ X(t) - (\mu_0 + (\mu_1 - \mu_0)Y(t)) \}$$
$$= (\rho_0 + (\rho_1 - \rho_0)Y(t)) \{ X(t-1) - (\mu_0 + (\mu_1 - \mu_0)Y(t)) \} + \epsilon(t). \quad (4.12)$$

The fit of this model can be achieved combining the method to fit the parameters of the discrete process with that for the Markov chains described in Section 3. The fit, presented in Table 10, appears much improved.

5. State Space Models for Hybrid Data

The state-space models are widely used in many fields of science and engineering. In these models, there are two sets of equations: one is a set of state

equations that defines the evolution of the process through time, and the other defines how the observed or the measured values are governed by the process. These processes are also known as latent process models, and, when the evolution of the process is Markov, they are often referred to as hidden Markov models. They have been used extensively for analysis of continuous-valued data (Ljung and Söderström (1987), Roweis and Ghahramani (1999) and Gharamani (2001)), and less so for discrete data (Camproux et al. (1996), Smith and Brown (2003) and Srinivasan and Brown (2007)).

When there is reason to believe that some implicit stimuli controls the spike train, a state space model is often fitted, as in Smith and Brown (2003) and Smith et al. (2003). The state space models for spike bursts usually assume that the conditional intensity function depends on some hidden latent process, which can be continuous (Smith and Brown (2003), Smith et al. (2003)), or discrete (Schneidman et al. (2006)). The hidden process is often assumed to have a Markovian structure. When the state space is assumed to be discrete, the usual procedure is to consider the number of hidden states to be finite, and estimate the number of states using some suitable criteria. The LFP is often thought of as a combination of the firing of multiple neurons in the background, and hence it is natural to assume it to be governed by the same stimuli controlling the spike bursts for single units. Therefore, in absence of any scientific evidence in favor of causal structures, a state space model seems justifiable for a hybrid process as considered in this paper.

5.1. Hidden Markov model fit for hybrid data

For modeling spike trains using state space models, the common practice is to consider a discrete set of time points for ease of notation for filtering, smoothing and application of the EM algorithm to fit the model (Smith and Brown (2003)), although continuous-time models could also be used. When modeling hybrid data, we prefer the discretization of the time for the same reasons. Further, note that the models used by Smith and Brown (2003) and Smith et al. (2003) do not model the refractory period, and hence, although these models are good for spike counts, they may not be very useful for the spike generation process unless additional assumptions are made to accommodate the refractory period, making the models complicated. Here, to motivate the usefulness of the state space models, we assume the state space to be discrete; this simplification helps us to allow for the refractory period in our model without making it complex. As is demonstrated later in this section, the fit of the state space model works well for the LFP and spike train data. The continuous state space models could also be extended to allow for the refractory period. We intend to address this in future work.
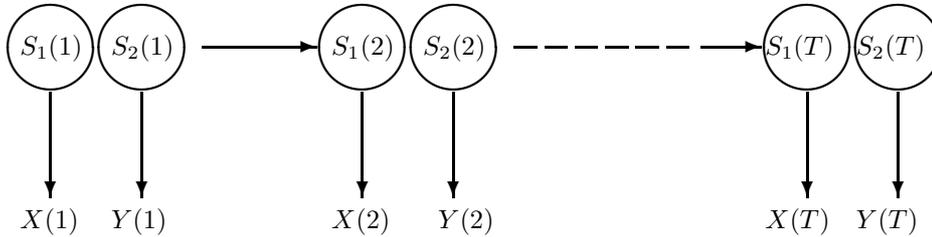
Figure 5. A general hidden Markov model: $S_1(t)$ and $S_2(t)$ are the hidden states corresponding to $X(t)$ and $Y(t)$ respectively.

We now describe a method for fitting hidden Markov models to the hybrid data set. We start by assuming that there are K_X possible different hidden states controlling the continuous process, and K_Y possible different hidden states for the discrete process. For any time instant t, we denote by the pair $(S_1(t), S_2(t)) :=$ $\mathbf{S}(t)$ the hidden states of the two processes at time t, respectively. See Figure 5 for an illustration. For the brain data example, the various hidden states can be thought of as the presence/absence of different stimuli from outside, or different parts of, the body. For a preliminary analysis, assume that X_t has a normal distribution with certain mean and variance depending on the latent process only through the present, i.e., is conditionally independent of the past values of X and Y the hidden states given the present state. Let us denote the mean and variances given state $\mathbf{S}(t)$ by $\mu_{S_1}(t)$ and $\sigma_{S_1}^2(t)$ respectively:

$$p(X(t)|\mathbf{S}(t)) = N(X(t); \mu_{S_1}(t), \sigma_{S_1}(t)). \tag{5.1}$$

Notice that there are K_X possibly different values of the means and the variances. We also assume that $Y(t)|\mathbf{S}(t)$ has an intensity function similar to (1.3),

$$\lambda(t) = s_{\mathbf{S}(t)}(t)\, r_{\mathbf{S}(t)}(t - w_N(t)),$$

where $s_{\mathbf{S}(t)}(t)$ is a firing function, and $r_{\mathbf{S}(t)}(\cdot)$ is a recovery function. As in Section 4.2, again put

$$r_{\mathbf{S}(t)}(t) = \begin{cases} 1 \text{ if } 0 \leq t \leq 2 \text{ ms}, \\ 0 \text{ otherwise}, \end{cases}$$

which effectively means that the recovery function is independent of the state, and hence can be denoted as $r(t)$. Assume farther that the firing function at time t is a constant given $S_2(t)$, i.e., $s_{\mathbf{S}(t)}(t) = s_{S_2(t)}$. Since we deal with discrete time, $\lambda(t)$ can be thought of as the probability $P(Y(t) = 1|\mathbf{S}(t))$.

For ease of notation, write the probability of the configuration $(\mathbf{S}(1), \ldots, \mathbf{S}(T);$ $X(1), \ldots, X(T); Y(1), \ldots, Y(T))$ as $P(\mathbf{S}, \mathcal{X}, \mathcal{Y})$, where $\mathbf{S} = (\mathbf{S}(1), \ldots, \mathbf{S}(T))$ etc.

We write the transition matrix for the Markov chain as $A := \{a(u,v) : u,v = 1,\ldots,K_X \times K_Y\}$, where $a(u,v) = P(\mathbf{S}(t+1) = v|\mathbf{s}(t) = u)$ independent of t. We also assume A to be aperiodic and irreducible, which ensures some nice asymptotic properties (Bickel, Ritov and Rydén (1988)), as well as numerical stability (Jordan (2002)). If the initial distribution is π, then

$$P(\mathbf{S}, \mathcal{X}, \mathcal{Y}) = \pi(\mathbf{s}(1)) \prod_{t=1}^{T-1} a(\mathbf{S}(t), \mathbf{S}(t+1)) \prod_{t=1}^{T} p(X(t)|S_1(t))p(Y(t)|S_2(t)). \quad (5.2)$$

As the states are hidden, the likelihood of the data set $(\mathbf{X}, \mathbf{Y})$ is of more actual use than (5.2), and is given by

$$P(\mathcal{X}, \mathcal{Y}) = \sum_{\mathbf{S}(1)} \cdots \sum_{\mathbf{S}(T)} \pi(\mathbf{S}(1)) \prod_{t=1}^{T-1} a(\mathbf{S}(t), \mathbf{S}(t+1)) \prod_{t=1}^{T} p(X(t)|S_1(t))p(Y(t)|S_2(t)).$$

The maximum likelihood estimators of the parameters can be calculated recursively using a suitable modification of the Baum-Welch algorithm for the hybrid data structure. The derivations are similar to those of Jordan (2002).

When the two processes are independent, one has

$$P((S_1(t+1), S_2(t+1))|(S_1(t), S_2(t))) = P(S_1(t+1)|S_1(t))\ P(S_2(t+1)|S_2(t))$$

for all $t = 1, T - 1$.

The full model is computationally intensive, as the transition matrix is of size $K_X \times K_Y$ and hence, even for moderate values of K_X and K_Y the model involves hundreds of parameters. A simpler approach takes the same hidden states for the two time series, when it is reasonable to assume that the same stimuli act as the "hidden states" for the spike trains as well as the LFP. Figure 6 summarizes this reduced model. Most of the illustrative results presented here use this model, referred to as a "conditional model"; the complete model described first is called the "full model".
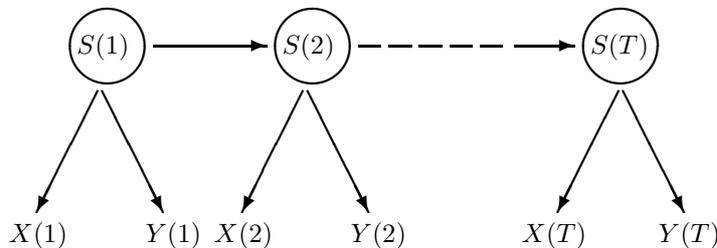

Figure 6. A representation of a hidden Markov model where the hidden states for both processes are the same.

5.2. Results and discussion

Analyses are performed for models corresponding to the restricted case $K_X = K_Y := K$. We perform and compare fits for values $K = 2, 3, 4$ and 5 for the various HMM models. The case $K = 2$ is found to perform much worse than the $K = 3$ case and hence we omit it from our discussion. The fits seem to 'improve' with larger K, but the improvement diminishes with increasing values of K. The full models have numbers of parameters of order K^4 (the exact number is $K^4 - 1 + 3K$), and hence the fit for the full models are computationally intensive; with five states the model takes about three hours to fit using a dual core 3.6 GHz workstation, when the convergence criteria used is that the change in loglikelihood values drop below 0.01. However, the numbers of parameters for both the independent and the conditional models are only of the order K^2, (the exact numbers are $2(K^2 - 1) + 3K$ and $K^2 - 1 + 3K$, respectively,) and are estimated very quickly. For example, models with $K = 10$ (fitted model not provided here) take only about ten minutes to fit for both the conditional and independent models when using the same machine.

Table 6. Fitted values for a three-state hidden Markov model where LFP and spikes are conditionally interdependent.

Parameter	State I	State II	State III
Restricted Conditional Model			
mean	-99.69	-1.97	98.35
sd	55.6	43.8	63.8
pr	0.0051	0.0027	0.0041
Independence case			
mean	-99.57	-1.83	98.57
sd	55.62	43.81	63.73
pr	0.0025	0.0037	0.0050

For the sake of brevity, we do not provide the fit for the full models here. The loglikelihood values for some fitted models are provided in Table 7. It is found that, in the independence case for any value of K, the fit for the spikes is not very good, and the fit for the LFP is very close to the fit for the conditional model; for an illustration refer to Table 6 where the fits for $K = 3$ are reproduced. The fits for the LFP are almost identical, but the fits for the spikes are very different. The transition matrices are not presented here, but they also follow the same pattern.

It may be argued that the simple model assumed here for the spikes does not work very well without added complexities, provided to a degree by the conditional models. It may also be seen from Table 7 that the full model with four

Table 7. A comparison of the fit for various hidden Markov models. The fit of the five state conditional model with conditionally AR(1) continuous component performs the best among the fitted models.

Model	$-2\times$ logli-kelihood	AIC	BIC
HMM(3 states each for LFP and spikes, full model)	568641.9	568819.9	569604.4
HMM(4 states each for LFP and spikes, full model)	562684.3	563218.3	565571.7
HMM(5 states each for LFP and spikes, full model)	558665.4	559943.4	565575.6
HMM(3 states each, LFP and spikes indep)	571764.4	571814.4	572034.8
HMM(4 states each, LFP and spikes indep)	567880.8	567964.8	568335
HMM(5 states each, LFP and spikes indep)	565490.1	565616.1	566171.4
HMM(3 states, conditional independence)	571756	571790	571939.8
HMM(4 states, conditional independence)	567867.4	567921.4	568159.4
HMM(5 states, conditional independence)	565472.9	565550.9	565894.6
HMM(3 states, conditional independence, AR(1))	558837.8	558873.8	559032.5
HMM(4 states, conditional independence, AR(1))	556378.4	556434.4	556681.2
HMM(5 states, conditional independence, AR(1))	555843	555923	556275.5

states performs better than the full model with five states in terms of BIC, although in terms of AIC the model with five states is best. One can also see that the conditional model with just six states, which takes about four minutes to fit, is actually better in terms of BIC than the "best" full model. Hence, as far as the three competing hidden Markov models are concerned, the conditional independence models seem best in this case. Among the conditional models, however, BIC improves with increasing number of states, which may be an indication that a model with a continuous latent process as in Smith and Brown (2003) would be more appropriate.

The standard error values for the estimates are not provided here. Theoretically, when the process $\{S(t), X(t), Y(t)\}$ is stationary, the standard error estimates could be based on the asymptotic normal distribution of the maximum likelihood estimates of the parameters, where the asymptotic variance is provided by the Fisher information matrix $\mathcal{I}$ (Bickel, Ritov and Rydén (1988)). However, due to the complexity of the algorithm used, direct estimation of the information matrix is hard. One can alternatively use replication-based estimates for the variance of parameters in that case. The methods are similar to those of Brillinger and Guha (2007).

It may be observed from Table 10 that the conditional model does not work as well as other simpler models for the given data set. We note that, except for

this model, all other models assume a direct dependence of $X(t)$ on $X(t-1)$, and hence we investigate the relatively more complex model with

$$p(X(t)|\mathbf{S}(t)) = N(X(t); \mu_{S_1}(t) - \rho(X(t-1) - \mu_{S_1}(t-1)), \sigma_{S_1}(t)). \qquad (5.3)$$

This model allows $X(t)$ to depend directly on $X(t-1)$, i.e., conditional on the hidden states, $X(t)$ is is assumed to be an autoregressive process of order 1 (AR(1)). In addition to the models discussed before, we also provide the fit for this model for $K = 3\text{-}5$ in Table 7. The fit for this model is much better, strengthening our claim that the LFP values have a strong lagged dependence structure.

6. Simulations

In this section, we provide the results from a simulation study to compare different models. In particular, we simulated six models: two hidden Markov models with three hidden states, with a conditional independence structure where, in one case, the continuous process has the lagged dependence structure of an AR(1) order (HMMAR) and, in the other case, is conditionally independent (HMMCI); the Markov chain model with $K = 3$ and $u = 2$; an IMI process; and two ARMA-type processes, one where the continuous process evolves following (4.5) (ARMA), and the other in which the continuous process evolves as in (4.11) (ARMA-JMLGS).

For all models, the simulations were carried out 1,000 times with parameters fitted to the mouse auditory cortex data from Section 2, and the observation number equal to that of the original data set, 49,722. In addition, we fit two existing models: a model similar to that of Smith and Brown (2003), referred to as the "Smith-Brown" (SB) model; and a JMLGS of order 1.

The simulations were repeated with a smaller sample size of 5,000, and all parameters as before except for a change in the parameters for the point process part; as for the original process the firing rate was very low (about 0.4%). The fits are assessed by AIC and BIC as before, and are reported in Tables 8 and 9. From them, it may be seen that both hidden Markov models perform significantly better than the rest when there is a hidden Markov structure, even if they are misspecified among themselves (in terms of whether the continuous process is conditionally IID or AR(1)). However, in all other cases they do not perform very well, especially the HMMCI. The ARMA model's performance is consistently the worst. One possible explanation is the simplified plug-in algorithm that we use; it might be improved by adapting the more efficient but time-consuming innovations algorithm (see Brockwell and Davis (1991)). The estimate we use can be regarded as the initial estimate of the innovations algorithm, and may be sub-optimal.

Table 8. A comparison of the fit based on 1,000 simulations for various types of models fitted to simulated data of length 49,722, and with parameters equal to those fitted to the original data. The AIC and BIC reported are averaged over the 1,000 simulations; the values in parentheses give the standard errors of the AIC values; the BIC standard errors are equal to that of the corresponding AIC standard errors.

Fitted Model Simulated Model		HMM:3 states cond.ind.	HMM:3 states AR(1)	Markov spikes K:3,u:2	IMI	ARMA (1,1)	ARMA-JMLGS	JMLGS	SB
HMM: 3 state	AIC	560544 (11.4)	564779 (11.3)	565126 (11.5)	565737 (11.3)	567213 (11.3)	565722 (11.7)	565717 (11.8)	565738 (11.8)
	BIC	560694	564938	565682	565790	567275	565801	565797	565791
HMM: 3 st AR(1)	AIC	560733 (13.4)	556927 (14.8)	559599 (14.1)	559614 (14.0)	560970 (14.0)	559599 (14.0)	559600 (14.0)	559615 (14.0)
	BIC	560883	557086	559837	559667	561032	559678	559679	559668
Markov K=3 u=2	AIC	574850 (11.1)	562421 (12.3)	557427 (11.2)	562394 (12.3)	567608 (12.0)	562364 (12.3)	562371 (12.3)	562395 (12.3)
	BIC	575000	562579	561900	562446	567669	562443	562450	562448
IMI	AIC	571450 (10.8)	562486 (10.6)	562470 (10.6)	562460 (10.7)	567680 (12.6)	562461 (10.7)	562460 (10.7)	562461 (10.7)
	BIC	571600	562644	562708	562512	5677412	562540	562539	562514
ARMA (1,1)	AIC	592842 (27.7)	587066 (10.9)	587047 (9.6)	587041 (9.6)	592267 (10.9)	587024 (9.6)	587036 (9.6)	587043 (9.6)
	BIC	592992	587224	587285	587093	592329	587103	587116	587096
ARMA JM-LGS	AIC	592925 (38.6)	587143 (9.4)	587113 (9.6)	587124 (9.6)	592337 (10.8)	587094 (9.6)	587109 (9.6)	587126 (9.6)
	BIC	593074	587301	587315	587177	592399	587174	587188	587179

The Markov chain model appears to be the most robust among all; it outperforms all misspecified models in most simulation, except for the IMI model with a small sample size, where the JMLGS performs the best. The two "existing models", SB and JMLGS, perform reasonably well, but are usually outperformed by the respective "improved versions": the IMI and the ARMA-JMLGS models. The improvements, especially from the SB model to the IMI model do not appear very significant at first glance, but the standard errors of the difference of AIC values for various models are of order 10^{-2}, and hence the improvements are significant. Further, keeping in mind that the only changes in these models are made for the point process part whose contribution to the likelihood function is very small, the importance of the modifications can be understood.

Note that we performed detailed simulation studies with several other parameter combinations and data lengths, and the findings were similar. Hence, we do not report them all.

Table 9. A comparison of the fit based on 1,000 simulations for various types of models fitted to simulated data of length 5,000 and with spike firing rate 10 times that of the original data, other parameters equal to those fitted to the original data. The AIC and BIC reported are averaged over the 1,000 simulations; the values in parentheses are the standard errors.

Fitted Model Simulated Model		HMM:3 states cond.ind.	HMM:3 states AR(1)	Markov spikes K:3,u:2	IMI	ARMA (1,1)	ARMA-JMLGS	JMLGS	SB
HMM: 3 state	AIC	57,450 (5.2)	57,886 (5.1)	57,928 (4.8)	57,994 (4.8)	58,140 (4.9)	57,978 (4.8)	57,978 (4.8)	58,003 (4.8)
	BIC	57,560	58,004	57,990	58,033	58,186	58,037	58,037	58,042
HMM: 3 st AR(1)	AIC	57,374 (6.2)	56,985 (6.2)	57,249 (6.1)	57,270 (6.3)	57,399 (6.2)	57,253 (6.3)	57,255 (6.3)	57,276 (6.4)
	BIC	57,485	57,102	57,425	57,309	57,444	57,312	57,314	57,315
Markov K=3 u=2	AIC	59,112 (5.1)	58,178 (5.4)	57,559 (5.0)	58,151 (5.4)	58,632 (5.1)	58,126 (5.4)	58,131 (5.4)	58,164 (5.5)
	BIC	59,222	58,295	58,092	58,190	58,678	58,185	58,189	58,203
IMI	AIC	59,596 (5.1)	58,829 (4.2)	58,810 (4.6)	58,801 (4.6)	59,384 (4.9)	58,807 (4.6)	58,805 (4.6)	58,855 (4.7)
	BIC	59706	58,946	58,986	58,840	59,430	58,866	58,863	58,894
ARMA (1,1)	AIC	61,044 (5.5)	60,445 (4.0)	60,417 (4.0)	60,424 (4.0)	60,954 (4.4)	60,410.7 (4.0)	60,411.4 (4.0)	60,446 (4.1)
	BIC	61155	60,563	60,593	60,463	61,000	60,469	60,470	60,485
ARMA JM-LGS	AIC	61,037 (5.9)	60,445 (4.0)	60,416 (4.0)	60,434 (4.0)	60,945 (4.3)	60,414 (4.1)	60,416 (4.1)	60,449 (4.1)
	BIC	61,148	60,563	60,565	60,473	60,990	60,473	60,475	60,488

7. Discussions of Results

In this article, we have discussed various methods of fitting parametric models to bivariate hybrid neuroscience data. The examples provided are mostly modifications of existing methodologies. We have provided examples of existing techniques to model hybrid structures when causal directions exist: a Markov model for spikes assuming the LFP depends on them, an IMI model and some mixed ARMA-type models assuming spikes depend on the LFP; and also assuming they do not, we fitted a state space model with a hidden latent process with discrete states. These are not necessarily the best models that could be fit to the particular dataset considered, for example, a non-parametric fit of the auto-intensity function showed an increased intensity near firing times, and this has not been modeled. The models that we show in this work are more like building blocks: they are simple, fast to fit, practical, and can be extended to more complex models with ease.

We have employed three model comparison statistics in the previous chapters: log likelihood, AIC and BIC. There are standard results that allow us to compare nested models, say for example the two IMI model fits with and without lagged spike predictors for spikes, where the difference in twice log likelihood follows an approximate χ^2-distribution with degrees of freedom equal to the difference in number of parameters (McCullagh and Nelder (1989)). Similar tests are available for hidden Markov models under certain regularity conditions, see Giudici et al (2000) for details. However, it is well known that, in general, log likelihood is not a good measure for comparing models, essentially non-nested ones. In Table 10 we only use AIC and BIC, commonly accepted as reasonable measures of goodness of fit of neuroscience models (Brown et al. (2003), Barbieri et al. (2001) and Brown, Barbieri Ventura, Kass and Frank (1998)). However, AIC and BIC also have their limitations, especially for point processes like spike trains, and various alternatives have been suggested, including an application of a time rescaling theorem for point processes (Brown, Barbieri Ventura, Kass and Frank (1998) and Kass, Ventura and Brown (2005)), and entropy-based measures (Nakahara, Amari and Richmond (2006) and Schneidman et al. (2006)). The former, at least in its present form, might not be very useful for hybrid processes, but the entropy-based methods could be extremely useful.

With the continuous time scale, unless some sort of conditional independence is assumed (for example see Smith and Brown (2003)), extension of the model for multiple spike trains becomes very complicated, see Brown, Kass and Mitra (2004). However, in many discrete time cases, the extensions, although computationally extensive when the number of neurons is large, would be immediate. For example, the Markov model for spikes used in Section 3 could be extended to multiple spike trains just by suitably increasing the number of values of the discrete variable: for example, when two neurons are present, define a discrete variable taking values $0, 1, 2$ or 3, respectively, when none, first, second or both neurons fire. It is true that this model becomes computationally difficult very quickly unless some sort of additional conditions are assumed, but nevertheless it is still a method that could be utilized.

One major drawback of using the AIC and BIC for hybrid processes is that the point processes receive very little weight. For example, look at the comparison of the IMI models with and without refractory period. The improvement of the model appears to be very small; due to the small contribution of the point process part to the likelihood, caused partly by the low firing rate of the spike train. However, the introduction of a refractory period does improve upon the existing model, and it might be argued that for a spike train with higher firing rate, the improvement would be more pronounced, as seen in the simulations. The simulations also indicate that this small improvement is statistically significant.

The simulations from Section 6 suggest that when the hidden Markov models hold, they outperform all the other models by far, otherwise they do not perform well. One other major problem with the hidden Markov models is that they often have a complicated likelihood function, more so when the number of states to be fitted is high, and often the iterated algorithm used to fit the model converges to a local maximum. This means that many initial values need to be tried before getting a good fit, a significant computational cost. This problem is more pronounced for the model with AR(1)-type complexity, also because the Baum-Welch algorithm needs to be modified for this case to accommodate the AR(1) structure, and the trade-off for the improved model is a slower convergence rate compared to the other conditional hidden Markov model. In comparison, the simple Markov models, even though they have large number of parameters, provide quick and good fits, and might be useful when an on-line model fit is needed.

To a lesser extent, the SB model suffers from the same problem as the hidden Markov model; it can diverge unless proper initial values are chosen. Also, considering the log link for the rate function (see (4.1)) means that, theoretically, the fitted rate could be more than 1, which cannot be allowed. Hence, using a logit link function appears to be a reasonable choice. Although for the fit to the original data the improvements in the modified IMI fit is not very pronounced, it is much more visible in simulations with a higher spike-firing rate.

The ARMA model fitted using the algorithm described in Section 4.2 suffers from some estimation bias, and does not perform well even when it is the true model. Applying a more complicated, EM-type iterative algorithm similar to the innovations algorithm (see Brockwell and Davis (1991)) improves the outcome, but it is computationally intensive. One other alternative is to fit the modified model of Section 4.4 (described in Table 10 as the ARMA-JMLGS model), whose fit is less computationally intensive, yet more accurate.

The Markov model with $K = 3$ and $u = 2$ seems to fit quite well. This model has the largest number of parameters but, unlike some others, the fit is achieved very quickly even using a large number of parameters. However, a model that is "best" in terms of AIC and BIC might not be the best hybrid model. The Markov chain structure assumed here is overly simplistic, and cannot be easily modified to allow for a randomized recovery function. The ARMA-type model can accommodate more complicated recovery functions; it allows the transition probabilities to be random, and to depend on the continuous process, which appears more reasonable in many situations.

In conclusion, it may be said that there is no single "best" model for all hybrid neuroscience processes, but the models discussed here, and suitable modifications, can be used in many situations that arise with real life neuroscience data.

Table 10. A comparison of the fit for various types of models fitted to the hybrid data. The last two are existing models.

Model	#para-meters	AIC	BIC
HMM(3 states, cond. ind.)	17	571,790	571,939.8
HMM(5 states, cond. ind.)	39	562,182.4	563,319.5
HMM(3 states, cond. ind., LFP AR(1))	18	558,873.8	559,032.5
HMM(5 states, cond. ind., LFP AR(1))	40	555,923	556,275.5
Markov spikes ($K = 3, u = 2$)	27	557,467.6	557,705.6
IMI model for spikes with AR(1) LFP	6	562,439.2	562,492.1
ARMA(1,1)-type model	7	567,641.4	567,703.1
ARMA(1,1)-JMLGS combined	9	562,404.2	562,483.5
JMLGS Model	9	562,404.6	562,483.9
Smith-Brown Model	6	562,441.3	562,494.2

References

Andrieu, C., Davy, M. and Doucet, A. (2003). Efficient particle filtering for jump Markov systems: application to time-varying autoregressions. *IEEE Trans. Signal Process.* **51**, 1762-1770.

Barbieri R, Quirk M. C., Frank L. M., Wilson M. A. and Brown E. N. (2001). Construction and analysis of non-Poisson stimulus response models of neural spike train activity. *J. Neuroscience Methods* **105**, 25-37.

Berry, M. J., Warland, D. K. and Meister, M. (1997). The structure and precision of retinal spike trains. *Proc. Natl. Acad. Sci. USA* **94**, 5411-5416.

Bickel, P. J., Ritov, Y. and Rydén, T. (1988). Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *Ann. Math. Statist.* **26**, 1614-1635.

Brillinger, D. R. (1988). The maximum likelihood approach to the identification of neuronal firing systems. *Ann. Biomedical Engineering* **16**, 3-16.

Brillinger, D. R. (1992). Nerve cell spike train data analysis: a progression of technique. *J. Amer. Statist. Assoc.* **87**, 260-271.

Brillinger, D. R., Bryant, H. L. J. and Segundo, J. P. (1976). Identification of synaptic interactions. *Biological Cybernetics* **22**, 213-228.

Brillinger, D. R. and Guha, A. (2007). Mutual information in the frequency domain, J. Statist. Plann. Inference **137**, 1076-1084.

Brillinger, D. R. and Segundo, J. P. (1979). Empirical examination of the threshold model of neuron firing. *Biological Cybernetics* **35**, 213-220.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series : Theory and Methods*. Springer-Verlag, New York.

Brown, E. N., Barbieri, R., Eden, U. T. and Loren M. .F. (2003). Likelihood methods for neural spike train data analysis. *Computational Neuroscience: A Comprehensive Approach*, 253-286. CRC Press, London.

Brown, E. N., Barbieri, R., Ventura, V., Kass, R. E. and Frank, L. M. (1998). The time rescaling theorem and its application to neural spike train data analysis. *Neural Computation* **14**, 325-346.

Brown, E. N., Frank, L. M., Tang, D., Quirk, M. C. and Wilson, M. A. (1998). A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal lace cells. *J. Neuroscience* **18**, 7411-7425.

Brown, E. N., Kass, R. E. and Mitra, P. P. (2004). Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nature Neuroscience* **7**, 456-461.

Camproux, A. C., Saunier, F., Chouvet, G., Thalabard, J. C. and Thomas, G. A. (1996). Hidden Markov model approach to neuron firing patterns. *Biophysical Journal* **71**, 2404-2412.

Chan, H. P. and Loh, W. L. (2007). Some theoretical results on spike train probability models. *Ann. Statist.* **6**, 2691-2722.

Chen Y. H., Bressler S. L., Knuth K. H., Truccolo W. A. and Ding M. Z., (2006). Stochastic modeling of neurobiological time series: Power, coherence, Granger causality, and separation of evoked responses from ongoing activity. *Chaos* **16**, 026113.

Davis, G. M. and Ensor, K. B. (2007). Multivariate time series analysis with categorical and continuous variables in an LSTR model. *J. Time Ser. Anal.* **28**, 867-885.

Doucet, A. and Andrieu, C. (2001). Iterative algorithms for state estimation of jump Markov linear systems. *IEEE Trans. Signal Process.* **49**, 1216-1227.

Gharamani, Z. (2001). An introduction to hidden Markov models and Bayesian networks. *Internat. J. Pattern Recognition* **15**, 9-42.

Giudici, P., Rydén, T. and Vandekerkhove, P. (2000). Likelihood-ratio test for hidden Markov models. *Biometrics* **56**, 742-747.

Johnson, D. H. and Swami, A. (1983). The transmission of signals by auditory-nerve fiber discharge patterns. *J. Acoustical Society of America* **74**, 493-501.

Jordan, M. I. (2002). An Introduction to Probabilistic Graphical Models. Lecture Notes.

Jørgensen, B., Lundbye-Christensen, S., Song, P.X. and Sun, L. (1996). State-space model for multivariate longitudinal data of mixed types. *Canad. J. Statist.* **24**, 385-402.

Kandel, E. R., Schwartz, J. H. and Jessell, T. M. (2000). *Principles of Neural Science*. 4 edition. Appleton & Lange, Norwalk, Connecticut.

Kass, R. E. and Ventura,V. (2001). A spike-train probability model. *Neural Computation* **13**, 1713-1720.

Kass, R. E., Ventura, V. and Brown, E. N. (2005). Statistical issues in the analysis of neuronal data. *J. Neurophysiology* **94**, 8-25.

Kiang, N. Y. S., Watanabe, T., Thomas, E. C., and Clark, L. F. (1965). *Discharge Patterns Of Single Fibers in the Cat'S Auditory Nerve*. M.I.T. Press, Cambridge, MA.

Krishnamurthy, V. and Evans, J. (1998). Finite-dimensional filters for passive tracking of Marov jump linear systems. *Automatica* **34**, 765-770.

Laubach, M. (2004). Wavelet-based processing of neuronal spike trains prior to discriminant analysis. *J. Neuroscience Methods* **134**, 159-168.

Ljung, L. and Söderström, S. (1987). *Theory and Practice of Recursive Identification.* MIT Press, Cambridge, MA.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models.* 2 edition. Chapman and Hall, New York.

McEchron, M. D., Weible, A. P. and Disterhoft, J. F. (2001). Aging and learningspecific changes in single-neuron activity in CA1 hippocampus during rabbit trace eyeblink conditioning. *J. Neurophysiology* **86**, 1839-1857.

Nakahara, H., Amari, S. I. and Richmond, B. J., (2006). A comparison of descriptive models of a single spike train by information-geometric measure. *Neural Computation* **18**, 545-568.

Noreña, A. and Eggermont, J. J. (2002). Comparison between local field potentials and unit cluster activity in primary auditory cortex and anterior auditory field in the cat. *Hearing Research* **166**, 202-213.

Quenet, B. and Horn, D. (2003). The dynamic neural filter: a binary model of spatiotemporal coding. *Neural Computation* **15**, 309-329.

Riehle, A., Grün, S., Diesmann, M. and Aertsen, A. (1997). Spike synchronization and rate modulation differentially involved in motor cortical function. *Science* **278**, 950-1953.

Roux, S. G., Cenier, T., Garcia, S., Litaudon, P. and Buonviso, N. (2007). A wavelet-based method for local phase extraction from a multi-frequency oscillatory signal. *J. Neuroscience Methods* **160**, 135-143.

Roweis, S. and Ghahramani, Z. (1999). A unifying review of linear Gaussian models. *Neural Computation* **11**, 305-345.

Schneidman, E, Berry, M. J., Segev, R. and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* **440**, 1007-1012.

Schoenberg, F. P. (2004). Testing separability in spatial-temporal marked point processes. *Biometrics* **60**, 471-481.

Smith A. C. and Brown E. N. (2003). Estimating a state-space model from point process observations. *Neural Computation* **15**, 965-991.

Smith, A. C., Frank, L. M., Wirth, S., Yanike, M., Hu, D., Kubota, Y.,Graybiel, A. M., Suzuki, W. A. and Brown, E. N. (2003). Dynamic analyses of learning in behavioral experiments. *J. Neuroscience* **24**, 447-461.

Srinivasan, L. and Brown, E. N., (2007). A state-space framework for movement control to dynamic goals through brain-driven interfaces. *IEEE Trans. Biomedical Engineering* **54**, 526-535.

Stiebler, I., Neulist, R., Fichtel, I. and Ehret, G. (1997). The auditory cortex of the house mouse: left-right differences, tonotropic organization and quantitative analysis of frequency representation. *J. Comparative Physiology* **181**, 559-571.

Valentine, P. A. and Eggermont, J. J. (2001). Spontaneous burst-firing in three auditory cortical fields: its relation to local field potentials and its effect on inter-area cross-correlations. *Hearing Research* **154**, 146-157.

Ventura, V., Carta, R., Kass, R. E., Olson, C. R. and Gettner, S. N. (2001). Statistical analysis of temporal evolution in single-neuron firing rates. *Biostatistics* **1**, 1-20.

Vere-Jones, D. (1995). Forecasting earthquakes and earthquake risk. *Internat. J. Forecasting* **11**, 503-538.

Vialatte, F. B., Martin, C., Dubios, R., Haddad, J., Quenet, B., Gervais, R. and Dreyfus, G. (2007). A machine learning approach to the analysis of time requency maps, and its application to neural dynamics. *Neural Networks* **20**, 194-209.

Villa, A. E., Hyland, B., Tetko, I. V. and Najem, A. (1998). Dynamic cell assemblies in the rat auditory cortex in a reaction-time task. *Biosystems* **48** , 269-277.

Wiener, M. C. and Richmond, B. J. (2003). Decoding spike trains instant by instant using order statistics and the mixture-of-Poissons model. *J Neuroscience* **23**, 2394-2406.

Willie, J. (1979). Analyzing relationships between a time series and a point process. Ph.D. dissertation, Department of Statistics, University of California, Berkeley.

Wilson, M. A. and McNaughton, B. L. (1993). Dynamics of the hippocampal ensemble code for space. *Science* **261**, 1055-1058.

Department of Statistics and Applied Probability, National University of Singapore, Singapore.

School of Mathematics, University of Birmingham, U.K.

E-mail: A.Guha@bham.ac.uk

Applied Statistics Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata – 700 108, India

E-mail: atanu@isical.ac.in