ELSEVIER

Journal of Neuroscience Methods 122 (2002) 25–42

JOURNAL OF NEUROSCIENCE METHODS

www.elsevier.com/locate/jneumeth

Detection of onset of neuronal activity by allowing for heterogeneity in the change points

Yaacov Ritov [a,c,*], A. Raz [b,c], H. Bergman [b,c]

[a] Department of Statistics, The Hebrew University of Jerusalem, 91905 Jerusalem, Israel
[b] Department of Physiology, The Hebrew University of Jerusalem, PO Box 12272, 91220 Jerusalem, Israel
[c] The Center for Neural Computation, The Hebrew University of Jerusalem, Jerusalem, Israel

Received 8 May 2002; received in revised form 27 August 2002; accepted 28 August 2002

Abstract

We consider situations in which there is a change point in the activity of a cell, that is, some time after an external event the firing rate of the cell changes. The change can occur after a random delay. The distribution of the time to change is considered unknown. Formally we deal with n random point processes, each of these is an inhomogeneous Poisson process, with one intensity until a random time, and a different intensity thereafter. Thus, the change point is not explicitly observed. We present both a simple estimator and the non-parametric maximum likelihood estimator (NPMLE) of the change point distribution, both having the same rate of convergence. This rate is proved to be the best possible. The extension of the basic model to multiple processes per trial with different intensities and joint multiple change points is demonstrated using both simulated and neural data. We show that for realistic spike train data, trial by trial estimation of a change point may be misleading, while the distribution of the change point distribution can be well estimated.

Keywords: Peri-stimulus time histogram; Neural data; Semiparametric models; Hidden Markov models; Mixture models; Empirical Bayes

1. Introduction

The synchrony between neural activity and external event is a major tool in neurophysiological studies of the brain. The data is typically analyzed using the Peri-stimulus time histogram (PSTH). Different trials are aligned with respect to the time of the external event and the average change over many trials of the intensity of the neural activity is observed. However, using this technique, one cannot distinguish between a smooth transition in any single trial between two regimes on one hand, and a sharp transition at each trial, but with a jitter in the transition times between the trials on the other.

Technically, we consider a situation in which copies of a multivariate point process on a fix interval are observed. The intensity of the process is not fixed along the interval but changes once or more. The time of the change points may vary from copy to copy.

In this manuscript the problem is treated in two different ways: both as a formal statistical problem and as a tool for the analysis of neural data. The biological question behind the statistical discussion is the extent to which the activity of a specific group of neurons in the monkey's brain is synchronized with the external behavior of the animal. Here the model is extended to more than one change point and to multivariate counting processes. The estimation procedure assumes that the change points actually exist, and that the different components of the multivariate process are, given the change points times, independent inhomogeneous Poisson processes. These assumptions are not necessarily valid in all situations, but we argue that they are reasonable for our examples. The different extensions of the model are applied to real and simulated data.

* Corresponding author. Tel.: +972-2-652-9929; fax: +972-2-652-9996
E-mail address: yaacov@mscc.huji.ac.il (Y. Ritov).

0165-0270/02/$ - see front matter
PII: S0165-0270(02)00275-3

Mathematically, we analyze a problem in which it is assumed that there is only one change point per trial, whose time is a random variable distributed according to a distribution function G. For each trial, we assume an inhomogeneous Poisson process that has a constant intensity λ_0 until the change point, and a constant intensity λ_1 thereafter. The actual time of the change point is not observed explicitly. The parameters G, λ_0 and λ_1 are not known. The information bounds and the efficient score functions are given. In a nutshell, we have an explicit expression for the score function only in a relatively trivial case. The maximum likelihood estimator as well as simple estimators are presented. In particular, the distribution function of the time to change can be estimated using a simple monotone regression estimator. The rates of convergence of these simple estimators are optimal.

The change point model for a single Poisson process (and a single trial) was discussed in a few earlier papers, e.g. Matthews et al. (1985), Akman and Raftery (1986) and the standard Bayesian analysis is discussed in Raftery and Akman (1986). The change point methodology was discussed in the context of neuron activity by Commenges and Seal (1985), where the change point was estimated for each trial separately, which may be difficult in some applications. A typical firing rate for a neuron is a spike every 20–200 ms on the average. Hence there is an error of a few hundred milliseconds in the estimation of a single change point. This is too crude for a typical behavioral task. In our simulation, we present an example, where the change point distribution can be estimated, reasonably well, while it is almost impossible to locate the individual change points. Note however, that the brain system observes many cells at the same time, and, therefore, can detect the change point exactly, even when it is not possible in the experimental setting where only one or at most a few cells can be observed simultaneously.

Our point of view is akin to hierarchical Bayes or, closer, to the empirical Bayes formulation of the problem. Previously, empirical Bayes models were employed by Joseph and Wolfson (1992) and Bélisle et al. (1998) in the context of change point detection for spike data. See a relevant recent discussion of empirical Bayes procedures in Efron (1996). Leaving philosophy aside, we consider the problem as a semiparametric mixture model, Bickel et al. (1993) and Robins and Ritov (1997). The typical neuronal experiment in which the activity of single cells is recorded involves a repeated task in which the animal is reacting to external cues. The experimenter tries to understand how the cells activity is related to different sensory and motor events. The distribution of the change point time may be interesting in particular in situations where it is not known a priori with which external event the neuronal activity is synchronized. In a typical experiment, an observed change in the neurons activity may be related to the visual cue that the monkey receives, to the eye movement that follows, or to the movement of his arm. The times of these events are recorded, and we may try to test to which of them the activity is better synchronized (Seal et al., 1983; Seal and Commenges, 1985; Schwartz et al., 1988; Montgomery, 1989; Crutcher and Alexander, 1990; Romo and Schultz, 1990). In this paper, a single change point was located for each neuron and task condition using a formal hierarchical Bayesian method. Our method was applied to other data set as presented in Ritov et al. (1997). The model analyzed by Bélisle et al. (1998) is similar to ours, except that it was analyzed using Bayesian tools, both in the model formulation and in the algorithms, they used Gibbs sampler, while we used a non-iterative simple estimator and the EM algorithm to calculate the maximum-likelihood estimator. Moreover, we extend their model to examples of multiple change point and multiple cells. Finally, we give the theoretical justification to the technique used.

2. Methods

Our empirical data were recorded from two awake vervet (green) monkeys (*Cercopithecus aethiops aethiops*). The monkeys were trained to perform a visual-motor task with two behavioral paradigms, see details in Raz et al. (2000). Briefly, the trials were as follows. Four seconds after the end of the previous trial the program started checking if the central key is touched. In most cases, the monkey would have touched the key during the inter-trial period. If not, the program waited until the key has been touched. Immediately (less than 1 ms) after it touched the key, the 'get ready' LED was turned on. After a variable delay (3–6 s), one of the two peripheral target keys was illuminated for 0.25 s, and the monkey got a trigger signal after another random delay of 1, 2, 4, or 8 s. At this point the monkey was supposed either to release the central key and touch the target key (the 'GO' paradigm), or to keep touching the central key (the 'NO-GO' paradigm). If the monkey did this, it was rewarded with 0.15 ml of juice. After four correct trials, there was a 4 s signal instructing the monkey to change paradigm from 'GO' to 'NO-GO' or vice versa. The monkey was fully trained before recording started. In each recording session, the activity of two to eight single cells in the basal ganglia was recorded. In a single recording session, a few hundred trials were recorded. Typically the number of valid records from any single cell is between a few tens to a few hundred trials.

The cells whose activity we analyze are from the external segment of the globus pallidus (GPe). These cells are characterized with a fast tonic rate (a few tens

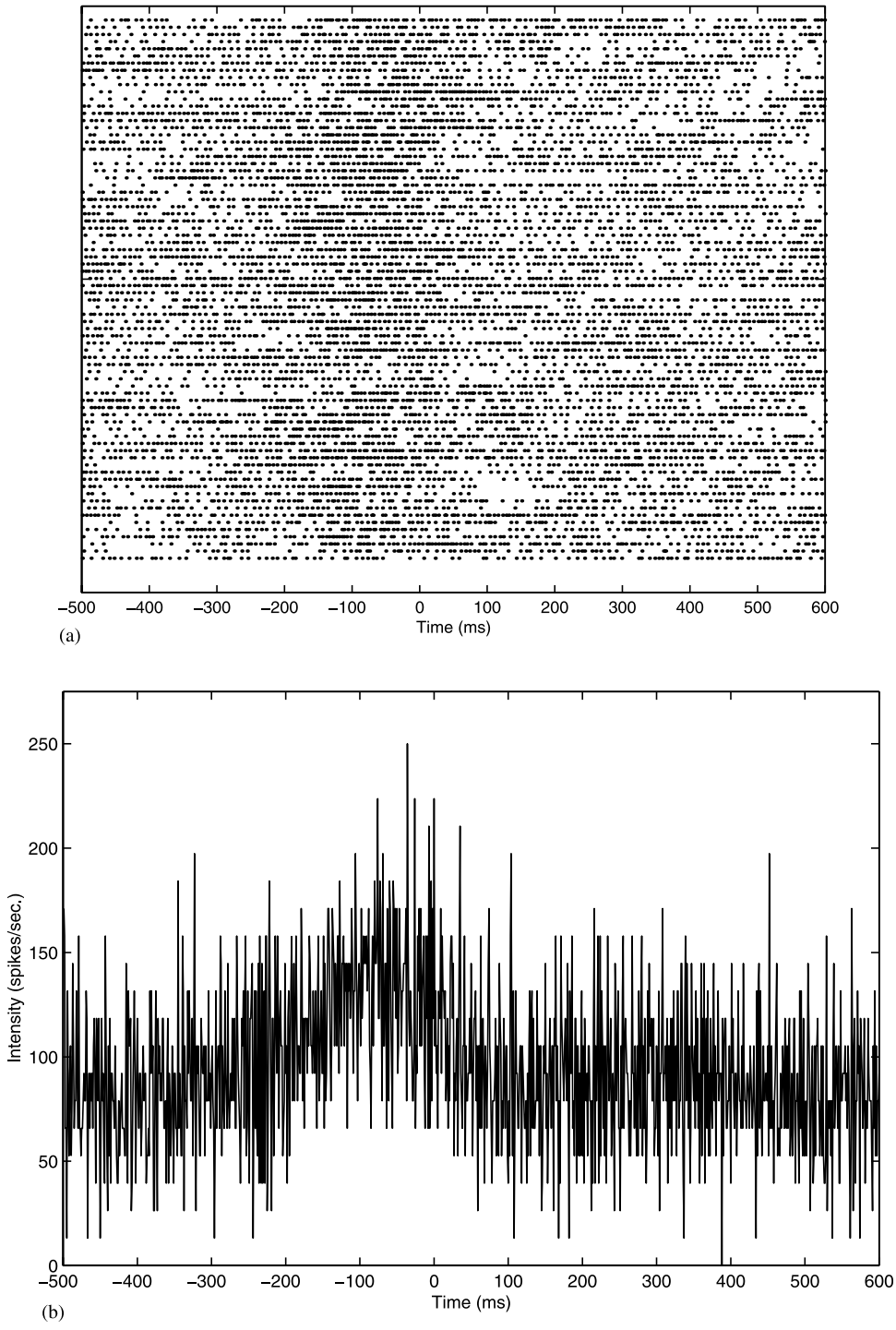


Fig. 1. The spiking activity of a GPe cell. The monkey released the central key at time 0. (a) The raster plot; (b) the PSTH; (c) the PSTH smoothed and a monotone regression estimate of intensity.

of spikes per s), and with unexplained short intervals in which they are silent (DeLong, 1971; DeLong and Georgopoulos, 1981). It looks as if the cells behave independently, Nini et al. (1995) and Raz et al. (2000). The exact function and the mode of activity of the basal ganglia are not known. We assume that a model in which some cells switch abruptly to a different mode is reasonable.

3. Results

3.1. The model

The data used for the analysis can be summarized as follows. We measure the activity of $K \geq 1$ cells during n trials. For each trial we record the activity of each of the

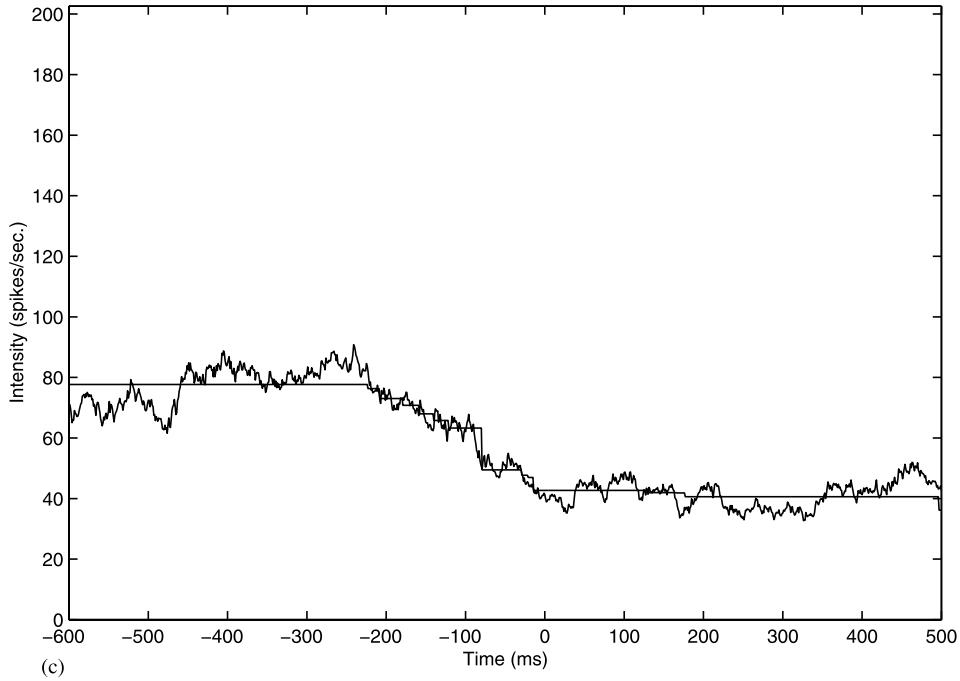


Fig. 1 (*Continued*)

cells during a window synchronized on a given activity of the monkey during the trial.

Formally, the observations are at discrete time, a, $a+1, \ldots, b$ for some a and b. We assume that for each trial i, $i=1, \ldots, n$, there are multiple change points $a < T_{i1} < \ldots T_{iM} < b$, for some $M \geq 1$. We observe for each trial K counting processes. The processes are independent homogeneous Poisson processes between the common change points. In other words, for all $i=1, \ldots, n$, $N_{i1}, \ldots, N_{iK}$ are independent given $T_{i1}, \ldots, T_{iM}$. The values $N_{ik}(t)$, $i=1, \ldots, n$, $k=1, \ldots, K$, $t=a, \ldots, b$, represent the total number of spikes fired by the k-th cell during the time interval from a to t: $N_{ik}(t) = \Sigma_{s=a}^{t} N_{ik}(s)$. We assume that the Bernoulli random variables $N_{ik}(a)$, $\ldots$, $N_{ik}(b)$ are independent given $T_{i1}, \ldots, T_{iM}$, and $P(N_{ik}(t)=1|T_{i1}, \ldots, T_{iM}) = 1-P(N_{ik}(t)=0|T_{i1}, \ldots, T_{iM}) = P_{km}$, $T_{im} \leq t < T_{i,m+1}$, $m=0, \ldots, M$, where, formally, $T_{i0} \equiv a$ and $T_{i,M+1} \equiv b+1$. In other words, the data are a collection of independent Bernoulli random variables, whose probability of success depend on the cell and the random time interval to which they belong. The latter is defined in terms of the change point times. The Bernoulli model is typically not valid, as the cells have refractory periods, however, it can be a valid approximation if the refractory period is much shorter than the mean inter spikes time.

We need to restrict the structure of the joint distribution of the change point times, because of statistical and computational considerations. We considered two alternative assumptions:

Assumption 1. $T_{i1}, \ldots, T_{iM}$ are independent with distribution functions $G_1 \ldots, G_M$, respectively. In particular, the supports of these distributions are mutually exclusive.

Or

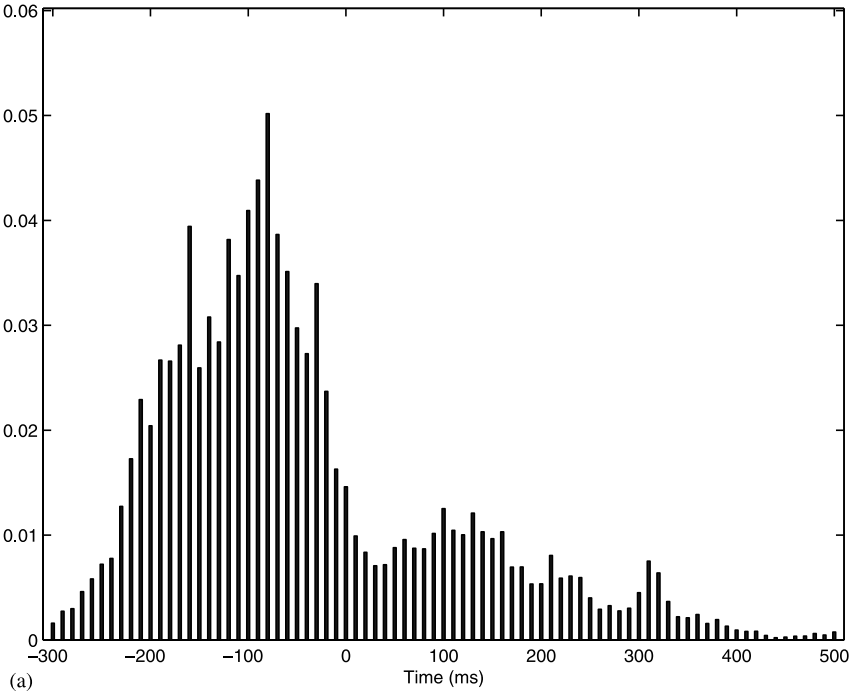
Assumption 2. $T_{i1}, T_{i2}-T_{i1}, \ldots, T_{iM}-T_{i,M-1}$ are independent with distribution functions $G_1 \ldots, G_M$, respectively.

Assumption 1 describes a situation in which all change points are relative to the synchronizing event, while Assumption 2, for $M=2$, describes a situation in which the cell reacts to the external event at time T_{i1} for a duration of $T_{i2}-T_{i1}$.

Practically, the algorithm for the second case was useful only for $M=2$: a random length first period, then an intermediate interval with a random length, and thereafter a final period. The algorithm can be too slow for any larger M.

The likelihood function to be maximized under Assumption 1 is:

$$L(\{g_m(t)\}_{m=1,\ldots,M,t=a,\ldots,b}, \{p_{km}\}_{k=1,\ldots,K,m=1,\ldots,M}) = \prod_{i=1}^{n} \sum_{a<t_1<\cdots<t_M<b} \prod_{m}^{M} g_m(t_m) \prod_{k=1}^{K}$$



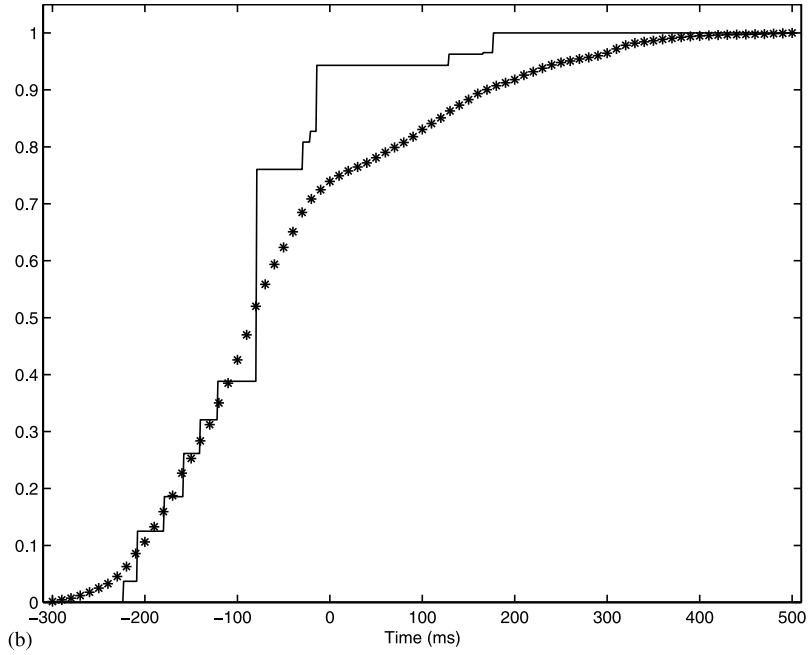


Fig. 2. Change point distribution of the single cell described in Fig. 1. (a) the MLE p.d.f.; (b) the MLE c.d.f. (stars) compared with the c.d.f. as estimated by the monotone regression of the PSTH (solid line).

$$\times \prod_{m}^{M+1} p_{km}^{\sum_{t=t_{m-1}}^{t_m} N_{ik}(t)} (1-p_{km})^{t_m - t_{m-1} + 1 - \sum_{t=t_{m-1}}^{t_m} N_{ik}(t)}$$

where $t_0 = a$ and $t_{M+1} = b$, and $g_m(t)$ is the point mass at t of the m-th distribution.

It is similar under Assumption 2:

$$L(\{g_m(t)\}_{m=1,\ldots,M,t=a,\ldots,b}, \{p_{km}\}_{k=1,\ldots,K,m=1,\ldots,M})$$

$$= \prod_{i=1}^{n} \sum_{a<t_1<\cdots<t_M<b} \prod_{m}^{M} g_m(t_m - t_{m-1}) \prod_{k=1}^{K}$$

$$\times \prod_{m}^{M+1} p_{km}^{\sum_{t=t_{m-1}}^{t_m} N_{ik}(t)} (1-p_{km})^{t_m - t_{m-1} + 1 - \sum_{t=t_{m-1}}^{t_m} N_{ik}(t)}$$

where, $t_0 = a$ and $t_{M+1} = b$.

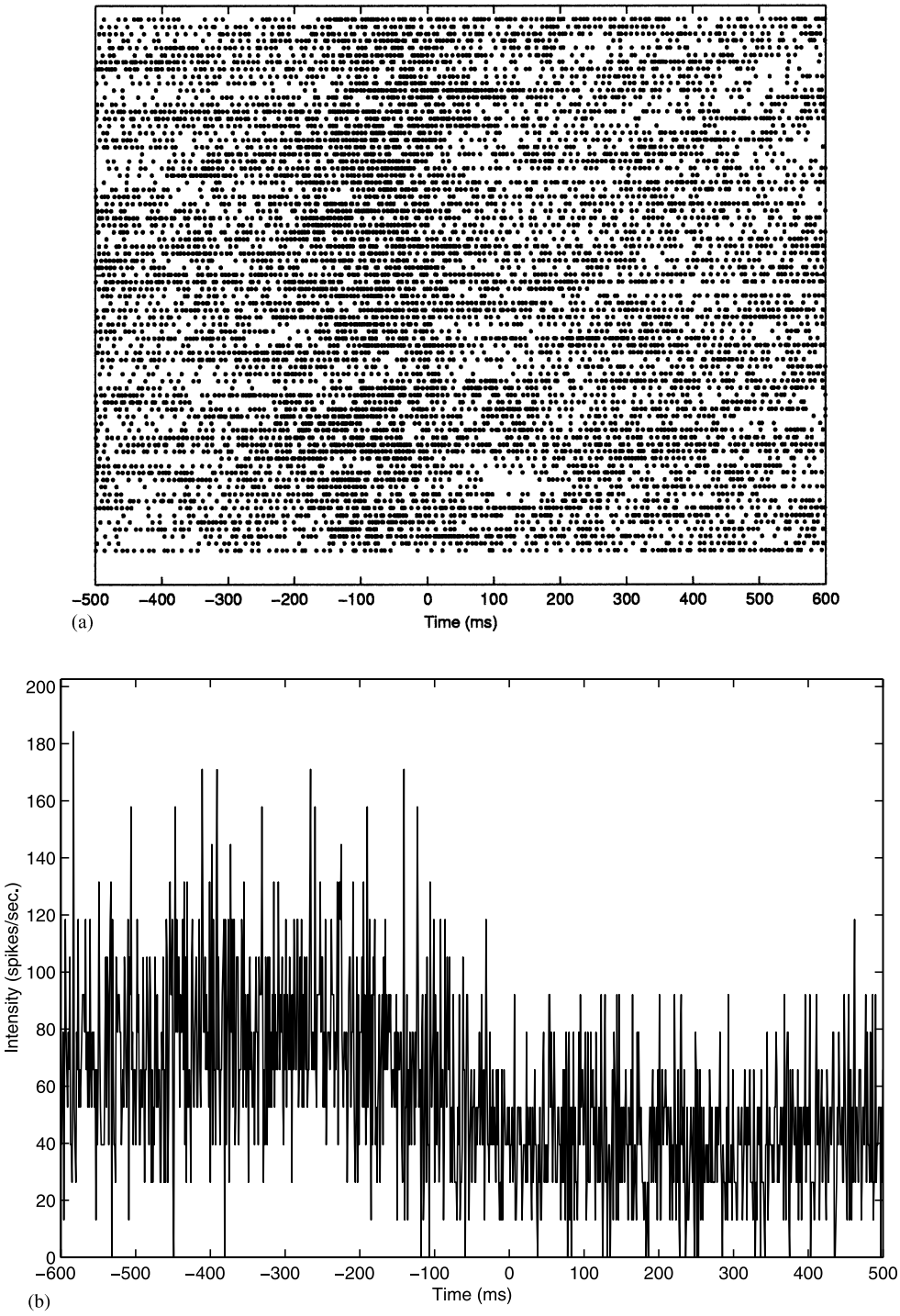


Fig. 3. One GPe cell with seemingly two change points. (a) Raster plot; (b) PSTH.

3.2. The algorithms

Suppose the change point times, T_{ij}, $i = 1, \ldots, n, j = 1, \ldots, M$, were observed. Then, the distribution of the time to the change points could be estimated easily by the empirical distribution of the corresponding variables, and the probabilities, $p_1, \ldots, p_M$ could be estimated by the corresponding means in the sample. This makes the model a typical missing data model. Note that by missing we do not necessarily mean that data was lost. It may that the model can be derived as a simpler model in which some of the variables are unobserved. A standard method to maximize the likelihood in models with missing data is the EM algorithm, Dempster et al.

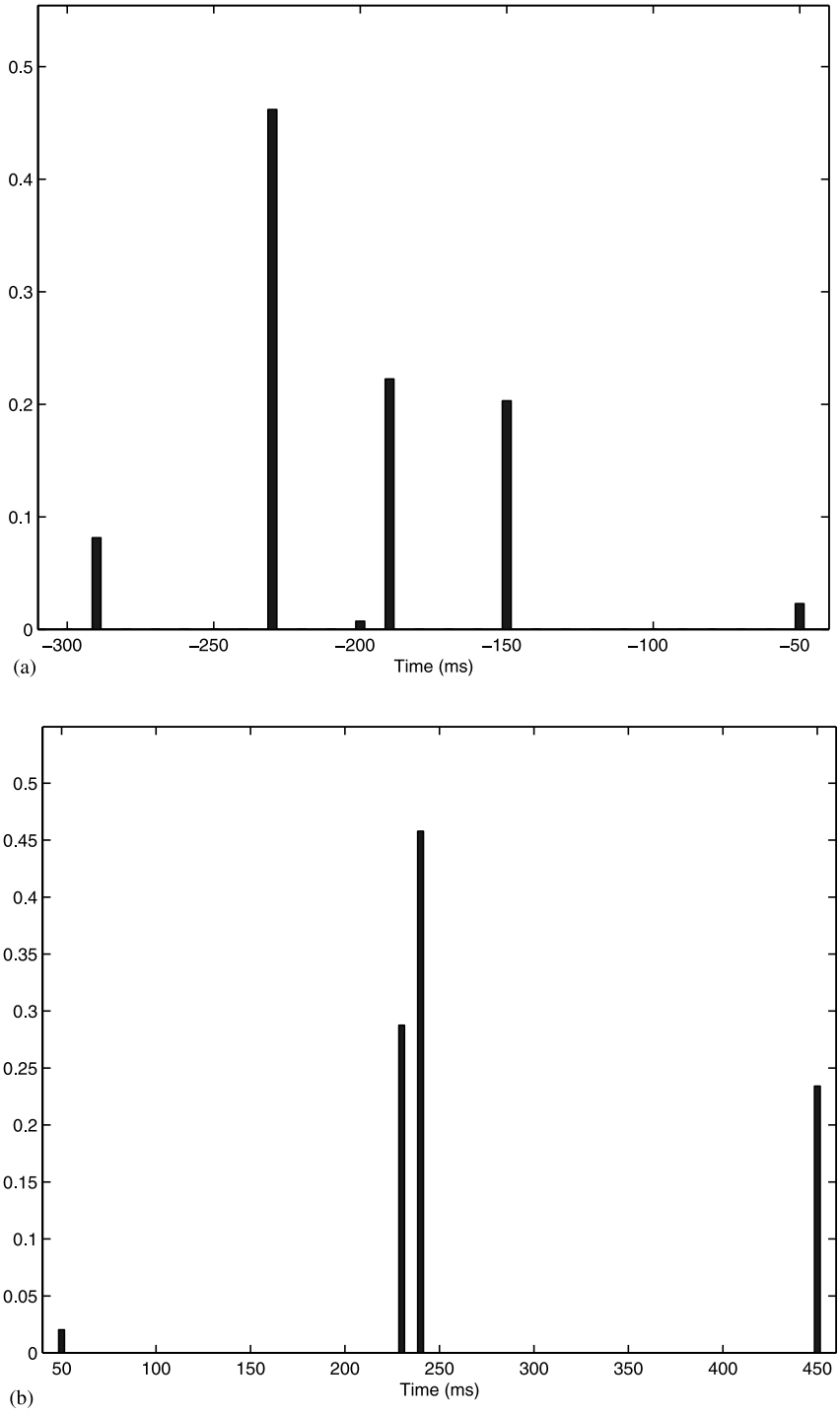


Fig. 4. One GPe cell with two dependent change points. (a) The estimated distribution of the first change point (0 is the RELEASE time); (b) estimated distribution of the time between the two change points.

(1977). Generally speaking, when the EM algorithm is used, it is assumed that besides the observed data there are unobserved data, and we iterate between computing expectation of the log-likelihood of the complete data over the conditional distribution of the unobserved random variables giving the observed ones (the E-steps), and maximizing this expectation (the M-steps).

We used two versions of the EM algorithm for computing the (approximate) maximum likelihood estimators (MLE) of the different parameters. The following notation is used. Hat above a parameter denotes an estimator. The distributions are approximated by discrete distributions have r_m, $m = 1, \ldots, M$ support points. Note that both M and the r_m's are prescribed

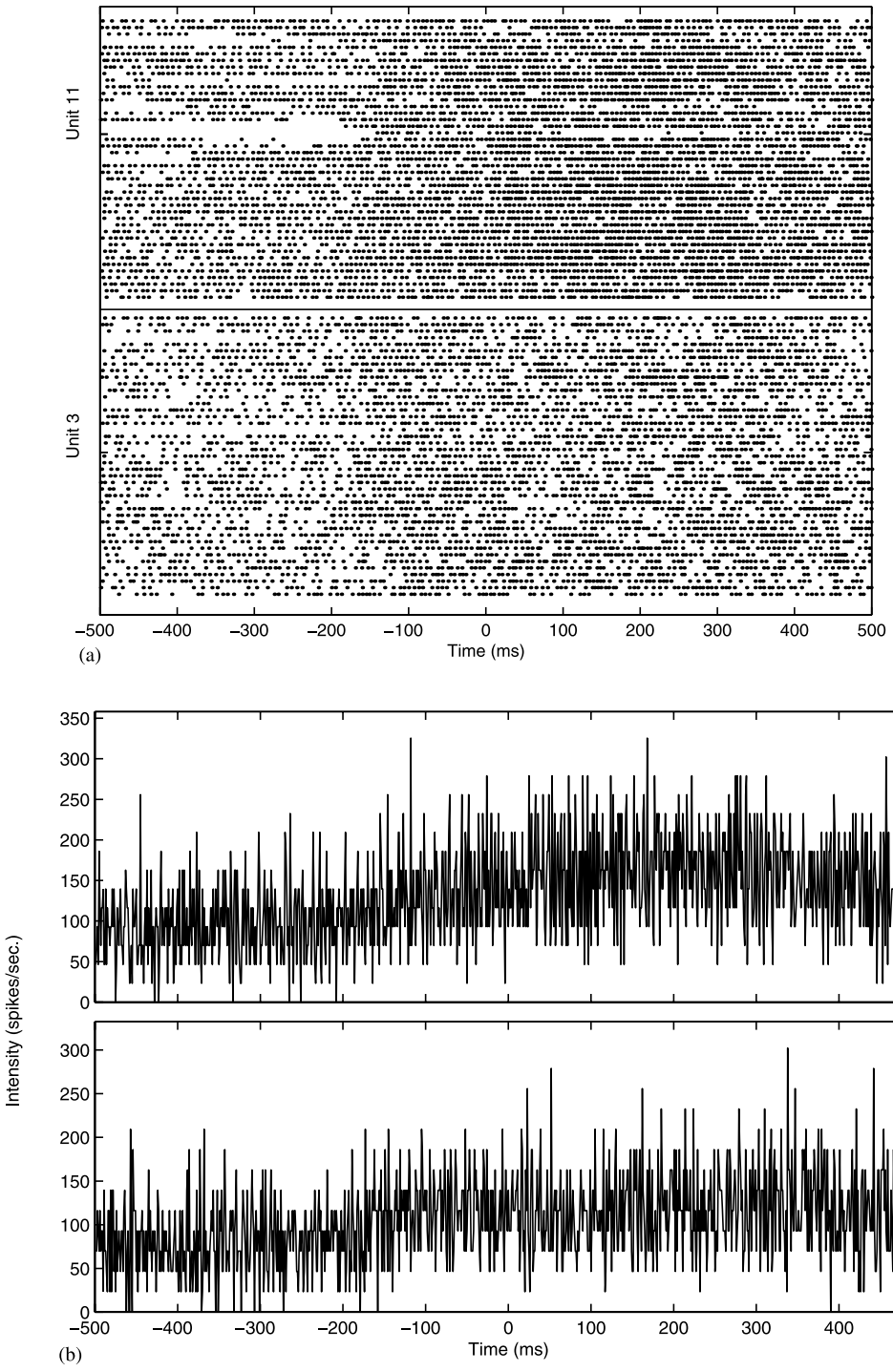


Fig. 5. Two GPe cells. (a) Raster plot; (b) the PSTH's for the two cells; (c) smoothed PSTH's.

by the user. The discrete distributions of the change point times are denoted by G, and their probability functions are denoted by g. The EM algorithm for multiple independent change points was as follows:

Algorithm 1.

1) **Initial step:** Set $l=0$, $\hat{g}_m^{(l)}(j)=1/r_m$, $j=1, \ldots, r_m$, $m=1, \ldots, M$, and $\hat{p}_{km}^{(l)}=(n(b-a+1))^{-1}\Sigma_{i=1}^{n} N_{ik}(b)$, $k=1, \ldots, K$, $m=0, \ldots, M$. Let $z_m(1), \ldots, z_m(r_m)$ be the support point of the distribution of the m-th change point, $m=1, \ldots, M$.
2) **E-step:** Compute the likelihood function that the changes of the i-th trial happened at $j_i, \ldots j_M$:

$$L_i(j_1, \ldots, j_M)$$

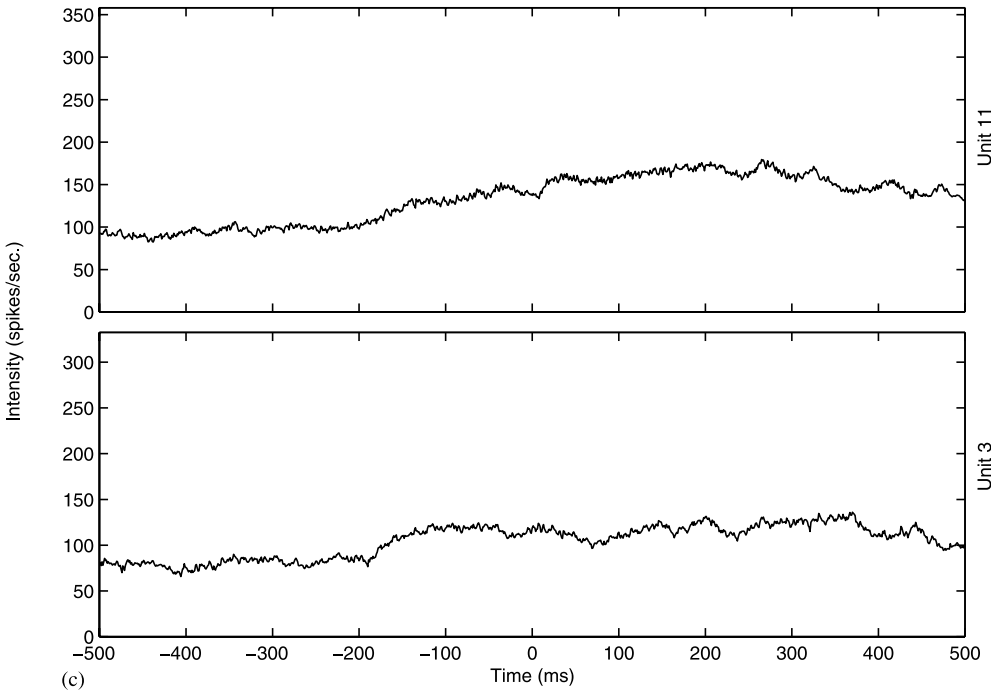


Fig. 5 (*Continued*)

$$= \prod_{k=1}^{K} \prod_{m=1}^{M} \left(\frac{\hat{p}^{(l)}_{k,m-1}(1-\hat{p}^{(l)}_{km})}{(1-\hat{p}^{(l)}_{k,m-1})\hat{p}^{(l)}_{km}} \right)^{N_{ik}(z_m(j_m))} \times \left(\frac{1-\hat{p}^{(l)}_{k,m-1}}{1-\hat{p}^{(l)}_{km}} \right)^{z_m(j_m)},$$

for $i=1, \ldots, n$, $1 \le j_m \le r_m$, $m=1, \ldots, M$. Compute the a posteriori probabilities for the vector of the i-th change point times:

$$P_i(j_1, \ldots, j_m) = \frac{L_i(j_i, \ldots, j_M) \prod_{m=1}^{M} \hat{g}^{(l)}_m(j_m)}{\sum_{j'_1=1}^{r_1} \cdots \sum_{j'_M}^{r_M} L_i(j'_1, \ldots, j'_M) \prod_{m=1}^{M} \hat{g}^{(l)}_m(j'_m)}$$

3) **M step:** Set $l=l+1$. Update $\hat{g}^{(l)}_1, \ldots, \hat{g}^{(l)}_M$ to be the marginal distributions of $n^{-1} \sum_{i=1}^{n} P(\cdot)$. Update for $k=1, \ldots, K$ and $m=0, \ldots, M$:

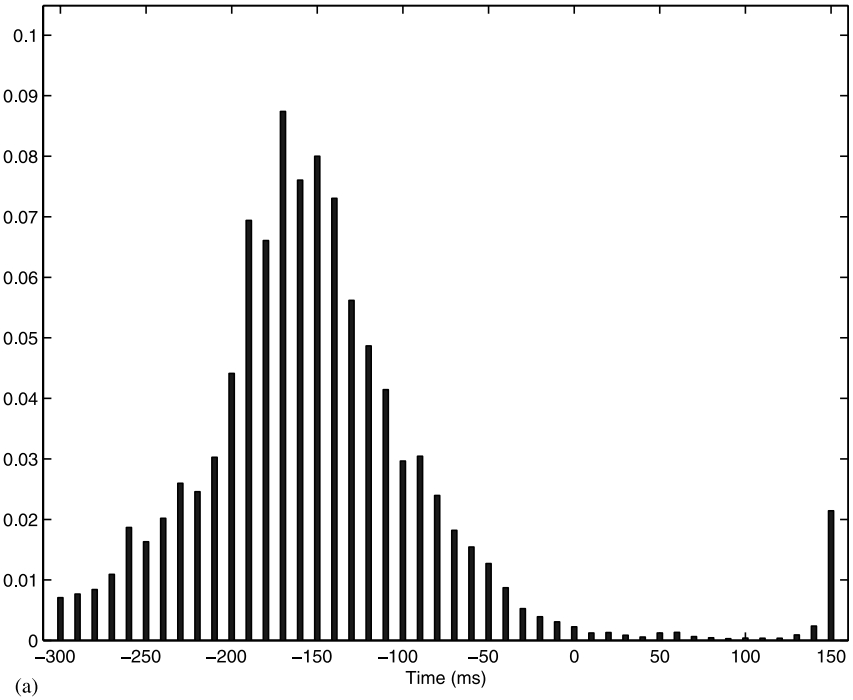
$$\hat{p}^{(l)}_{km} = \frac{\sum_{i=1}^{n} \sum_{j_1=1}^{r_1} \cdots \sum_{j_M=1}^{r_M} P_i(j_1, \ldots, j_M)(\mathbb{N}_{ik}(z_{m+1}(j_{m+1})) - \mathbb{N}_{ik}(z_m(j_m)))}{\sum_{j=1}^{n} \sum_{j_1=1}^{r_1} \cdots \sum_{j_M=1}^{r_M} P_i(j_1, \ldots, j_M)(z_{m+1}(j_{m+1}) - z_m(j_m))} \tag{3.1}$$

4) **Convergence check:** Stop if the number of iterations exceeds the pre-decided tolerable number or the convergence criterion error below was less than the pre-decided value. Otherwise return to the E-step.

The algorithm for dependent change points (or independent start and duration of an intermediate period) was as follows:

Algorithm 2.

1) **Initial step:** Like the initial step of Algorithm 1 with $M=2$.
2) **E-step:** Let $M=2$. Let $z_1(1), \ldots, z_1(r_1)$ be the support of the distribution of the first change point, and let $t_2(1), \ldots, t_2(r_2)$ be the support of the distribution of the time between the two change points. Compute the likelihood function for the i-th trial:



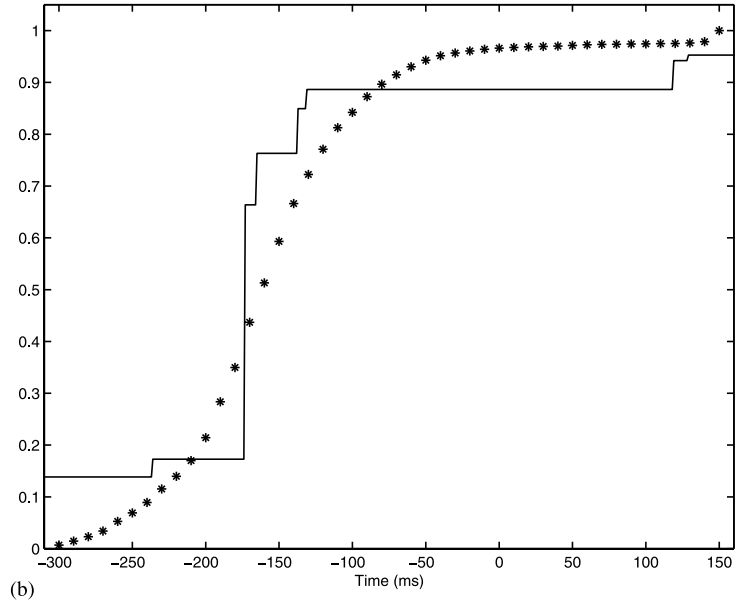


Fig. 6. Two GPe cells with seemingly one change point. (a) MLE of p.d.f. of the change point; (b) cell 9, MLE of the c.d.f., and the estimated based on the monotone regression; (c) cell 13, MLE of the c.d.f., and the estimated based on the monotone regression.

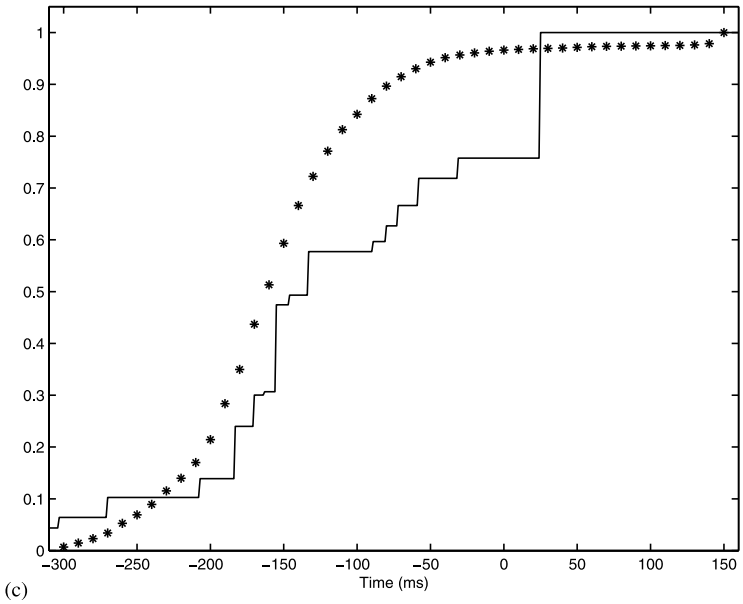
$$L_i(j_1, j_2) = \left(\frac{\hat{p}_{k0}^{(l)}(1-\hat{p}_{k1}^{(l)})}{(1-\hat{p}_{k0}^{(l)})\hat{p}_{k1}^{(l)}}\right)^{N_{ik}(z_1(j_1))} \times \left(\frac{\hat{p}_{k1}^{(l)}(1-\hat{p}_{k2}^{(l)})}{(1-\hat{p}_{k1}^{(l)})\hat{p}_{k2}^{(l)}}\right)^{N_{i,k}(z_1(j_1)+z_2(j_2))} \times \left(\frac{1-\hat{p}_{k0}^{(l)}}{1-\hat{p}_{k1}^{(l)}}\right)^{z_1(j_1)} \left(\frac{1-\hat{p}_{k1}^{(l)}}{1-\hat{p}_{k2}^{(l)}}\right)^{z_1(j_1)+z_2(j_2)}$$

Define P_i as in Eq. (3.1).

3) **M-step:** Set $l = l+1$. Update $\hat{g}_1^{(l)}$, $\hat{g}_2^{(l)}$ to be the two marginals of $n^{-1}\ \Sigma_{i=1}^{n}\ P()$. Update for $k = 1, \ldots, K$;

$$\hat{p}_{k0}^{(l)} = \frac{\sum_{i=1}^{n}\ \sum_{j_1=1}^{r_1}\ \sum_{j_2=1}^{r_2}\ P_i(j_i,\ j_1)\mathbb{N}_{ik}(z_1(j_1))}{\sum_{i=1}^{n}\ \sum_{j_1=1}^{r_1}\ \sum_{j_2=1}^{r_2}\ P_i(j_i,\ j_1)(z_1(j_1)-a)}$$

$$\hat{p}_{k1}^{(l)} = \frac{\sum_{i=1}^{n}\ \sum_{j_1=1}^{r_1}\ \sum_{j_2=1}^{r_2}\ P_i(j_i,\ j_1)(\mathbb{N}_{ik}(z_1(j_1)+z_2(j_1)) - \mathbb{N}_{ik}(z_1(j_1)))}{\sum_{i=1}^{n}\ \sum_{j_1=1}^{r_1}\ \sum_{j_2=1}^{r_2}\ P_i(j_i,\ j_1)z_2(j_2)}$$



(c)

Fig. 6 (*Continued*)

$$\hat{p}_{k2}^{(l)} = \frac{\sum_{i=1}^{n}\sum_{j_1=1}^{r_1}\sum_{j_2=1}^{r_2} P_i(j_i, j_1)(\mathbb{N}_{ik}(b) - \mathbb{N}_{ik}(z_1(j_1) + z_2(j_2)))}{\sum_{i=1}^{n}\sum_{j_1=1}^{r_1}\sum_{j_2=1}^{r_2} P_i(j_i, j_1)(b - z_1(j_1) - z_2(j_2))}$$

4) **Convergence check:** Stop if the number of iterations exceeds the pre-decided tolerable number or the convergence criterion error below was less than the pre-decided value. Other wise return to the E-step.

In general, the EM algorithm may be very slow. In our simulation it was reasonably fast for the single change points examples. It took around 1 min on a 133 MHz PC. It was quite slow for some of our extensions where there were more than one change point.

The stopping time was defined as the minimum between $l = 500$ and the first l such that the first time the difference between the estimates in two consecutive iterations is measured by:

$$\frac{\sum_{m=1}^{M}\sum_{k=1}^{K}|\hat{p}_{km}^{(l)} - \hat{p}_{km}^{(l-1)}|}{\sum_{m=1}^{M}\sum_{k=1}^{K} p_{km}} + \sum_{j=1}^{r}|\hat{g}_j^{(l)} - g_j^{(l-1)}| < 4 \times 10^{-6}, \quad (3.2)$$

where $\hat{p}_{km}^{(l)}$ and $\hat{g}_j^{(l)}$ are the estimators after l cycles of the algorithm.

In the first example the algorithm converged after ten iterations. It converges after nine in the case of the third example. On the other hand, it stopped in the second example after 500 iterations, when the first term of error was equal to 2.5×10^{-7} and the second term was equal to 1.4×10^{-4}.

3.3. Examples

3.3.1. One cell and one change point

The first record we discuss is of a GPe cell. We consider the interval starting 600 ms before the time the monkey released the central key (RELEASE) and ending 500 ms after this event. The raster plot is given in Fig. 1a, where each horizontal line of dots represents a single trial, and each dot denotes a spike at the trial and time relative to the synchronizing event, as given by its coordinates. The same data is summarized in Fig. 1b by the Peri stimulus time histogram (PSTH). Here we plot the total number of spikes (over all the trials) in each 1 ms interval, relative to the RELEASE time. The vertical axis is scaled to denote the intensity (in spikes per s). The average intensity is 57.8, or, on the average, a spike every 17.3 ms. The PSTH is smoothed in Fig. 1c with a Gaussian kernel with bandwidth of 4 ms. On the same graph, the PSTH is smoothed also by a monotone regression estimator. We can observe from these figures that the intensity in decreasing. The transition from high to low average intensity, as can be observed from the PSTH, is smooth. A more detailed observation of the raster plot shows that in each trial the transition between the two periods is quite abrupt, but the change time varies from trial to trial. One possible interpretation of Fig. 1b or c, is that a change is happening before or at time −200 ms, and there after intensity is decreasing during approximately 200 ms before it stabilized again. However, this interpretation of the PSTH is wrong in view of the raster plot. Our analysis will model this exactly. In this example and in the other two examples, there is no need for a sophisticated test to verify that the intensity is not constant. Formally, we

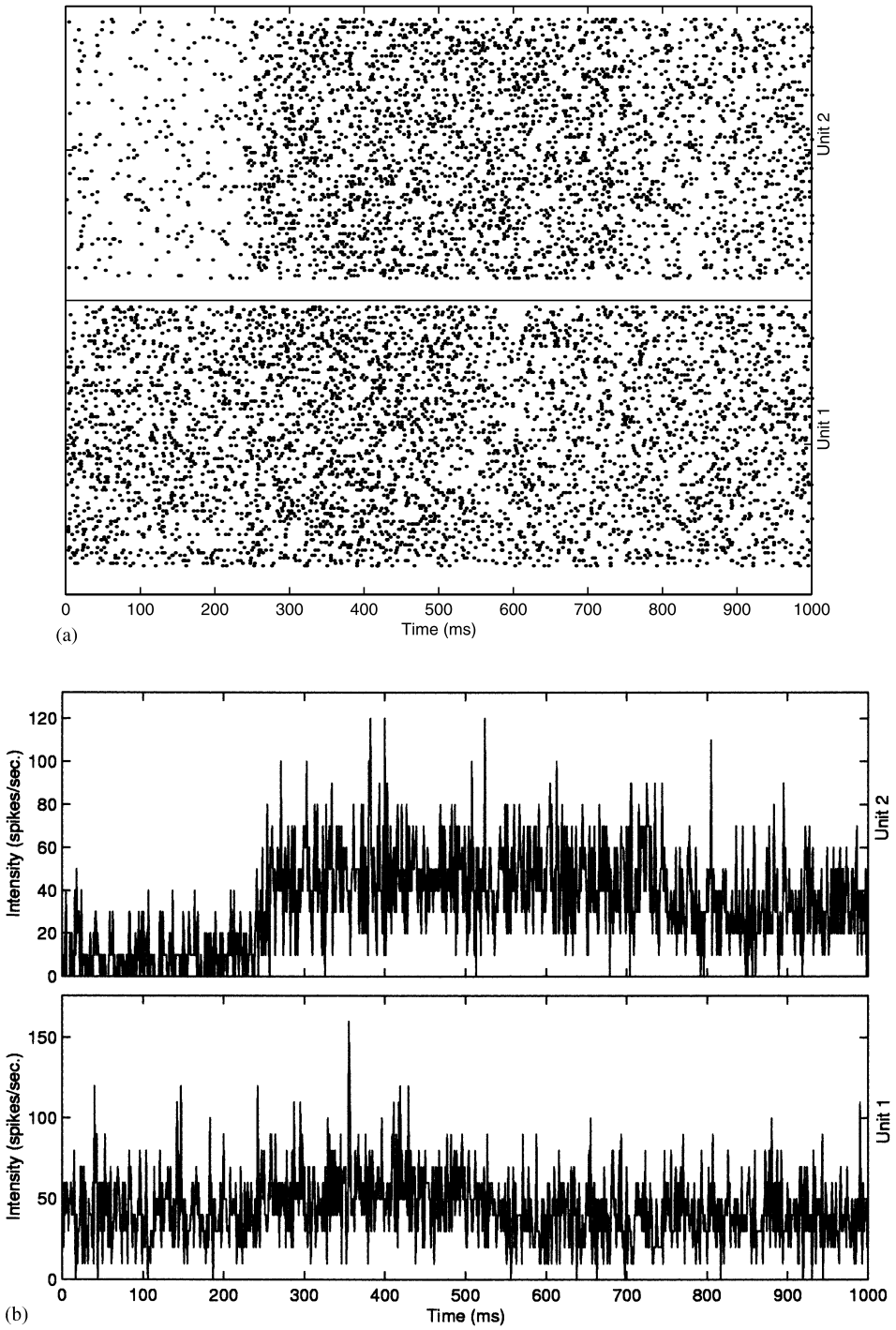


Fig. 7. Simulated data: (a) the Raster plot; (b) PSTH; (c) smoothed PSTH.

considered a t-test for comparing the total count in the first 550 ms to the total count of the second half. The t-statistic has a value of 14.3 ($P < 0.001$).

The estimated p.d.f. and c.d.f. of the change point distribution are shown in Fig. 2. It can be observed that the lower intensity period starts in most trials before the actual release, but the actual time varies between trial to trial.

3.3.2. *Two change points*

We consider now a second GPe cell from the same recording session as the cell analyzed above. The raster plot and the PSTH are given in Fig. 3. As can clearly be seen from the PSTH, a simple change point model cannot fit the data. However, we can try to fit a model with two change points. The fact that there is an intermediate period with higher intensity can be verified

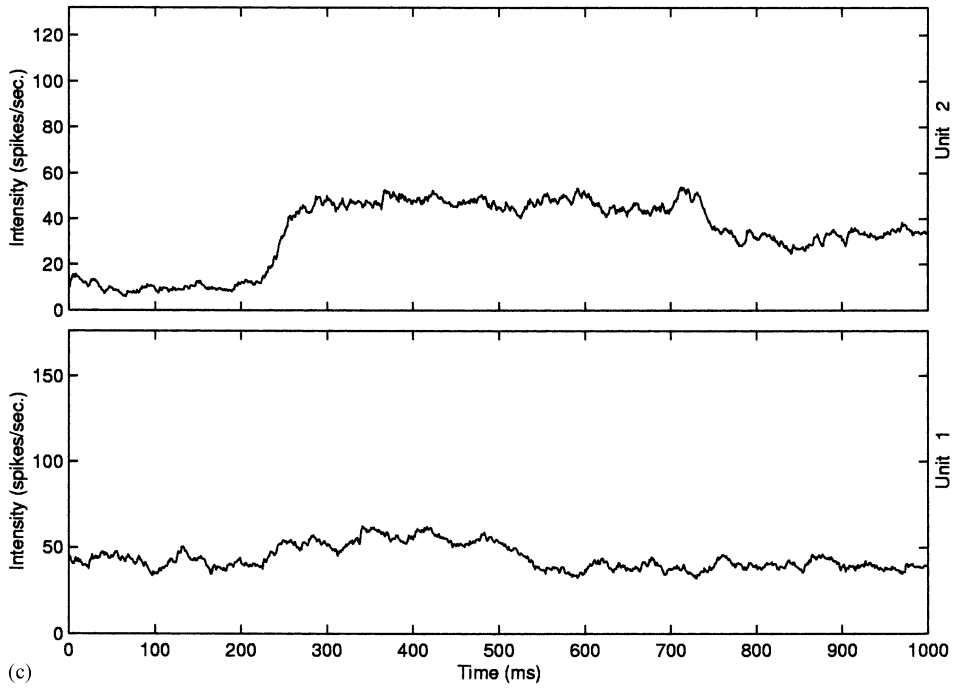


Fig. 7 (*Continued*)

simply by considering the interval of length 1200 ms around RELEASE. We divided the interval into three equal parts and counted the number of spikes in each sub-interval. The P-value of the t-test that compare the two extreme sub-intervals is 0.7, while the t-test that compares the first and second sub-interval has an apparent P-value of 10^{-8}.

A model in which the width of the interval is independent of its initial time was fitted to the data. That is, we assumed that there are i.i.d. pairs (T_{i1},T_{i2}), $i=1,\ldots,n$, such that T_{i1} and $T_{i2}-T_{i1}$ are independent, and the intensity of the process is λ_1, λ_2 and λ_3, for $t \leq T_{i1}$, $t_{i1} \leq T_{i2}$, and $t > T_{i2}$, respectively. The support of the distribution was fitted by eye, to be as wide as possible. Thus the support of the first change point was in the range of -300 to -50 ms, while the width of the second period was restricted to be in the range of 50–450 ms. The results obtained from applying Algorithm 2 to these data are given in Fig. 4. Note that the p.d.f. is 0 at most points. The first change point is mainly distributed between 250 and 150 ms before RELEASE. The width of the high intensity period was found to be mostly around 240 ms, but with probability of approximately 0.25 it got the maximal value that was permitted, as if some in a quarter of the trial the second change point is missing. Note that the probability mass assigned by the estimator is quite negligible on most of the permitted support points. In fact, only in six out of the 26 support points of the first distribution, and in four out of the 41 support points of the second distribution, the algorithm assigned a probability larger than 0.001.

To check the reliability of the estimation procedure, we introduced a 100 ms jitter. That is, each spike train was shifted by a random time distributed uniformly between -50 and 50 ms. The spread of the distribution of the first change point was increased (although, less than could be expected), while the distribution of the second change point remained almost the same as expected (since the time between the two change points was not expected to change by the random shift). One can judge from the shape of the estimated distribution and the effect of introducing the jitter, that either the change point model is not appropriate to this cell, or a much larger sample is needed for a stable estimator.

3.3.3. *Two cells with one change point*

We consider now a record of two GPe cells of the second monkey. (This monkey was trained for somewhat simplified experiment with only the 'GO' paradigm). We observe two cells around the RELEASE time. The data is exhibited in Fig. 5. There are 44 valid trials.

It seems that the two cells behave similarly. Clearly, the processes are not homogeneous. Formally, we calculated a t-statistic that compared the number of spikes in the first half with the number in the second half of the segment. We obtained the values of 9.1 and 7.2 ($P < 0.001$), respectively, for the two cells. Marginally, for each cell we assume the same model as above. We assume, however, that the two processes have the same change point, this change point may vary from trial to trial, and the cells are independent given this change

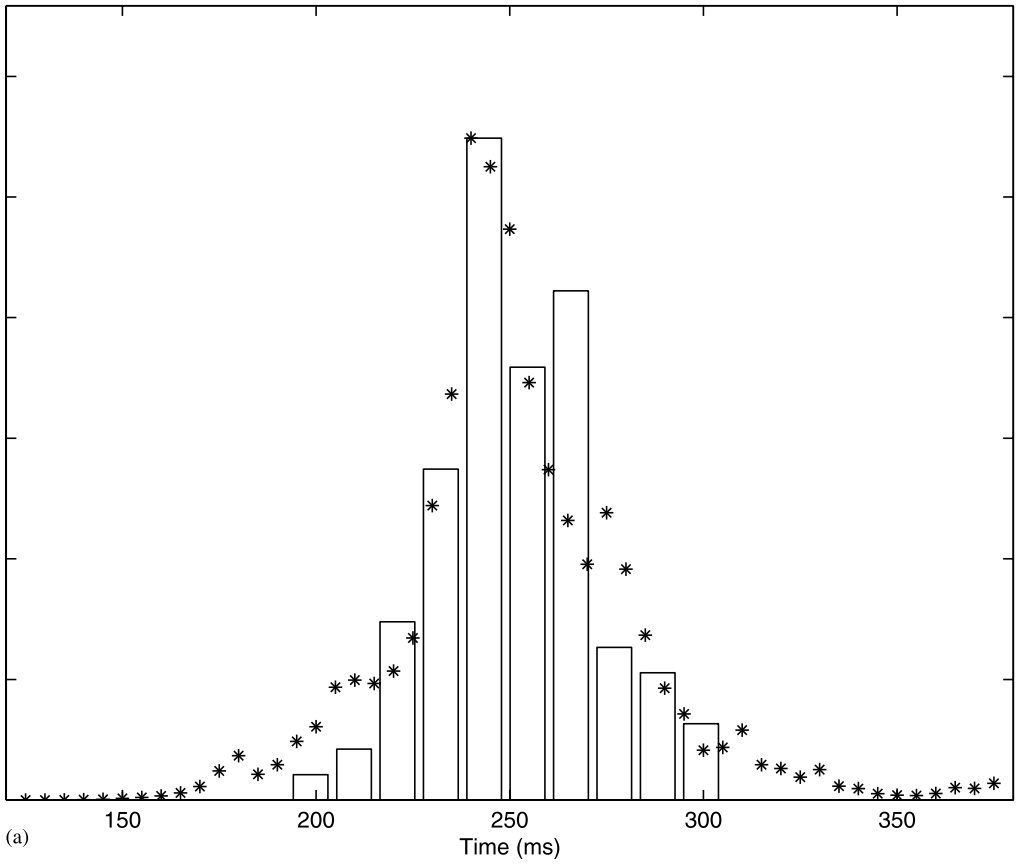


Fig. 8. Simulated data, histogram of the actual time of the change points and their estimates (stars). (a) First change point; (b) second change point; (c) third change point.

point. This assumption seems to be plausible: we applied the algorithm to the two neurons independently. The root-mean-squares distance between the two estimated p.d.f.'s (one for each cell) was 0.07. The correlation coefficient between the two vectors of a posteriori expected values of the change point times was 0.46.

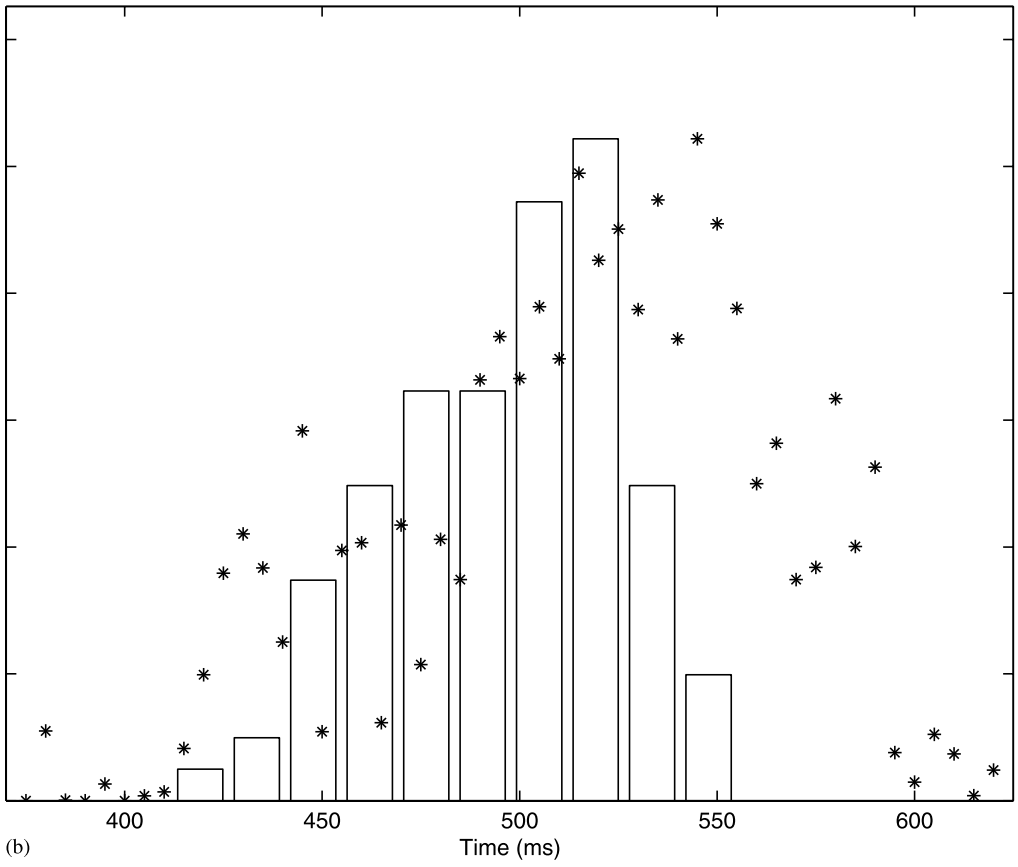
The estimator is given in Fig. 6. In Fig. 6b and c we compare the MLE estimate of the distribution function of the joint change point to the separate estimates based on the monotone regression of the corresponding PSTHs. These graphs show that the assumption of the existence of a joint change point is reasonable.

3.3.4. Simulation: multiple processes and change points

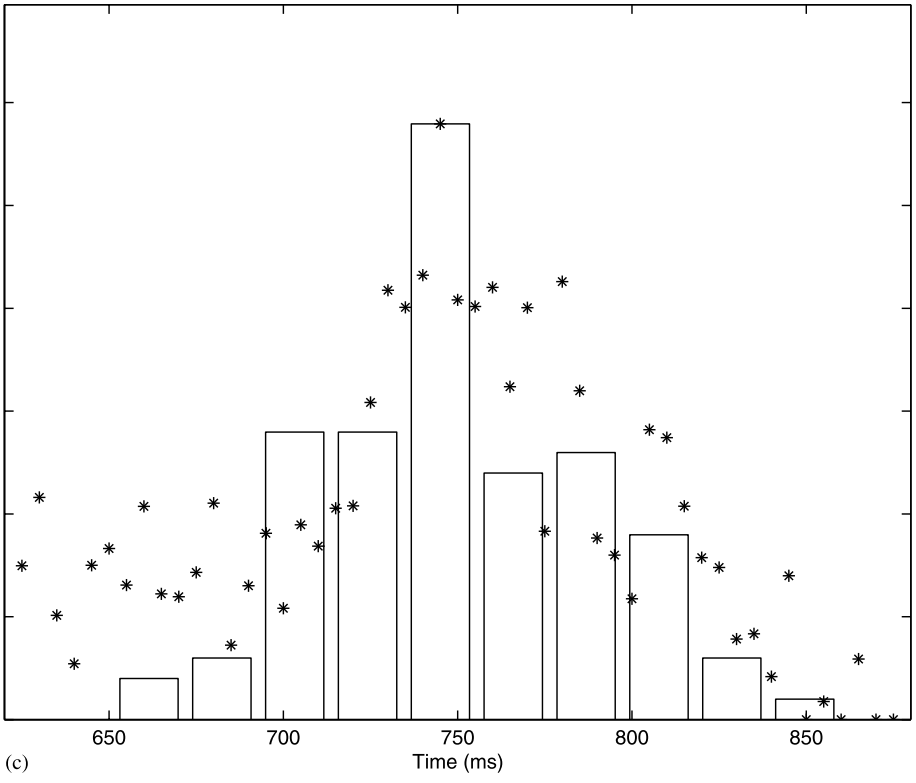
We continue in our generalization. This time we simulated 100 trials in which two neurons are observed. The two cells have the same change points, but are independent otherwise. Three change points were simulated. The change point times were independent and with different supports. The distributions of the change points were gamma with a scale parameter 2 and shape parameters 125, 250 and 375, respectively (and hence the mean times were 250, 500 and 750, while the standard deviations are 22.36, 31.63, 38.73, respectively). The time scale was chosen to be similar to the biological data, so the whole interval was considered as having 1000 ms length. The distribution was truncated to the intervals (125, 375), (375, 625), and (625, 875), respectively. The intensities of the two processes were (40, 10) (in the units of spikes per s) before the first change point, (60, 50) between the first and the second change points, (40, 50) after the second and (40, 30) after the last change point. The raster plot is given in Fig. 7a, and the PSTH of these data is given in Fig. 7b and c.

The first change can be observed nicely, the other change points can be observed but less clearly. We looked for a single change point in each of the intervals (125, 375), (375, 620), and (625, 875). The starting point was a homogeneous Poisson process and uniform change point distribution on the grid of 5 ms in each of the intervals. In Fig. 8 the estimated densities of the change points are plotted together with the histogram of the actual 'unobserved' times. As can be expected from the raster plot, the distribution of the first change point was well estimated. The two other distributions were estimated better than we expected, but not as good as the first. The estimates of the intensities are given in Table 1.

In Fig. 9 we ordered the trials according to the time of the first change point, and plotted the a posteriori



(b)



(c)

Fig. 8 (*Continued*)

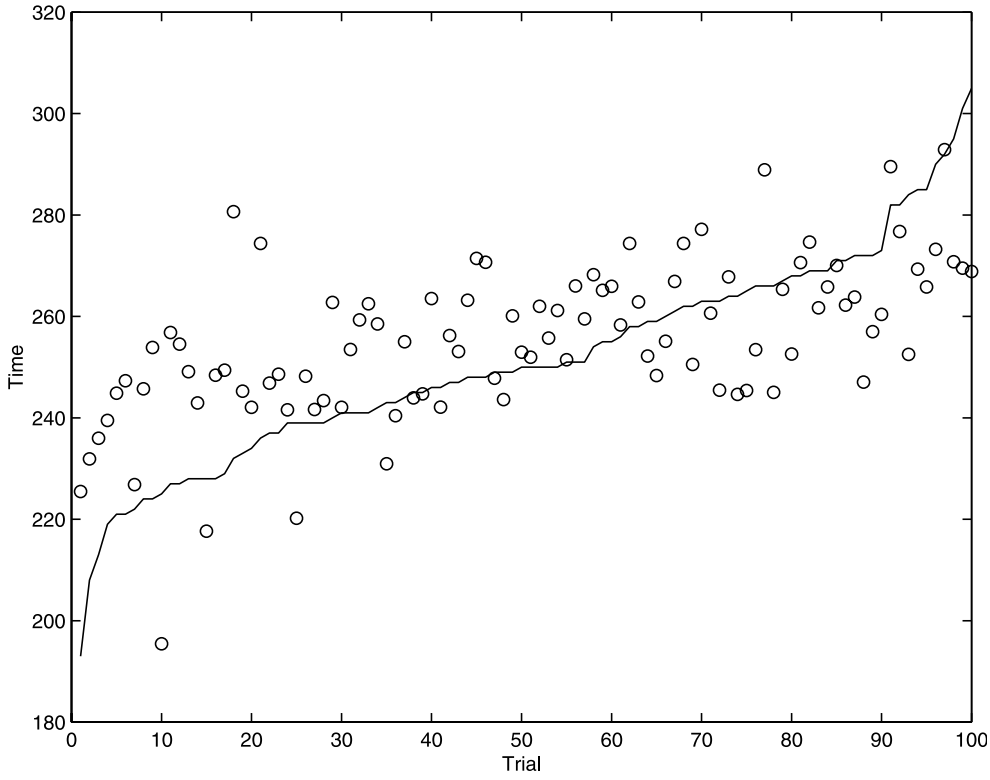


Fig. 9. Simulated data, actual times of the change point times and their a posteriori mean. The trials are ordered according to the time of the change point. The actual times are given by the solid time as function of the trial number. The opened circles are the a-posteriori expectation.

expectation and the actual time of the first change point as against trial number. That is, for each trial we computed the a-posteriori distribution of the change point, as in the E-step of the algorithm, and calculated the expectation of this distribution. It can be observed that although the distribution was estimated quite well the individual times were not. Of course, it can be expected that the Bayes estimator will shrink towards the mean. The estimator seems to depend mainly on the a priori distribution. In the introduction we argued that estimators which are based on the trial by trial estimation of the change point may yield a poor estimator of the change point distribution. Fig. 9 proves our case.

3.4. Mathematical background

In the appendix of an extended version of the paper, (Ritov et al., 2002) we give a rigorous analysis of a mathematical model of the problem. Unlike the model discussed above, the theoretical model considered has one change point and the observed process is of inhomogeneous Poisson process. We discuss in the extended version the information bound for estimating the intensities, and show that they can be estimated in the parametric $\sqrt{n}$ rate. The distribution function, on the other hand, can be estimated only in the rate of $n^{1/3}$, a much slower rate than the $n^{1/2}$ which is attainable with direct observations on the change point times. A bound on the achievable rate is established by presenting pairs of distributions which are $n^{-1/3}$ apart but the Neyman–Pearson tests between them have sum of errors bounded away from 0. That this bound is actually achievable is proved by exhibiting a simple non-iterative estimator that actually achieves the optimal rate. This estimator is based on monotone smoothing of the PSTH which was used above. See Figs. 1(c), 2(b), 5(c), 6(b and c), 7(c). We also show that under some conditions the maximum-likelihood is rate optimal.

4. Discussion

We applied the empirical Bayes change point methodology to neural data. Using empirical examples and simulated data we showed that this technique can be used to obtain a sound understanding of the nature of the synchronization between an external event and the cells activity.

Table 1
Simulation: estimates and true intensities (in spikes per s)

Interval	$(0, \tau_1)$	(τ_1, τ_2)	(τ_2, τ_3)	$(\tau_1, 1)$
First process	41.5 (40)	56.2 (60)	37.6 (40)	39.7 (40)
Second process	9.9 (10)	47.9 (50)	48.0 (50)	31.4 (30)

The theoretical statistical discussion was restricted to point processes in continuous time, while the algorithms were restricted to 0–1 processes in discrete time. Both are approximations of reality. In practice, the cells operate in continuous time while the output of the experimental system is a discrete one. Moreover, the spikes are not points in time, but have a duration of the order of 1 ms. So we preferred to use the convenient model for the given discussion.

Another statistical method that was used for similar data is that of the hidden Markov model (HMM), (Radons et al., 1994; Abeles et al., 1995; Gat et al., 1997; Ver Hoef and Cressie, 1997). This model presumes that the recorded cells are behaving as a Markov process with a finite state space. These states are not observed directly. Instead, each state is characterized by a different vector of cell intensities, the hidden mechanism. The above papers suggest different algorithms to estimate the parameters of the model, and show that the inferred states may have a biological meaning.

The change point model suggested in this paper may seem more restricted than the HMM. Practically, the number of possible states was restricted to two or three, with a prescribed transition order. However, by definition, the HMM assumes that the brain stays at each state an exponential time. In theory, this can be bypassed by assuming many pseudo-states. Actually, any stationary process can be weakly approximated by (not necessarily simple) HMM, see Kunsch et al. (1995). However, for this we may need many more states than it would be practical to assume.

The change point model does not suffer from this problem. Any distribution function for the time of change can be assumed. Therefore, the change point model is preferred to the HMM, whenever we assume that the number of change points is small, and the distribution of the time to the changes is of interest.

In this paper we consider a non-parametric model for the time to the change in the intensity. We could assume a parametric model, such as a gamma distribution with one or two unknown parameters. However, such a parametric assumption is restricting and without the usual benefits in terms of speed of convergence and simplicity of the estimation procedure. The algorithm will be very much the same, and the rate of convergence will not be much different.

We considered the testing of the existence of a change point versus the hypothesis that no change occurs. We intend to discuss elsewhere the more difficult problem of the existence of a relatively sharp change at a random time versus graduate change.

In this paper we considered a mathematical model of neurons which react to external event after a random delay. The reaction is an abrupt change in the firing intensity, but whose time is different in different trials. The main take home message of our analysis is that in 'real cases' of neuronal data, the distribution of the change point can be estimated, while single trial estimation of the specific value of the change point is prone to a large error. However, proper estimation of the distribution of the change point can reliably help in the discrimination between two plausible physiological scenarios. In the first scenario there is a smooth transition of the discharge rate in all single trials, whereas in the second scenario there are sharp transitions with jitter of their timings. This discrimination can not be done by the classical PSTH analysis, and the present manuscript provides a quantitative (rather than the subjective raster plot display) method for the discrimination between these two possible physiological settings.

Acknowledgements

This study was supported in part by the Israeli Academy of Science. We like to thank the reviewers for their comments which improved the quality of the paper.

References

Abeles M, Bergman H, Gat I, Meilijson I, Seidermann E, Tishby N. Cortical activity flips among quasi-stationary states. Proc Natl Acad Sci USA 1995;92:8615–20.

Akman VE, Rafrery AE. Asymptotic inference for a change-point poisson process. Ann Stat 1986;14:1583–90.

Bélisle P, Joseph L, MacGibbon B, Wolfson DB, du Berger R. Change-point analysis of neuron spike train data. Biometrics 1998;54:113–23.

Bickel P, Klaassen C, Ritiv Y, Wellner J. Efficient and Adaptive Estimation for Semiparametric Models. Baltimore: John Hopkins University Press, 1993.

Commenges D, Seal J. The analysis of neural discharge sequence: change-point estimation and comparison of variences. Stat Med 1985;4:91–104.

Crutcher MD, Alexander GE. Movement-related neural activity selectively coding either direction or muscle pattern in three motor areas of the monkey. J Neurophysiol 1990;64:151–63.

DeLong MR. Activity of palladial neurons during movement. J Neurophysiol 1971;34:414–27.

DeLong M, Georgopoulos A. Motor functions of the basal ganglia. In: Brookhart JM, Mountcastle VB, Brooks VB, Geiger SR, editors. The Nervous System. Motor Control. Volume II Pt. 2 of Handbook of Physiology. American Physiological Society Bethesda, 1981:1017–61.

Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the em algorithm (with comments). JRSS-B 1977;39:1–37.

Efron B. Emprical Bayes methods for combining likelihoods. J Am Stat Assoc 1996;91:538–50.

Gat I, Tishby N, Abeles M. Hidden markov modelling of simultaneously recorded cells in the associative cortex of behaving monkeys. Networks Comp Neural 1997;8:297–322.

Joseph L, Wolfson DB. Estimation in multi-path change-point problems. Comm Stat A Theory Methods 1992;21:897–913.

Kunsch H, Geman S, Kchagias A. Hidden markov random fields. Ann Appl Probab 1995;5:577–602.

Matthews DE, Farewell VT, Pyke R. Asymptotic Score-statistics processes and test for constant hazard against a change-point alternative. Ann Stat 1985;13:538–91.

Montgomery EJB. A new method for relating behavior activity in performing monkeys. J Neurosci Methods 1989;28:197–204.

Nini A, Feingold A, Solvin H, Bergman H. Neurons in the globus pallidus do not show correlated activity in the normal monkey, but phase-locked oscillations appear in the mptp model of parkinsonism. J Neurophysiol 1995;74:1800–5.

Radons G, Becker JD, Dulfer B, Kruger J. Analysis, classification, and coding of multielectrode spike trains with hidden markov models. Biol Cybern 1994;71:359–73.

Raftery AE, Akman VE. Bayesian analysis of a poisson process with a change-point. Biometrika 1986;73:85–9.

Raz A, Vaadia E, Bergman H. Firing patterns and correlations of spontaneous discharge of palladial neurons in the normal and tremulous 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine vervet model of parkinsonism. J Neurosci 2000;20:8559–71.

Ritov Y, Raz A, Bergman H. An empirical Bayes change point problem with application to neurological data. In: The International Symposium on Contemporary Multivariate Analysis and Its Applications, Hong Kong, 1997.

Ritov Y, Raz A, Bergman H. Detection of onset of neuronal activity by allowing for hetrogeneity in the change points. Technical Report 622 Dept. of Stat., UCB, 2002 http://www.stat.berkely.edu/tech-reports/index.html.

Robins JM, Ritov Y. Towards a curse of dimensionality appropriate (coda) asymptotic theory for semiparametric models. Stat Med 1997;17:285–319.

Romo R, Schultz W. Dopamine neurons of the monkey midbrain: contingencies of responses to active tough during self-initiated are movements. J Neurophysiol 1990;63:592–606.

Schwartz AB, Kettner RE, Georgopoulos AP. Primate motor cortex and free arm movements to visual targets in three-dimensional space. I. Relations between single cell discharge and direction of movement. J Neurosci 1988;8:2913–27.

Seal J, Commenges D. A quantitative analysis of stimulus- and movements-related responses in the posterior parietal cortex of the monkey. Exp Brain Res 1985;58:144–53.

Seal J, Commenges D, Salamon R, Bioulac B. A statistical method for the estimation of neuronal response latency and its functional interpretation. Brain Res 1983;278:382–6.

Ver Hoef JM, Cressie N. Using hidden markov chains and empirical Bayes change-point estimation for transect data. Environ Ecol Stat 1997;4:247–64.