

# Statistical smoothing of neuronal data

Robert E Kass, Valérie Ventura and Can Cai

Department of Statistics, Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA

Received 4 July 2002, in final form 25 November 2002

Published 17 January 2003

Online at [stacks.iop.org/Network/14/5](http://stacks.iop.org/Network/14/5)

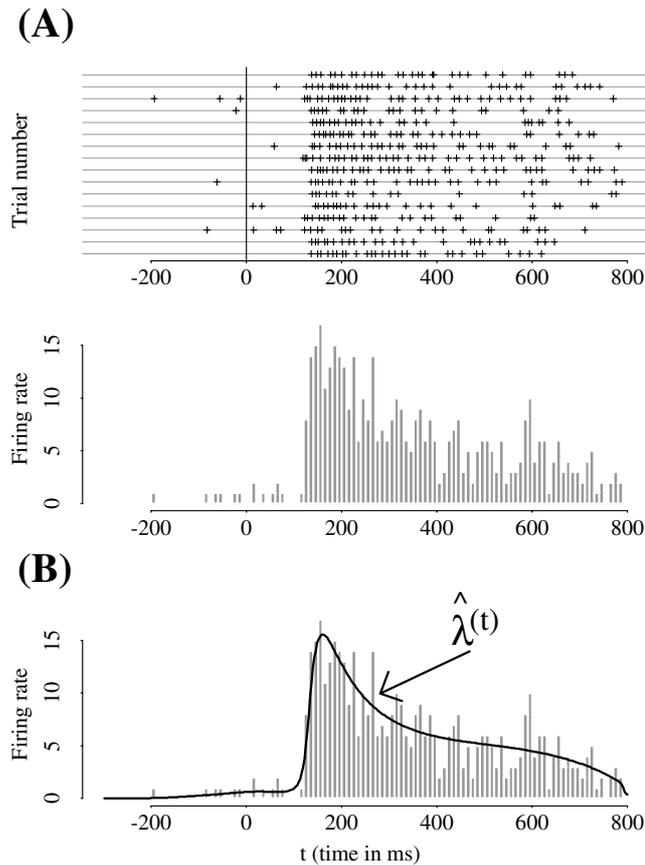
## Abstract

The purpose of smoothing (filtering) neuronal data is to improve the estimation of the instantaneous firing rate. In some applications, scientific interest centres on functions of the instantaneous firing rate, such as the time at which the maximal firing rate occurs or the rate of increase of firing rate over some experimentally relevant period. In others, the instantaneous firing rate is needed for probability-based calculations. In this paper we point to the very substantial gains in statistical efficiency from smoothing methods compared to using the peristimulus–time histogram (PSTH), and we also demonstrate a new method of adaptive smoothing known as Bayesian adaptive regression splines (DiMatteo I, Genovese C R and Kass R E 2001 *Biometrika* **88** 1055–71). We briefly review additional applications of smoothing with non-Poisson processes and in the joint PSTH for a pair of neurons.

## 1. Introduction

The raster and peristimulus–time histogram (PSTH) plot is a wonderful graphic. As in figure 1(A), the raster displays the complete set of spike times for all trials on a single neuron under particular experimental conditions, while the PSTH accumulates these to show the way the firing rate varies over time. One reason the PSTH works well is that our eye is able to smooth the PSTH so that we see the temporal evolution of the firing rate. Once we articulate the goal of estimating the firing rate, however, it is possible to improve the PSTH by smoothing (often called filtering), as shown in figure 1(B). In this paper we review the physiological motivation for smoothing and both old and new methods for carrying it out. We will focus, especially, on a method called Bayesian adaptive regression splines (BARS) (DiMatteo *et al* 2001) that produces nice smooth curves such as the one shown in figure 1(B).

From the outset we would like to be clear that when we talk of ‘estimating the firing rate’ we mean that we will use the data to produce an estimate of the *instantaneous* firing rate, which we write as  $\lambda(t)$ , at each time  $t$ , where  $t$  varies across the whole range of experimentally interesting values. In other words, we are interested in estimating the *curve* described by  $\lambda(t)$ . In statistics, we usually write an estimate obtained from data with a hat, so  $\hat{\lambda}(t)$  should really

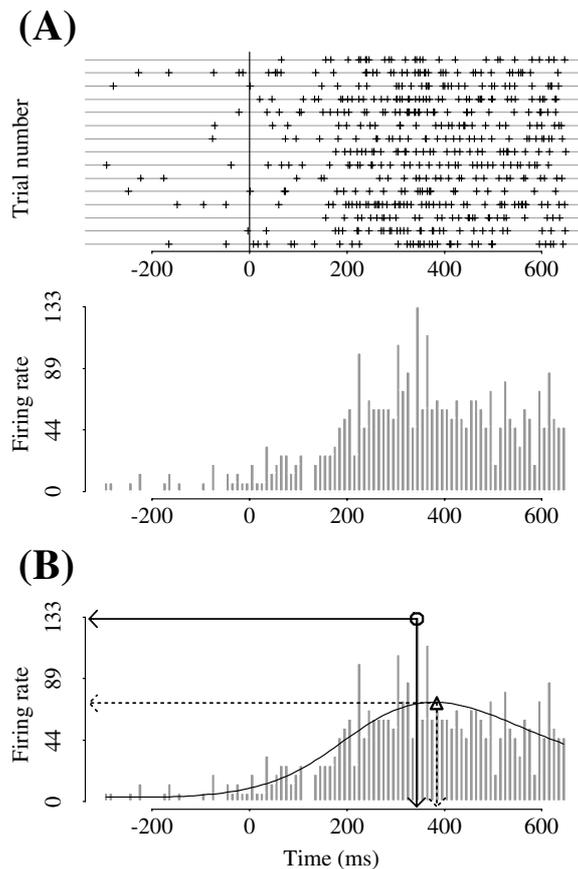


**Figure 1.** (A) Raster plot and PSTH for a neuron, and (B) the smooth PSTH obtained with BARS (DiMatteo *et al* 2001).

be written as  $\hat{\lambda}(t)$ . We therefore label the fitted firing-rate curve as  $\hat{\lambda}(t)$  in figure 1(B). This emphasizes the distinction between an unknown ‘true’ curve and an estimate of it.

Why do we care about the instantaneous firing rate? Sometimes, questions of interest require it. For example, a study in the laboratory of Carl Olson, our colleague at the Center for the Neural Basis of Cognition in Pittsburgh (Olson *et al* 2000), examined neurons in the supplementary eye field (SEF) when a monkey moved its eyes in response to either an explicit external cue (the point to which the eyes were to move was illuminated) or an internally generated translation of a complex cue (a particular pattern at the fixation point determined the location to which the monkey was to move his eyes). We were interested in the time at which maximal firing rate was achieved, and the delay of this maximum for the internally generated cue compared to the external cue. The raster and PSTH are given in figure 2(A).

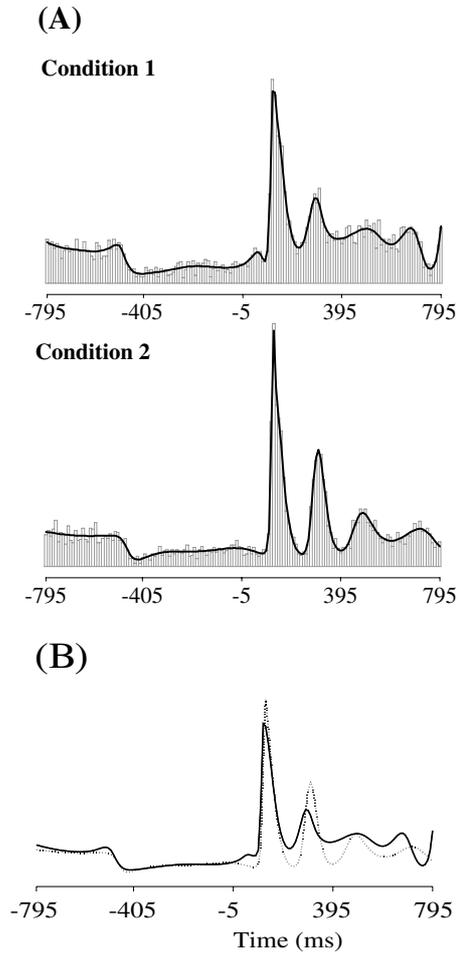
It would be possible to use the PSTH to estimate the time at which the maximal firing rate is achieved: find the highest peak and then the time at which it occurs. However, as can be seen in figure 2(B), that estimate would be noisy. A more accurate method is to first fit a smooth curve, then repeat the process for the fitted curve. This produces a different value, which better represents the time of the peak firing rate under the rather natural assumption that the firing rate is varying smoothly.



**Figure 2.** (A) Raster and PSTH for a neuron in SEF from Olson *et al* (2000), and (B) maximum firing rate and time of maximum estimates obtained from the raw and smoothed PSTHs.

A second example again comes from Carl Olson's lab, this time involving neurons from inferotemporal (IT) cortex (Olson and Rollenhagen 1999). As shown in figure 3, in one condition a stimulus was displayed alone, while in the second condition a pair of distractors were also illuminated. For the second condition the damped oscillatory response was more pronounced. One way to quantify this is to measure the peak-to-trough differences. Again, these values may be computed from the fitted curves. It is also worth pointing out that, as figure 3(B) shows, it can be easier to compare two curves than to compare two histograms; there, the two curves are overlaid on top of each other. Using the curves we have great flexibility in quantifying the distinctions we perceive. When we quantify the contrast between peak-to-trough differences in the two conditions, we must include a standard error to indicate the variability in our estimate. The statistical methods we discuss provide such standard errors and can be applied to essentially any feature of the firing-rate curve we wish to analyse.

A second reason for caring about the instantaneous firing rate is that probability models make efficient use of the data, and the function  $\lambda(t)$  is needed in writing down a probability model. Under the Poisson assumption (Cox and Miller 1965), the probability density for the



**Figure 3.** (A) Raw and smoothed PSTHs in the two conditions for an IT neuron in Olson and Rollenhagen (1999), and (B) the two smoothed curves overlaid.

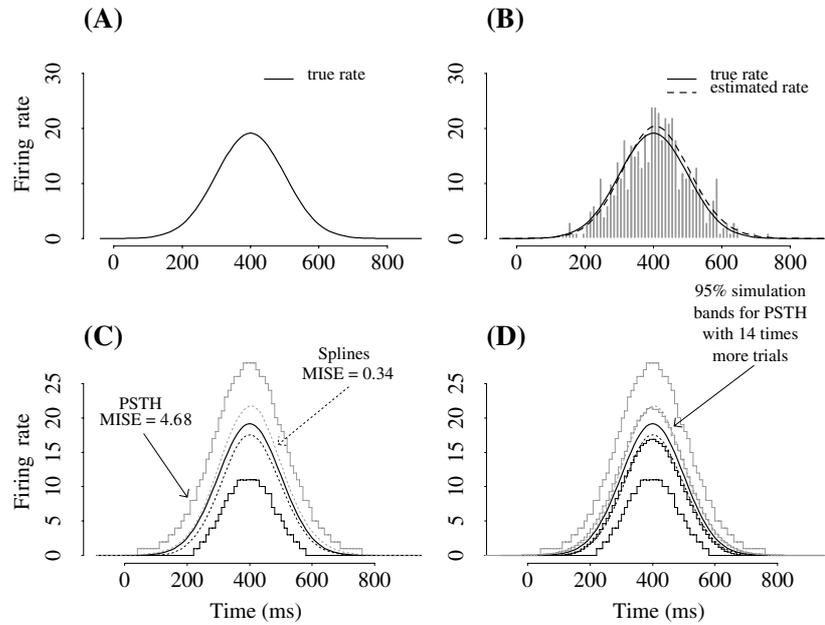
set of spikes  $s_1, \dots, s_n$  is

$$p(s_1, \dots, s_n) = \exp\left(-\int_0^T \lambda(t) dt\right) \prod_{k=1}^n \lambda(s_k). \quad (1)$$

(In the non-Poisson case  $\lambda(t)$  is generalized; we return to this in what follows.) This kind of model may be fitted using either the maximum likelihood or Bayesian methods, which often provide nearly identical fits, both methods being optimal in the sense of statistical efficiency. The issue of efficiency is particularly important when studying trial-to-trial variation (and its relation to effects such as response time) or when making real-time predictions (decoding) from multiple neurons recorded simultaneously.

## 2. Statistical efficiency of smoothing

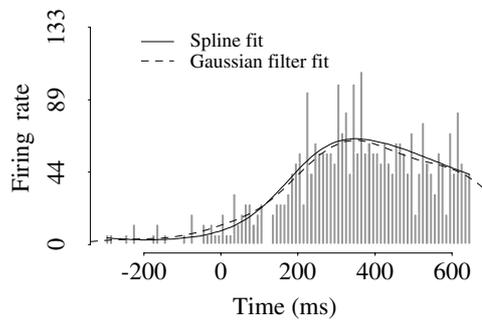
To illustrate the statistical efficiency of smoothing, consider data simulated (on the computer) from the true firing rate function shown in figure 4(A). When we simulate 16 trials (the number



**Figure 4.** (A) True rate from which 16 trials are simulated; their PSTH is shown in (B), with true and estimated firing rates overlaid. (C) The true rate and 95% simulation bands obtained from smoothed and unsmoothed PSTHs. (D) The same curves as (C), as well as 95% simulation bands obtained from unsmoothed PSTHs with  $16 \times 14$  trials instead of 16.

of trials in the SEF data shown earlier) we get the PSTH shown in figure 4(B). We then use the spline-based smooth fit, also shown in figure 4(B), as the ‘estimated rate.’ Note that the estimated rate is close to the true rate from the simulation, but it misses by a small amount due to the small number of trials we used in the simulation. A standard way to quantify this deviation is to use the ‘integrated squared error,’ or ISE, which is simply the sum of squared deviations of the estimated curve from the true curve, summed across all time values.

If we repeat this a large number of times (in what follows, we simulated 1000 data sets) we can compute the mean value of the ISE, or MISE. The MISE is a standard statistical measure of the accuracy of an estimate of a curve, averaged across repeated samples of data. We can also compute 95% bands, within which fall 95% of the estimated curves. We carried out these procedures, computing the MISE and 95% bands for the spline-based estimate from 1000 such simulated data sets (with 16 trials in each, the estimate being computed a total of 1000 times). We also computed the MISE and 95% bands for the PSTH, viewed as an estimate of the firing-rate curve. Figure 4(C) shows the two pairs of bands, now labelled with the two MISE values: the spline-based estimate has a MISE of 0.34 (in spikes  $s^{-2}$ ) while the PSTH has a MISE of 4.68, which is 14 times larger. This means that when the PSTH is used to estimate the firing rate, 14 times as much data are needed to achieve the same level of accuracy. Similarly, the 95% bands for the PSTH are much further from the true firing-rate curve than the bands for the spline-based estimate. In figure 4(D) we add a pair of 95% bands obtained from the PSTH when 224 trials are used rather than 16 (because  $224 = 14 \times 16$ ). This is another way of showing that the accuracy in estimating the firing rate using spline smoothing based on 16 trials is the same as the accuracy using the PSTH based on 224 trials. Clearly it is very much better to use smoothing when estimating the instantaneous firing rate.



**Figure 5.** PSTH of an SEF neuron from Olson *et al* (2000) with Gaussian filter and spline fits overlaid.

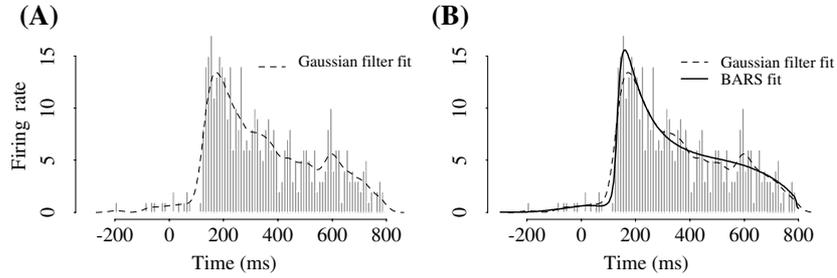
### 3. Spline-based smoothing

There are many ways to accomplish the smoothing or filtering. We happen to like spline-based methods, which we will explain briefly in a moment. In figure 5 we return to the SEF data and show a PSTH together with two smooth versions of it, one based on a spline (Venables and Ripley 1999) and the other based on a Gaussian filter (Silverman 1986), or, what in the statistical literature is called a kernel density estimate. The two methods give very similar fitted curves. The reason the two methods agree closely in this example is that the firing rate is in this case varying quite slowly.

Data from another experiment on IT neurons conducted in Carl Olson's lab (Baker *et al* 2002) are summarized in figure 6. In this experiment, the monkey maintained fixation for 300 ms, a visual stimulus was presented for 500 ms, and after a further 300 ms of fixation following the offset of the stimulus, reward was given. In figure 6, time zero corresponds to stimulus onset. Figure 6(A) shows the case in which the Gaussian filter has difficulty. Note the wiggleness of the fitted curve it produces, together with its inability to track the quick increase in firing rate. The difficulty here is due to the uneven variation of the function: it begins relatively flat, then at around 100 ms after time 0 it has a very steep increase, and then has a relatively smooth decline. The Gaussian filter, which is linear, has a fixed amount of smoothing it applies throughout the time range. If we increase the bandwidth of the filter we can remove the wiggleness in the curve, but will also remove much of the sharp increase in the process; conversely, if we decrease the bandwidth we can eventually pick up the sharp rise but at the cost of greatly increasing the wiggleness. What is needed, instead, is what we might call an *adaptive* method, which will vary the amount of smoothing (nonlinearly) across time. In figure 6(B) an adaptive method called BARS is shown. It succeeds in capturing what we believe are the essential features of the firing rate curve.

We now briefly outline what BARS is, and why we like it. BARS uses cubic splines. A cubic spline is a piecewise cubic polynomial: several cubic pieces are joined at locations called 'knots'. The pieces are constrained so that the resulting curve is smooth (it is twice continuously differentiable). When we use a spline to fit a curve to some data we must begin by choosing how many knots to use and where these knots will go. This determines how smooth the curve will be in each interval of time. From a statistical point of view, choosing the knots is the hardest part of the problem.

BARS, which stands for Bayesian adaptive regression splines, was defined and studied in a recent paper by DiMatteo *et al* (2001). In the statistical literature, this kind of application of splines is called 'regression splines' because the fitting procedure uses linear regression and



**Figure 6.** (A) PSTH of an IT neuron from Baker *et al* (2002) with a Gaussian filter fit overlaid, and (B) with the BARS fit of DiMatteo *et al* (2001).

its generalizations. BARS uses a Bayesian Monte Carlo method to search through the space of possible numbers of knots and their locations, and to provide an optimally fitted curve based on this search. It also provides assessments of statistical uncertainty, which are crucial for the sort of applications we have outlined above. For example, in the distractor and non-distractor conditions compared in the IT neuron shown earlier, DiMatteo *et al* (2001) computed both an estimated difference in peak minus trough heights of 50 and also a standard error of 20.8 spikes  $s^{-1}$  to indicate uncertainty in the estimate.

Formally, BARS uses a posterior probability distribution on the number of knots and their locations, and we take the expectation of the unknown function of time to be our fitted curve. This results in a fitted curve that is not itself a spline but rather a weighted combination of splines, with the weights being the posterior probabilities assigned to each spline. DiMatteo *et al* showed in a simulation study that BARS reduces the mean integrated squared error (MISE) of the fitted curve substantially compared to other methods presented in the literature (see also Kass and Wallstrom 2002).

#### 4. Non-Poisson models

We would like to mention, briefly, the way smoothing can be used when the non-Poisson nature of neural spiking activity may play an important role in data analysis. According to the Poisson process model the probability of a spike at time  $t$  is  $\lambda(t) dt$ . Please note that when some people refer to a Poisson model they mean a *constant, time-invariant* firing rate, but here we use the term in its more general *time-varying* or *inhomogeneous* form, which specifically allows for a time-varying firing rate (as we observe in experimental data). The beauty of the Poisson model is that the firing rate depends *only* on time. To generalize, we must allow the firing rate to depend also on past spikes; hence the probability of a spike at time  $t$  is  $\lambda(t|s_1, \dots, s_k) dt$ , where  $s_1, \dots, s_k$  are spike times prior to  $t$ . This general form, however, is too general to be useful: we now have a  $\lambda$  function that depends not only on time  $t$  but also on the values of all the spike times prior to  $t$ , and to fit this firing rate empirically from the data is an impossibly difficult task. To simplify things several authors have introduced what Kass and Ventura (2001) called inhomogeneous Markov interval (or IMI) processes (see Johnson 1996, Kass and Ventura 2001 for additional references). The simplification here is that the firing rate now depends on only two variables: the time  $t$ , and the time since the last spike prior to  $t$ , so that the probability of a spike at time  $t$  is  $\lambda(t, t - s_*(t)) dt$ , where  $s_*(t)$  is the time of the previous spike. These various models are summarized in table 1.

The simplification from the general to the IMI models allows the same kind of empirical fitting as we showed previously, using spline methods, now complicated in a manageable way

**Table 1.** Forms for the firing rate function used in three alternative probability models.

Model	Probability spike at time $t$
Poisson	$\lambda(t) dt$
General	$\lambda(t s_1, \dots, s_k) dt$ <span style="margin-left: 2em;"><math>s_1, \dots, s_k</math> are spike times prior <math>t</math></span>
IMI	$\lambda(t, t - s_*(t)) dt$ <span style="margin-left: 2em;"><math>s_*(t)</math> is the time of the previous spike</span>

by involving a function of two variables rather than one. Kass and Ventura (2001) noted that standard software could be used to fit this non-Poisson probability model.

With the possibility of fitting several models to the data comes the question of which should be used. A method is provided by Brown *et al* (2002), who show how the time-rescaling theorem is used to assess the goodness of fits of these and other models.

## 5. Smoothing the joint PSTH

Our final application of spline-based smoothing is to the study of correlated firing in a pair of neurons. Here we examine the simultaneous firing of two neurons, with simultaneity defined as coincident spikes occurring within a specified window of time such as 1 ms, or within 1 ms of a fixed lag. The statistical issues are, first, whether there is any evidence that the coincident firing is due to sources other than chance, i.e., whether the independence model is inadequate and, second, if so, what the departures from independence may be.

The first issue can be resolved by assessing if the pair of spike trains show significant departures from the independence model. That is, if  $p^1(t)$  and  $p^2(t)$  denote the spike probabilities at time  $t$  for neurons 1 and 2 respectively, and  $p^{12}(t)$  the probability of simultaneous spike at time  $t$ , we want to test the null hypothesis

$$H_0 : p^{12}(t) = p^1(t)p^2(t).$$

This is usually done by testing if each cell of the adjusted joint peristimulus–time histogram (JPSTH; Aertsen *et al* 1989) deviates substantially from zero, or by comparing the observed JPSTH to its expectation under the hypothesis of independence via a  $\chi^2$  test. The JPSTH is a matrix of joint spike counts, with generic cell  $(u, t)$  recording the number of trials on which neuron 1 fired at time  $u$  and neuron 2 at  $t$ ; figure 7(G) shows such a matrix, with counts replaced by a grey scale. These tests are typically performed directly on the JPSTH, but smoothing  $p^1(t)$ ,  $p^2(t)$  and  $p^{12}(t)$  first would make the test more efficient.

Time-varying departures from independence can be assessed by proposing an alternative model that defines a functional form for these departures. Specifically, the independence model could be tested against the alternative, more general, model

$$H_A : p^{12}(t) = p^1(t)p^2(t)\zeta(t), \quad (2)$$

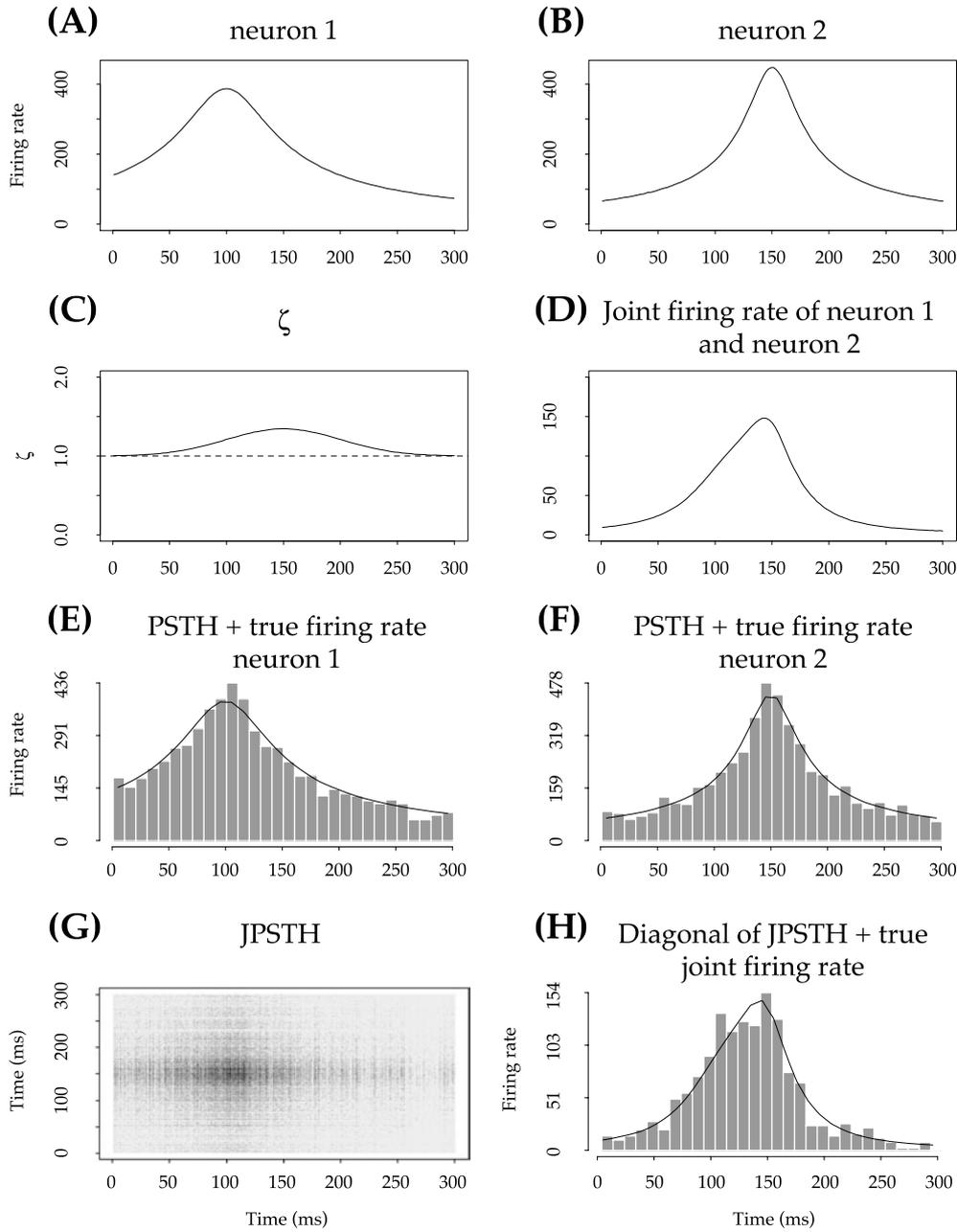
where  $\zeta(t)$  models the time-dependent change in the simultaneous firing rate *beyond* which would be predicted by the independence model. Then testing  $H_0$  against  $H_A$  reduces to testing

$$H_0 : \zeta(t) = 1 \quad \text{for all } t$$

against

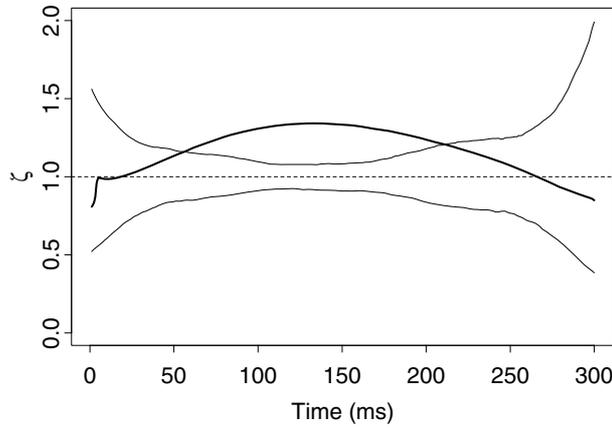
$$H_A : \zeta(t) \neq 1 \quad \text{for some } t,$$

while a smooth estimate of  $\zeta(t)$  describes the departures from the independence model, as a function of time.



**Figure 7.** (A)–(C) True components  $p^1(t)$ ,  $p^2(t)$  and  $\zeta(t)$  in model (2), and (D) the product  $p^{12}(t) = p^1(t)p^2(t)\zeta(t)$ . (E) PSTH for neuron 1 and (A) overlaid. (F) PSTH for neuron 2 and (B) overlaid. (G) JPSTH and (H) main diagonal of the JPSTH and (D) overlaid.

Before we illustrate this, let us first show how model (2) summarizes the JPSTH. Again we use simulated data, with true model components shown in figure 7(A)–(C), where  $\zeta(t)$  represents excess coincident spiking. If  $\zeta(t)$  were equal to the constant 1, the independence model would hold. Figure 7(D) plots the products of the curves in figures 7(A)–(C), which



**Figure 8.** Spline fitted  $\hat{\zeta}(t)$  (bold) and 95% simulation bands for  $\zeta(t)$  under the hypothesis that  $\zeta(t) = 1$  for all  $t$ .

is  $p^{12}(t) = p^1(t)p^2(t)\zeta(t)$  in (2). Figure 7(G) shows the JPSTH of 50 simulated joint spike trains from this model, figures 7(E) and (F) the marginal PSTHs for neurons 1 and 2, obtained by summing over rows and columns of the JPSTH, and figure 7(H) is the main diagonal of the JPSTH, which is used to assess synchrony. Overlaid on figures 7(E) and (F) are the true firing rates shown in figures 7(A) and (B), and on figure 7(H) the joint firing rate shown in figure 7(D); the good agreement we see is no surprise: we designed model (2) to summarize the JPSTH in this very way. Having formalized the problem this way we may now apply spline-based methods to estimate  $\zeta(t)$  from the data.

Figure 8 shows the spline-fitted  $\hat{\zeta}(t)$  from the simulated data set in figure 7. This is obtained by dividing the smoothed diagonal of the JPSTH by the product of the two individual smoothed PSTHs (where all three smooth curves were produced using BARS). It does a reasonably good job of uncovering the true time-varying excess in coincident spiking, shown in figure 7(C). Also plotted are 95% simulation bands for  $\zeta(t)$  under the null hypothesis that  $\zeta(t)$  is equal to 1, i.e., that there is no time-dependent increase in coincident firing above what is implied by the independence model. The fitted curve clearly falls outside these bands for some values of  $t$ , which indicates that there is an increase in coincident firing over an interval of time. These bands were obtained by bootstrap methods, which we will describe in detail in a future paper. In addition, we found that this degree of variation above the simulation bounds would occur by chance with a probability of  $p < 0.0005$ .

Although the data were simulated, our intention here is to illustrate the sort of results that may be obtained by spline-based fitting combined with modern statistical inference methods such as the bootstrap method. More generally, replacing the JPSTH by a statistical model makes available all the statistical tools we know to make inference, so questions of interest, such as testing for synchrony, are conceptually easier to handle. Moreover, the model can be very easily extended to account, for example, for lagged correlations, and/or for sources of correlations other than stimulus induced correlation or joint spiking, for example trial to trial or latency effects (see Brody 1999). We are thereby able to resolve several distinct features of coincident firing that are otherwise confounded in the joint PSTH and cross-correlogram: we can separate the effects that act to produce an excess (or deficit) in coincident spiking globally across time (trial to trial or latency effects) from those that produce an excess (or deficit) in a time-dependent fashion (the  $\zeta(t)$  function). This work will be described in a future paper.

## 6. Summary and further remarks

In this paper we have suggested that statistical smoothing of neuronal data is advisable whenever the instantaneous firing rate is needed. In a small simulation study smoothing decreased the amount of data needed to obtain statistical accuracy comparable to the PSTH by a factor of 14 (16 trials rather than 224 trials).

Numerous smoothing methods are available, including Gaussian filters and cubic splines. Both these methods are easy to use in their simplest forms, but become challenging when one allows them to adapt to the data, decreasing the degree of smoothing at times when the neuron's firing rate is changing quickly. In the case of splines, the problem is to determine how many knots to use and where to put them. The BARS method does this automatically and currently appears to be the most powerful method available for this purpose.

In addition to smoothing the PSTH, spline-based methods are more generally useful for fitting temporal evolution of firing rates with minimal parametric assumptions. We gave two examples. The IMI model can be used to describe non-Poisson data, and splines used to provide an estimate of the instantaneous firing rate, conditional on the recent past. We also showed how a statistical model can be used to smooth the JPSTH, and estimates of the model components used to test for synchronous neuronal activity. More generally, BARS is useful in a wide variety of applications, and has been applied already to fMRI and EEG analyses. One drawback of BARS is that it requires substantial computational time. We are currently developing faster implementations of the basic BARS strategy by incorporating the methods of Hansen and Kooperberg (2002).

## References

- Aertsen A M H J, Gerstein G L, Habib M K and Palm G 1989 Dynamics of neuronal firing correlation: modulation of 'effective connectivity' *J. Neurophys.* **61** 900–17
- Baker C I, Behrmann M and Olson C R 2002 Impact of learning on representation of parts and wholes in monkey inferotemporal cortex *Nat. Neurosci.* **5** 1210–16
- Brody C D 1999 Correlations without synchrony *Neural Comput.* **11** 1537–51
- Brown B N, Barbieri R, Ventura V, Kass R E and Frank L M 2002 The time-rescaling theorem and its applications to neural spike train data analysis *Neural Comput.* **14** 325–46
- Cox D R and Miller H D 1965 *The Theory of Stochastic Processes* (London: Methuen)
- DiMatteo I, Genovese C R and Kass R E 2001 Bayesian curve-fitting with free-knot splines *Biometrika* **88** 1055–71
- Hansen M H and Kooperberg C 2002 Spline adaptation in extended linear models *Stat. Sci.* **17** 2–51 (with discussion)
- Johnson D 1996 Point process models of single-neuron discharges *J. Comput. Neurosci.* **3** 275–99
- Kass R E and Ventura V 2001 A spike-train probability model *Neural Comput.* **13** 1713–20
- Kass R E and Wallstrom G 2002 Comment on Hansen and Kooperberg's paper *Stat. Sci.* **17** 24–30
- Olson C R, Gettner S N, Ventura V, Carta R and Kass R E 2000 Neuronal activity in macaque supplementary eye field during planning of saccades in response to pattern and spatial cues *J. Neurophys.* **84** 1369–84
- Olson C R and Rollenhagen J E 1999 Low-frequency oscillations in Macaque IT cortex during competitive interactions between stimuli *Soc. Neurosci. Abstr.* **25** 916
- Silverman B W 1986 *Density Estimation for Statistics and Data Analysis* (London: Chapman and Hall)
- Venables W N and Ripley B D 1999 *Modern Applied Statistics with S-PLUS* 3rd edn (New York: Springer)
- Ventura V, Carta R, Kass R E, Olson C R and Gettner S N 2002 Statistical analysis of temporal evolution in single-neuron firing rates *Biostatistics* **1** 1–20