Biostatistics (2000), **1**, 1, *pp.* 1–36
Printed in Great Britain

Statistical Analysis of Temporal Evolution in Single-Neuron Firing Rates

By VALÉRIE VENTURA[1], ROBERTO CARTA[1], ROBERT E. KASS[1,2],

CARL R. OLSON[2], and SONYA N. GETTNER[2]

[1] *Department of Statistics and the* [2] *Center for the Neural Basis of Cognition*

Carnegie Mellon University, Baker Hall 132, Pittsburgh PA 15213

e-mail: vventura@stat.cmu.edu

Summary

A fundamental methodology in neurophysiology involves recording the electrical signals associated with individual neurons within brains of awake behaving animals. Traditional statistical analyses have relied mainly on mean firing rates over some epoch (often several hundred milliseconds) that are compared across experimental conditions by Analysis of Variance. Often, however, the time course of the neuronal firing patterns is of interest, and a more refined procedure can produce substantial additional information. In this paper we compare neuronal firing in the supplementary eye field of a macaque monkey across two experimental conditions. We take the electrical discharges, or "spikes," to be arrivals in a inhomogeneous Poisson process and then model the firing intensity function using both a simple parametric form and more flexible splines. Our main interest is in making inferences about certain characteristics of the intensity, including the timing of the maximal firing rate. We examine data from 84 neurons individually and also combine results into a hierarchical model. We use Bayesian estimation methods and frequentist significance tests based on a nonparametric bootstrap procedure. We are thereby able to conclude that a substantial fraction of the neurons exhibit important temporal differences in firing intensity across the two conditions, and we quantify the effect across the population of neurons.

Some key words: Bayesian methods, Bootstrap hypothesis testing, Functional data analysis, Kernel smoothing, inhomogeneous Poisson process, Regression splines.

1. Introduction

One of the most important methodologies in neuroscience has been the recording of electrical activity of individual neurons in laboratory animals. In this technique, a thin electrode is inserted into the animal's brain and action potentials (spikes) generated by a single neuron are measured. Using semi-automated methods these brief events, lasting about 1 millisecond (ms), are identified, and the times at which they occur are recorded (see Olson, Gettner, Ventura, Carta, and Kass, 2000, and the references cited there). More frequent spikes indicate increased activity of the neuron, and the firing rate is taken to be a mechanism by which neurons transmit information that controls behavior. (See, for example, p. 13 of Hubel, 1995.) A particularly important aspect of this paradigm is that the animal, often a monkey possessing substantial cognitive abilities, may be awake and carrying out some complicated task while the neuronal recording is taking place. In many experiments the timing of increased neuron firing rate is of central interest. The purpose of this paper is to describe and illustrate statistical methods for drawing inferences about the temporal evolution of neuron firing rates in the context of an experiment contrasting two types of cues used to instruct a monkey where to focus attention. We apply both Poisson regression splines (with maximum likelihood) and Gaussian kernel density estimation (with the bootstrap) to obtain smoothed versions of what are often called a peristimulus time histograms (see Figure 2 in Section 3), then draw conclusions based on features extracted from these smooth representations.

The data we analyze here come from 84 neurons in the supplementary eye field (SEF) of a macaque monkey, recorded while he watched images appear on a screen in front of him (Olson, *et al.*, 2000). The SEF is thought to be involved in generating eye movements in response to stimuli. In this experiment the monkey was shown a visual instructional cue to indicate the direction in which he was to look next, the choices being up, down, left, or right, compared to the central point on which the monkey otherwise fixed his eyes. There were two instructional cues: the "spatial" cue was a white light that flashed at the location toward which the monkey was to look; the "pattern" cue was a particular pattern instead shown in the center of the visual field, and the monkey had to associate the pattern with a direction (previously learned) and move his eyes accordingly. Data from one of the neurons are displayed in Figure 2 and a graphical summary of the pooled data from all 84 neurons appears in Figure 3. One important phenomenon that apparently emerges from Figure 2 is the tendency for the cell to reach its maximal firing rate later in the pattern cue task than in the spatial cue task. This phenomenon also occurs at the population level, as can be seen in Figure 3. The goal of the work described here was to determine quantitatively and thereby compare formally how the temporal patterns of neuronal activity were affected by the nature of the instructional cue.

In Section 2 we give further details of the experiment and its purpose, in Section 3 we briefly explain the data displays and nonparametric fits appearing in Figures 2 and 3. In Section 4 we present several variations on fitting inhomogeneous Poisson process models, and in Section 5 we check the fits of our models. All of this might be considered background to the central elements of our work, the inferential methods, which are described in Section 6. We apply both Bayesian and frequentist techniques. We begin in Section 6.1 with a general method for comparing firing intensities in the spirit of functional data analysis (Ramsay and Silverman, 1997). In Section 6.2 we indicate how Bayesian estimation may be used to answer substantive questions via either maximum likelihood-based Normal approximations to the posterior or full Markov chain Monte Carlo simulation. In Section 6.3 we show how bootstrap-based significance tests may be used to address the substantive issues. In Section 6.4 we construct a hierarchical model to analyze the variability of firing intensity characteristics across the population of neurons. We organize our results in Section 7 and offer some concluding remarks in Section 8.

2. The Experiment

One of the chief goals of cognitive neuroscientific research is to clarify the neural mechanisms by which attention is attracted to locations. The experiment that produced the data analyzed here examined the temporal evolution of neuron firing in a part of the brain affected by attention, the SEF. The use of the spatial and pattern cues described in Section 1 was based on a substantial literature that differentiates what is called exogenous cueing from endogenous, the latter here referring to the internally-generated process of translating the pattern cue into an explicit instruction of where to move the eyes. Following an endogenous cue, attention is allocated more slowly but lingers longer at the instructed location. The aim of this experiment was to ascertain whether neural activity in the SEF followed a similar time course.

At the beginning of each day's session, an electrode is inserted through a recording chamber into the cerebral cortex. Several neurons may be examined each day, with measurements from a particular neuron being made within a single continuous block of time. Standard software is used to convert the electrical signal into a sequence of event codes marking the times of spikes. Details may be found in Olson, *et al.* (2000) and the references therein. The experimental paradigm for both the pattern and spatial cue tasks is summarized in Figure 1; the general sequence of events is as follows:

1. Central target on at $t = -650$ ms; the monkey starts fixation on the central target, and will have to maintain fixation until the signal to move on Step 5.

2. Four peripheral targets on (and stay on) at $t = -250$ ms.

3. At $t = 0$, a cue (spatial or pattern) appears.

4. Cue off at $t = 100$ ms.

5. Central target off at a random time $t \in [650, 850]$ ms: signal to move the eyes to the target.

6. The monkey makes an eye movement to the peripheral target associated with the cue, and maintains fixation on the correct target until reward is delivered.

Note that the times in Figure 1 indicate the duration in ms of particular trial epochs rather than the absolute times.

The monkey was trained over a period of several months before the experimental sessions took place. The portion of the experiment of interest here begins 300 milliseconds (ms) before the cue and ends at the signal to move (fixation spot off), which occurs at a random time between 650 and 850 ms after presentation of the cue. Each neuron was recorded on a number of trials ranging from 11 to 42 with a median of 16 and quartiles 14 and 16. Exhibiting a commonly-encountered "directional tuning," each cell tended to fire most rapidly when the eye movement was in a particular one of the four possible directions, which we then referred to as the *preferred direction* for that cell. In this analysis we retained the data only in each neuron's preferred direction, which effectively reduced the number of experimental conditions from eight conditions (two instructional tasks $\times$ four directions of movement) to two: the spatial and pattern tasks. See Olson *et al.* (2000) for details.

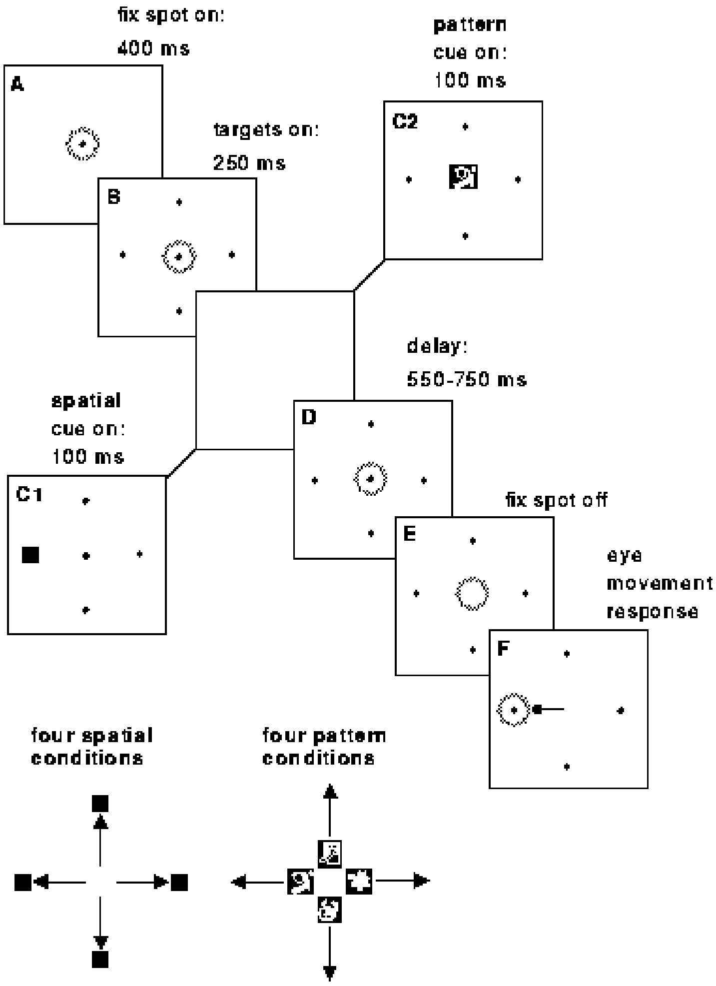

Fig. 1. Experimental paradigm for the pattern and spatial tasks.

3. Data Display

The data for one neuron are shown in Figure 2. The two raster plots display the observed spikes for each of the trials on separate lines, with 15 trials for each instructional cue task. Below these the data are pooled across trials and then displayed both as counts of spikes occurring within 10

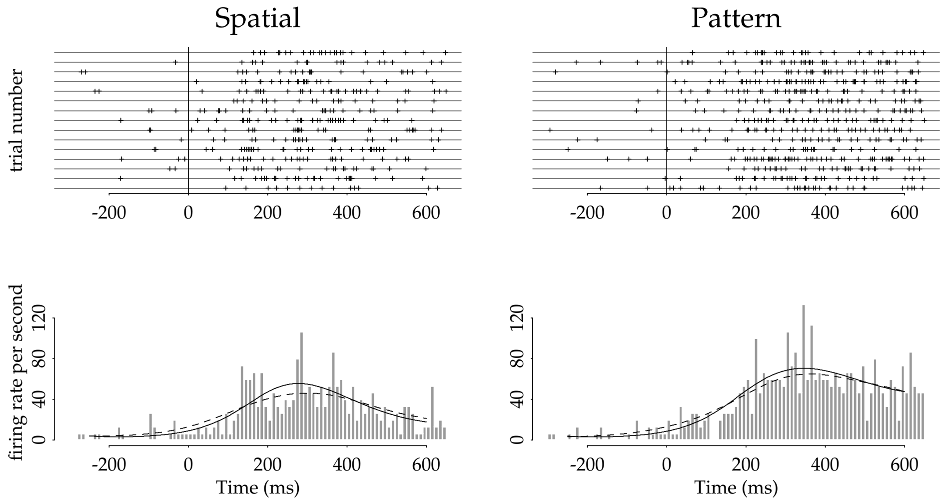

Fig. 2. Raster plots and histograms for neuron PK96c.1. The raster plots (top panels) show the observed spikes for each of the trials on separate lines. The lower panels show the observed spikes pooled across trials, displayed as count spikes occurring within 10 ms intervals, and as kernel smoothed (dotted curves) and spline fitted (solid curves) versions. Time $t = 0$ corresponds to presentation of the cue.

ms intervals and as smoothed versions of the counts obtained with a kernel density estimator, and with the regression splines described in Section 4.2. These both estimate the instantaneous mean firing rate, or intensity, $\lambda(t)$ and the units on the y-axis are spikes per second per trial. (Formally, for a counting process $N(t)$ we would write $\lambda(t) = E(dN(t))$.)

The display of counts is often called a peristimulus time histogram (PSTH). The PSTH is not, in general, a true histogram because we are not observing independent and identically distributed data. However, for an inhomogeneous Poisson process, conditionally on the number of spikes, the individual spike times are distributed as independent identically distributed observations from the density $f(t) = \lambda(t)/\int \lambda(t)dt$. That is, after conditioning on the total number of spikes, the inhomogeneous Poisson process likelihood function is equal to the likelihood function that would be obtained by drawing n observations from the density $f(t)$. Thus, any estimate $\hat{f}(t)$ provides a corresponding estimate $\hat{\lambda}(t) = n \cdot \hat{f}(t)$. (This fact may be exploited for bandwidth selection; see Diggle and Marron, 1988.)

For a particular neuron and task, let $\{X_i\}_{i=1,\ldots,M}$ denote the point process of firing times aggregated over the N available repeats. A kernel estimator of the varying intensity $\lambda(t)$ for that neuron in that task is given by

$$\hat{\lambda}(t) = N^{-1}h^{-1}\sum_{i=1}^{M} K\{(t - X_i)/h\}, \tag{1}$$

where K is a kernel function and h denotes the bandwidth, and M is the number of spikes aggregated over the N trials. We began with the Gaussian kernel and the "second generation" bandwidth selection rule of Sheather & Jones (1991),

$$h = \left[\frac{\int K^2}{n\int(\lambda'')^2\{\int t^2 K\}^2}\right]^{\frac{1}{5}} \tag{2}$$

obtained by minimizing an asymptotic expansion of the mean integrated square error (MISE) with respect to h. Calculation of (2) was implemented for the Gaussian kernel by Venables and Ripley (1994, p. 183). Although the time period of interest is $[-300, 600]$ ms, we fitted $\lambda(t)$ by kernel smoothing for $t \in [-650, 750]$ ms to avoid making a correction for edge effects. The bandwidths in (2) for the 84 neurons ranged from 35 to 113 ms, with quartiles 52, 61, and 70 ms. These produced intensity estimates as versions of the peristimulus time histograms, shown in Figure 7, that were not as smooth as we expect the underlying neuronal firing intensities to be. We therefore decided to use a larger bandwidth of 100 ms; the more satisfactory results using this bandwidth are shown in Figure 8.

It is also of interest to understand the behavior of the population of neurons considered as a whole. One simple representation is a smoothed version of the aggregated data, shown in Figure 3, where the smoothing is accomplished with smoothing splines. The remainder of our work is devoted to quantifying the qualitative appearances gleaned from Figures 2 and 3.

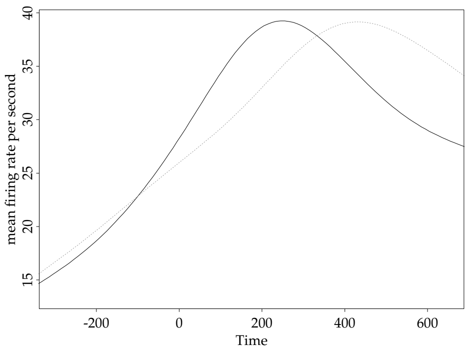

Fig. 3. Smoothed aggregate firing rate of the 84 neurons. The solid line corresponds to the spatial task, the dotted line to the pattern task.

4. Poisson Process Models

It is evident from Figures 2 and 3 (as well as from our examination of plots like Figure 2 for the rest of the neurons, which can be found in the Appendix in Figures 8 and 9) that, for most neurons, and on average in the neuron population, the intensity increases gradually up to some maximum and then declines somewhat until the end of the time interval of interest. This observation motivates our modeling efforts.

The simplest model for the sequence of single-neuron spike times $t_1, t_2, \ldots, t_n$ in an interval $(0, T]$ is a inhomogeneous Poisson process with intensity $\lambda(t)$. We will parameterize the intensity function, in ways to be specified below, generically using a multidimensional parameter θ, so that $\lambda(t) = \lambda(t; \theta)$. The likelihood function for a single trial is

$$L(\theta) = e^{-\int_0^T \lambda(u;\, \theta) du} \prod_{j=1}^{n} \lambda(t_j;\ \theta)\cdot$$

The likelihood for a given experimental task is the product of the likelihoods for each of the trials recorded for that task:

$$L(\theta) = \prod_{i=1}^{N} \left\{ e^{-\int_0^{T_i} \lambda(u;\theta) du} \prod_{i=j}^{n_i} \lambda(t_{ij}; \theta) \right\} \tag{3}$$

where t_{ij} is the jth spike time and $(0, T_i]$ is the total time interval for the ith trial, $i = 1, \ldots, N$. This, of course, assumes that the trials were independent. A referee has asked about this. In long experiments one may sometimes see obvious signs of change from the beginning to the end of the data collection period, for example, a shift in baseline rate or response strength. However such problems typically occur after more than an hour of recording whereas the present experiment can be completed in less than ten minutes. Furthermore, the eight experimental conditions were interleaved at random, so that the trials for each neuron in each condition were rarely adjacent in time. Thus, it would have been extremely surprising to have found any temporal dependence across trials and we did not examine this possibility. Aside from the random end times T_i that vary from trial to trial, we assume the trials are true replications.

It is important to keep in mind that our inhomogeneous Poisson assumption concerns the data *after aggregating across trials*. General limit theory (Daley and Vere-Jones, 1988, Theorem 9.2.V) makes it very plausible that data pooled across substantial numbers of replicated trials will approximately follow an inhomogeneous Poisson process. We return to this assumption in Section 5.

4.1. A Parametric Model of the Intensity Function

Based on substantive intuition and preliminary examination of the data, we first modeled the intensity function as a smooth hump-shaped function with constant initial and final segments, as represented in Figure 4 . That is, the intensity should be nearly constant for some period of time preceding the cue, then rise beginning when the monkey anticipates the cue, and finally decline to some other roughly constant level. Suppressing the dependence on θ, an analytic expression of such a function is:

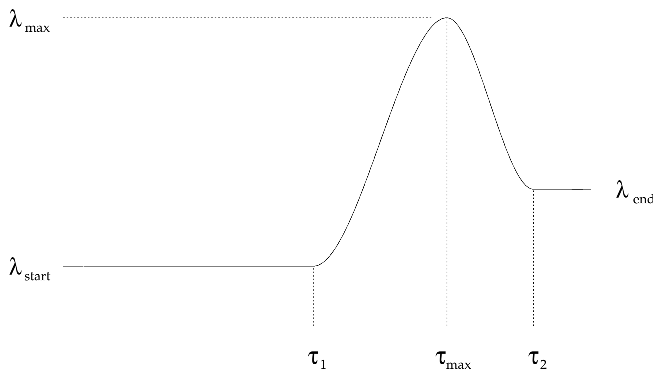

Fig. 4. Mean of the intensity function for the parametric model of Section 4.1.

$$\lambda(t) = \begin{cases} \lambda_{start}, & t \leq \tau_1 \\ \lambda_1(t) = a_3 t^3 + a_2 t^2 + a_1 t + a_0, & \tau_1 < t \leq \tau_{max} \\ \lambda_2(t) = b_3 t^3 + b_2 t^2 + b_1 t + b_0, & \tau_{max} < t \leq \tau_2 \\ \lambda_{end}, & t > \tau_2 \end{cases}$$

where λ_{start} is the initial level, τ_1 is the point in which the mean start to increase, $\lambda_1(t)$ and $\lambda_2(t)$ are 3−rd degree polynomials representing the increasing and decreasing parts, τ_{max} is the time of maximum firing rate λ_{max}, and τ_2 is the time at which the function returns to a final constant level λ_{end}. Assuming that $\lambda(t)$ has continuous first derivative, together with zero derivative at τ_1 and τ_2, this imposes a total of seven equality constraints, leaving $\lambda(t)$ with only six free parameters, which we take to be $\theta = (\lambda_{start}, \lambda_{end}, \lambda_{max}, \tau_1, \tau_{max}, \tau_2)$.

To fit the model we rewrote $\lambda(t)$ in terms of θ by solving the constraint equations symbolically using MAPLE (1981), then estimated θ via both maximum likelihood and Bayesian inference with independent diffuse Normal priors on each component of θ (via random-walk Metropolis-Hasting posterior simulation) to estimate θ. We determined that a Normal distribution based on maximum likelihood provided a very good approximation to the posterior by following the suggestions of Kass and Slate (1992). Specifically, we computed the global measure $(\bar{\theta}-\hat{\theta})^T\hat{\Sigma}^{-1}(\bar{\theta}-\hat{\theta})$ where $\bar{\theta}$ is the posterior mode, $\hat{\theta}$ is the MLE, and $\hat{\Sigma}$ is the approximate covariance matrix based on observed information, and we found the marginal Pearson skewness for each component; we also evaluated the correct posterior probability assigned to the putatively .05 tails determined from the approximate Normal distribution. These diagnostics indicated a satisfactory approximation. This model fit well for the large majority of an initial subsample of neurons (Carta, 1998), but we encountered computational and inferential difficulties for some neurons. These came from two sources. First, the random time delay between the cue and the signal to move meant that for some neurons there was very little data at the end of the experimental time period and thus little information with which to estimate the last change-point τ_2. Second, some neurons exhibited firing at the end of the period that did not conform to the model. Possible anticipation of the signal to move contaminated the supposedly constant ending firing rate. In any case, some neurons appeared to have a slightly elevated firing intensity at the very end of this period. We therefore omitted τ_2 from the model. We also simplified the procedure by binning the data and applying regression splines. This alternative approach is relatively easy to implement. It is described in Section 4.2.

4.2. Poisson Nonparametric Regression with Splines

Binning the count data of the inhomogeneous Poisson process is a useful simplification that, for small bin widths, loses very little information. Suppose we aggregate the spikes into bins B_k of width δ centered at t_k^*, $k = 1, \ldots, K$. We write the number of spikes in the k-th bin as y_k.

The Poisson regression likelihood function is

$$L^*(\theta) = e^{-\delta N \sum_{k=1}^{K} \lambda(t_k^*)} \prod_{k=1}^{K} \lambda(t_k^*)^{y_k}.$$

That $L^*(\theta)$ approximates $L(\theta)$ when δ is small follows from

$$\log \lambda(t_k^*) \doteq \tfrac{1}{y_k} \sum_{\{t_{ij} \in B_k\}} \log \lambda(t_{ij})$$

(so that $\lambda(t_k^*)^{y_k} \doteq \prod_{\{t_{ij} \in B_k\}} \lambda(t_{ij})$) and

$$\tfrac{1}{N} \sum_{i=1}^{N} \int_0^{T_i} \lambda(u) du \doteq \delta \cdot \sum_{k=1}^{K} \lambda(t_k^*).$$

We have used a bin width of $\delta = 10$ ms, as in Figure 2. Experimentation with different values of δ showed that there was little sensitivity to this choice.

The model in Section 4.1 provided a good fit to a large number of neurons using the three knots defined there. After a little experimentation we determined that two knots were generally adequate and the data could be fit by assuming the intensity had the cubic spline form

$$\log \lambda(t) = \beta_0 + \beta_1 (t - \xi_1)_+ + \beta_2 (t - \xi_1)_+^2 + \beta_3 (t - \xi_1)_+^3 + \beta_4 (t - \xi_2)_+^3$$

where $\xi_1 = -250$ ms and $\xi_2 = 200$ ms. Note that this particular form assumes the intensity is constant until 250 ms prior to the cue, which was as expected and also empirically verified. To implement the Poisson regression we created appropriate basis vectors, and applied `glm` in S-PLUS (Becker, Chambers and Wilks, 1988). Because the trials ended at varying times, we pooled all counts after 580 ms into a single bin and then adjusted for this by weighting the resulting count by (the reciprocal of) the number of 10 ms intervals included in this last interval.

5. Model Checking

In Section 4 we assumed the data pooled across trials came from an inhomogeneous Poisson process with intensity $\lambda(t)$. Although, as we indicated, general limit theory makes this a plausible

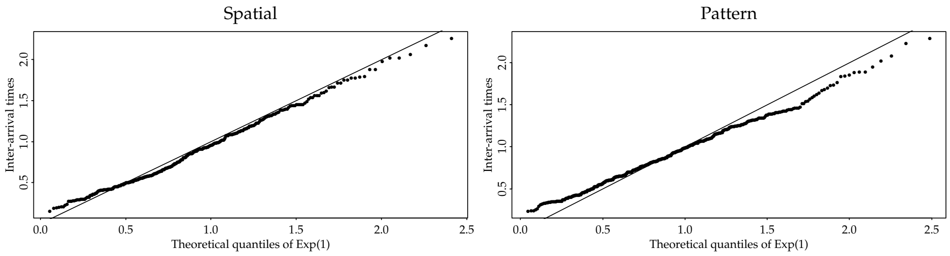

Fig. 5. Q-Q plots of the square-roots of the inter-firing times on the transformed time scale $\int_0^t \lambda(u)du$ against the square-roots of Exp(1) quantiles for neuron PK96c.1.

assumption after the data are pooled across trials, there is substantial evidence that cortical neurons in behaving animals often have non-Poisson spike times within trials (for discussion and references see Barbieri, Quirk, Frank, Wilson and Brown 1999; Gabbiani and Koch, 1998; Shadlen and Newsome, 1998; and Kass and Ventura, 2000) and in this experiment the number of trials was not very large. In any case, it is certainly good practice to check the Poisson assumption, which we do using a Q-Q plot. We also check the fit of the intensity function using Pearson's chi-squared goodness-of-fit criterion.

5.1. The Poisson assumption

Under the Poisson assumption of the data pooled across trials, the series of events on the transformed time scale $\tau(t) = \int_0^t \lambda(u)du$ is a Poisson process of constant unit rate (this is not hard to verify; see Cox and Lewis, 1966, p. 29). Hence the transformed inter-firing times are Exp(1), which can be checked via a Q-Q plot, as in Ogata (1988), Barbieri, *et al.* (1999), and Carta (1998). A technical complication arises because the spike times are actually recorded only to the nearest 1 ms: when the data are aggregated this produces local structure in the Q-Q plot. Ignoring this artifact, examination of the data revealed no systematic departures from the

Poisson assumption.

Here we have also gone further and checked the stronger assumption that the data follow an inhomogeneous Poisson process even within trials. In addition to examining the transformed inter-spike firing times from the pooled data we have formed the Q-Q plot by transforming the inter-spike firing times within trials, then aggregating these across trials. Again, $\lambda(t)$ is unknown; we estimate it via kernel smoothing so that the Poisson assumption will be assessed without relying on the regression spline model. Figure 5 displays such a Q-Q plot for neuron PK96c.1; we use the square-root scale to make deviations from linearity more readily visible for small quantiles. Although in most of its range the Q-Q plot is approximately linear, a departure from this relationship is apparent in a slight shortage of small inter-spike intervals, especially in the pattern task. This is likely to be due, at least in part, to the well-known phenomenon that immediately after each spike there is a refractory period during which the neuron does not fire or is much less likely to fire. The effect may also be exaggerated because the data are recorded to 1 ms accuracy. In any case, the departure from Poisson firing remains relatively small, especially when we are interested in data pooled across roughly 15 trials, which will smooth out mild deviations from the Poisson process assumption. The corresponding Q-Q plots for all 84 neurons are shown in Figure 10 in the Appendix. The large majority of neurons do not show serious departures from the Poisson assumption; only a few are questionable. We return to this point briefly in Section 8. Note that these Q-Q plots were insensitive to the choice of bandwidth in $[75, 125]$ ms.

5.2. Fit of the firing intensity

Under the Poisson assumption, the statistic $\frac{O_i - E_i}{\sqrt{E_i}}$ should be approximately $N(0,1)$, where O_i is the number of firing times in the i-th bin of a subdivision of the time axis, and E_i is the expected number in the same bin based on the maximum likelihood fit. This was examined informally by a normal probability plot, which indicated no systematic deviations from the model.

As a summary statistic, $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$ may be used. We compared the observed p-value to the reference distribution χ^2_{n-p}, with n the number of bins, and p the number of parameters in the spline model of Section 4.2. For neuron PK96c.1, this approximate test gave p-values of 4% and 97% for the two tasks ($n - p = 85$ with 10 ms bins). (The p-values do not change appreciably with a bin width in $[10, 30]$ ms.) We carried out the goodness of fit test for all 84 neurons and found that in the spatial and pattern tasks, respectively, ten and seven neurons were significant at a significance level of 5%, while three and six were significant at level 1%. This is somewhat above what would be expected under the null hypothetical model (the probabilities of observing these or more extreme events were 0.3%, 6%, 1% and 0.0002%, respectively). While this indicates that the model does not fit all neurons well, we were unable to identify a systematic departure that could be modelled. As an additional check, we therefore also carried out inferences using the kernel-based intensity estimator described in Section 3.

6. Inference

In the previous sections, we characterized the firing intensity quantitatively, and checked the assumptions of our model. We now compare formally the temporal evolution of the firing rates in the two tasks.

We examine three quantities: λ_{max}, the maximal firing rate; τ_{max}, the time at which this maximum occurs; and λ_{end} the mean end firing rate, that is the firing rate averaged over the interval $[500, 600]$ ms after presentation of the cue; these parameters are depicted in Figure 4. Superscripts s or p will be used to distinguish their values for the two tasks. Thus, $\phi^s = (\lambda^s_{max}, \tau^s_{max}, \lambda^s_{end})$ and $\phi^p = (\lambda^p_{max}, \tau^p_{max}, \lambda^p_{end})$ denote the vectors of features of interest in the spatial and pattern tasks, respectively. From Figures 2, 3, 8, and 9 it appears that for most neurons $\tau^p_{max} - \tau^s_{max} > 0$ and $\lambda^p_{end} - \lambda^s_{end} > 0$. We consider these possibilities and also check whether $\lambda^p_{max} - \lambda^s_{max} > 0$ using parametric and nonparametric methods. Although ratios or other comparisons could have been considered, we used $\phi^p - \phi^s$ because of its simplicity in terms of the familiar scales of firing rate and time.

We first analyze each neuron separately in Section 6.1 and then, in Section 6.2, we estimate the mean and covariance of the three quantities of interest among the population of neurons as represented by our sample of 84. The procedures are described below, and the results appear in Section 7.

Note that we will perform multiple tests without any adjustment for multiple inferences, which would be important if conclusions about specific neurons were of interest. Here it is of no scientific importance whether or not we find specific neurons to be erroneously significant. We will instead compare the number of significant neurons to its expectation under the null hypothesis, as we did in Section 5.2.

6.1. Individual neuron analysis

In the preceding paragraph we defined the three characteristics of the firing rate we used to contrast the neuronal responses under the two instructional cue tasks. These were selected after some preliminary examination of the data and one might wish to check for a difference between the two firing intensity functions with a conservative "omnibus" procedure, analogous to using an F-test in Analysis of Variance prior to examining particular contrasts. We give a simple procedure for doing so, and then move on to MLE and bootstrap-based methods of comparing the selected characteristics under the spatial and pattern tasks.

6.1.1. Functional Comparison of Two Curves

For each neuron we have used the same spline basis, described in Section 4.2, to fit the firing intensity functions $\lambda^s(t)$ and $\lambda^p(t)$. Thus, letting $\boldsymbol{\beta}^s$ and $\boldsymbol{\beta}^p$ be the vectors of spline coefficients, a test of $H_0 : \boldsymbol{\beta}^s = \boldsymbol{\beta}^p$ is, in fact, a test of $\lambda^s(t) = \lambda^p(t)$ for all t.

Let $\hat{\boldsymbol{\beta}}^s$ and $\hat{\boldsymbol{\beta}}^p$ denote the MLEs of $\boldsymbol{\beta}^s$ and $\boldsymbol{\beta}^p$, and $\boldsymbol{\Sigma}^s$, $\boldsymbol{\Sigma}^p$ the corresponding variance-covariance matrices (obtained from the information matrices). We noted in Section 4.1 that we are in a large-sample situation, so that we may take the MLEs to be Normally distributed. Under the null hypothesis the multivariate T^2 (Hotelling) statistic

$$T^2 = (\hat{\boldsymbol{\beta}}^s - \hat{\boldsymbol{\beta}}^p)'(\boldsymbol{\Sigma}^s + \boldsymbol{\Sigma}^p)^{-1}(\hat{\boldsymbol{\beta}}^s - \hat{\boldsymbol{\beta}}^p),$$

will therefore follow a χ^2_k distribution, where $k = \dim(\boldsymbol{\beta}^p) = 5$. This is a conservative test in the sense that it guards against chance discrepancies of arbitrary form between the two intensity functions; we specify methods of examining the discrepancies of interest, $\phi^s - \phi^p$, in the subsequent subsections.

For neuron PK96c.1, the observed value of T is 13.36, corresponding to p-value 0.02: we conclude that the intensity functions are different in pattern and spatial tasks. The p-values for the 84 neurons are reported in Section 7. From a Bayesian point of view, we could define the discrepancy

$$\Delta = (\boldsymbol{\beta}^s - \boldsymbol{\beta}^p)'(\boldsymbol{\Sigma}^s + \boldsymbol{\Sigma}^p)^{-1}(\boldsymbol{\beta}^s - \boldsymbol{\beta}^p),$$

and then compute posterior tail areas. This would yield similar results.

6.1.2. Component Tail Areas

To test $\phi^p_j - \phi^s_j = 0$ versus $\phi^p_j - \phi^s_j > 0$, for $j = 1, 2, 3$, we can use the fact that, asymptotically,

$$\hat{\phi}^p - \hat{\phi}^s \sim N_3(\phi^p - \phi^s, \hat{\Sigma}),$$

where $\hat{\phi}^s$ and $\hat{\phi}^p$ are the MLEs of ϕ^s and ϕ^p, and $\hat{\Sigma}$ is the variance-covariance matrix of $\hat{\phi}^p - \hat{\phi}^s$ based on the delta method, i.e., a linear transformation of (β^s, β^p) based a first-order Taylor series expansion, which is easily carried out by difference approximations to the derivatives. This yields tail areas that may be interpreted as either p-values or posterior probabilities.

6.1.3. Bootstrap Tests of Significance

In the previous two subsections we have used parametric methods to look for differences in the firing intensity functions for the two tasks. The resulting inferences depend on the validity of the Poisson assumption, of the spline model, and of the asymptotic distributions of the maximum likelihood estimators. For robustness of inference to these assumptions, we now turn to bootstrap resampling methods (e.g., Davison and Hinkley, 1997).

Note that the null hypothesis $H_0 : \phi^s = \phi^p$ does not constrain the distributions enough to determine a resampling plan under H_0. Instead we test the stronger null hypothesis

$$H_0 : \text{The point processes for the pattern and spatial tasks are the same.}$$

Under this hypothesis we may combine all the trials (assumed to be i.i.d. replications) for the two tasks then sample them at random with replacement, assign the first n_1 to the spatial task and the remaining to the pattern task, where n_1 is the original number of trials in the spatial task. This is the nonparametric bootstrap, which we apply here.

Other resampling plans are possible; see for example Cowling, Hall and Phillips (1996).

6.2. Population analysis

Our data set consists of a sample of 84 neurons from a large ensemble of similar neurons. As one might expect, these cells vary considerably with respect to the firing intensity characteristics we are investigating. To summarize the population effect while quantifying this variability, we begin with the asymptotic normality of the MLE $\hat{\phi}^p - \hat{\phi}^s$ already used Section 6.1.2 and add the second stage of a normal hierarchical model

$$\phi^p - \phi^s \sim N_3(\mu, D), \tag{4}$$

where μ is the population mean discrepancy between spatial and pattern tasks. We then use standard Bayesian methods to make population-level inferences about μ and D. That is, we introduce diffuse conjugate priors and obtain samples from the posterior distribution using BUGS (Gilks, Richardson, and Spiegelhalter 1994; Spiegelhalter, Thomas, Best and Gilks, 1996), and we check whether the results are sensitive to the choices of hyperparameters in the priors, which can be important (Natarajan and Kass, 2000). The sample is drawn haphazardly (through multiple insertions of the electrode) and there is no special ordering, nor any other apparent reason for departure from exchangeability among the measured neurons. We do, however, also check whether our assumption of i.i.d. Normality appears appropriate.

7. Results

We first carried out the Hotelling T test described in Section 6.1.1. We found that 75% of the 84 neurons have significantly different firing intensities in the two tasks with $p < 0{\cdot}05$, 66% with $p < 0{\cdot}01$, and 57% with $p < {\cdot}001$. For a large number of neurons there is clear evidence of a difference between the firing intensities in the two tasks.

Next, we verified that kernel and maximum likelihood estimates of $\phi^p - \phi^s$ were comparable: the values of the components differed by less than 5% among all but two neurons (neurons 41 and 72), for which τ^p_{max} and τ^s_{max} were substantially different. Examination of Figures 8 and 9 reveals that for these two neurons, the firing rate is very flat in at least one task, hence the time at its maximum ill-defined. This is an additional indication of the good fit of our model and it

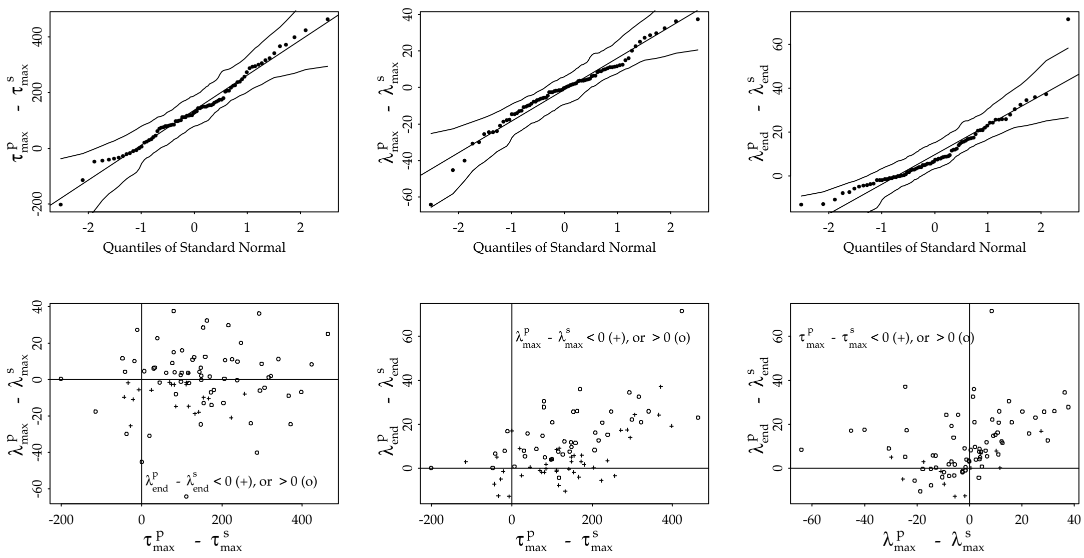

Fig. 6. Normal probability plots and scatter plots of the 84 three-dimensional MLEs. The straight lines added to the Normal probability plots have intercepts and slopes equal to the population posterior means and SDs found in Table 1. The confidence bands have estimated joint coverage of 85%.

also suggests that our inferences are likely to be consistent across estimation methods.

Figure 6 shows normal probability plots of the MLEs of the components of $\phi^p - \phi^s$, and how they vary and covary across the sample of neurons. The straight lines added to the normal probability plots have intercepts and slopes equal to the population posterior means and SDs found in Table 2. The curved lines are joint confidence bands obtained by bootstrap simulation, as in Davison and Hinkley (1997, pp.152-154); the envelope has estimated joint coverage of 85%. The MLEs (and therefore kernel estimates) do not deviate substantially from Normality, which makes the Normal assumption (4) in Section 6.1.2 reasonable. These plots indicate that the mean maximum firing rate λ_{max} is about the same for the two tasks, whereas the end rate λ_{end} and time of maximal firing rate τ_{max} are quite different; there is considerable variability and, for a

Table 1. *Percentage of neurons (out of 84) that have significant differences at the stated significance level. For each analysis method and each statistic, the percentages of significant neurons is presented as P / S, where P (S) indicates the percentage of neurons for which the statistic is significantly greater than zero, i.e. pattern is greater than spatial (smaller than 0).*

	at 5 % significance		
analysis	$\tau^p_{max} - \tau^s_{max} > 0 \ / < 0$	$\lambda^p_{max} - \lambda^s_{max} > 0 \ / < 0$	$\lambda^p_{end} - \lambda^s_{end} > 0 \ / < 0$
regr. spline	48.8 % / 3.6 %	25 % / 29.8 %	52.4 % / 10.7 %
bootstrap	48.8 % / 7.1 %	11.9 % / 21.4 %	39.3 % / 6.0 %
	at 1 % significance		
analysis	$\tau^p_{max} - \tau^s_{max} > 0 \ / < 0$	$\lambda^p_{max} - \lambda^s_{max} > 0 \ / < 0$	$\lambda^p_{end} - \lambda^s_{end} > 0 \ / < 0$
regr. spline	40.5 % / 2.4 %	19 % / 21.4 %	44.0 % / 6.0 %
bootstrap	23.8 % / 2.4 %	4.8 %/ 13.1 %	29.8 % / 1.2 %

majority of neurons, the estimates of $\lambda^p_{end} - \lambda^s_{end}$ and of $\tau^p_{max} - \tau^s_{max}$ are positive. Judging from these figures, firing in the pattern task appears to be delayed, and stronger at the end of the experiment for most neurons, as was suggested by Figures 2 and 3.

We now focus on the significance of each neuron individually. Table 1 gives the percentages of neurons that are significant, using the methods described in Section 6. The line “Bootstrap” in Table 1 corresponds to using the bandwidth of 100 ms. (Results using instead the bandwidth of (2) were generally similar, though the agreement with ML was not as good.)

Using ML estimation for our regression splines we found that for 41 (49 %) of the 84 neurons the pattern cue produced a statistically significant delay in maximal firing rate ($p < \cdot 05$) as compared with 3 (3.6 %) for which the spatial cue produced a delayed maximal firing rate. The bootstrap analysis similarly produced 41 (49 %) and 6 (7.1 %). For 44 (52 %) neurons the pattern cue produced a statistically significant greater firing rate during 500-600 ms after the cue ($p < \cdot 05$) as compared with 9 (10.7 %) for which the spatial cue produced a greater end firing rate, using ML estimation, and the bootstrap analysis produced 33 (39 %) and 5 (6 %). In contrast, for the maximal firing rate, roughly equal numbers of neurons had statistically significant greater maximal firing rates in pattern or spatial tasks, with the numbers being smaller than those for which pattern produced a significant delay or elevated end firing rate. (Here, the Bootstrap numbers were somewhat different from those obtained with regression splines, but even there we do not see an extremely lopsided distribution of differences in λ_{max} as we did for τ_{max} and λ_{end}.)

The individual neuron results must be considered in the context of Figure 6. From that figure it is clear that there is a continuum in the neuron population regarding the features of interest, which may be summarized by the population analysis of Section 6.2. Table 2 gives the posterior means and SDs of μ and the posterior means of the population standard deviations, i.e., the square-roots of the diagonal elements of D in (4), obtained by MCMC (Section 6.1.2). Note that the empirical population means and standard deviations for the MLEs are quite close to those presented in Table 2, as shown in Figure 6. The population mean delay to maximum firing intensity is 137 ± 17 ms with population standard deviation 140 ms and the mean elevation in end firing rate is $9{\cdot}7 \pm 1{\cdot}6$ spikes per second with standard deviation 14 ms.

8. Discussion

We have shown how statistical methods may be used to answer questions about the temporal evolution of neuron spike times. The main substantive contributions of this work are to document and quantify the delayed build-up to maximum firing rate among neurons in the SEF for the endogenous pattern cue as compared to the exogenous spatial cue; we found a large variation

Table 2. *Posterior summaries for population parameters in (4).*

	$\tau^p_{max} - \tau^s_{max}$	$\lambda^p_{max} - \lambda^s_{max}$	$\lambda^p_{end} - \lambda^s_{end}$
units	ms	rate per sec	rate per sec
posterior mean of μ	137.1	-1.15	9.67
posterior SD of μ	17.7	2.06	1.59
posterior mean of $\sqrt{D_{ii}}$	140.1	18.31	14.08

among neurons with about 49% having significant delays ($p < \cdot 05$), the population mean delay being 137 ($SE = 18$) ms with a population standard deviation of 140 ms, and 52% having significantly elevated end firing rate ($p < \cdot 05$), the population mean elevation in firing rate being 9.7 ($SE = 1{\cdot}6$) spikes per second with a population standard deviation of 14 spikes per second. See Olson, *et al.* (2000) for detailed substantive discussion. Our methods were based on an assumption that the spike times pooled across trials followed a inhomogeneous Poisson process, which appeared justifiable according to our model checks but we also checked the results using a nonparametric Bootstrap significance test.

We began this work by using a parametric form for the intensity function, and carried out statistical inference using ML estimation and MCMC. This was effective for a large majority of the neurons we initially examined with this approach, but failed to fit many neurons (partly due to the random interval length T_i in the experimental design). This kind of model may be useful in similar problems, but the easier and more flexible regression spline approach is very appealing. An important issue in applying regression splines, of course, is the choice of knots. In this experiment, as in most neuron recording studies in alert animals, the data set includes a fairly large number of responsive neurons. Thus, some exploratory fitting with a few neurons prior to systematic use in the whole sample may be informative and is unlikely to cause inferential difficulties. Results using kernel smoothers (or smoothing splines) were very similar to those with regression splines for this data set. In general, however, the inability of kernel smoothers and smoothing splines to adapt to varying degrees of smoothness over time may be problematic. Free-knot splines (DiMatteo, Genovese, and Kass, 2000) can provide useful and more automatic alternatives.

The general point of view we have adopted here is consistent with the functional data analysis perspective of Ramsay and Silverman (1997). In particular, we believe functional significance tests of the kind we displayed in Section 6.1.1 should be useful supplements to the usual t-tests and ANOVAs typically performed on neuron firing rates. There, and elsewhere, we relied on the asymptotic normality of MLEs (or of the MLE-based normal approximation to the posterior). Results in Slate (1994) combined with Kass and Slate (1992) suggest that for Poisson regression with a log (canonical) link the normal approximation is adequate when there at least 8 observations per parameter; here we had 5 parameters and typically more than 150 spikes per neuron. Our checks confirmed the prediction of this rule of thumb.

We also introduced a formal hierarchical model to summarize the population mean effects and variability while accounting for the uncertainty due to estimation of the intensity functions. We provided individual-neuron significance tests in order to make clear that the estimation uncertainty was generally small compared to the size of the effects of interest. In doing so we also considered corrections for multiple comparisons (Carta, 1998) but did not present these because they add complexity and did not materially contribute to our substantive conclusions.

An advantage of our treatment using binned data, pooling across trials, is the substantially increased applicability of the assumption that the data follow an inhomogeneous Poisson process. We judged this approximation adequate for our work, with our bootstrap checks showing that the conclusions here remain justified even in the presence of small departures from non-Poisson variability. In general, however, the assumption is restrictive and, in fact, our within-trial Q-Q plots revealed some small departures from the Poisson assumption. Thus, within-trial analyses will typically have to account for non-Poisson behavior. In subsequent work (Kass and Ventura, 2000) we have used regression splines to fit a conditional intensity and thus have provided a framework for generalization of the methods applied here.

Our work supports the rather general statistical conclusion that parametric modeling with Bayesian inference and the nonparametric bootstrap can complement each other effectively: the Bayesian parametric machinery is powerful under specific assumptions, while potential conflicts with the bootstrap would signal deviations from those assumptions. We are inclined toward the regression-spline results that produced intensity functions similar to those of the model we presented in Section 4.1. We expect the intensity functions to vary slowly, and the shape modeled parametrically in Section 4.1 captures key interpretable characteristics of the firing rate. On the other hand, our use of bootstrap methods provided indispensible support for what otherwise would have been potentially worrisome conclusions.

References

Bair, W., Koch, C., Newsome, W. and Britten, K. (1994). Power spectrum analysis of bursting cells in area MT in the behaving monkey. *J. Neurosci.* **14** 2870–2892.

Barbieri, R., Quirk, M.C., Frank, L.M., Wilson, M.A. and Brown, E. (1999). Non-Poisson stimulus-response models of neural spike train activity. *Submitted*

Becker, R.A., Chambers, J.M. and Wilks, A.R. (1988). *The NEW S Language.* New York: Chapman and Hall.

de Boor. C. (1978). *A Practical Guide to Splines.* Berlin: Springer-Verlag.

CARTA, R. (1998). A parametric model of neuronal firing rate in behaving monkeys. Advanced Data Analysis Report, Department of Statistics, Carnegie Mellon University.

COWLING, A., HALL, P. and PHILLIPS, M.J. (1996). Bootstrap Confidence regions for the intensity of a Poisson point process. *Journal of the American Statistical Association* **91** 1516–1524.

COX, D.R. and LEWIS, P.A.W. (1966). *The Statistical Analysis of Series of Events.* London: Methuen.

DALEY, D.J. and VERE-JONES, D. (1988). *An Introduction to the Theory of Stochastic Processes.* New York: Springer-Verlag.

DAVISON, A.C. and HINKLEY, D.V. (1997). *Bootstrap Methods and Their Applications.* Cambridge University Press.

DIGGLE, P. and MARRON, J.S. (1988). Equivalence of smoothing parameter selectors in density and intensity estimation. *Journal of the American Statistical Association* **83** 793–800.

DIMATTEO, I., GENOVESE, C. and KASS, R.E. (2000). Bayesian curve fitting with free-knot splines. *Submitted.*

GABBIANI, F. and KOCH, C. (1998). Principles of spike train analysis, in *Methods in Neuronal Modeling: From Ions to Networks.*, edited by C. Koch and I. Segev, Cambridge MA: MIT.

GILKS, W.R., RICHARDSON, S. and SPIEGELHALTER, D.J. (1996). *Markov Chain Monte Carlo in Practice.* London: Chapman & Hall.

HASTIE, T. and TIBSHIRANI, R. (1990). *Generalized Additive Models.* London: Chapman & Hall.

HUBEL, D.H. (1995). *Eye, Brain, and Vision.* New York: Scientific American Library.

IYENGAR, S. and LIAO, Q. (1997). Modeling neural activity using the generalized inverse Gaussian distribution. *Biol. Cyber.* **77**: 289-295.

KASS, R.E. and SLATE, E.H. (1992). Reparametrization and Diagnostics of Posterior Non-Normality, in *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting*, edited by J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith. Oxford: Claredon Press.

KASS, R.E. and VENTURA, V. (2000). A spike train probability model. *Neural Comput.*, to appear.

OGATA, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association* **83** 9–27.

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. Neurophysiology*, **84**, 1369–1384.

MAPLE (1981). Maple V Release 5. Copyright (c) 1981-1997 by Waterloo Maple Inc.

RAMSAY, J.O. and SILVERMAN, B.W. (1997). *Functional Data Analysis.* Springer-Verlag.

RICHMOND, B.J., OPTICAN, L.M. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. II. Quantification of response waveform. *J. of Neurophysiol.* **57** 147–161.

SHALDEN, M.N. and NEWSOME, W.T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of Neuroscience* **18**:3870-3896.

SHEATHER, S.J. and JONES, M.C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society, Series B* **53** 683–690.

SILVERMAN, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman & Hall.

SPIEGELHALTER, D.J., THOMAS, A., BEST, N.G. and GILKS, W.R. (1996). *BUGS: Bayesian inference Using Gibbs Sampling, Version 0.5.* Cambridge: Medical Research Council Biostatistics Unit.

VENABLES, W.N. and RIPLEY, B.D. (1994). *Modern Applied Statistics with S-Plus.* New York: Springer-Verlag.

Appendix

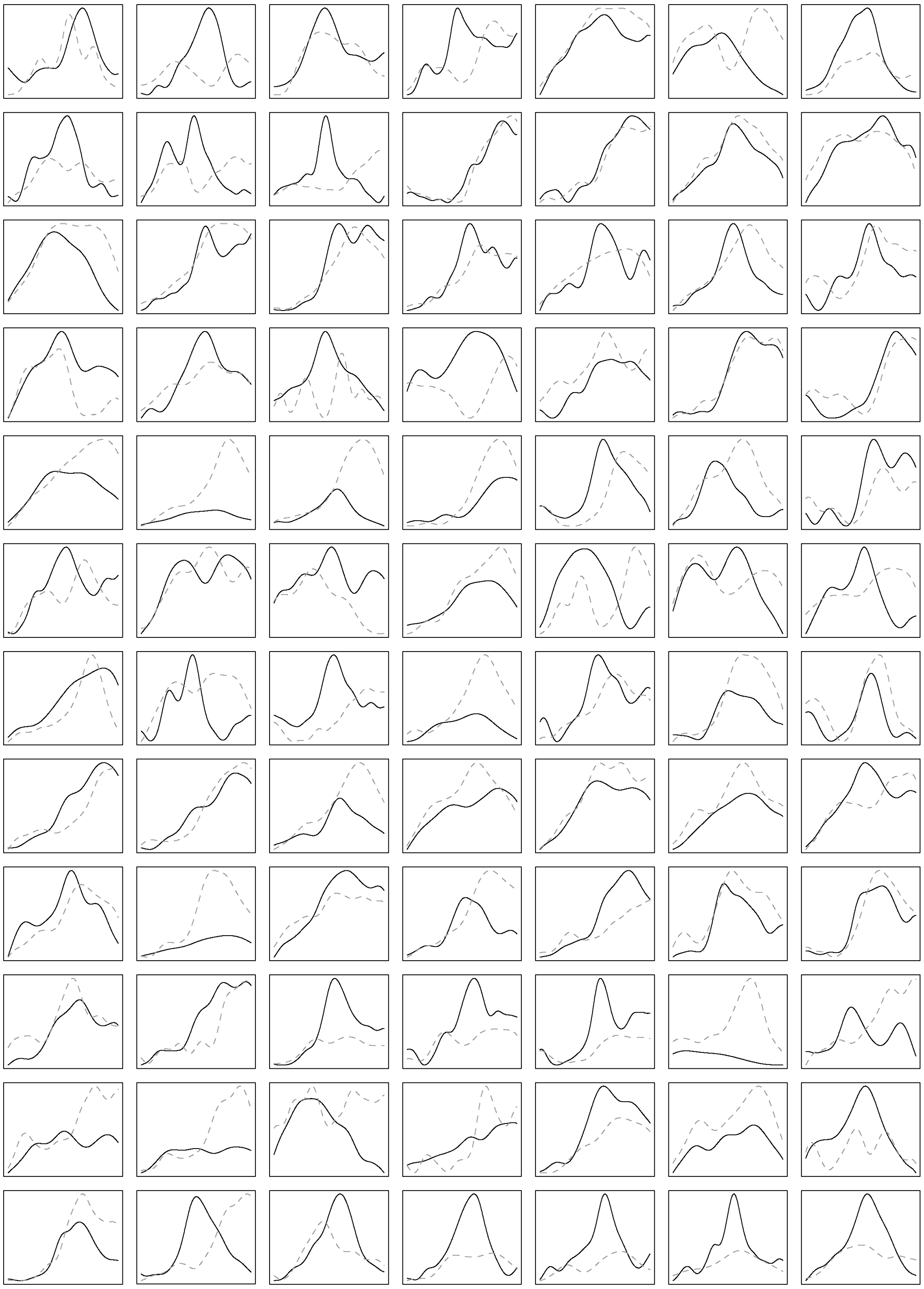

Fig. 7. Kernel smoothing fits using bandwidth (2) to the firing rates of the 84 neurons, as in the lower panels of Figure 2. The solid curves are for the spatial task, the dotted for the pattern task.

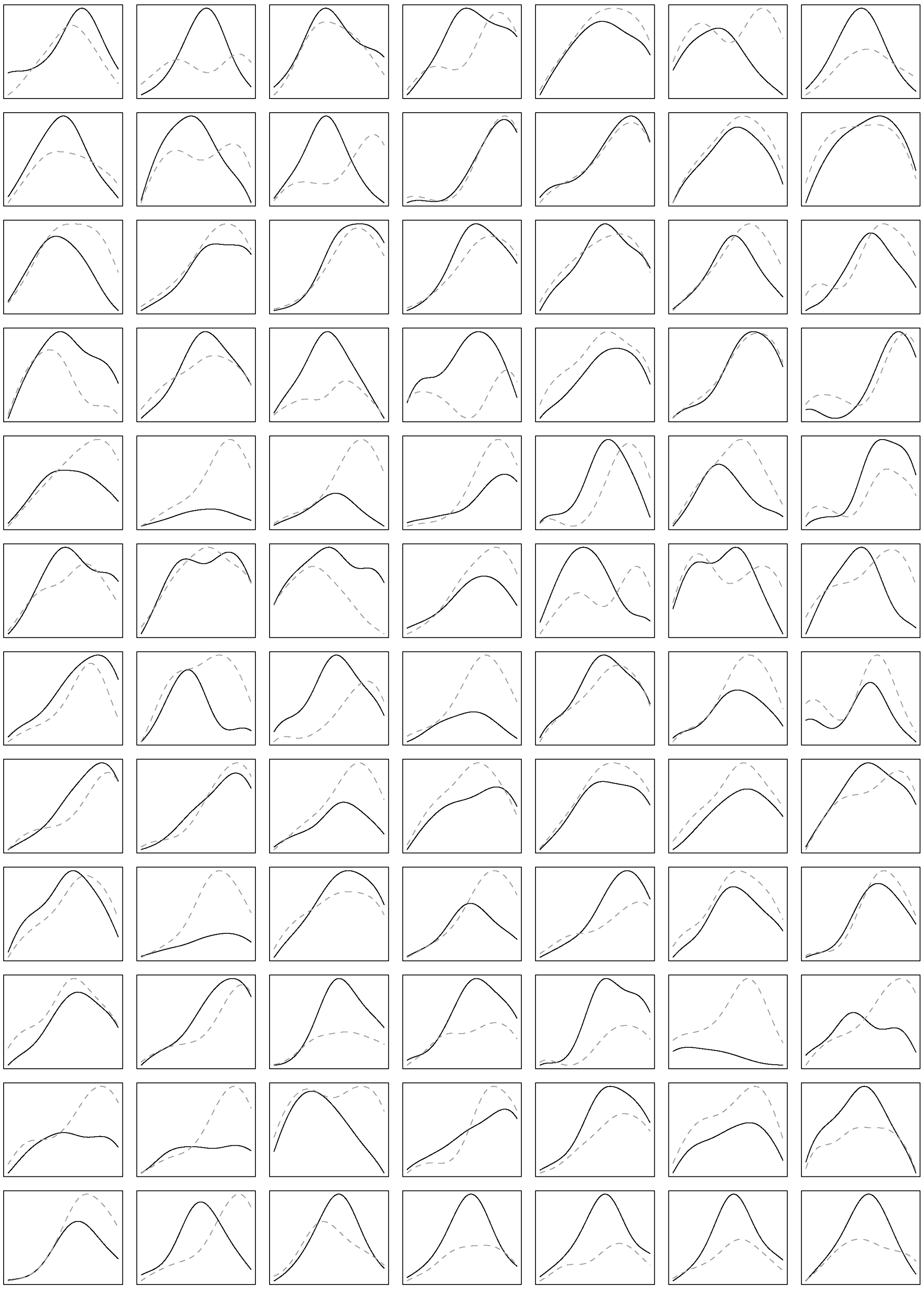

Fig. 8. Kernel smoothing fits using bandwidth $h = 100$ ms to the firing rates of the 84 neurons, as in the lower panels of Figure 2. The solid curves are for the spatial task, the dotted for the pattern task.

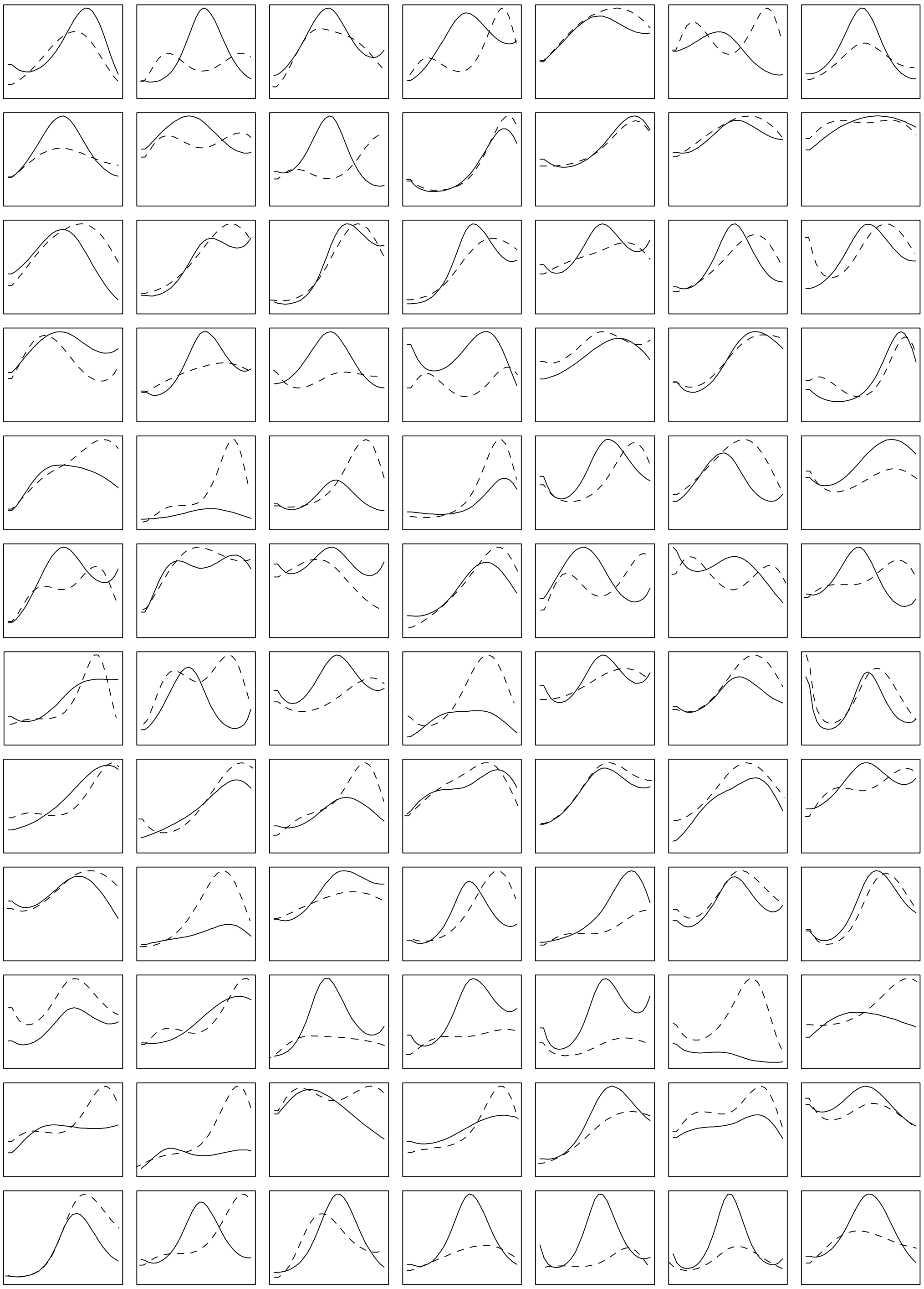

Fig. 9. Smoothing spline fits to the firing rates of the 84 neurons, as in the lower panels of Figure 2. The solid curves are for the spatial task, the dotted for the pattern task.

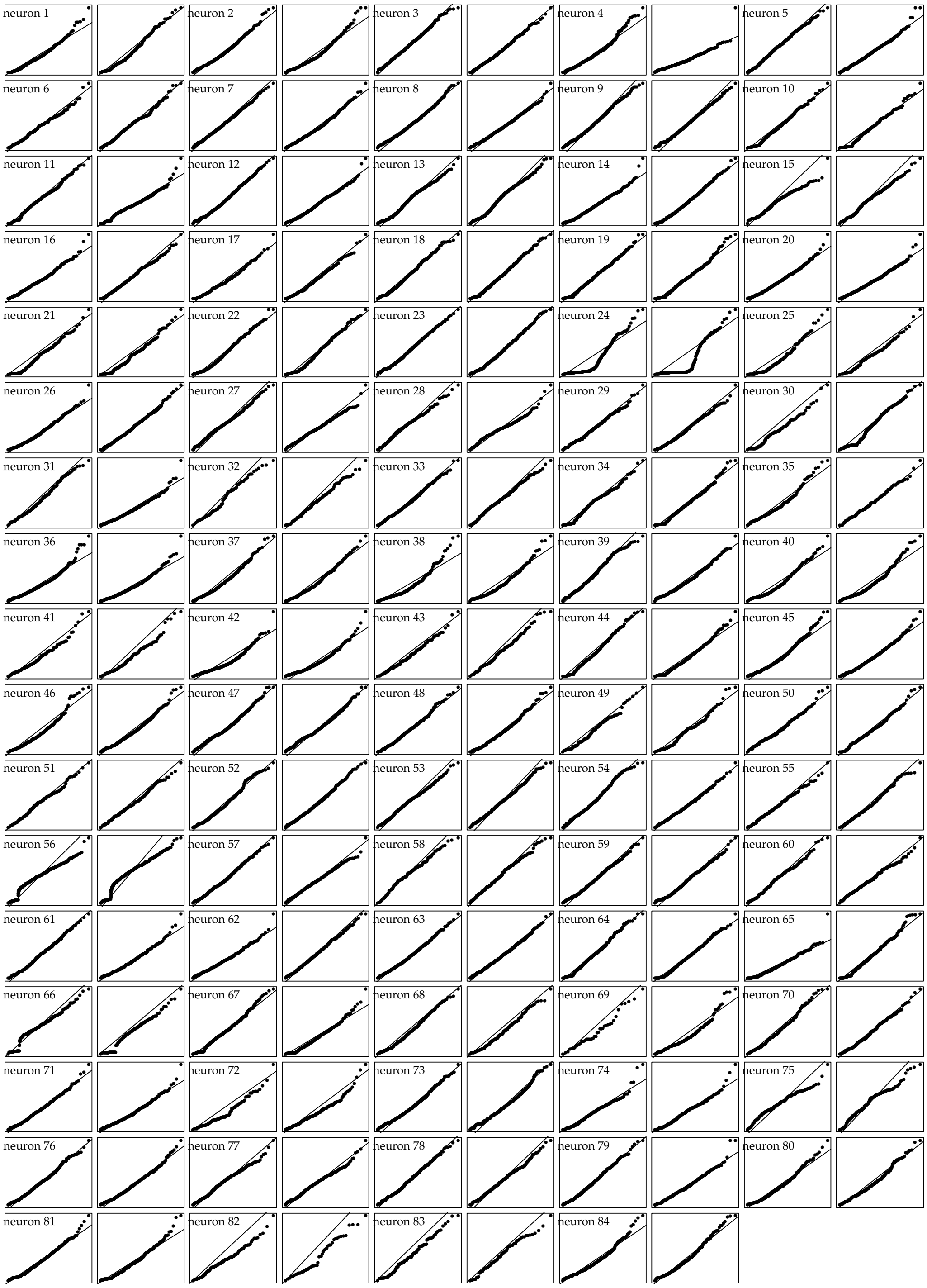

Fig. 10. Q-Q plots of the square-roots of the inter-firing times on the transformed time scale $\int_0^t \lambda(u)du$ against the square-roots of the quantiles of an Exp(1) distribution. Note that these plots are insensitive to the choice of bandwidth within $[75, 125]$ ms.

List of Tables

LIST OF FIGURES

Figure 10: Q-Q plots of the square-roots of the inter-firing times on the transformed time scale $\int_0^t \lambda(u)du$ against the square-roots of the quantiles of an Exp(1) distribution. Note that these plots are insensitive to the choice of bandwidth within $[75, 125]$ ms.

[*Received*]