

Testing equality of several functions: Analysis of single-unit firing-rate curves across multiple experimental conditions

Sam Behseta^{1,*}, Robert E. Kass², David E. Moorman³ and Carl R. Olson⁴

¹*Department of Mathematics, California State University, Bakersfield, CA 93311, U.S.A.*

²*Department of Statistics and Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.*

³*Department of Neurosciences, Medical University of South Carolina, Charleston, SC 29425, U.S.A.*

⁴*Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.*

SUMMARY

We consider the problem of comparing noisy functions, here trial-averaged neuronal firing-rate curves, across multiple experimental conditions. Of interest are comparisons both within neurons and also among populations of individually recorded neurons. We propose likelihood ratio tests to perform comparisons either pointwise or globally over the entire experimental time. A simulation study of power demonstrates the strength of these tests even for moderate sample sizes. We implement these tests on a group of 233 neurons recorded from primate frontal oculomotor cortex, first, to screen for condition-related differential activity and, second, to search for neurons displaying interesting time-locked features that vary with condition. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: BARS; functional data analysis; multiple curve fitting; neuronal data; PSTH

1. INTRODUCTION

Electrophysiological studies often examine *in vivo* the activity of individual neurons following various stimuli or preceding various movements. These investigations typically attempt to characterize differential responsiveness across experimental conditions (the stimuli or movements), in order to demonstrate that a certain type of information is available at a particular point in the neural processing stream. In some cases [1–3], simple summaries of spike counts, together with well-established statistical procedures such as *t*-tests or ANOVA, are sufficient to document the results—this occurs when timing is considered to be extraneous, and results are unlikely to depend

*Correspondence to: Sam Behseta, Department of Mathematics, California State University, Bakersfield, CA 93311, U.S.A.

†E-mail: sbhseta@csu.edu, statistician@gmail.com

Contract/grant sponsor: National Institute of Health; contract/grant number: RO1 MH064537-04

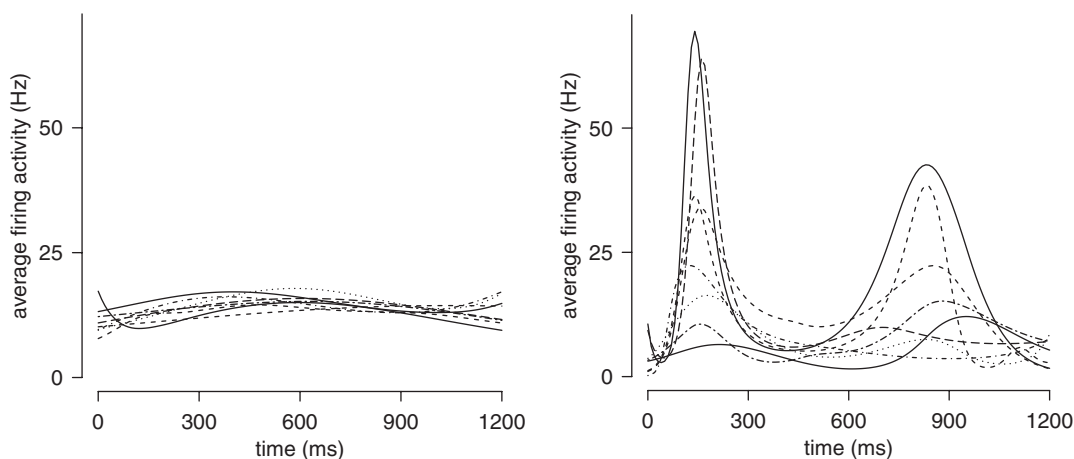


Figure 1. Fitted firing-rate functions for eight-directional eye movement conditions, in each of two SEF neurons. For the neuron on the left, the firing-rate functions were not significantly different according to the likelihood ratio test of Section 2.2.

much on the time interval chosen for the analysis. When it is important to focus on temporal response, it would remain possible to analyze spike counts obtained from many different time windows. Such an approach might be effective, but it would beg the question of how the time windows should be chosen. Furthermore, it does not take account of the smoothness of neuronal firing rates across time. An alternative is to begin with smooth estimates of firing-rate functions [4, 5], and to devise tests based on these. Fitted functions across eight experimental conditions, for each of two neurons, are shown in Figure 1. The figure suggests that the hypothesis of equality among firing-rate functions would be rejected for the second neuron, but not for the first. This illustrates the approach we take here: we provide simple modifications of standard methods from multivariate analysis to produce ANOVA-type tests of null hypotheses of equality among firing-rate functions across experimental conditions. We discuss both pointwise and global hypothesis tests.

The procedures we define and study are natural generalizations of those discussed in [6] for examining differences in firing-rate curves across two conditions. They are, in addition, applicable in any statistical setting that presents the problem of estimating multiple curves $f_1(t)$, $f_2(t)$, \dots , $f_J(t)$ and then testing the hypothesis

$$H_0: f_1(t) = f_2(t) = \dots = f_J(t) \quad (1)$$

either pointwise, for each t in a finite interval $[A, B]$, or globally, meaning simultaneously for all $t \in [A, B]$. These procedures thus provide generalizations of those in [7], which emphasized a spline-based fitting method called BARS [8]. The usual approach to testing $H_0: f_1(t) = f_2(t)$ is to assume time to be discrete and then to apply multivariate analysis. To be concrete, suppose we have, for each of $J=2$ experimental conditions, $j=1, 2$ (subsequently to become J conditions $j=1, \dots, J$), a vector of estimated function values $(\hat{f}_j(t_1), \dots, \hat{f}_j(t_p))$ at times t_1, \dots, t_p . In the case of neuronal firing rates, estimates $\hat{f}_j(t)$ may be fitted, say, every 10 ms. If we define $Y_j = (\hat{f}_j(t_1), \dots, \hat{f}_j(t_p))$, it may often be assumed, with good justification (e.g. for sample sizes typically found in neurophysiological studies), that Y_j is p -dimensional multivariate

normal. We write this as $Y_j \sim N_p(\mu_j, \Sigma_j)$, where Σ_j represents the uncertainty in estimating $(f_j(t_1), \dots, f_j(t_p))$. However, the usual Hotelling's T^2 test from multivariate analysis cannot be applied because (i) the matrices Σ_1 and Σ_2 are unequal; (ii) the matrices Σ_1 and Σ_2 are not estimated by sample covariance matrices; and (iii) the matrices Σ_1 and Σ_2 are singular (non-invertible). Furthermore, (iv) the number of time points p may be large, so that the test may not be very powerful (for example, see [9]). As discussed in [7], problems (i)–(iii) are easily fixed, while the use of BARS drastically reduces the dimensionality of the problem by fitting splines with a relatively small number of knots, thereby gaining power. We use an analogous approach here, modifying the usual multivariate ANOVA. In the applications reported here, we have used BARS to fit the curves initially and then applied likelihood ratio (LR) tests based on the asymptotic distribution of BARS fits. For brevity we omit a description of BARS and refer readers to [7] and the references cited there. The techniques we describe could be applied with any smoothing method.

We also generalize by considering populations of individuals (here, populations of neurons) indexed by $i = 1, \dots, I$. We then have functions $f_j^i(t)$ and wish to examine whether there is evidence of differences among conditions, on average, across the population. We may conceptualize the problem in this case as analogous to two-way ANOVA and MANOVA. An example of the kind of results obtained from a population-level analysis is given in Figure 2.

In Section 2, we first present methods for analyzing individual neurons, and then discuss population analyses. Each of these two situations includes both 'pointwise' analyses, which provide significance tests at every value of discretized time, and 'global' analyses, which evaluate the statistical significance of the totality of differential responses across time. Our point of view here is that the global analyses provide protection against false discovery of pointwise differences. Thus, one would first conduct global analyses and then, when significant, move on to pointwise analyses. In Section 3, we give results from a small simulation study of these approaches, where, to save

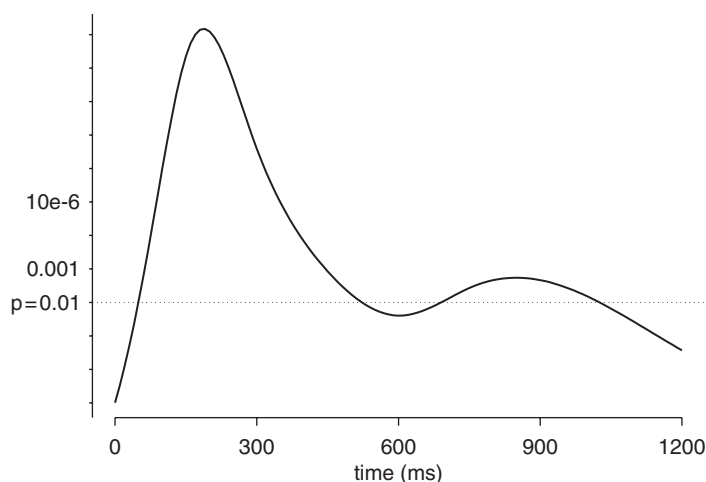


Figure 2. Pointwise likelihood ratio test for the population of neurons. The log-likelihood is plotted as a function of time (0–1200 after appearance of the target). The dotted horizontal line identifies significance at the 0.01 level. This test indicates differential firing roughly 100–500 ms post target, and again roughly 700–1000 ms post target.

computing time, we have used Normal kernel smoothing (Gaussian filtering) instead of BARS. In Section 4 we describe an analysis of data from 233 neurons recorded in the supplementary eye field (SEF) of two rhesus monkeys. We show how the methods developed here provide, first, a rigorous approach to identifying differences across conditions among firing-rate functions that evolve over time and, second, a way of identifying neurons that are involved in differing networks under differing experimental conditions. Section 5 contains discussion.

2. METHODS

We will present methods for the four cases separately: (i) pointwise analysis for a single neuron; (ii) global analysis for a single neuron; (iii) pointwise analysis for a population of neurons; and (iv) global analysis for a population of neurons. Before doing so, we introduce notation and discuss the general approach of applying the LR test. All of the discussion is stated in terms of neuronal data, although, as we emphasized at the outset, the methodology is general.

We index neurons by $i = 1, \dots, I$, and for each neuron we assume that we have spike train data from multiple trials $r = 1, \dots, R_i$ for each of several conditions $j = 1, \dots, J$. We begin by smoothing the Peri-Stimulus Time Histogram (PSTH) (using BARS or some other method) to get fitted vectors $Y_j^i = (Y_j^i(t_1), \dots, Y_j^i(t_p))$. We assume (as is the case with BARS) that the smoothing method also gives us corresponding covariance matrices Σ_j^i which, throughout, are taken to be fixed (known). That is, any uncertainty introduced by variance estimation is assumed to be of little importance and is ignored. The functions $f_j(t)$ appearing in the null hypothesis of equality (stated in the Introduction) are, in this context, trial-averaged firing-rate intensity functions.

We assume Y_j^i is multivariate normal, as it typically would be in practice, to a good approximation (with at least several trials per condition, and at least a handful of spikes per trial). We note, however, that Σ_j^i is typically (for large p) singular. For example, when smoothing with a B-spline basis having k knots, there would be only $k + 2$ degrees of freedom and the variance matrix would, therefore, have rank at most $k + 2$, which would typically be much less than the number p of grid points at which the functions are evaluated. We also assume that observations obtained under differing conditions are statistically independent.

The LR test has the general form

$$\text{LR} = \frac{\max_{\theta \in \Omega_0} f_{x|\theta}(x|\theta)}{\max_{\theta \in \Omega} f_{x|\theta}(x|\theta)} \quad (2)$$

in which θ is the vector of parameters generally lying in a parameter space Ω but restricted to lie in Ω_0 under H_0 . The density $f_{x|\theta}(x|\theta)$ is the probability density function of the data x . We derive LR tests in our four situations below, writing first the numerator of (2) and then its denominator. To obtain a p -value for testing H_0 , we then have various possibilities, depending on the four cases considered. In cases (i) and (iii), a χ^2 distribution may be used. In cases (ii) and (iv), we suggest rough choices of degrees of freedom for χ^2 , but in all cases parametric or nonparametric Bootstrap procedures could be used instead. For completeness, we provide some details.

2.1. Pointwise analysis for single neurons

Our pointwise analysis will occur across a sequence of time points, but by ‘pointwise’ we mean that each time t_k (for $k = 1, \dots, p$) will be considered separately. In addition, in this subsection,

we focus on an individual neuron. Thus, we drop the superscript i , and let Y_j be a single, arbitrary component of the vector $(Y_j(t_1), \dots, Y_j(t_p))$.

Under the assumption $Y_j \sim N(\mu_j, \sigma_j^2)$, the hypothesis H_0 is that of one-way ANOVA, but the within-group variances are assumed unequal and known. (Weerahandi [10] discusses the unequal-variance case in some detail.) In the numerator of the LR statistic, we obtain the maximum likelihood estimator (MLE)

$$\hat{\mu}_0 = \frac{\sum_{j=1}^J y_j w_j}{\sum_{j=1}^J w_j}$$

where $w_j = 1/\sigma_j^2$, while in the denominator we have $\hat{\mu}_j = y_j$. Therefore, we get

$$\text{LR} = \frac{\prod_j \left[\frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(\frac{-(y_j - \hat{\mu}_0)^2}{2\sigma_j^2}\right) \right]}{\prod_j \left[\frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(\frac{-(y_j - y_j)^2}{2\sigma_j^2}\right) \right]} \quad (3)$$

which simplifies to

$$-2 \log \text{LR} = \sum_{j=1}^J \left(\frac{y_j - \hat{\mu}_0}{\sigma_j} \right)^2 \quad (4)$$

Under H_0 and the approximate normality of the fitted values Y_j , the distribution of this statistic is approximately χ^2 with $J - 1$ degrees of freedom.

2.2. Global analysis for a single neuron

To treat this case, we write $Y_j^i = (Y_j^i(t_1), \dots, Y_j^i(t_p))$ and then, because we are still discussing a single neuron, we drop the superscript and assume $Y_j \sim N_p(\mu_j, \Sigma_j)$. The null hypothesis $H_0: \mu_1 = \dots = \mu_J$ is that of the usual multivariate ANOVA (MANOVA) setting, except that the variance matrices Σ_j are unequal and known. They are also not of full rank, and we therefore project the observed data vectors y_j to a lower-dimensional space where the projections do have full-rank variance matrices. We write the projections as $Q_j y_j$, where Q_j is the matrix made up of eigenvectors of Σ_j such that

$$\Lambda_j = Q_j * (\Sigma_j) * (Q_j)^T \quad (5)$$

is diagonal and full rank, as in [7]. Then, we can write $Q_j Y_j \sim N_{q_j}(Q_j \mu_j, \Lambda_j)$, where q_j is the rank of Λ_j .

In Appendix A, we show that the numerator of the LR has the form of a product (over j) of $N_{q_j}(Q_j \hat{\mu}_0, \Lambda_j)$ densities, where

$$\hat{\mu}_0 = \left[\sum_{j=1}^J Q_j^T \Lambda_j^{-1} Q_j \right]^{-1} \left[\sum_{j=1}^J Q_j^T \Lambda_j^{-1} Q_j y_j \right] \quad (6)$$

and the denominator has the form of a product of $N_{q_j}(Q_j \hat{\mu}_j, \Lambda_j)$ densities, where $\hat{\mu}_j = y_j$. In (6), $[\sum Q_j^T \Lambda^{-1} Q_j]^-$ is the generalized inverse (the ‘Moore-Penrose’ generalized inverse; e.g. [11, p. 56]) of $\sum Q_j^T \Lambda^{-1} Q_j$. We obtain

$$-2 \log(\text{LR}) = \sum_{j=1}^J (Q_j y_j - Q_j \hat{\mu}_0)^T \Lambda_j^{-1} (Q_j y_j - Q_j \hat{\mu}_0)$$

If the matrices Q_j were equal, and we had $q = q_1 = \dots = q_J$, then $-2 \log \text{LR}$ would have an approximately χ^2 distribution on $q(J - 1)$ degrees of freedom. We might therefore take $-2 \log \text{LR}$ to have roughly a χ^2 distribution on $\bar{q}(J - 1)$ degrees of freedom, where $\bar{q} = (1/J) \sum q_j$. In several simulations we have found the rough approximation to be quite adequate for purposes such as screening neurons for task-related activity. For example, simulating from $J = 3$ conditions with 20 trials, we observed an agreement between the quantiles of the empirical distribution of the simulated LR values and their counterparts obtained from a χ^2 distribution with $\bar{q}(J - 1)$ degrees of freedom, as shown in Figure 3.

An alternative is to apply a nonparametric Bootstrap procedure by resampling the trials while enforcing H_0 . Specifically, if there are n_1, \dots, n_J trials among the J conditions,

1. For $g = 1, \dots, G$
 - Resample the $n_1 + n_2 + \dots + n_J$ trials, then assign the resampled trials to J conditions, the first n_1 resampled trials being assigned to condition 1, the next n_2 to condition 2, etc.

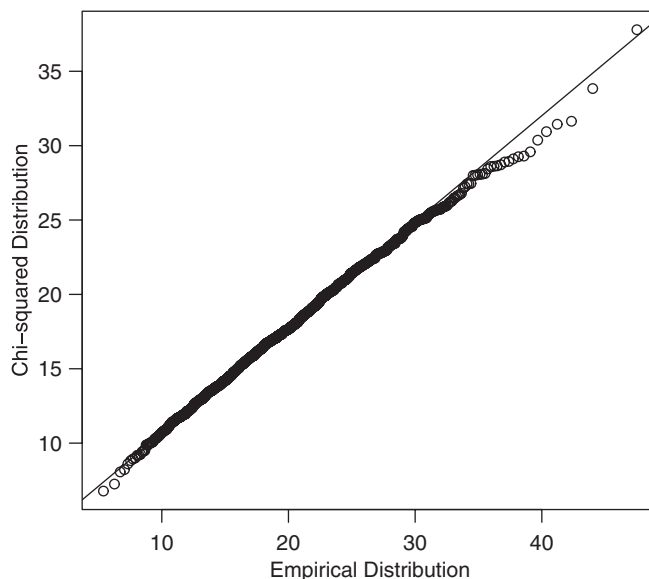


Figure 3. The qq -plot between the quantiles of the empirical distribution of the simulated likelihood ratio values with their counterparts obtained from a χ^2 distribution with $\bar{q}(J - 1)$ degrees of freedom. In this example, $J = 3$ conditions were considered according to the true firing-rate curves as shown in Figure 4. In this case, $\bar{q} = 7$.

- For each condition, form the PSTH and smooth it to obtain vectors $Y_1^{(g)}, \dots, Y_J^{(g)}$ in place of the original vectors.
 - Compute $-2 \log(\text{LR})$ based on $Y_1^{(g)}, \dots, Y_J^{(g)}$ instead of the original vectors.
2. Set r equal to the proportion of values g for which $-2 \log(\text{LR})$ based on $Y_1^{(g)}, \dots, Y_J^{(g)}$ was larger than $-2 \log(\text{LR})$ based on the original data. It should be noted that r is the Bootstrap estimate of the p -value.

2.3. Pointwise population analysis

We consider the set of fitted values $Y_j^i(t_k)$ at an arbitrary time point t_k and drop the reference to t_k , writing instead Y_j^i . The usual two-way ANOVA model may be written in the form

$$E(Y_j^i) = \mu + \alpha_i + \beta_j$$

but in this situation $V(Y_j^i) = (\sigma_j^i)^2$ is known. As a result, to maximize the likelihood function, we obtain the MLEs for μ , α_i , and β_j , as weighted least-squares estimates with weights $w_{ij} = 1/(\sigma_j^i)^2$ [12]. We have

$$\hat{\mu} = \frac{\sum_{i=1}^I \sum_{j=1}^J w_{ij} Y_j^i}{\sum_{i=1}^I \sum_{j=1}^J w_{ij}}$$

while α_i can be estimated by solving the equation

$$\hat{\alpha} = A^{-1}b$$

in which $A = \text{Diag}(w_{1+}, \dots, w_{I+}) - C$ and the ik th element of C is $c_{ik} = \sum_{j=1}^J w_{ij} w_{kj} / w_{+j}$, followed by replacing the last row of matrix A by w_{1+}, \dots, w_{I+} ; and the i th element of b is given by $b_i = \sum_{j=1}^J w_{ij} y_j^i - \sum_{i=1}^I (w_{ij} / w_{+j}) \sum_{k=1}^I w_{kj} y_j^k$, where $w_{i+} = \sum_{j=1}^J w_{ij}$ and $w_{+j} = \sum_{i=1}^I w_{ij}$. The mirror solution is applied to obtain estimates for β_j :

$$\hat{\beta} = B^{-1}d$$

where $B = \text{Diag}(w_{+1}, \dots, w_{+J}) - C$ and the jk th element of C is $c_{jk} = \sum_{i=1}^I w_{ij} w_{ik} / w_{i+}$, followed by replacing the last row of matrix B by w_{+1}, \dots, w_{+J} ; the j th element of d is given by $d_j = \sum_{i=1}^I w_{ij} y_j^i - \sum_{i=1}^I (w_{ij} / w_{i+}) \sum_{k=1}^J w_{ik} y_k^i$.

In the numerator of the LR , we have $\beta_j = 0$ for all j . Therefore, we get

$$\text{LR} = \frac{\prod_i \prod_j \left[\frac{1}{\sqrt{2\pi\sigma_j^i}} \exp \left(\frac{-(y_j^i - \hat{\mu} - \hat{\alpha}^i)^2}{2(\sigma_j^i)^2} \right) \right]}{\prod_i \prod_j \left[\frac{1}{\sqrt{2\pi\sigma_j^i}} \exp \left(\frac{-(y_j^i - \hat{\mu} - \hat{\alpha}^i - \hat{\beta}_j)^2}{2(\sigma_j^i)^2} \right) \right]}$$

which simplifies to

$$-2 \log \text{LR} = \sum_{i=1}^I \sum_{j=1}^J w_j^i (\hat{\beta}_j^2 - 2\hat{\beta}_j(y_j^i - (\hat{\mu} - \hat{\alpha}^i)))$$

Under H_0 and the approximate normality of the fitted values Y_j^i , the distribution of this statistic is approximately χ^2 with $J - 1$ degrees of freedom.

2.4. Global population analysis

In this case, we write $Y_j^i = (Y_j^i(t_1), \dots, Y_j^i(t_p))$ and assume $Y_j^i \sim N_p(\mu_j^i, \Sigma_j^i)$. We have $E(Y_j^i) = \mu_j^i$ and write

$$\mu_j^i = \mu + \alpha_i + \beta_j$$

while $V(Y_j^i) = \Sigma_j^i$. Using the dimensionality reduction projections [7] we have

$$Q_j^i Y_j^i \sim N_p(Q_j^i \mu_j^i, \Lambda_j^i)$$

where $\Lambda_j^i = Q_j^i * (\Sigma_j^i) * (Q_j^i)^T$. The null hypothesis is $H_0 : \beta_j = 0$, for all j .

To calculate the LR, first we obtain the maximum likelihood estimates $\hat{\mu}$, $\hat{\alpha}_i$, and $\hat{\beta}_j$ (see Appendix B for details).

Now, defining

$$\hat{l}_0 = \sum_{i=1}^I \sum_{j=1}^J [(Q_j^i y_j^i - Q_j^i(\hat{\mu} + \hat{\alpha}_i))^T (\Lambda_j^i)^{-1} (Q_j^i y_j^i - Q_j^i(\hat{\mu} + \hat{\alpha}_i))]$$

and

$$\hat{l} = \sum_{i=1}^I \sum_{j=1}^J [(Q_j^i y_j^i - Q_j^i(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j))^T (\Lambda_j^i)^{-1} (Q_j^i y_j^i - Q_j^i(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j))]$$

we get

$$-2 \log \text{LR} = -\hat{l}_0 + \hat{l}$$

The distribution of this statistic can be roughly approximated *via* a χ^2 with $\bar{q}(J - 1)$ degrees of freedom, where $\bar{q} = (1/J) \sum q_j$. We have compared this with a nonparametric Bootstrap in several simulated examples and judge the rough approximation, adequate except in borderline circumstances, where a Bootstrap procedure should be used. The Bootstrap is similar to that in Section 2.2, except that here the trials must be resampled separately for each neuron. Specifically, if there are n_1^i, \dots, n_J^i trials among the J conditions for the i th neuron.

1. For $g = 1, \dots, G$
 - For neurons $i = 1, \dots, I$,

resample the $n_1^i + n_2^i + \dots + n_J^i$ trials, then assign the resampled trials to J conditions; the first n_1^i resampled trials are assigned to condition 1, the next n_2^i to condition 2, etc. For each condition, form the PSTH and smooth it.

This produces vectors $Y_1^1(g), \dots, Y_J^1(g), \dots, Y_1^I(g), \dots, Y_J^I(g)$, in place of the original vectors.

- Compute $-2 \log(\text{LR})$ based on $Y_1^{(g)}, \dots, Y_J^{(g)}$ instead of the original vectors.
2. Set p equal to the proportion of values g for which $-2 \log(\text{LR})$ based on $Y_1^{(g)}, \dots, Y_J^{(g)}$ was larger than $-2 \log(\text{LR})$ based on the original data.

3. SIMULATION STUDY

We assessed the power of the LR tests under case (ii), the global test for single neurons. We did so under two scenarios, each with three distinct true firing-rate curves as indicated in Figures 4 and 5. The firing-rate curves were based roughly on the observed firing-rate function from an inferotemporal neuron displayed in [4].

For each firing-rate curve, we generated Poisson data in 10 ms bins for repeated trials. We implemented the simulation by considering a Bernoulli random variable $W(t)$ (taking the values 0 or 1), for $t = 1, \dots, 1000$. We generated a spike with $pr(W(t) = 1) = f(t)/100$, where $f(t)$ is the value of each curve in Figure 4. Finally, we smoothed the data using kernel estimators (formula 1 in [6]), with the choice of Gaussian kernel. The choice of bandwidth was based on the method of [13] (see [6, formula 2]; the R-algorithm of this method can be found in [11, p. 183]).

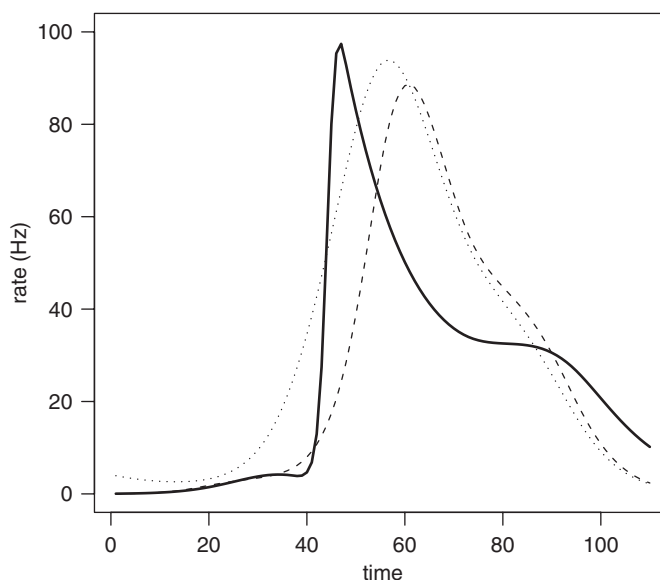


Figure 4. Curves used for power study under the first simulation scenario. In this case the curves are close to each other and, to distinguish them reliably, a fairly large number of trials are needed.

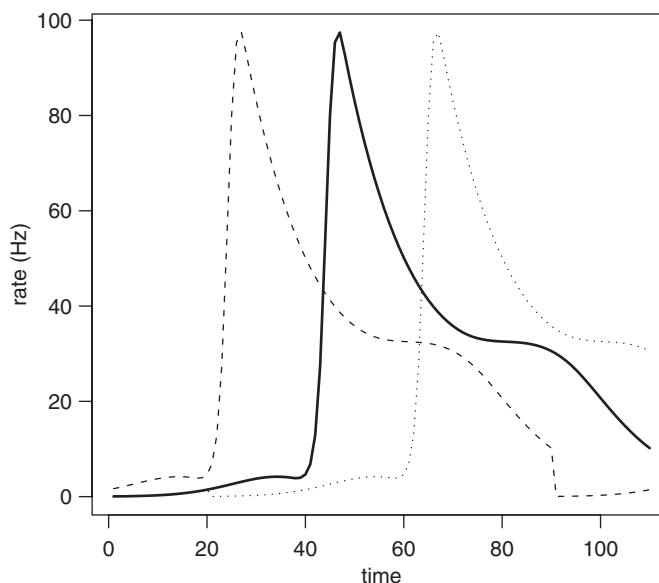


Figure 5. Curves used for power study under the second simulation scenario. In this case the curves are quite far apart, and relatively little data are needed to distinguish them reliably.

Table I. Simulation study of power under two scenarios for the alternative hypothesis.

Type I error	Scenario 1	Scenario 2
0.05	28	7
0.01	61	15

The minimum number of trials required to achieve a power of 0.80 is shown. For example, 28 trials are needed to achieve power of 0.80 under scenario 1 with a test having type I error of 0.05.

We display results in two ways. First, the minimum number of trials needed to achieve a power of 0.80 (a widely applied standard) is shown in Table I for type I error set to each of 0.05 and 0.01. The power based on 15, 30, and 60 trials is shown in Table II. It is apparent that the method is able to detect even relatively small differences with a moderate number of trials, and larger differences with a quite modest number of trials.

4. ANALYSIS OF SEF DATA

4.1. Data and scientific issues

The macaque SEF, an area on the dorsomedial shoulder of the frontal lobe, is involved in the generation of saccadic eye movements. Characterizations of SEF activity are discussed in [14] and the references therein. The data analyzed here come from a study conducted by Moorman and

Table II. Power of tests under the two scenarios.

Type I error	Scenario 1			Scenario 2		
	15	30	60	15	30	60
0.05	0.598	0.750	0.890	0.793	0.890	1.00
0.01	0.473	0.662	0.798	0.670	0.799	1.00

Results are shown as power achieved for 15, 30, and 60 trials. For example, in the first scenario with 15 trials and type I error of 0.05, the power is only 0.598.

Olson [15]. Our analysis is a demonstration of the capabilities of our methodology rather than a documentation of the properties of SEF neurons. For this purpose, we focus only on the spatial response of SEF neurons, and only on the ‘dot’ component of the ‘bar-dot’ task used by Moorman and Olson. A summary of this component of the task appears in Figure 6.

In the data considered here, SEF neuronal responses were obtained in $J = 8$ spatial conditions, each condition corresponding to a direction of eye movement. It is known that many SEF neurons are spatially tuned, so a given neuron will tend to fire most rapidly for movement in a particular direction—called the ‘preferred direction’ for that neuron. A range of 8–16 trials (repetitions) were obtained for each condition, for each neuron. The data for our analysis consist of the counts, accumulated across trials, in 10 ms bins, beginning 0 ms and ending 1200 ms after appearance of the target. We thus have data vectors of length 120. In our analysis, data from 233 SEF neurons were examined.

There are two issues we wish to address. First, there is the identification of task-related activity. In the setting we are considering here, this becomes identification of neurons for which there is differential directional response. Second, among directionally responsive neurons, we wish to identify those where there is a different temporal response in the preferred direction than in the anti-preferred direction. A neuron showing a distinct temporal pattern of responses would be involved in different functional networks for differing movement conditions.

4.2. Analysis

For each of $I = 233$ neurons in $J = 8$ directional eye movement conditions, we have a vector $Y_j^i = (Y_j^i(t_1), \dots, Y_j^i(t_p))$ of spike counts, accumulated across trials, with $p = 120$. One would expect many SEF neurons to be directionally tuned, with a maximal firing rate in a particular direction, called the cell’s ‘preferred direction’. With only eight-directional movements we cannot identify the preferred direction very precisely, and it is not our purpose here to provide a detailed model of directional tuning. (For methods of doing so, see [16].) However, the directional conditions do need to be defined in terms of the preferred direction; that is, in principle, we would want to define experimental conditions in terms of their rotation, say, clockwise, relative to the preferred direction. For this purpose, we began by identifying each neuron’s direction of maximal firing rate. We then defined this direction as Condition 1, and defined Conditions 2–8 as successive directions found by rotating clockwise from Condition 1. Thus, for every neuron, Condition 1 was defined as the direction of maximal firing rate; Condition 2 was defined as the direction obtained by rotating 45° clockwise from Condition 1, etc.

We applied the global analysis for single neurons to all 233 neurons, and identified those neurons for which the firing-rate functions across the eight-directional conditions were significantly

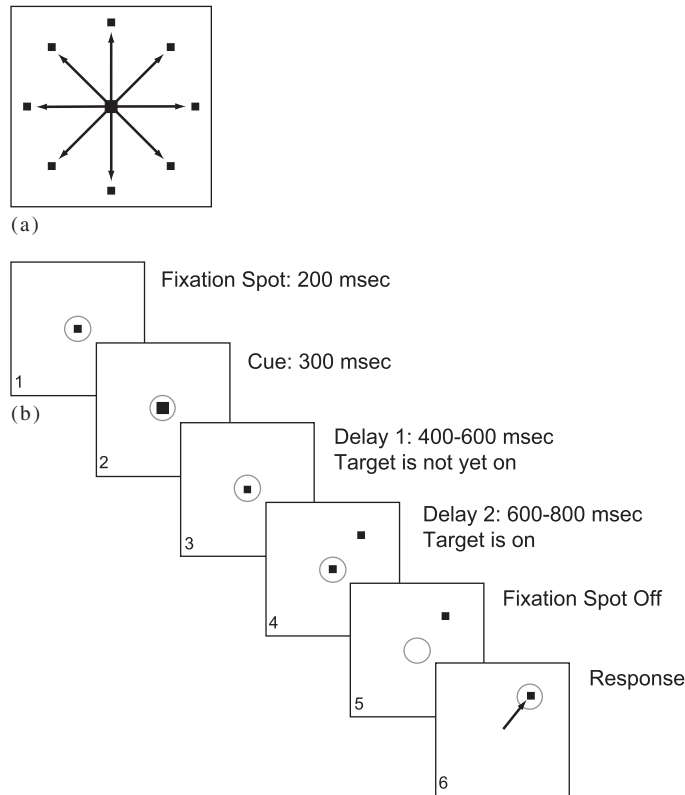


Figure 6. Directional visually guided saccade task. Part A of the figure indicates there were eight conditions in which the saccade target was a dot, each condition corresponding to a different direction. Part B shows the sequence of events during the trial. (1) A white fixation spot appeared at the center of the screen and the monkey achieved foveal fixation. (2) After 200 ms, the fixation spot was replaced by a gray target which instructed the monkey to prepare a saccade to the upcoming dot target. (3) A delay ensued. (4) The target display appeared and remained on during an ensuing delay period. (5) Offset of the fixation spot served as an imperative cue. (6) The monkey was required to make a saccade directly to the target.

different, with $p < 0.01$. Examples of fitted firing-rate functions for two neurons, one where they are significantly different and the other where they are non-significant, are shown in Figure 1. This eliminated 75 non-significant neurons, leaving us with 158 neurons having condition-dependent firing-rate functions.

We also applied a pointwise population analysis to identify time regions where there tend to be differential firing rates across neurons. Figure 2 shows the log-LR, plotted across time. A dotted horizontal line is drawn to indicate the 0.01 significance level. This figure indicates that most of the population-level differential firing occurs 100–500 ms post target, but there is also considerable differential response surrounding the eye movement, roughly 700–1000 ms post target.

The population-level differential response in Figure 2 could arise in at least two different ways. It could come from a population of firing-rate curves that tend to exhibit both early and late response, as in the second neuron in Figure 1, where the essential distinction among directions

is simply that each neuron tends to fire more rapidly in its preferred direction, with the overall firing rate diminishing as the directions move away from the preferred direction. In other words, the firing-rate functions might differ by additive or multiplicative constants. Another possibility is that the temporal response itself might change with direction. For example, for some neurons there may be a substantial response related to eye movement mainly in the preferred direction, which is suppressed for movement in the anti-preferred direction.

To search for such differential temporal responses that would not depend on the overall firing rate, we normalized the fitted firing-rate functions before subjecting them to analysis: we divided by the pointwise average firing rate, averaged across the eight conditions; that is, for each time t_k , we created observations $W_j^i(t_k) = Y_j^i(t_k)/\bar{Y}^i(t_k)$, the mean $\bar{Y}^i(t_k)$ being found separately for each neuron. We also reduced the directional data somewhat. Because the preferred direction was not exactly at Condition 1, but rather somewhere between Condition 1 and Condition 2 or between Condition 1 and Condition 8, and also because Conditions 2 and 8 typically had responses that were similar to those in Condition 1, we contrasted Conditions 8, 1, and 2 against Conditions 4, 5, and 6. We did this for each neuron by creating one vector of length $360 = 3 \times 120$, obtained by concatenating the data for Conditions 8, 1, and 2; and then we created a second vector of length 360 by concatenating the data for Conditions 4, 5, and 6. We then applied a global analysis of single neurons for $J = 2$ conditions (which reduces to a modified Hotelling's T^2 , as in [7]). This produced 20 neurons that were significant at the 0.01 level. The fitted firing-rate functions for these 20 neurons were examined by eye, and nine shared a particularly interesting pattern. As shown in Figure 7, these neurons demonstrated both early and late response at or near the preferred direction, but only early response at or near the anti-preferred direction.

5. DISCUSSION

The purpose of this article was to provide methods for testing equality among functions across multiple conditions. In the setting of neuronal data analysis, we have illustrated these methods by identifying, within a substantial sample, neurons that have condition-related differences among firing-rate functions; population-level differences across conditions; and neurons that exhibit interesting patterns of temporal difference across conditions. The methods are easily applied in high-level software such as R or MATLAB.

In our population-level analyses, we applied fixed-effects tests. An alternative would be to assume, as in [17], that each function is a Gaussian process and then, instead, use tests based on random-effects models. We explored this alternative. That approach requires additional effort and more specialized software. Unless assessments of variability among functions were of immediate interest (as in [17]), we do not think the extra effort would yield much useful additional information. The relative simplicity of the tests used here is part of their appeal.

A referee asked us to be more specific in justifying approximate multivariate normality for data sets such as those discussed in Section 4. Our confidence in the suitability of approximate normality is based on experience in this and related situations (see [18]). To demonstrate the accuracy of the approximation within the context of our BARS-based smoothing, we examined the posterior of the vector $\mu = (f_j(t_1), \dots, f_j(t_p))$, which is estimated by $y = (\hat{f}_j(t_1), \dots, \hat{f}_j(t_p))$. Specifically, we calculated the exact posterior probability of sets of the form $\{\mu : (\mu - y)^T \Sigma^{-1} (\mu - y) < c\}$ where c is the 0.95 quantile of the χ^2 distribution that would result if the posterior were exactly normal; if the posterior were highly non-normal, the probability of these sets would be different

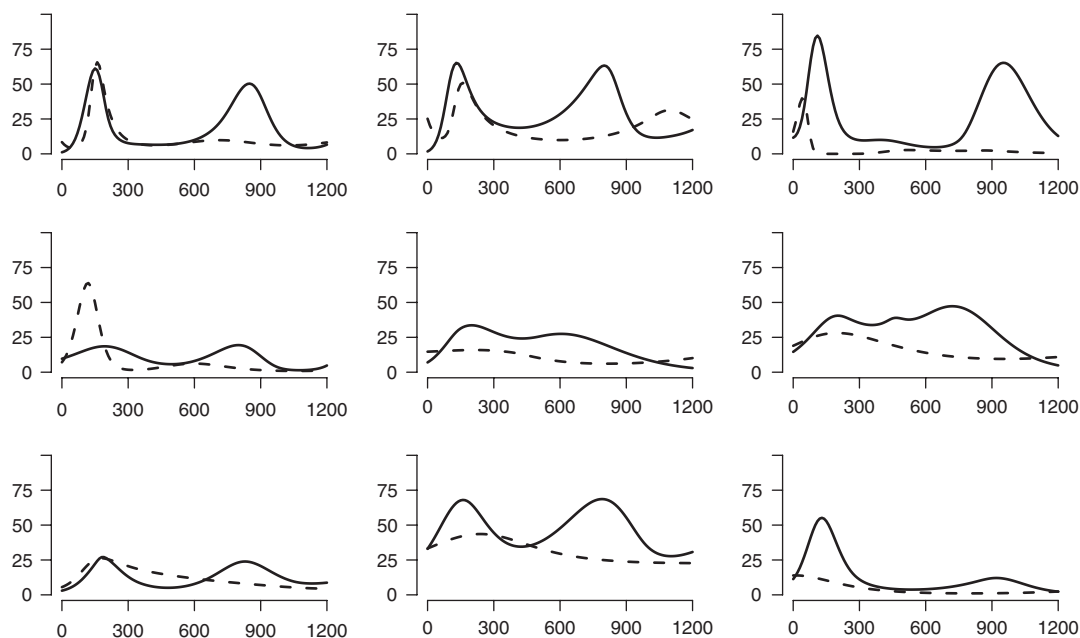


Figure 7. Firing-rate functions for nine neurons in preferred (solid) and anti-preferred (dashed) directions. The x -axis is time measured in milliseconds, and the y -axis represents the firing rate in Hertz. These neurons had normalized firing-rate functions that were significantly different ($p < 0.01$) at or near preferred *versus* at or near anti-preferred directions. At or near preferred directions these neurons responded both early (with target-related activity) and late (with movement-related activity), while at or near the anti-preferred direction the late response was suppressed.

than 0.95. (Posterior normality of μ is not the same as normality of Y , but posterior and frequentist Normality tend to go hand in hand, and the latter is harder to evaluate for BARS-based procedures.) We obtained results for five sets of data obtained from two neurons among those described in Section 4. The first neuron had 88 trials, and the second 10 trials. This gave us two of our data sets. We then randomly sampled 10 and 5 trials from the first neuron and 5 trials from the second neuron, so that we had two cases with 5 trials, two cases with 10 trials, and one case with 88 trials. The resulting posterior probabilities were obtained from 5 long Markov Chain Monte Carlo runs, having a simulation standard error of 0.005 (i.e. two-digit accuracy in the posterior probabilities). The 88-trial data set gave a posterior probability of 0.95 (in agreement with the putative figure to two digits), while all the other four data sets gave a posterior probability 0.94. In our experience, close agreement between exact and approximate central probabilities is common. We would expect that, with firing rates of at least several Hz and five or more trials, it would be unusual to obtain highly inaccurate p -values for the resulting multivariate tests.

In multi-trial neurophysiological data, there are sometimes trends across trials, as when learning occurs. In such situations, trial-averaged firing rates may be misleading. The methodology we have described should be of interest when condition-related temporal responses may reasonably be examined by averaging across trials. More specifically, from a theoretical point of view, in settings such as the one described in Section 4, the functions $f_j(t)$ have the form $f_j(t) = E(\lambda_{jr}(t|H_t))$,

where $\lambda_{jr}(t|H_t)$ is a conditional intensity function that describes the spiking behavior of a given neuron on trial r and the expectation is taken across trials. One situation in which this is clearly of interest is when there is no excess trial-to-trial variability, that is, no variability beyond that produced by a conditional intensity that contains no trial-specific effects [19]. However, analysis of trial-averaged firing-rate functions does not assume trial-to-trial effects to be absent; it merely assumes that the average behavior across trials is informative. This assumption is analogous to assumptions made in many diverse statistical contexts involving random-effects models or aggregate response estimation based on generalized estimating equations, and is imbedded in much neurophysiological investigation.

APPENDIX A: DERIVATION OF LIKELIHOOD RATIO FOR GLOBAL ANALYSIS WITH A SINGLE NEURON

Letting $MVN(y; \mu, \Sigma)$ denote the multivariate normal density with mean μ and variance Σ evaluated at y , we write the numerator of the LR statistic as follows:

$$\begin{aligned} L_{H_0}(\mu_0, \dots, \mu_0) &= \prod_{j=1}^J MVN(Q_j y_j; Q_j \mu_0, Q_j \Sigma_j Q_j^T) \\ &= c \prod_{j=1}^J |\Lambda_j|^{-\frac{1}{2}} \times \prod_{j=1}^J \exp \left\{ \frac{-1}{2} [(Q_j y_j - Q_j \mu_0)^T \Lambda_j^{-1} (Q_j y_j - Q_j \mu_0)] \right\} \end{aligned}$$

Therefore, the log-likelihood is as follows:

$$l_0 = c + \sum_{j=1}^J \ln |\Lambda_j|^{-\frac{1}{2}} - \sum_{j=1}^J \frac{1}{2} [(Q_j y_j - Q_j \mu_0)^T \Lambda_j^{-1} (Q_j y_j - Q_j \mu_0)]$$

so, the quadratic form can be written as follows:

$$\begin{aligned} U &= \sum [(Q_j y_j)^T \Lambda_j^{-1} (Q_j y_j) - (Q_j y_j)^T \Lambda_j^{-1} (Q_j \mu_0) - (Q_j \mu_0)^T \Lambda_j^{-1} (Q_j y_j) \\ &\quad + (Q_j \mu_0)^T \Lambda_j^{-1} (Q_j \mu_0)] \end{aligned}$$

therefore,

$$\frac{\partial U}{\partial \mu_0} = \sum_{j=1}^J [-Q_j \Lambda_j^{-1} (Q_j y_j) + Q_j^T \Lambda_j^{-1} (Q_j \mu_0)] = 0$$

ends up being

$$\hat{\mu}_0 = \left[\sum_{j=1}^J Q_j^T \Lambda_j^{-1} Q_j \right]^{-1} \left[\sum_{j=1}^J Q_j^T \Lambda_j^{-1} Q_j y_j \right]$$

Similarly, to obtain the denominator of the LR statistic, we write

$$\frac{\partial U}{\mu_j} = [-(Q_j \Lambda_j^{-1} (Q_j y_j) + Q_j^T \Lambda_j^{-1} (Q_j \mu_j))] = 0$$

which ends up producing

$$\hat{\mu}_j = y_j$$

So, the LR will become:

$$LR = \frac{\prod_j \text{MVN}(Q_j y_j; Q_j \hat{\mu}_0, \Lambda_j)}{\prod_j \text{MVN}(Q_j y_j; Q_j y_j, \Lambda_j)}$$

Therefore,

$$\log(LR) = \sum_{j=1}^J \left[\frac{-1}{2} (Q_j y_j - \hat{\mu}_0)^T \Lambda_j^{-1} (Q_j y_j - \hat{\mu}_0) \right]$$

APPENDIX B: DERIVATION OF THE MLES OF THE GLOBAL-POPULATION PARAMETERS

Letting $\text{MVN}(y; \mu, \Sigma)$ denote the multivariate normal density with mean μ and variance matrix Σ evaluated at y , the likelihood function is

$$\begin{aligned} L &= \prod_{i=1}^I \prod_{j=1}^J \text{MVN}(Q_j^i y_j^i; Q_j^i \mu_j^i, \Lambda_j^i) \\ &= c \prod_{i=1}^I \prod_{j=1}^J |\Lambda_j^i|^{-\frac{1}{2}} \times \prod_{i=1}^I \prod_{j=1}^J \exp \left\{ \frac{-1}{2} [(Q_j^i y_j^i - Q_j^i \mu_j^i)^T (\Lambda_j^i)^{-1} (Q_j^i y_j^i - Q_j^i \mu_j^i)] \right\} \end{aligned}$$

Consequently, the log-likelihood function, l , becomes

$$\begin{aligned} l &= \sum_{i=1}^I \sum_{j=1}^J \frac{-1}{2} [(Q_j^i y_j^i)^T (\Lambda_j^i)^{-1} (Q_j^i y_j^i) - (Q_j^i y_j^i)^T (\Lambda_j^i)^{-1} (Q_j^i (\mu + \alpha_i + \beta_j)) \\ &\quad - (Q_j^i (\mu + \alpha_i + \beta_j))^T (\Lambda_j^i)^{-1} (Q_j^i y_j^i) + (Q_j^i (\mu + \alpha_i + \beta_j))^T (\Lambda_j^i)^{-1} (Q_j^i (\mu + \alpha_i + \beta_j))] \end{aligned}$$

To maximize this function, we extend the argument of Anand and Weerahandi [12] to the multivariate setting, differentiating l with respect to μ , α_i , and β_j and equating the result to zero. We obtain the following system of equations:

$$\sum_{i=1}^I \sum_{j=1}^J [(Q_j^i)^T (\Lambda_j^i)^{-1} (Q_j^i) (y_j^i - \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j)] = 0$$

$$\sum_{i=1}^I [(Q_j^i)^T (\Lambda_j^i)^{-1} (Q_j^i) (y_j^i - \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j)] = 0$$

$$\sum_{j=1}^J [(Q_j^i)^T (\Lambda_j^i)^{-1} (Q_j^i) (y_j^i - \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j)] = 0$$

By letting $w_{ij} = (Q_j^i)^T (\Lambda_j^i)^{-1} (Q_j^i)$, $w_{i+} = \sum_{i=1}^I w_{ij}$, $w_{+j} = \sum_{j=1}^J w_{ij}$, we simplify the above and obtain the solutions:

$$\hat{\mu} = \left[\sum_{i=1}^I \sum_{j=1}^J w_{ij} \right]^{-1} \left[\sum_{i=1}^I \sum_{j=1}^J w_{ij} y_j^i \right]$$

$$w_{i+} \hat{\alpha}_i - \sum_{k=1}^I \hat{\alpha}_k \left[\sum_{j=1}^J [w_{+j}]^{-1} w_{ij} w_{kj} \right] = \sum_{j=1}^J w_{ij} y_j^i - \sum_{j=1}^J [w_{+j}]^{-1} w_{ij} \sum_{k=1}^I w_{kj} y_k^j$$

$$w_{+j} \hat{\beta}_j - \sum_{k=1}^J \hat{\beta}_k \left[\sum_{i=1}^I [w_{i+}]^{-1} w_{ji} w_{jk} \right] = \sum_{i=1}^I w_{ij} y_j^i - \sum_{i=1}^I [w_{i+}]^{-1} w_{ij} \sum_{k=1}^J w_{ik} y_k^i$$

REFERENCES

1. Roesch MR, Olson CR. Neuronal activity related to reward value and motivation in primate frontal cortex. *Science* 2004; **304**(5668):307–310.
2. Nieder A, Diester I, Tudusciuc O. Temporal and spatial enumeration processes in the primate parietal cortex. *Science* 2006; **313**(5792):1431–1435.
3. Freedman DJ, Assad JA. Experience-dependent representation of visual categories in parietal cortex. *Nature* 2006; **443**(7107):85–88.
4. Kass RE, Ventura V, Cai C. Statistical smoothing of neuronal data. *Network: Computation in Neural Systems* 2003; **14**:5–15.
5. Kass RE, Ventura V, Brown EN. Statistical issues in the analysis of neuronal data. *Journal of Neurophysiology* 2005; **94**:8–25.
6. Ventura V, Carta R, Kass RE, Gettner SN, Olson CR. Statistical analysis of temporal evolution in single-neuron firing rates. *Biostatistics* 2002; **1**:1–20.
7. Behseta S, Kass RE. Testing equality of two functions using BARS. *Statistics in Medicine* 2005; **24**:3523–3534.
8. DiMatteo I, Genovese CR, Kass RE. Bayesian curve-fitting with free-knot splines. *Biometrika* 2001; **88**: 1055–1071.
9. Fan J, Lin S. Test of significance when data are curves. *Journal of the American Statistical Association* 1998; **93**:1007–1021.
10. Weerahandi S. ANOVA under unequal variances. *Biometrics* 1995; **51**:589–599.
11. Venables WN, Ripley BD. *Modern Applied Statistics with S-plus*. Springer: New York, 1997.
12. Ananda MA, Weerahandi S. Two-way ANOVA with unequal cell frequencies and unequal variances. *Statistica Sinica* 1997; **7**:631–646.
13. Sheather SJ, Jones MC. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society, Series B* 1991; **53**:683–690.
14. Olson CR, Gettner SN, Ventura V, Carta R, Kass RE. Neuronal activity in macaque supplementary eye field during planning of saccades in response to pattern and spatial cues. *Journal of Neurophysiology* 2000; **84**(3):1369–1384.
15. Moorman DE, Olson CR. Combination of neuronal signals representing object-centered location and saccade direction in macaque supplementary eye field. *Journal of Neurophysiology* 2007; **97**(5):3554–3566.

16. Kaufman CG, Ventura V, Kass RE. Spline-based nonparametric regression for periodic functions and its application to directional tuning of neurons. *Statistics in Medicine* 2005; **24**:2255–2265.
17. Behseta S, Kass RE, Wallstrom G. Hierarchical models for assessing variability among functions. *Biometrika* 2005; **92**:419–434.
18. Kass RE, Slate EH. Reparameterization and diagnostics of posterior non-normality (with discussion). In *Bayesian Statistics*, vol. 4, Berger JO, Bernardo JM, Dawid AP, Smith AFM (eds). Oxford University Press: Oxford, 1992; 289–305.
19. Ventura V, Cai C, Kass RE. Trial-to-trial variability and its effect on time-varying dependence between two neurons. *Journal of Neurophysiology* 2005; **94**:2928–2939.