

## THE MAXIMUM LIKELIHOOD APPROACH TO THE IDENTIFICATION OF NEURONAL FIRING SYSTEMS

David R. Brillinger

Statistics Department  
University of California  
Berkeley, CA 94720

*The concern of this work is the identification of the (nonlinear) system of a neuron firing under the influence of a continuous input in one case, and firing under the influence of two other neurons in a second case. In the first case, suppose that the data consist of sample values  $X_t, Y_t, t = 0, \pm 1, \pm 2, \dots$  with  $Y_t = 1$  if the neuron fires in the time interval  $t$  to  $t + 1$  and  $Y_t = 0$  otherwise, and with  $X_t$  denoting the (sampled) noise value at time  $t$ . Suppose that  $H_t$  denotes the history of the process to time  $t$ . Then, in this case the model fit has the form*

$$\text{Prob}\{Y_t = 1 | H_t\} = \Phi(U_t - \theta)$$

where

$$U_t = \sum_{u=0}^{\gamma_t-1} a_u X_{t-u} + \sum_{u=0}^{\gamma_t-1} \sum_{v=0}^{\gamma_t-1} b_{u,v} X_{t-u} X_{t-v}$$

where  $\gamma_t$  denotes the time elapsed since the neuron last fired and  $\Phi$  denotes the normal cumulative. This model corresponds to quadratic summation of the stimulus followed by a random threshold device. In the second case, a network of three neurons is studied and it is supposed that

$$U_t = \sum_{u=0}^{\gamma_t-1} a_u X_{t-u} + \sum_{u=0}^{\gamma_t-1} b_u Z_{t-u}$$

with  $X_t$  and  $Z_t$  zero-one series corresponding to the firing times of the two other neurons. The models are fit by the method of maximum likelihood to Aplysia californica data collected in the laboratory of Professor J.P. Segundo. The paper also contains some general comments of the advantages of the maximum likelihood method for the identification of nonlinear systems.

---

This work was partially supported by the National Science Foundation Grant DMS-8316634.

Professor J.P. Segundo introduced me to the problems of the analysis of neuron spike trains, provided the data sets discussed and generally offered much help and encouragement. I cannot thank him enough. Ettore Marchetti prepared the noise data for computer analysis. The referees made many helpful and perceptive comments.

Address correspondence to David R. Brillinger, Statistics Department, University of California, Berkeley, CA 94720.

*Keywords*—Causal connections, Maximum likelihood, Model fitting, Neuronal firing, Neuronal networks, Parameter estimation, Point process, Quadratic kernel, Spike trains, System identification, Threshold element.

## INTRODUCTION

Following the seminal work of Wiener (15) and Lee and Schetzen (10), the fitting of nonlinear models to physiological systems has generally been implemented by kernel expansions and the method of moments. Theoretical and practical aspects of this approach are described in detail in Marmarelis and Marmarelis (11). This present paper is concerned with fitting instead by the method of maximum likelihood. Advantages that result include: a more efficient use of the data in a broad variety of circumstances and, in the case of principal concern, a procedure well adapted to systems with point process output. In one case studied, the model fit by maximum likelihood also includes a quadratic kernel.

A descriptive model for the firing process of a neuron involves “summation” of input, followed by firing when the “sum” exceeds a threshold level. In Brillinger and Segundo (5), a model involving linear time invariant filtering of a continuous input, with memory back to the time of last firing, was fit. Briefly, this present paper is concerned with a similar model but one including quadratic terms in the input in a first example, and including the spike train outputs of two associated neurons in a second example.

## METHODS

### *Statistical Methods*

Maximum likelihood has proven to be one of the longlasting effective estimation techniques in statistics. This method does have limitations, but these are generally understood. The procedure is simply described. One sets down the likelihood function for a data set and model of concern, the likelihood being the probability function (or probability density function) of the random variables that will be measured. This expression typically involves some unknown parameters. In the method of maximum likelihood one chooses as estimates of the unknown parameters, the values maximizing the realized likelihood. Expressions are available for the approximate standard errors of these estimates and for formally testing whether some of the parameters have prespecified values (e.g., zero). The method is efficient when applied to large data sets in the sense that, for a given stochastic model and assuming certain regularity conditions, the distribution of the maximum likelihood estimate is more concentrated about the true parameter value than the distribution of any other estimate. One reference to the general properties of maximum likelihood is the book by Rao (13).

In the random process situation, the setting down of the likelihood can be complicated by the statistical dependencies present amongst the variates present; however, once the likelihood has been derived, one proceeds in a direct fashion. To form the likelihood, one can sometimes proceed via conditional distributions given the history of the process up to the present. Specifically, if  $Y_t$  denotes the random variate of concern at time  $t$  and  $H_t$  denotes the collection of variates  $\{Y_u, u < t\}$  and  $f_t$  denotes

the conditional distribution of  $Y_t$  given  $H_t$ , then the probability distribution of the data is given by

$$\prod_t f_t.$$

This approach is employed with the models of this paper.

### Neurophysiological Methods

For the data to be studied, there were experiments of two basic types: some with a “noise” input and others with spike train inputs. For the first case, the experiments are described in Bryant and Segundo (7). For the particular case addressed in this paper, the neuron L5 of *Aplysia californica* was isolated and impaled with two electrodes. One electrode was for injecting a continually varying current. The second was for observing the intracellular membrane potential. Specifically, the current injected was generated to be band-limited Gaussian white noise. For numerical analysis the input was sampled at 32 Hz. (In other experiments the sampling rate was 50 Hz., but the results obtained were not qualitatively different.) The output series for analysis was the corresponding sequence of times at which the neuron was observed to fire. All told, there were 81700 input time series values and the neuron was seen to fire 2017 times. The details of the method of intracellular recording are given in Bryant *et al.* (6). An analysis of this data set was previously presented in Brillinger and Segundo (5). A brief stretch of the data analysed is presented in Fig. 1. It seems impossible to discern the “rule” the neuron is employing in deciding whether or not to fire via direct examination of this figure. Input processes of other types might be equally or even more informative for the model fit in this paper. It is worth emphasizing however that when the present data were collected there were no thoughts of fitting models of the type to be discussed here. Rather, it was felt that data collected with noise input would be rich enough for the analysis of many classes of models. This attitude is dominant in Marmarelis and Marmarelis (11).

For the second case, of a neuron receiving input from two other neurons, the basic

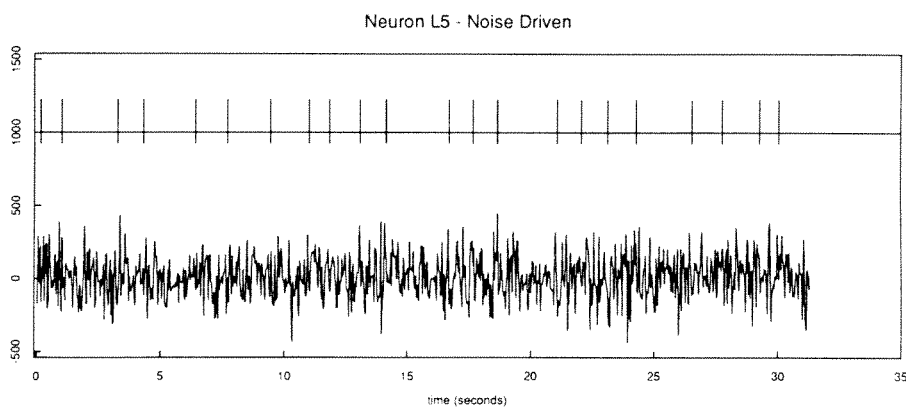


FIGURE 1. The lower trace provides a stretch of noise input to the *Aplysia* neuron L5 and the upper trace the corresponding times at which the neuron fired.

series of experiments is described in Bryant *et al.* (6). The neurons again were from *Aplysia*. The neuron L10 was impaled with neurons L2 and L3. Microelectrodes were inserted in order to measure the individual electrical activity of all three of the cells. L10 was allowed to fire spontaneously, that is with no deliberate stimulation. In the data set analysed, it was seen to fire at more or less regular intervals, (i.e., as a pacemaker). L2 and L3 were known to be postsynaptic to L10; however, it was not known whether there were any direct (causal) connections of L2 and L3 with each other. Focusing on L2, causal models that might be studied include:  $L3 \rightarrow L2$ ,  $L10 \rightarrow L2$  and  $L3, L10 \rightarrow L2$ . A brief display of some of the data from this experiment is given in Fig. 2. It again seems virtually impossible to infer relationships on the basis of a direct examination of the data alone. The number of firings in this data set were 767, 539, and 741 for L10, L2, and L3 respectively. It is the data analysed in Fig. 15 a,b,c of Bryant *et al.* (6) and in Fig. 6 of Brillinger *et al.* (3).

### THE BIOLOGICAL AND STOCHASTIC MODELS

The internal process by which a neuron fires has been described as follows. A time fluctuating potential exists within the neuron. This potential results from a combination of internal and external mechanisms. When the potential level, at a special location within the neuron (the trigger zone), exceeds a threshold level, the neuron fires and the potential is reset to its extant value. General references to the phenomenon of neuron firing are Aidley (1) and Segundo (14). Evidence for the resetting of the potential may be found in Eccles (8), Chapter 2.

The firing process may be described formally as follows. It will be assumed that the threshold level at time  $t$  has the form  $\theta_t = \theta + \epsilon_t$  with the  $\epsilon_t$  independent standard normal variates,  $t = 0, \pm 1, \pm 2, \dots$ . It will be assumed that the neuron fires at time  $t$  if the internal potential  $U_t$  at time  $t$  exceeds  $\theta_t$ . In the case of noise input it will be assumed that the internal potential has the form

$$U_t = \sum_{u=0}^{\gamma_t-1} a_u X_{t-u} + \sum_{u=0}^{\gamma_t-1} \sum_{v=0}^{\gamma_t-1} b_{u,v} X_{t-u} X_{t-v} \quad (1)$$

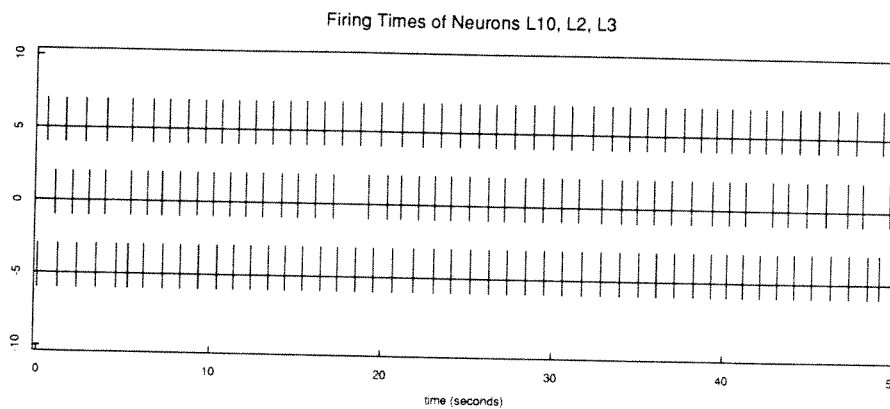


FIGURE 2. The corresponding times at which the *Aplysia* neurons L2, L3, L10 fired in an experiment. The top graph is the L10, the middle the L2 and the bottom the L3 spike train.

with  $\gamma_t$  the time elapsed since the neuron last fired and with  $X_t$  the input signal level at time  $t$ . In the second case it will be assumed that  $U_t$  has the form

$$U_t = \sum_{u=0}^{\gamma_t-1} a_u X_{t-u} + \sum_{u=0}^{\gamma_t-1} b_u Z_{t-u} + \theta_1 \gamma_t + \theta_2 \gamma_t^2 \quad (2)$$

with  $X_t$  and  $Z_t$  respectively defined to equal 1 or 0 depending on whether or not L2 or L3 respectively fired or not at time  $t$ . The presence of  $\gamma_t$  in the upper limits of the expressions here has the effect of introducing feedback from the system output to the input and so negates the usual kernel fitting procedures. The terms in  $\theta_1$  and  $\theta_2$  in Eq. 2 allow the neuron to fire spontaneously in the case of no input.

The functions  $a_u, b_{u,v}$  in Eq. 1 and  $a_u, b_u$  in Eq. 2 reflect spatio-temporal summation involved in bringing the influences of the driving cell(s) to the trigger zone of the dependent neuron. In the first case the summation is quadratic. In the second it is linear. The expression in Eq. 1 is somewhat like that of the Volterra model of the Wiener-Lee-Schetzen approach to systems, see Marmarelis and Marmarelis (11) for example; however a crucial distinction results from the fact that  $U_t$  is not observed in the experiments of concern. Therefore, the Wiener-Lee-Schetzen identification techniques cannot be applied directly. The second crucial distinction results from the variable numbers of terms  $a_u$  and  $b_{u,v}$  appearing in Eq. 1. This means that even were the system linear in  $U_t$ , which it is not, the usual identification techniques would lead to biased estimates. This occurrence cannot be handled by dimension estimation techniques, such as that of Akaike, because the number of terms appearing at time  $t$  depends on  $t$ .

The term  $\theta_1 \gamma_t + \theta_2 \gamma_t^2$  has several effects. It can be seen as allowing the distribution of the (random) threshold to be changing with the time since last firing. It can be seen as handling the refractoriness of the neuron. Further its functional form might be altered to, say,  $\exp\{-\theta \gamma_t\}$ , to be in accord with analytic models that have been suggested previously. Its form might be further altered to allow the presence of an absolute refractory period. The power of the maximum likelihood method is that it can handle such changes of model directly.

Let  $H_t$  denote the history of the process, that is the collection of variates  $\{Y_u, u < t, X_u, u \leq t\}$ . Then

$$Prob\{Y_t = 1 | H_t\} = Prob\{U_t > \theta + \epsilon_t | H_t\} = \Phi(U_t - \theta)$$

under the assumptions made, with  $\Phi$  the normal cumulative. The model is seen to contain two sources of randomness: one derived from the threshold and the second based on whether or not the neuron fires given the threshold level. Writing

$$p_t = \Phi(U_t - \theta)$$

the likelihood function of the data is given by

$$\prod_t p_t^{Y_t} (1 - p_t)^{1 - Y_t} \quad (3)$$

In the method of maximum likelihood the unknowns,  $a_u, b_{u,v}, \theta, b_u, \theta_1, \theta_2$  are estimated by maximizing this expression as a function of the unknowns. It may be remarked that the likelihood of Eq. 3 is not a function of the cross-moments of the data values, reflecting the essential distinction of the proposed procedure from that of Wiener-Lee-Schetzen.

Some earlier firing models, including references to such, are indicated in Brillinger (4). The seminal reference is McCulloch and Pitts (12).

## RESULTS

### *Single Cell With Continuous Input*

This is the case of a system with a continuous input process and with a point process output.

The model set down above, involving both linear and quadratic terms, was fit to the data by maximizing the likelihood of expression, Eq. 3. Figure 3 is a graph of  $\hat{a}_t$ , the estimate of the linear kernel and  $\pm$  two standard error values graphed about 0 (to allow easy assessment of the significance of the coefficients.) There is seen to

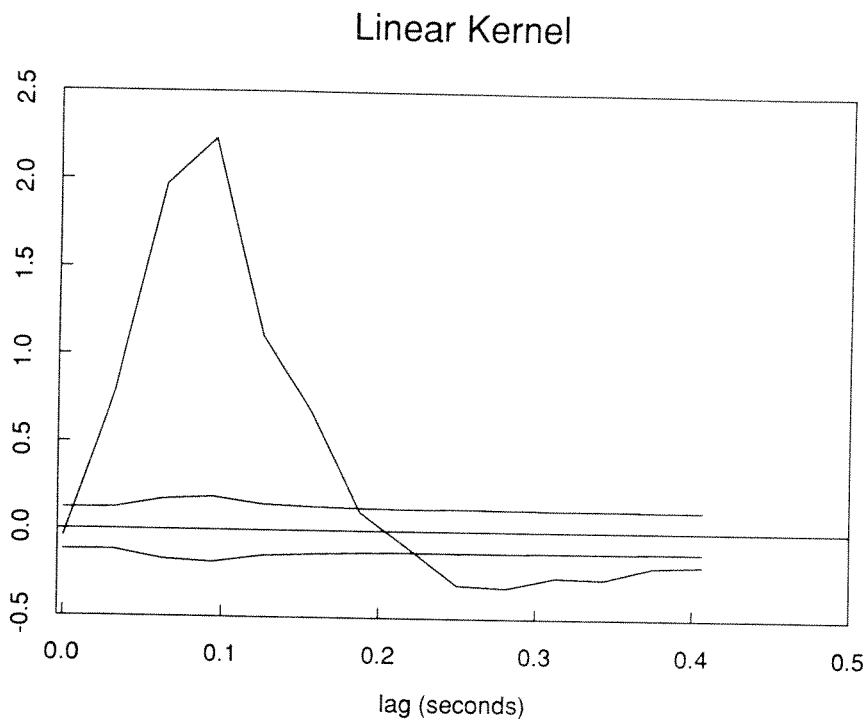


FIGURE 3. An estimate of  $a_u$ , the linear kernel of Eq. (1). The kernel was estimated at 14 lags. The solid lines give  $\pm$  two standard error values about 0.

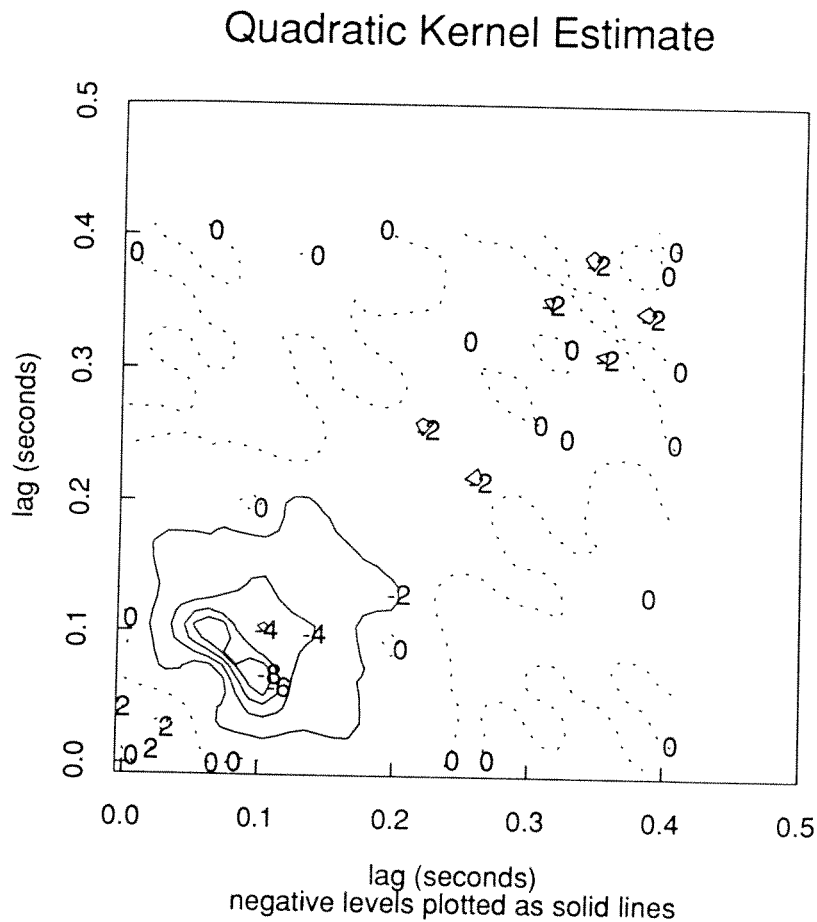


FIGURE 4. A contour plot of an estimate of the quadratic kernel,  $b_{u,v}$  of Eq. (1). The solid lines correspond to contours at negative levels.

be a latency of approximately .1 second and a later negative swing. Figure 4 is a contour plot of  $\hat{b}_{u,v}$ , the quadratic kernel. Figure 5 is the corresponding perspective plot. It is seen from these figures that the quadratic kernel appears to be concentrated near the diagonal and to have a negative bulge at lags corresponding to the location of the positive peak in the linear kernel. Through its presence it is compensating for overshoot on the part of the linear kernel. The maximum likelihood analysis does give standard errors for the individual quadratic kernel values, but there seems no need to present them here, rather we indicate the likelihood ratio test statistic for the hypothesis that the quadratic kernel is identically zero. The value of this test statistic is 1894.3 on 840 degrees of freedom (The 840 here comes from the 105 distinct quadratic coefficients fit and the fact that the data were segmented into 8 stretches for analysis.) It is virtually impossible for this value to have arisen by chance if in fact the quadratic kernel is identically zero.

## Quadratic Kernel Estimate

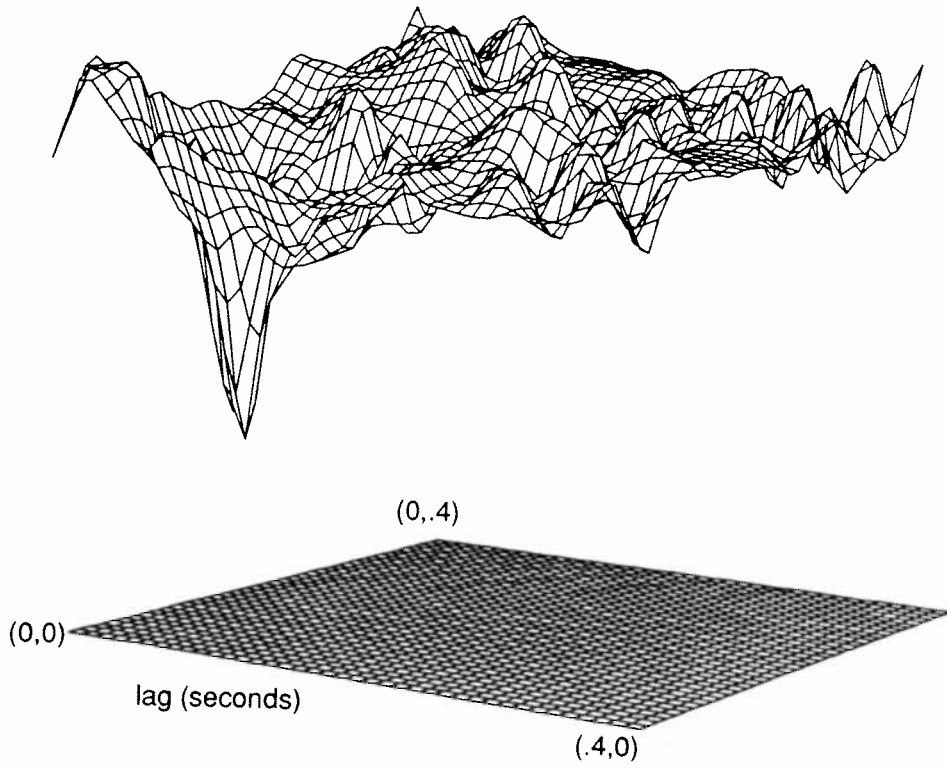


FIGURE 5. A perspective plot corresponding to Fig. 4.

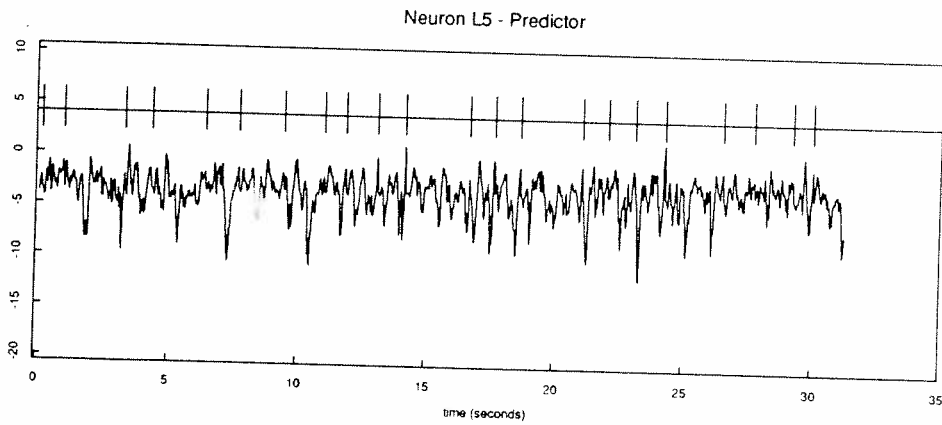


FIGURE 6. A graph of an estimate of  $U_t$ , the predictor, of Eq. (1) and corresponding times at which the neuron fired. One can note large values of  $\hat{U}_t$  at firing times.



Figure 6 provides a plot of  $\hat{U}_t$ , the estimate of the predictor, and corresponding firing times for the segment of data graphed in Fig. 1. Close examination shows that  $\hat{U}_t$  generally takes on high values at firings as the model implies should be the case.

Figure 7 presents the results of attempts to assess the goodness of fit of the model. The top graph is a plot of the proportion of times the neuron fired when the value of  $\hat{U}_t$ , the predictor, lay in a small interval near  $u$ , the corresponding abscissa. Formally, the points plotted are given by

$$\#\{Y_t = 1 \text{ with } u - h < \hat{U}_t < u + h\} / \#\{t \text{ with } u - h < \hat{U}_t < u + h\}$$

for small  $h$  and selected  $u$ . In essence one is forecasting the probability that the neuron will fire. The smooth curve is the assumed Gaussian cumulative,  $\Phi(u)$ . The fit is very good. In fact, the top graph was prepared employing only the data for the first segment (10000 observations) of the data, i.e., the fit was assessed for the data of that segment using the estimates based on the data of that same segment. A more powerful test procedure is to employ the estimates derived from one segment of the data with firings from another segment. The second graph of Fig. 7 presents the performance of the predictor estimate based on the first 7 (of 8) segments of the data when applied to the final segment. The fit is not as good, but does seem reasonable.

### Three Cell Network

Figures 8 and 9 provide the results of fitting the threshold model to the output of cell L2 with linear time invariant summation of input and a term  $\theta_1 \gamma_t + \theta_2 \gamma_t^2$ , to allow for spontaneous firing of the cell. Figure 8 refers to the problem of explaining the firing of L2 by the firings of L3 alone. The solid lines correspond to  $\pm$  two standard error limits about 0. There is seen to be a significant, negative-valued, summation function. The lower graph is a plot of the estimated  $\theta_1 \gamma_t + \theta_2 \gamma_t^2$  and  $\theta$ . Figure 9 refers to explaining the L2 firing by the firings of both L3 and L10. The solid lines again correspond to  $\pm$  two standard error limits for the L3 kernel. It is interesting to note that the L3 firings do not enter significantly when the L10 firings are included. It would seem that L10 is driving both L2 and L3 and that the apparent connectivity of L2 and L3 is simply through their joint influence by L10. The effect of L10 is seen to be inhibitory, but there are apparently later rebounds. The lower graph provides the estimate of  $\theta_1 \gamma_t + \theta_2 \gamma_t^2$  for this fit.

The lack of influence of the neuron L3 in the presence of L10 may be examined formally by a likelihood ratio test of the hypothesis that its coefficients are identically 0. One proceeds by seeing the difference between the minus twice log likelihood values including just L3 and including both L3 and L10. This difference is 13, with 13 degrees of freedom. There is no evidence that the coefficients of L3 differ from 0.

There are other examples of this type of analysis in Brillinger (4), where figures of the type of Fig. 7 are also presented allowing assessment of the validity of the model. The fits all appeared good.

This model and fitting procedure extend quite directly to the case of moderately large numbers of neurons interacting with each other in possibly complicated fashions.

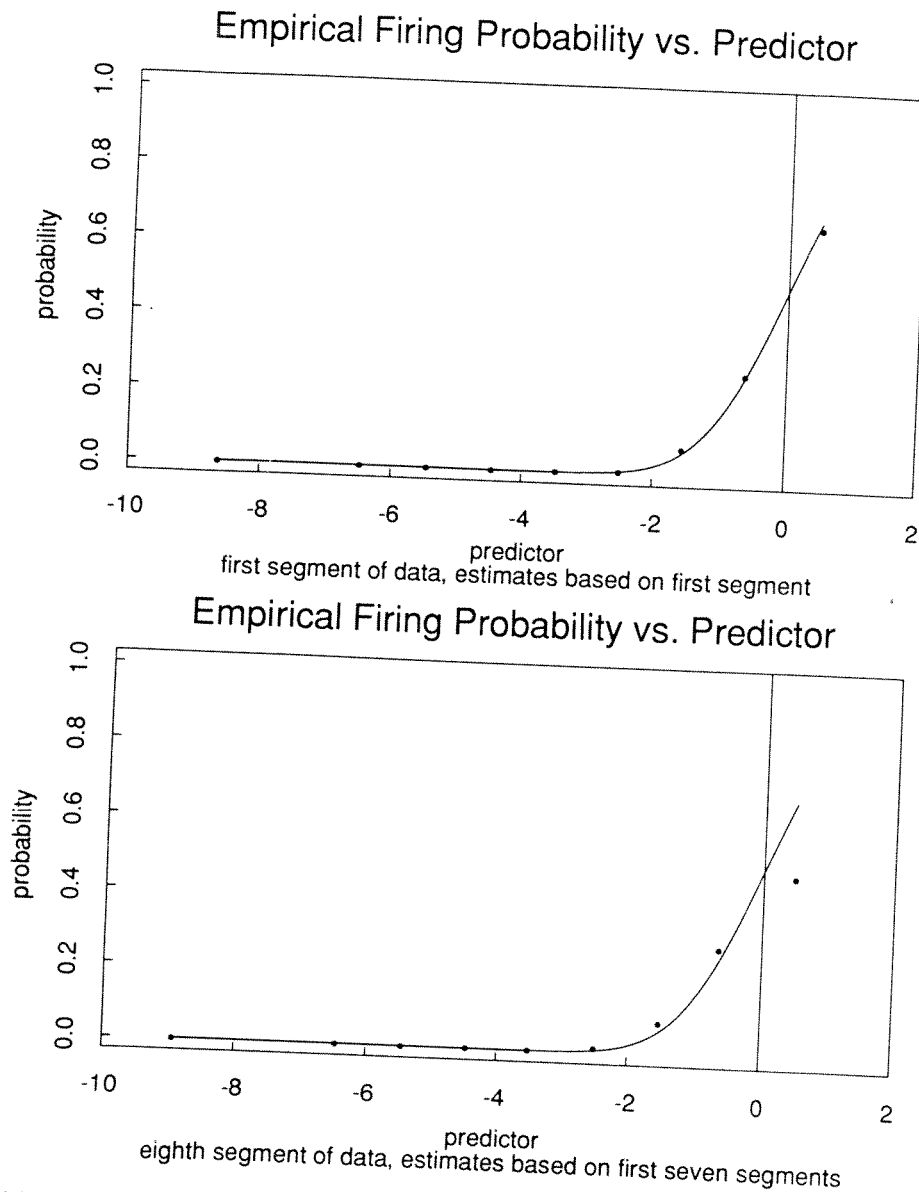
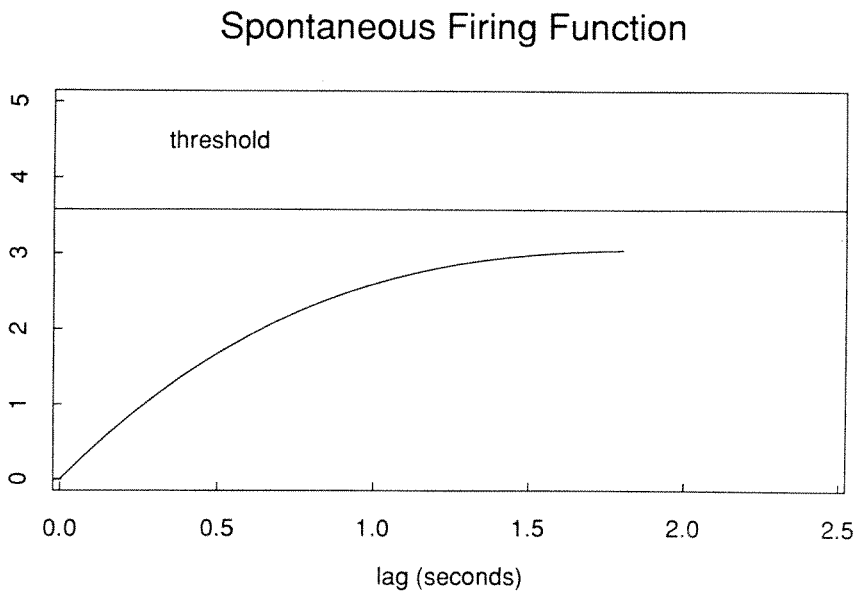
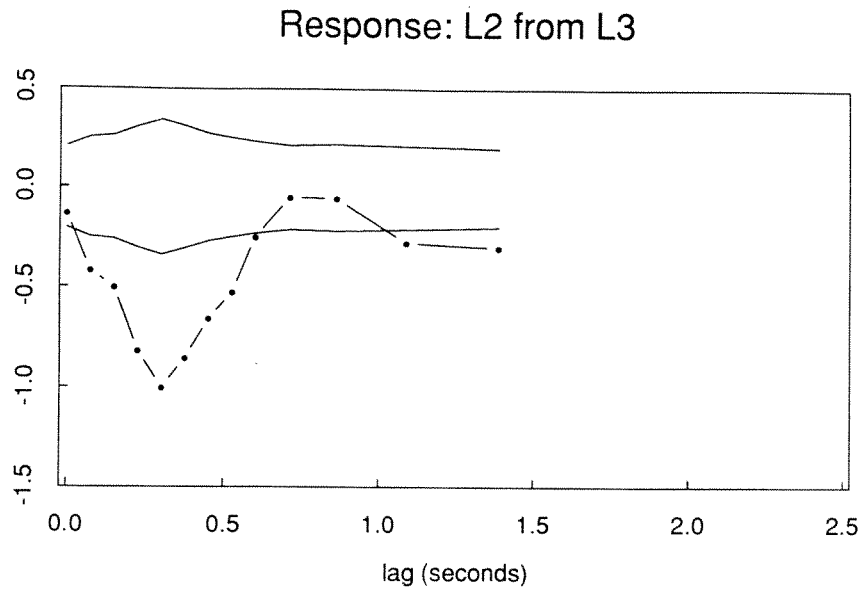


FIGURE 7. The top graph is a plot of the obtained empirical probability of firing, as a function of the estimated predictor  $\hat{U}_t$ , computed for the first 1000 points of the data set and a plot of the corresponding theoretical curve,  $\Phi(u)$ . The bottom graph shows the result of employing a predictor with coefficients based on the first 70000 data points to forecast the firings of an unrelated 10000 data points.

### DISCUSSION

The purposes of this paper were severalfold: firstly to present some new empirical results, (a) the linear and quadratic kernels for a particular neuron model involving a threshold device, and (b) inferences re the connectivities in a network of three

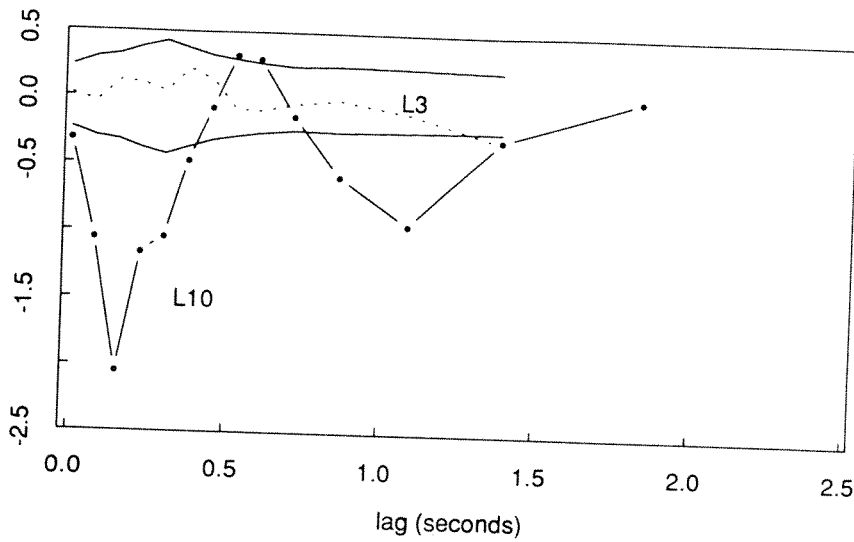


**FIGURE 8.** This figure to an analysis of spike trains from neurons L2 and L3. The linked points provide the estimated linear summation function for the firing of L2 as "driven" by L3 alone. The solid lines give  $\pm$  two standard error limits about 0. The lower graph gives the estimated spontaneous firing term  $\theta_1\gamma_i + \theta_2\gamma_i^2$  and the estimated threshold level,  $\theta$ .

neurons, again with a model including a threshold device. Secondly a purpose was to describe the character of and to list some advantages of carrying out system identification by the method of maximum likelihood.

In the case involving linear and quadratic kernels and a fit by maximum likeli-

## Responses: L2 from L3 and L10



## Spontaneous Firing Function

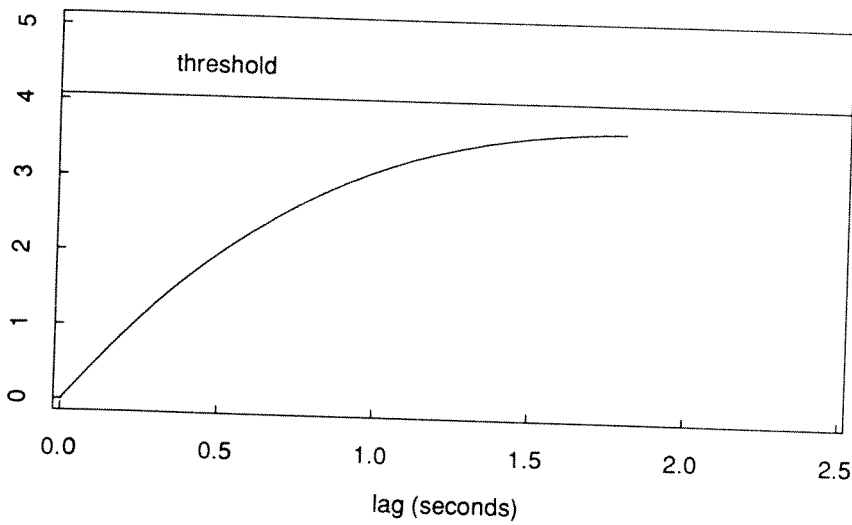


FIGURE 9. For the experiment of Fig. 8, the top graph here provides estimates of the summation functions,  $a_u$  and  $b_u$ , of Eq. (2) corresponding to the L3 and L10 inputs respectively. The solid lines give  $\pm$  two standard errors about 0 for the L3 estimate. L3 is no longer having any apparent effect. The lower graph gives the estimated spontaneous firing term as in Fig. 8.

hood, it was found that after assuming a Gaussian distributed threshold, a linearly increasing dependence on the input signal was apparently too strong. This conclusion came from the negative values of the quadratic kernel estimate around lags (.1, .1) seconds and the concentration near the diagonal.

In the case of the three cell network, maximum likelihood and modeling proved an effective tool for inferring the connectivities. There appeared to be no direct L2-L3 connections. Direct examination of the data, even in this case of pacemaker input, seemed hopeless.

We end by listing various of the advantages of the maximum likelihood approach. Some hold broadly and some are particularly pertinent to the present circumstance. The maximum likelihood approach is highly flexible. Uncertainty measures of the parameter estimates are directly available. Formal tests of the values of parameters appearing are also available. In the particular case at hand, the parameters of the models are directly estimated and can be chosen to be biologically interpretable. The statistics of the input signals are of no concern to the analysis; specifically, there is no need for white noise-type input. (Of course in the matter of design it will pay to think about what is an efficient input.) Feedback, as present in the  $\gamma_t$  in the model, caused no difficulty in the analysis. (It is unclear just how to modify the kernel approach to handle this phenomenon.) Other terms can be included in the model as necessary, e.g., a spontaneous firing term  $\theta_1\gamma_t + \theta_2\gamma_t^2$ . A term in  $t$  could also have been included to handle a cell's slowly dying.

### SOME DETAILS OF THE COMPUTATIONS

All of the computations were carried out on a Sun 3/50 Workstation. For the noise input case the data were broken up into eight segments of 10,000 points and the individual maximum likelihood estimates of the segments averaged to obtain the final estimates. (This procedure had the added advantage that the stationarity of the relationship could be examined. There was no evidence of time trends in the estimated values.) The maximization of the likelihood was carried out by the routine va09a of the Harwell subroutine library, Hooper (9). The length of running time depended on the aptness of the initial values employed in the optimization, but generally was of the order of 5-15 minutes. In the case of the three cell network, the computations were carried out via the statistical package GLIM, by Baker and Nelder (2). The running times were 10-15 minutes. An example of a GLIM program to carry out such computations is provided in Brillinger (4).

### REFERENCES

1. Aidley, D.J. *The Physiology of Excitable Cells*. Cambridge, MA: Cambridge Press; 1978.
2. Baker, R.J.; Nelder, J.A. *The GLIM System*. Numerical Algorithms Group, 7 Banbury Road, Oxford, England, 1978.
3. Brillinger, D.R.; Bryant, H.L.; Segundo, J.P. Identification of synaptic interactions. *Biol. Cybernetics* 22:213-228; 1976.
4. Brillinger, D.R. Maximum likelihood analysis of spike trains of interacting nerve cells. Technical Report No. 80. Statistics Dept., Berkeley, CA: University of California; 1986.
5. Brillinger, D.R.; Segundo, J.P. Empirical examination of the threshold model of neuron firing. *Biol. Cybernetics* 35:213-220; 1979.
6. Bryant, H.L.; Ruiz-Marcos, A.; Segundo, J.P. Correlations of neuronal spike discharges produced by monosynaptic connexions and by common inputs. *J. Neurophysiol.* 36:205-225; 1973.
7. Bryant, H.L.; Segundo, J.P. Spike initiation by transmembrane current: A white-noise analysis. *J. Physiol.* 260:279-314; 1976.
8. Eccles, J.C. *The Physiology of Nerve Cells*. Baltimore, MD: Johns Hopkins Press; 1957.
9. Hooper, M.J. Harwell Subroutine Library. Atomic Energy Research Establishment, Harwell, England; 1980.

10. Lee, Y.W.; Schetzen, M. Measurement of the kernels of a non-linear system by cross-correlation. *Int. J. Control* 2:237-254; 1965.
11. Marmarelis, P.Z.; Marmarelis, V.Z. *Analysis of Physiological Systems*. New York, NY: Plenum, 1978.
12. McCulloch, W.S.; Pitts, W.H. A logical calculus of ideas immanent in nervous activity. *Bull. Math. Biophys.* 5:115-133; 1943.
13. Rao, C.R. *Linear Statistical Inference and its Applications*, 2nd ed. New York, NY: John Wiley; 1973.
14. Segundo, J.P. *La Neurofisiologica: Algunos Supuestos y Bases, Recovecos e Implicaciones*. Mexico: Universidad Nacional Autonoma de Mexico; 1985.
15. Wiener, N. *Nonlinear Problems in Random Theory*. Cambridge, MA: MIT Press; 1958.