LETTER Communicated by Loren M. Frank

A General Likelihood Framework for Characterizing the Time Course of Neural Activity

Michael J. Prerau
prerau@nmr.mgh.harvard.edu
Graduate Program in Neuroscience, Boston University, Boston, MA 02215, U.S.A.

Uri T. Eden
tzvi@bu.edu
Graduate Program in Neuroscience, and Department of Math and Statistics, Boston University, Boston, MA 02215, U.S.A.

We develop a general likelihood-based framework for use in the estimation of neural firing rates, which is designed to choose the temporal smoothing parameters that maximize the likelihood of missing data. This general framework is algorithm-independent and thus can be applied to a multitude of established methods for firing rate or conditional intensity estimation. As a simple example of the use of the general framework, we apply it to the peristimulus time histogram and kernel smoother, the methods most widely used for firing rate estimation in the electrophysiological literature and practice. In doing so, we illustrate how the use of the framework can employ the general point process likelihood as a principled cost function and can provide substantial improvements in estimation accuracy for even the most basic of rate estimation algorithms. In particular, the resultant kernel smoother is simple to implement, efficient to compute, and can accurately determine the bandwidth of a given rate process from individual spike trains. We perform a simulation study to illustrate how the likelihood framework enables the kernel smoother to pick the bandwidth parameter that best predicts missing data, and we show applications to real experimental spike train data. Additionally, we discuss how the general likelihood framework may be used in conjunction with more sophisticated methods for firing rate and conditional intensity estimation and suggest possible applications.

1 Introduction

The estimation of firing rates from neural spike train data is useful for relating patterns of discrete spikes to continuous correlates of neural activity.

Michael Prerau is currently in the Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Charlestown, MA.

 © 2011 Massachusetts Institute of Technology

Firing rate is a helpful mathematical construct rather than a biologically observable signal, so it is impossible to empirically verify the accuracy of a rate estimate. It is therefore important to derive principled computational approaches to select the parameters and methodologies used in rate estimation so that the choice of the degree of smoothing is not unfounded or misleading. Inappropriate estimates of rate can lead to spurious results in higher-level analyses. Herein, we describe how parameter selection based on maximizing the likelihood of missing data can significantly increase the accuracy of rate estimation algorithms. We illustrate the ability of the general framework to improve simple, commonly used smoothing procedures and discuss how it can be used in the development of more sophisticated methods.

Since the early findings that neural firing rates can be related to sensory stimuli (Adrian & Zotterman, 1926), numerous methods have been developed to define continuous measures of spiking intensity (Cunningham, Gilja, Ryu, & Shenoy, 2009). In the electrophysiological literature, the most commonly employed methods for calculating neural firing rates are the time histogram (Gerstein & Kiang, 1960) and kernel smoother (Parzen, 1962; Sanderson, 1980; Nawrot, Aertsen, & Rotter, 1999). Discussions of these methods can be found in the appendix and in Dayan and Abbott (2001). A common feature of these methodologies is the need to select a bandwidth parameter, which affects the temporal range over which the smoothing algorithm incorporates information.

The selection of the bandwidth parameter imposes particular assumptions about the physiology of the neuron producing the spike train. Specifically, the bandwidth size defines the time course over which the neuron can change its firing rate, with a large bandwidth implying a slowly changing firing rate and a small bandwidth implying a rapidly changing rate. Mathematically, the effects of the bandwidth can be described as governing the maximum of the absolute value of the second derivative of the rate process. Thus, beyond the signal processing definition, we will use the term *bandwidth* to describe the intrinsic factors governing the expected rate of change in neural spiking activity for a given neuron.

Many approaches to bandwidth selection in kernel estimators have been developed in the density estimation literature (Turlach, 1993). Based on these kernel bandwidth estimation methods, specific nongeneralized algorithms have been developed for neural data involving user-defined cost functions (Shimazaki & Shinomoto, 2007a, 2007b), or empirically derived heuristics (Nawrot et al., 1999). Kernel smoother bandwidth has also been defined adaptively as a function of spike frequency (Richmond, Optican, & Spitzer, 1990). In the electrophysiological literature, however, the values of the bandwidth parameters are almost exclusively determined ad hoc, based solely on experimenter preference.

Model-based approaches for calculating the instantaneous spiking probability have also been shown to be useful in characterizing the

firing properties of neural ensembles. Techniques such as generalized-linear (McCullagh, 1984; Truccolo, Eden, Fellows, Donoghue, & Brown, 2005) and state-space (Brown, Frank, Tang, Quirk, & Wilson, 1998; Brown, Nguyen, Frank, Wilson, & Solo, 2001; Eden, Frank, Barbieri, Solo, & Brown, 2004; Czanner et al. 2008; Kulkarni & Paninski, 2008) modeling relate neural spiking activity to other observed experimental correlates such as the spiking history, spatial location, or other experimental or biologically relevant factors. Parametric modeling is powerful because parameter estimates are fit to the data, providing statistical justification for the degree of smoothing. The model structure itself often provides biologically relevant explanations of firing rate–related phenomena. However, these parametric methods rely heavily on modeling assumptions and often require additional data from correlates of spiking activity to fit the model. Bayesian methods (DiMatteo, Genovese, & Kass, 2001; Kass, Ventura, & Cai, 2003; Cunningham, Yu, Shenoy, & Sahani, 2008; Endres & Oram, 2009) have also been effectively used to capture firing rate and to work as smoothers of a peristimulus time histogram (PSTH).

In general, most firing rate bandwidth estimation procedures choose bandwidth based on the optimization of a cost function, designed to penalize bandwidths that oversmooth or undersmooth the data. An issue with these methods is that the selection of the cost functions is often highly subjective and that different cost functions can produce disparate results. In striving toward a more principled cost function for spike smoothing, we develop a framework based on the likelihood of a general point process, a natural means of assessing goodness-of-fit for models of spiking activity. The application of the likelihood to firing rate bandwidth estimation, however, is nontrivial. Since reducing the size of the bandwidth will produce a rate better able to capture the individual spikes, the likelihood of the spike train will tend to increase as the bandwidth size approaches 0. Consequently, maximum likelihood (ML) procedures on the bandwidth parameter will always choose small bandwidth sizes with rates that greatly undersmooth the data.

As a solution, we develop an estimation procedure that sequentially treats each data point as unknown and selects the bandwidth that best predicts the missing data. To do this, we compute the likelihood of the missing data points given all the remaining spiking activity. This cross-validated likelihood is the tool that makes maximum likelihood bandwidth estimation possible.

While cross-validation schemes and likelihoods have been used with regard to general parametric model fitting (Harris, Csicsvari, Hirase, Dragoi, & Buzsáki, 2003; Itskov, Curto, & Harris, 2008), to our knowledge, neither the general point process likelihood nor cross-validation schemes have thus far been applied to bandwidth selection in the calculation of neural firing rates, for the reasons we have set out.

This general likelihood procedure is centered around operations on an estimated instantaneous firing rate, the common output of all rate estimation procedures. Consequently, the framework is independent of the actual rate estimation procedure and can be applied to any of the many algorithms developed in the literature for firing rate or conditional intensity estimation. (For useful discussions of the myriad of rate estimation methods and a comparison thereof, see Cunningham et al., 2009.)

As a demonstration of the many possible applications of the general framework and the resultant benefits, the framework is adapted to the time histogram and the kernel smoother, the two most common approaches to rate estimation used in the experimental literature. We discuss the specific innovations necessary to adapt the methods to the likelihood framework, as well as optimizations for efficiently computing the cross-validated likelihood. The goal here is not to uncover a single optimal rate estimation procedure, but rather to illustrate the vast reduction in estimation error that can be achieved when the general likelihood framework is applied to even the most basic means of spike smoothing. We present a simulation study as well as applications to real data, in which we compare the standard methods, using ad hoc parameter selection, to their cross-validated incarnations, for which the bandwidth is selected based on the cross-validated likelihood procedure. We discuss possible future applications for the framework and potential improvements in estimation accuracy achieved by doing so.

2 Materials and Methods

2.1 Calculating Firing Rate Using Smoothers. Firing rate, $r(t)$, is most generally defined as the instantaneous probability of spiking at time t,

$$r(t) = \lim_{\Delta t \to 0} \frac{\Pr\left[\text{spike in } (t, t+\Delta t]\right]}{\Delta t} = \lim_{\Delta t \to 0} \frac{E\left[\Delta N_{(t,t+\Delta t]}\right]}{\Delta t}, \tag{2.1}$$

where $\Delta N_{(a,b]}$ is the spike count between times a and b.

The instantaneous probability of a neuron firing at a given time is typically unknown, so the expected number of spikes in a small interval must be estimated. Numerous different rate estimates can be computed for the same neural data (see Figure 1) depending on the smoothing methodology, model, or parameters used in the estimation of $E\left[\Delta N_{(t,t+\Delta t)}\right]$.

In Figures 1a through 1d, four different firing rate estimates are shown for the same 5 s simulated spike train, along with the spike raster. For this spike train, 250 ms (see Figure 1a) and 800 ms (see Figure 1b) time histograms, as well as 100 ms (see Figure 1c) and 1500 ms (see Figure 1d) Hanning kernel smoothers (described in detail in the appendix), are computed. Smoothers with a small bandwidth, like those in Figures 1a and 1c, combine information from highly localized temporal regions and

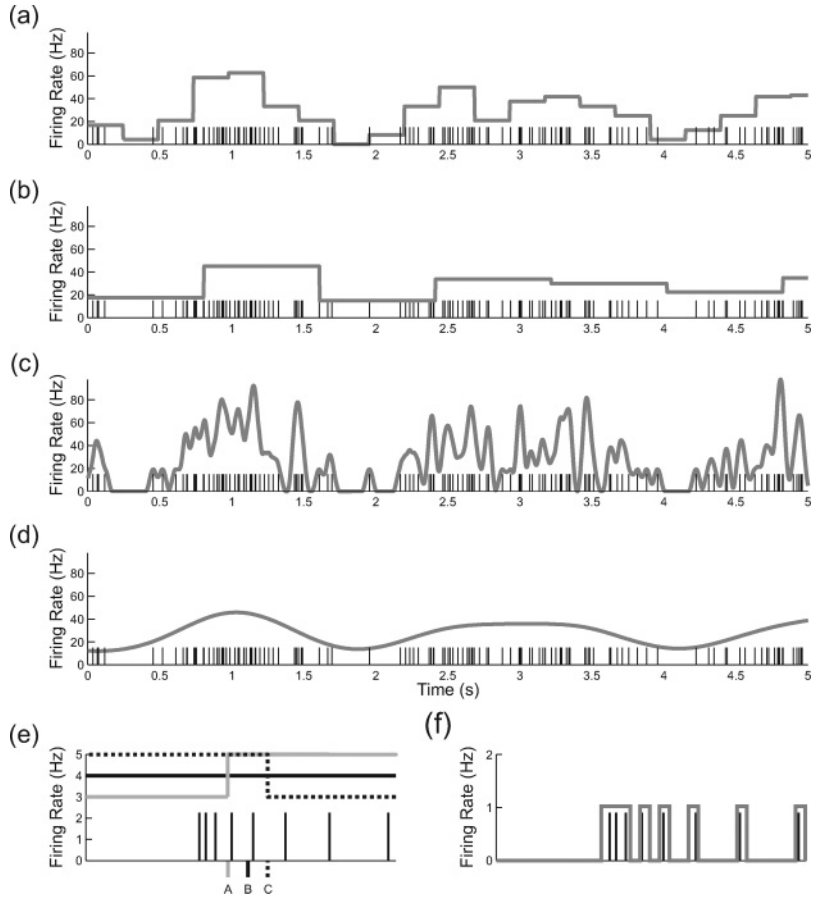

Figure 1: Variation in rate estimation due to parameter selection for spike train. For a 5s stimulated spike train, 250 ms (a) and 800 ms (b) time histograms, as well as 100 ms (c) and 1500 ms (d) Hanning kernel smoothers are computed. (e) A theoretical spike train from multiple trials from a sparsely firing neuron showing frequency adaptation can be shown to have increasing (gray curve), constant (solid black curve), or decreasing (dashed black curve) rates based on the position of the bin boundary. (f) An undersmoothed rate estimate (gray curve) is shown for a similar theoretical spike train.

produce fluctuating estimates with a short time course for changes in firing rate. Smoothers with a large bandwidth, like those in Figures 1b and 1d, combine information from wide temporal regions and produce smooth rate estimates with a long time course for changes in firing rate.

In practice, an accurate evaluation of the bandwidth of neural firing is important when making context-based comparisons of firing activity or computing correlations between firing rate and simultaneously observed experimental measurements. In the electrophysiological literature, bandwidth is almost exclusively chosen ad hoc, without statistical justification. In the absence of a principled approach for selecting bandwidth, parameters are often chosen so that the bandwidth of the smoothed firing rate matches that of a stimulus that is presupposed to be correlated with the spiking data. Oversmoothing or undersmoothing the data due to inappropriate bandwidth selection can affect statistical analyses and lead to incorrect conclusions about the neural system.

Figures 1e and 1f show a schematic of the superposition of very sparsely spiking activity over multiple trials, exhibiting a decreasing spike rate similar to a neuron with phasic response properties. In Figure 1e, three time histograms are shown, based on different bandwidths, that affect bin boundaries. If the boundary of a bin fell at point A, there would be three spikes in the first bin and five spikes in the second bin—describing increasing spiking activity (gray curve). If the bandwidth was increased slightly so that bin the boundary fell at point B, there would be four spikes in each bin—describing constant spiking activity (black curve). If the bandwidth was increased slightly more so that the bin boundary fell at point C, then five spikes would fall in the first bin and three in the second bin—describing decreasing spiking activity (dashed curve). It is a prevailing intuition that picking a small bandwidth is always safe because it allows greater flexibility in estimates. However, undersmoothing the data can obfuscate larger, more important trends. In Figure 1f, a small bandwidth is used in the time histogram rate calculation and provides little information about the larger trends in firing rate.

A more striking example of the dangers of incorrect bandwidth selection is illustrated by the spiking activity shown in the example in Figures 1a to 1d, which was generated using a physiological model of neural firing based on an external stimulus. While the firing rates shown in Figures 1a to 1d may all appear to be reasonable approximations of the neural spiking activity, we shall see that two of the four rate estimates produce incorrect conclusions about the relationship between the spiking data and the stimulus. Thus, it is vital to choose bandwidth in such a way as to preserve correlation when it exists and prevent spurious correlation estimates when it does not.

It should be noted that herein, all presented spiking data can be viewed as a spike train from a single trial, the superposition of the spikes from multiple trials, or the aggregate spike count across trials. While multitrial data are most prevalent in the neuroscience literature, there are many instances within neuroscience in which single-trial data are routinely used, such as the decoding procedures found within in neural prosthetics algorithms or the calculation of correlations between neural firing rates and other

stimuli. Thus, the examples may be viewed from the standpoint of either paradigm.

2.2 The Cross-Validated Likelihood of the Missing Data. As a general strategy, we would like to estimate a firing rate with a bandwidth parameter derived from the underlying temporal structure of the spike train data using a likelihood-based, rather than adhoc, cost function. A well-known method for fitting model parameters is the maximum likelihood (ML) approach, in which parameter values are chosen such that they maximize the likelihood of the data given the model estimate. However, ML-based bandwidth estimation procedures for firing rate will tend to select small bandwidths; they describe the data well but do not generalize to future data.

In order to prevent overfitting, we borrow from the nonparametric regression (Wong, 1983; Picard & Cook, 1984; Silverman, 1984; Racine, 1993; Walk, 2002; Leung, 2005) and density estimation literature (Hall, Racine, & Li, 2004; Duong & Hazelton, 2005) and use cross-validation techniques to select parameters for a given smoothing procedure. Cross-validation (Efron & Gong, 1983), often used in neuroscience for model parameter selection (Harris et al., 2003), is a statistical procedure to minimize overfitting in model parameter selection by separating the data used to fit the model from the data used to evaluate the model. While there are many variations on cross-validation schemes, the form most often used within neuroscience separates repeated trials into groups of training sets, on which a model is fit, and test sets, on which the model is verified.

Cross-validation is more challenging when the data set is a time series, as it must be performed so as to retain the temporal structure of the data. To do this, a subset of the data is removed, and an estimate of the model values at the missing points is calculated from the remaining data. The parameter value that optimizes a specified error measure between the missing data and the model estimates is computed. In our case, we select the bandwidth parameter that maximizes the likelihood of observing the missing data given the remaining data. We present a general framework for a leave-one-out cross-validated estimate of firing intensity and then show the illustrative example of adapting the kernel smoother for use within this framework.

2.2.1 Leave-One-Out Cross-Validation for Spike Train Time-Series Data. To create a cross-validated estimate of firing rate, we first adapt the concept of time-series cross-validation to spike train data. Figure 2a shows a schematic of a cross-validation procedure applied to a hypothetical spike train in discrete time, with the spike count displayed in each time bin below. The seventh data point (shown as a question mark) is a missing data point that is removed as part of the cross-validation procedure. Given the spikes and this missing data point, we illustrate two hypothetical rate estimates: one from a smoother with a large bandwidth (see Figure 2a, black curve) and one from a smoother with a small bandwidth (see Figure 2a, gray curve).

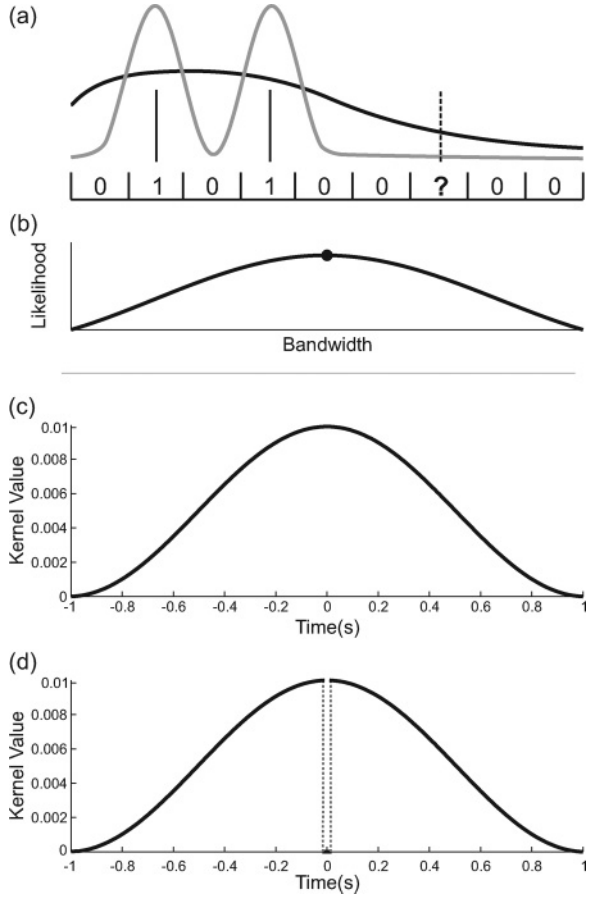

Figure 2: Schematic of the spike train cross-validation process for kernel smoothers. (a) A theoretical example of a spike train in discrete time with large-bandwidth (black curve) and small-bandwidth (gray curve) rate estimates. The spikes are shown above the discretized spike count, with missing data represented by the dashed spike over the question mark. (b) A theoretical likelihood curve (black curve) that displays the likelihood of the rate given the spikes as a function of bandwidth. The maximum likelihood bandwidth value is marked with the black dot. Examples of a Hanning kernel function (c) and its notch filter formulation (d) are shown. Both windows have a period width of 2 seconds.

Since the firing rate is directly proportional to the probability of firing in a specified time interval, the event that a spike occurs in that bin is more likely given a large bandwidth, while the event that a spike does not occur in that bin is more likely under a small bandwidth.

To capture the firing dynamics of the entire spike train, we perform a complete leave-one-out cross-validation procedure. At each step, we remove the spike data from a single temporal bin, compute a rate estimate within that bin from the remaining data, and calculate the likelihood of the missing data point given that rate estimate. The logarithms of the missing data likelihoods are then summed to find the leave-one-out cross-validated log likelihood. By combining the cross-validated likelihoods from every temporal bin, we are able use all of the data to find the bandwidth that best accounts for the entire spike train and will not overfit to a specific data set. For example, if the missing data point in the example in the seventh bin were truly a spike, the large bandwidth rate, which falls slowly toward 0 Hz, would predict the spike well, yet provide a poor prediction of the nonspike times surrounding it. The small bandwidth rate, which falls quickly toward 0 Hz, would predict the nonspike times well yet provide a poor prediction of the missing spike. We therefore want to use all the data to achieve a balance between underpredicting spikes and overpredicting nonspikes. For a given spike train, we calculate the complete cross-validated likelihood over a wide range of bandwidths (see Figure 2b) and choose the bandwidth that maximizes the cross-validation likelihood.

2.3 A General Likelihood Framework for Firing Rate Estimation. We define a spike train in discrete time with N_T observations as the sequence of spike counts,

$$s_i = \Delta N_{(t_i, t_{i+1}]}, \tag{2.2}$$

where $\Delta N_{(a,b]}$ is the spike count between times a and b, and

$$t_i = i\,\Delta t \tag{2.3}$$

given a fixed time resolution, Δt, and an integer $i \in \{1, \ldots, N_T\}$.

A feature common to all rate estimation algorithms is that they produce a rate or intensity function at each point at time, which can be used to estimate the instantaneous probability of spiking. This rate or intensity may vary as a function of time, past spiking history, and other covariates and will depend on the choice of a bandwidth parameter. We notate the output of the rate estimation algorithm at time t_i as $\lambda(t_i|H_i, \theta)$, where H_i is past spiking history from time 0 to t_i and θ is the bandwidth parameter. We define $\lambda^-(t_i|H_i, \theta)$ to be the result of the algorithm when s_i, the observed number of spikes in the interval $(t_{i-1}, t_i]$, is unknown.

Using the log likelihood of a general point process (Daley & Vere-Jones, 2003), we can calculate $\log[L_{cv}(\theta)]$, the leave-one-out cross-validated log likelihood of the spike train from the rate or intensity estimate resulting

from any algorithm,

$$\begin{aligned}\log\left[L_{cv}\left(\theta\right)\right] &= \sum_{i=1}^{N_T} s_i \log\left[\lambda^{-}\left(t_i|H_i,\theta\right)\Delta t\right] \\ &\quad - \lambda^{-}\left(t_i|H_i,\theta\right)\Delta t - \log\left[s_i!\right].\end{aligned} \tag{2.4}$$

We can then compute the maximum likelihood estimate of θ,

$$\theta_{\max} = \arg\max_{\theta} \log\left[L_{cv}\left(\theta\right)\right], \tag{2.5}$$

which is the bandwidth parameter that maximizes the cross-validated likelihood, and therefore best predicts missing spike data.

The general likelihood framework is summarized in this way:

A General Likelihood-Based Bandwidth Selection Procedure for Spike Trains

1. For each of a broad range of values for bandwidth parameter θ:
   a. Calculate the leave-one-out rate estimate, $\lambda^{-}\left(t_i|H_i,\theta\right)$, at each time point
   b. Calculate the leave-one-out cross-validated log likelihood using equation 2.4.
2. Compute $\theta_{\max}$, the bandwidth that maximizes the full cross-validated log likelihood, using equation 2.5.
3. Calculate the full rate, $\lambda\left(t_i|H_i,\theta_{\max}\right)$.

As cross-validation is a method of model selection orthogonal to the method of rate calculation itself, the framework is algorithm independent given that a reasonable $\lambda^{-}\left(t_i|H_i,\theta\right)$ can be computed. Thus, while automatic bandwidth estimation algorithms have been developed for several specific smoothing procedures, this general framework can be broadly applied to any of the multitude of established methods for firing rate or conditional intensity estimation.

We next show two illustrative examples of the application of this general framework to the estimation methods most commonly presented in the literature. The goal here is not to describe a particular optimal smoother but rather to show the benefits reaped from this computational strategy, even when applied to the simplest of rate estimation methods.

2.4 Likelihood-Based Bandwidth Selection for the Time Histogram. To demonstrate the use of the general framework, we first apply the likelihood framework to the method for rate estimation most often used in electrophysiological literature: the time histogram. A full mathematical

derivation of the time histogram and a reiteration of its associated caveats can be found in the appendix. The time histogram, or peristimulus time histogram (PSTH) when time is defined in relation to an event, breaks up time ranging from 0 to t_{N_T} into K equal and nonoverlapping histogram bins of size $\Delta b = t_{N_t}/K$, and computes $r(t)$, the mean spike count per temporal band (defined by equations A.3 and A.4 in the appendix, respectively).

2.4.1 Handling Missing Data. In order to apply the likelihood framework, we must first adapt the time histogram to allow for missing data. Conceptually we can achieve a leave-one-out rate estimate for a given histogram bin by excluding the missing data time segment and its spike count from the rate computation. This gives us an estimate of the rate within the histogram bin without using any of the information from within the missing time bin.

Explicitly, $r^-(t_m, K)$, the leave-one-out time histogram rate for a bandwidth K and missing data at t_m, is defined as

$$r^-(t_m, K) = \frac{\Delta N_{(k\Delta t,(k+1)\Delta t]} - \Delta N_{(m\Delta t,(m+1)\Delta t]}}{\Delta b - \Delta t}, \tag{2.6}$$

where k is an integer ranging from 0 to $K-1$, which denotes the bin number into which t falls. The numerator $\Delta N_{(k\Delta t,(k+1)\Delta t)} - \Delta N_{(k\Delta t,(k+1)\Delta t)}$ is the difference of total spike count within the entire histogram bin and the spike count within the missing data time bin. The denominator is the difference between the total time of the histogram bin, Δb, and the duration of the time bin, Δt.

2.4.2 The Complete Cross-Validated Likelihood. Once the missing data firing rate estimate is calculated at every temporal bin using equation 2.6 given a particular bandwidth, we calculate the leave-one-out cross-validated likelihood of the spikes given that estimate,

$$\begin{aligned}\log[L_{cv}(K)] &= \sum_{i=1}^{N_T} \log[L(s_i, K)] \\ &= \sum_{i=1}^{N_T} s_i \log\left[r^-(t_i, K)\,\Delta t\right] - r^-(t_i, K)\,\Delta t - \log[s_i!],\end{aligned} \tag{2.7}$$

which is simply equation 2.4, where r^- is substituted for λ^-.

To determine the optimal cross-validated bandwidth, we select a broad range of bandwidth values and calculate the leave-one-out cross-validated loglikelihood for each bandwidth. We then choose the bandwidth value that maximizes equation 2.7, which we call $K_{\max}$. To generate the final rate estimate, we perform a standard time histogram full-time histogram using equation A.4, where $K = K_{\max}$ in equation A.3.

We summarize the entire process as follows:

The Likelihood-Based Bandwidth Estimation Procedure for Spike Trains Using the Time Histogram

1. If a PSTH is to be computed with data from multiple trials, use the sum of the spike counts across all trials within a temporal bin as data, and divide the final rate estimate by the number of trials.
2. For each of a broad range of histogram bandwidths (bin widths) K:
   a. Calculate the leave-one-out rate estimate at all time points using equation 2.6.
   b. Calculate the full leave-one-out cross-validated log likelihood using equation 2.7.
3. Choose the histogram bandwidth $K_{\max}$ that maximizes the full cross-validated likelihood, using equation 2.7.
4. Calculate the final rate estimate using a standard time histogram from equation A.4, with bandwidth $K = K_{\max}$ in equation A.3.

2.5 Efficient Likelihood-Based Bandwidth Selection for the Kernel Smoother. The kernel smoother is another simple, commonly used method for rate estimation. While the kernel smoother does not include certain useful aspects found in other more sophisticated models, such as history dependence or dynamic bandwidth, it is in many ways far superior to the time histogram. The kernel smoother uses a sliding window for rate estimation, which produces a smooth estimate without discontinuities and avoids the arbitrary partitioning pitfalls of the fixed-bin time histogram. The kernel smother also allows a temporal weighting of the spiking data, whereas the time histogram weights all spikes within a temporal bin equally. Given the kernel selected, purely causal filters may be computed. The derivation of the kernel smoother is described detail in the appendix.

2.5.1 Handling Missing Data. In order to apply the likelihood framework, we must first adapt the kernel smoother to allow missing data. In a standard convolution, as defined in equation A.5, all of the temporal bins surrounding and including the current bin are combined in a weighted average as defined by the kernel function. If we use a kernel that excludes the current bin from the weighted average calculation, we can perform a convolution that incorporates only information from surrounding bins and treats the current temporal bin as missing data.

We accomplish this by defining a notch kernel $w^-(t, K)$, which is any kernel function that is a function of time t, has a bandwidth parameter of value K, and for which $w^-(0, K) = 0$. Given a notch kernel w^-, we can then calculate the leave-one-out rate estimate r^-, which is defined as

$$r^-(t_m, K) = \sum_{i=1}^{N_T} \frac{w^-(t_m - t_i, K)}{\nu} s_i \Delta t, \tag{2.8}$$

where t_m is the time of the missing data point, N_T is the number of discrete temporal bins in the spike train, and ν is a normalization factor,

$$\nu = \sum_{i=-\infty}^{\infty} w^{-}(t_i, K), \tag{2.9}$$

which ensures that w^{-}will always integrate to 1.

Herein, the kernel we have chosen is a variant of the Hanning function (see Figure 2c),

$$w(t, K) = \begin{cases} 0.5\left(1 + \cos\left(\dfrac{2\pi t}{K-1}\right)\right) & \text{for } -\dfrac{K}{2} < t \le \dfrac{K}{2} \\ 0 & \text{otherwise} \end{cases}, \tag{2.10}$$

where K is a number that defines the period size. We selected the Hanning kernel over the more traditional gaussian kernel because in discrete time, the bandwidth parameter, K, is directly proportional to the number of samples used in the kernel, and thus easy to interpret, though either kernel would function comparably well.

The notch kernel form of the Hanning kernel (see Figure 2d) is defined as

$$w^{-}(t, K) = \begin{cases} 0.5\left(1 + \cos\left(\dfrac{2\pi t}{K-1}\right)\right) & \text{for } -\dfrac{K}{2} < t \le \dfrac{K}{2}, \text{ and } t \neq 0, \\ 0 & \text{otherwise} \end{cases} \tag{2.11}$$

where K is a number that defines the magnitude of the notch kernel bandwidth. This is simply the same equation as equation 2.10, but with the function taking on a value of 0 at $w^{-}(0, K)$. The notch kernel can be calculated likewise for any viable kernel function, such as the gaussian.

Convolving the spike train with the notch filter is a computationally efficient way of calculating the value of leave-one-out smoother (Leung, 2005) at every temporal bin in the data set. By using leave-one-out convolution, we incorporate all the surrounding data in the weighted average while excluding the current bin, thus adapting the kernel smoother to handle missing data in the cross-validation procedure. If we perform the leave-one-out convolution for an entire time-series data set, we get the missing data rate estimate for every point at every time. This is the first step in the leave-one-out cross-validation procedure, which we efficiently compute in a single pass.

2.5.2 The Complete Cross-Validated Likelihood. Once the missing data firing rate estimate is calculated at every temporal bin with a notch filter of a particular bandwidth, we calculate the leave-one-out cross-validated likelihood using equation 2.7 as was done with the time histogram, this time using equation 2.8 for $r^{-}(t_m, K)$. Again, we select a broad range of bandwidth values and calculate the leave-one-out cross-validated log likelihood for each bandwidth. We then choose the bandwidth value that maximizes equation 2.7, which we call $K_{\max}$. To generate the final rate estimate, we perform a convolution using the full Hanning kernel (see Figure 2a, and equation 2.10) with a bandwidth equal to $K_{\max}$.

We summarize the entire process as follows:

The Likelihood-Based Bandwidth Estimation Procedure for Spike Trains Using the Kernel Smoother

1. For each of a broad range of kernel bandwidths K:
   a. Calculate the leave-one-out rate estimate on the full spike train using equations 2.8 and 2.11.
   b. Calculate the full leave-one-out cross-validated log likelihood using equation 2.7.
2. Choose the kernel bandwidth $K_{\max}$ that maximizes the full cross-validated likelihood using equation 2.7.
3. Calculate the final firing rate estimate using a standard convolution, equation A.5, and the full kernel, equation 2.10, with $K_{\max}$ as the bandwidth parameter.

2.6 Parameter Confidence. To calculate confidence bounds on the bandwidth estimates from any method adapted to the framework, we can compute the observed Fisher information, which in this case is the inverse of the second derivative of the full cross-validated log-likelihood, equation 2.7, as a function of bandwidth. The confidence interval for the bandwidth parameter is

$$CI = K_{\max} \pm 2\left[-\frac{\partial^2 \log\left[L_{cv}\left(K_{\max}\right)\right]}{\partial K^2}\right]^{-\frac{1}{2}}. \tag{2.12}$$

We use this confidence interval in our calculations as a measure of uncertainty in our bandwidth estimate.

2.7 Simulated Spike Train Data. We next demonstrate the ability of the general likelihood framework to greatly reduce the estimation error of the kernel smoother relative to ad hoc parameter estimation. Since the firing rate is not directly observable, we must use spike trains generated from known

simulated "true rates" to viably assess estimation error. To generate these known rates, we use spline interpolation to fit curves that pass through control points that are equally spaced in time and have magnitudes that vary uniformly from 2 Hz to 110 Hz (e.g., see Figures 3a and 3c, red curves). By changing the number of control points used in the interpolation, the bandwidth of the true rates can be adjusted, which produces rates with a wide range of bandwidths. We tested the ability of the cross-validated kernel smoother to estimate firing rate from spike trains generated by simulating an inhomogeneous Poisson process based on the true rate r_{true}. To do this, we generated a spike count at each discrete time bin by simulating a Poisson random variable at a time resolution $\Delta t = 33.3$ ms with a rate parameter of

$$\lambda(t) = r_{true}(t) \tag{2.13}$$

given the true rate curve r_{true}.

2.8 Simulation Study. To verify the consistent accuracy of the cross-validated kernel smoother, we performed a large-scale simulation study using many spike trains generated from simulated random spline rates with a wide range of bandwidths. We created a set of known true rate random spline curves with 5, 10, 20, 30, or 50 possible control points. For each number of control points, we generated 200 curves with control point magnitudes ranging uniformly from 2 Hz to 110 Hz, resulting in 1000 total simulated rates. From each of these 1000 rate curves, spike trains were generated using the true rate and equation 2.13. From these spike trains, we then reestimated the true rate by using the cross-validated kernel smoother, as well as by using many different Hanning kernel smoothers with a broad range of kernel widths.

In order to evaluate our estimated rates, we then calculated the mean-squared error (MSE) between each of the cross-validated smoother estimates and the corresponding true rates. To illustrate how estimation error changes as a function of bandwidth, we calculated the mean MSE for each set of rates generated from a given number of control points. We computed confidence intervals about the mean MSE using a nonparametric bootstrap to build a distribution of the mean statistic at each control point. This was done by sampling with replacement from the observed MSEs many times and calculating the mean at each iteration. From these estimates of the mean statistic, we took the 2.5th and 97.5th percentiles of the computed distribution to serve as confidence intervals about the mean MSE.

3 Results

3.1 Rate Reconstruction. Figures 3a to 3d shows the output of the cross-validated kernel smoother for two spike trains generated from two

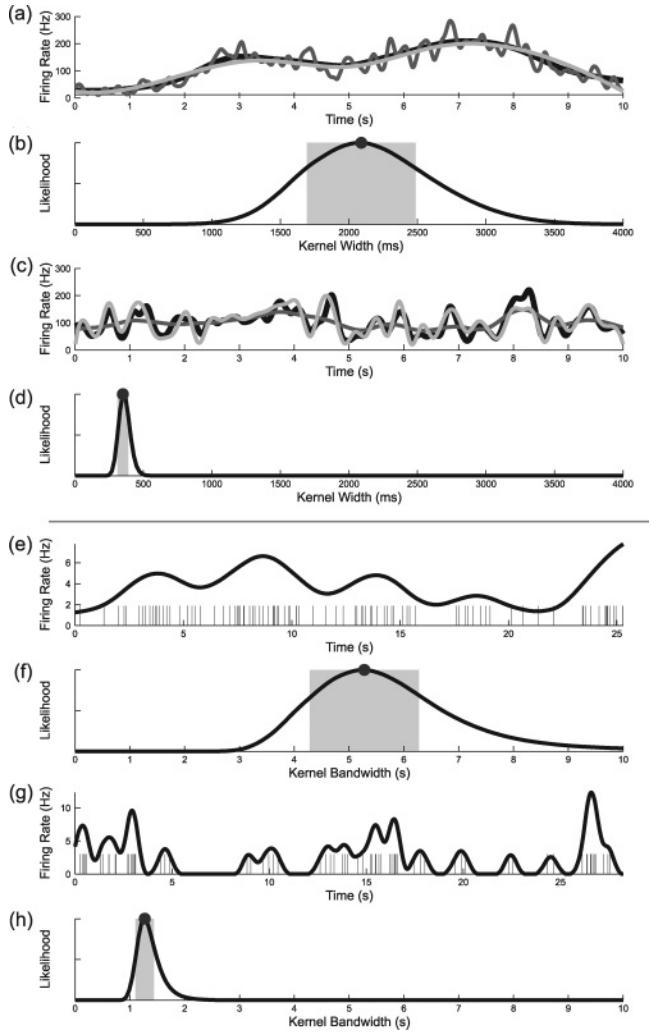

Figure 3: The cross-validated kernel smoother applied to simulated and experimental data. (a, c) The rate estimates for the cross-validated kernel smoother (black curves) from single spike trains (not shown) generated from a true rate (thick light gray curves). For comparison, estimates from Hanning kernel smoothers with widths 250 ms (a, dark gray curve) and 1510 ms (c, dark gray curve) are shown. (e, g) The rate estimates for the cross-validated kernel smoother (black curves) for spike trains from two neurons in the rat MEC during single trials of a T-maze spatial alternation task. (b, d, f, h) The cross-validation likelihood against the notch kernel width, along with the maximum likelihood values (gray dots) with 95% confidence (light gray regions) as calculated by the observed Fisher information matrix.

simulated true rates with differing bandwidths. Figure 3a shows a large bandwidth rate generated using seven control points. The true firing rate used to generate the spikes is shown as the thick, light gray curve. Using spikes generated from this true firing rate, we calculated the estimates using the cross-validated kernel (black curve), as well as a Hanning kernel smoother with a fixed width of 250 ms (dark gray curve) for comparison. Figure 3b shows the leave-one-out cross-validated likelihood plotted against the Hanning notch kernel width in milliseconds. The maximum likelihood estimate of kernel width was calculated to be 2090 ms (gray dot) with a confidence interval (light gray region) ranging from 1690 ms to 2490 ms, as determined by equation 2.12. The cross-validated estimate accurately reproduces the smooth bimodal structure of the true rate and has a root-mean-squared error (RMSE) of 10.89 Hz. In comparison, the kernel smoother with an overly small bandwidth produces a rate with a much shorter time course of change than the true rate and has an RMSE of 28.20 Hz.

Figure 3c shows a small bandwidth rate generated using 50 control points. Using spikes generated from this true firing rate, we calculated the estimates using the cross-validated kernel (black curve), as well as a Hanning kernel smoother with a fixed width of 1500 ms (dark gray curve) for comparison.

Figure 3d shows the leave-one-out cross-validated likelihood plotted against the Hanning notch kernel width. The maximum likelihood estimate of kernel width was calculated to be 350 ms (gray dot) with a tight confidence interval (light gray region) ranging from 310 ms to 390 ms, as determined by equation 2.12. The cross-validated estimate tracks the overall structure of the true rates, capturing many of the numerous peaks. The RMSE of the cross-validated kernel smoother is 23.55 Hz. In comparison, the kernel smoother with an overly large bandwidth produces a rate with a much longer time course of change than the true rate and tracks only the general trend of the data. As a result, the large bandwidth kernel smoother has an RMSE of 36.62 Hz.

In both cases, the cross-validated kernel smoother accurately captures the bandwidth of the true process from the single trial of spikes. The ability of cross-validated kernel smoother to estimate the bandwidth enables us to have an appropriate data-based representation of firing rate, regardless of the underlying time course of change.

3.2 Simulation Study of Cross-Validated Kernel Smoother Performance. The results of our simulation study are shown in Figure 4. Figure 4a illustrates how the error of rate estimates change with respect to the bandwidth of the true rate. The mean MSE is plotted against the number of control points used in generating the spikes. The mean MSE (dots) and its bootstrapped 95% confidence (shaded regions) are shown for cross-validated kernel smoother and selected Hanning smoothers with fixed

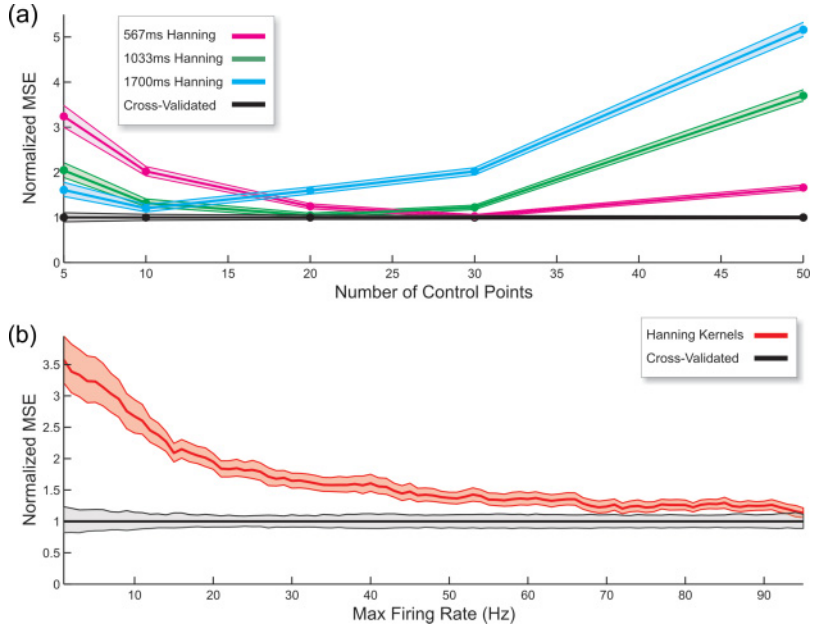

Figure 4: Simulation study results for cross-validated kernel smoother performance. In (a) the normalized mean MSE (dots) and 95% bootstrapped confidence (shaded regions) are shown for the cross-validated kernel smoother (black), as well as for Hanning kernel smoothers with small (17 bin, magenta), medium (31 bin, green), and large (51 bin, blue) bandwidths. (b) The mean MSE (curves) and 95% bootstrapped confidence (shaded regions) for the cross-validated kernel smoother (black) and the small Hanning kernel smoother (red) plotted against the maximum firing rate of true rates used to generate the simulated spike trains.

bandwidths. In order to easily compare the relative performance of the cross-validated kernel smoother to the different-sized Hanning kernel smoothers, all values were normalized by the mean MSE of the cross-validated smoother at each number of control points. When the bandwidth of the kernel is set to a fixed value, the mean MSE of the resulting convolution performs poorly when the kernel bandwidth is smaller than the rate bandwidth, gradually improves until the bandwidth of the rate is equal to that of the cross-validated smoother, and then increases in error when the kernel bandwidth becomes larger than the rate bandwidth. This is evident in the selected Hanning kernels with bandwidths of 17 bins (567 ms, magenta), 31 bins (1033 ms, green), and 51 bins (1700 ms, blue) (odd bin sizes ensure that the kernel convolution result is the same size as the original data set, without the use of zero padding). These bandwidth selections

were chosen to show the range changing accuracy across the given control points used in the simulation. The large bandwidth (1700 ms) kernel smoother performed well for spike trains with a slow time course of change, performing maximally at 10 control points, with a marked increase in MSE as the rate bandwidth increased. The medium (1033 ms) bandwidth Hanning kernel smoother showed the full extent of the performance curve over our range of control points, with decreasing MSE from 5 to 10 control points, increasing error from 30 to 50 control points, and the best performance at 20 control points. The small bandwidth (567 ms) Hanning kernel smoother showed decreasing MSE from 5 to 20 control points, increasing error from 30 to 50 control points, and the best performance at 30 control points. These curves are in contrast to the MSE of the cross-validated kernel smoother (black), which performs consistently equal to or better than the fixed kernel smoothers regardless of the bandwidth of the true rate.

To analyze the change in performance based on the sparseness of data, we calculated the mean MSE and bootstrapped 95% confidence intervals for the cross-validated kernel smoother as well as Hanning smoothers, which we plotted against the maximum firing rate of the true rate from which the spikes were simulated. In order to easily compare the relative performance of the cross-validated kernel smoother to the different sized Hanning kernel smoothers, all values were normalized by its mean MSE across the maximum firing rate. For any Hanning kernel width, the error in estimate increased relative to the cross-validated kernel smoother as the maximum firing rate decreased. We performed this analysis using kernels with numerous different bandwidths, and the overall structure of the curve of the MSE as a function of number of spikes looked virtually identical regardless of the bandwidth used, changing only in scale. Figure 4b compares the simulation results for the cross-validated kernel smoother (black) and a Hanning kernel smoother of width 567 ms (red). This result indicates that a proper choice of bandwidth matters the most for neurons with low firing rates. Thus, the cross-validated kernel smoother is potentially the most useful for sparsely firing neurons, such as those found in hippocampal and parahippocampal regions.

3.3 Analysis of Spike Trains from Experimental Data. To illustrate our method in the context of real experimental data, we applied the cross-validated kernel smoother to spike trains from Lipton, White, and Eichenbaum (2007), in which tetrode recordings of neural activity from the hippocampus and entorhinal cortex were recorded from rats performing a spatial alternation task on a T-maze. These neurons are known to fire sparsely and would thus benefit from the cross-validation bandwidth estimation procedure, given the results of the second simulation. Figures 3e to 3h shows the output of the cross-validated kernel smoother for spike trains from two neurons in the rat MEC during single trials of a spatial alternation task. Both neurons are from the same animal. Figure 3e shows the firing

rate estimate from the cross-validated kernel (black curve) for the first spike train, and Figure 3f shows the leave-one-out cross-validated likelihood plotted against the kernel width in milliseconds. The maximum likelihood estimate of the kernel width was calculated to be 5278 ms (see Figure 3f, red dot) with a confidence interval (see Figure 3f, green region) ranging from 4284 ms to 6271 ms. The large bandwidth structure of the estimate implies that this neuron has a relatively slow time course of rate change during this trial.

Figure 3g shows the firing rate estimate from the cross-validated kernel (black curve) for the other spike train, and Figure 3h shows the leave-one-out cross-validated likelihood plotted against the Hanning notch kernel width in milliseconds. The maximum likelihood estimate of kernel width was calculated to be 1272 ms (see Figure 3h, red dot) with a confidence interval (see Figure 3h, green region) ranging from 1103 ms to 1440 ms. The small bandwidth multimodal structure of the estimate implies that this neuron has a relatively fast time course of rate change during this trial.

3.4 Analysis of Firing Rate Correlation to Behavioral Data for a Simulated Motor Trajectory Task. To illustrate how assumptions about the bandwidth of a firing rate estimate translate to higher-level conclusions, we examined a simulated motor neuron in the primary motor cortex of a monkey during an arm trajectory task. To generate the arm trajectory, we simulated the x and y components of the arm velocity using two independent AR(100) models with parameters fit from real primate experimental data (Eden, Truccolo, Fellows, Donoghue, & Brown, 2004) and then calculated the magnitude, $v(t)$ in cm/s, and the angle, $\phi(t)$ in degrees, of the arm velocity. We simulated a spike train from a cosine tuning model of motor neuron activity (Moran & Schwartz, 1999) where the firing rate $r(t)$ is defined as

$$r(t) = \max\left(\alpha + \beta_1 v(t)\cos\left[\phi(t)\right] + \beta_2 v(t)\sin\left[\phi(t)\right], 0\right), \tag{3.1}$$

where, in this case, $\alpha = 25,\ \beta_1 = 3,\ \beta_2 = 3$. The trajectory was simulated for 5 s with a Δt of 4 ms, and a spike train was generated from $r(t)$ (see Figure 5a, gray curve) using an inhomogeneous Poisson process model.

From the spike train, we calculated the cross-validated kernel smoother estimate of firing rate (see Figure 5a, black curve), which followed the general trend of $r(t)$ over time. To show the effects of bandwidth on correlation analyses, we calculated the Hanning kernel smoother and the time histogram firing rate estimates for all possible bandwidth sizes that were multiples of Δt from 4 ms to 5000 ms. We computed the least-squares linear regression between the movement trajectory and estimated firing rate for each bandwidth value and plotted the resulting R^2 value against bandwidth

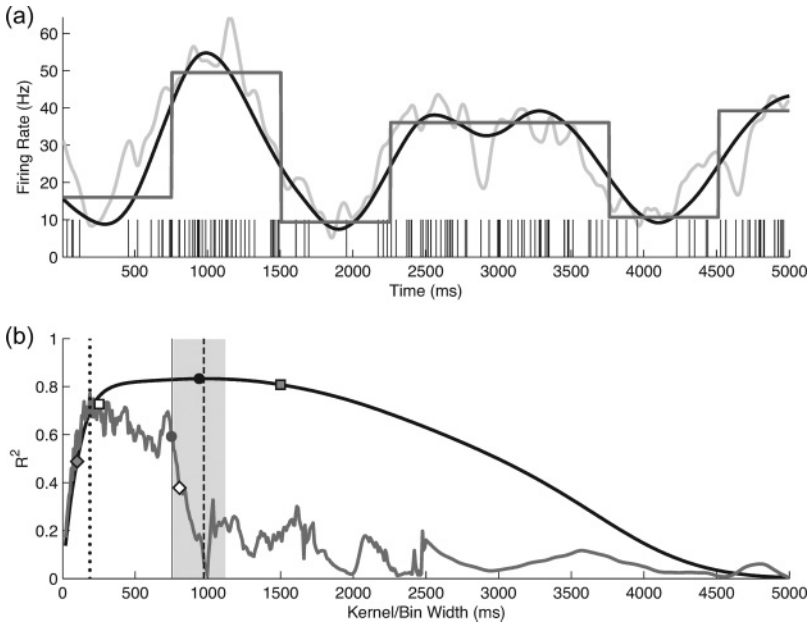

Figure 5: A correlation analysis of simulated data from the motor cortex of a monkey during a motor trajectory task using a cosine tuning model of neural activity with parameters fit from actual experimental data. (a) The true rate given by the model (light gray curve) is plotted against time histogram (dark gray curve) and cross-validated kernel smoother (black curve) estimates of firing rate, calculated from spikes generated from the true rate using an inhomogeneous Poisson process. (b) The values of R^2 for a Hanning kernel smoother (black curve) and time histogram methods (dark gray curve) versus bandwidth, which was calculated using least-squares linear regression with the generating model. The kernel and bin widths that maximize R^2 for the Hanning kernel smoother (b, dashed line) and the time histogram (b, dotted line) are shown. Also displayed are the computed bandwidths and corresponding R^2 values for the cross-validated kernel smoother (b, black dot) and time histogram (b, gray dot), along with their confidence bounds (light gray region, thin dark gray region, respectively). Additionally, the corresponding R^2 values for the 800 ms (b, white square) and 250 ms (b, white diamond) time histograms, as well as 1500 ms (b, gray square) and 100 ms (b, gray diamond) Hanning kernel smoothers are shown.

for both the Hanning kernel smoother and the time histogram. Since we know that the spikes are truly correlated with the movement data, we expect to find high correlation for the bandwidths that accurately capture the data structure. The R^2 curve for the Hanning kernel smoothers (see Figure 5b, black curve) is smooth and rises sharply from a kernel bandwidth of 0 to

300 ms before declining slowly after a maximum R^2 of 0.83 for a kernel width of 972 ms (see Figure 5b dashed line). In contrast, the R^2 curve for the histogram-based rate estimate (see Figure 5b, gray curve) fluctuates rapidly and rises sharply from 0 ms to its maximum R^2 of 0.77135 for a bin width of 188 ms. The R^2 curve then stays at a noisy plateau from 150 ms to 750 ms, before it drops drastically from 750 ms to 1000 ms and remains low and noisy for larger bin sizes. These results indicate that the use of a kernel smoother offers a higher degree of correlation as well as a broader margin of error in bandwidth selection than does the time histogram. However, even slight changes in bandwidth can result in conflicting conclusions about the correlation of the spiking activity to the behavioral data. For the time histogram, a change in bandwidth of only 250 ms, from 750 ms to 1000 ms, changes the R^2 from about 0.7 to an R^2 of about 0.01. For both the time histogram and the kernel smoother, the R^2 plummets from about 0.75 toward 0 for bandwidths of 200 ms and below. This is particularly troublesome, as often the intuition is that very small bandwidths are safe choices because they will quick fluctuations in firing variation as well as the macroscopic structure of the firing rate. In actuality, these small bandwidths can inject variability into the calculations and result in incorrect conclusions about the data. The structure of these results persisted across the multiple iterations of this simulation and is not specific to this spike train.

These results connect directly with the rates shown in Figures 1a to 1d, as the spike train used in those examples was the same one used in this iteration of the simulation (see Figure 5a). The smaller-bandwidth 250 ms time histogram (see Figure 1a and Figure 5b, white square) and the larger-bandwidth 1500 ms kernel smoother (see Figure 1d and Figure 5b, gray square) had R^2 values of 0.71 and 0.81, respectively, which indicate a strong correlation with the arm trajectory. The larger-bandwidth 800 ms time histogram (see Figure 1b and Figure 5b, white diamond) and the smaller-bandwidth 100 ms kernel smoother (see Figure 1c and Figure 5b, gray diamond) had R^2 values of 0.38 and 0.49, respectively, which indicate a weak correlation with the arm trajectory. Thus, while all four estimates may appear to be reasonable representations of the spiking data, the differences in bandwidth can result in varying and opposite conclusions about the data. If one were to choose the bandwidths and methods for the rates in Figures 1a or 1d, the regression analysis would conclude that the neural spiking activity is strongly related to the motion of the primate's arm. If, on the other hand, one were to choose the bandwidths and methods for the rates in Figures 1b or 1c, the regression analysis would conclude that the neural spiking activity is unrelated to the motion of the primate's arm. It should also be noted that while in this case, inaccurate bandwidth estimates led to erroneously decreased correlation between spiking and the covariate, it is also possible that inaccurate bandwidths biased toward the timescale of the presupposed correlate may find spurious increased correlation between spiking and the covariate. Therefore, a purely subjective determination

of bandwidth, even if many parameter values are explored, can result in widely misleading conclusions, even if the resultant rate estimates do not look unreasonable.

We then applied the cross-validated kernel smoother to the same simulated data. The kernel width chosen by the cross-validated kernel smoother was 940 ms (see Figure 5b, black dot) with confidence bounds on the kernel width of ±108 ms (see Figure 5b, light gray region). The chosen cross-validated kernel of 940 ms yielded an R^2 of 0.83. This value is virtually identical to the maximum possible correlation, with the maximum correlation falling well within the confidence interval. The cross-validated time histogram selected a bandwidth of 752 ms (see Figure 5b, gray dot) with a confidence of 6 ms (see Figure 5b, thin gray region), an R^2 of 0.59, which falls before the major drop in R^2 as bandwidth increases. These results suggest that the ability of a cross-validated rate estimation method to accurately determine the timescale for changes in firing rate can directly translate into high degrees of correlation for those signals with which the neuron is actually correlated. Assuming the neuron's rate of change in activity is directly correlated to the driving stimulus, an accurate estimate of bandwidth can be highly useful in determining to which of many possible signals the neuron may be related.

4 Discussion

The study of electrophysiological data often necessitates the smoothing of neural spiking data in order to produce an estimate of neural firing that is continuous in time. Herein, we develop an algorithm-independent, general framework for a likelihood-based approach to parameter selection in firing rate or conditional intensity models using leave-one-out cross-validation and maximum likelihood. The basis for this framework is that all spike-smoothing algorithms produce a rate or intensity function at each point at time, which provides an estimate of the instantaneous probability of spiking. The estimation of this probability allows us to calculate the likelihood of the spikes given the firing rate and intensity, which can act as a natural cost function for the smoother when used in conjunction with a leave-one-out cross-validation procedure. We show how this procedure can be used to select the model parameters that best predict missing data. This framework thus removes the need for ad hoc parameter selection in many common rate estimation procedures without having to develop specialized new techniques for each method.

As a simple, illustrative example, we applied the general framework to the time histogram and kernel smoother, the most common methods of rate estimation in practice within the electrophysiological literature. Both methods have a single bandwidth parameter, which implicitly imposes specific assumptions about the time course over which the neuron can change its activity onto the rate estimates. We demonstrated that the application

of the framework to the kernel smoother yields a vast reduction in estimation error to an ad hoc selection of bandwidth and that the benefits are the greatest when the firing rate is low. In our correlation analysis, we demonstrated that in practice, the use of such a procedure can provide a way of determining the correlation between a smoothed spike train and external correlates that is not reliant on subjective choices of bandwidth, which may result in incorrect conclusions about the relationship between the spiking and the correlates. Thus, the general framework can provide substantial improvements to even simple methods of rate estimation and can create an internally consistent approach for identifying the timescale on which neurons transmit information.

An accurate characterization of bandwidth can provide insight into neuronal functionality, as well as suggest plausible signals with which neural activity may be associated, which vary on the same timescale. This characterization is especially useful when the function of the neuron is not well known or when it is variable between experimental contexts. Firing rate variability itself may be thought of as an additional feature of neural activity to be collected, and it can be used in much the same way as any other neural measurement. A parameter-based metric of the time course of rate change can be used to classify and differentiate neurons, and analyses can be performed attempting to link changes in bandwidth to context, behavior, or other correlates. Furthermore, an analysis of bandwidth in populations can show the variety of timescales on which an ensemble may operate and provide information regarding network functionality and oscillatory behavior and further our understanding of the neural code.

There are many obvious limitations of the cross-validated kernel smoother, shown herein as an example of an application of the general framework. The cross-validated kernel smoother is based on the assumption that spiking is an inhomogeneous Poisson process with a constant bandwidth. It has been shown, however, that in actual neural systems, neural spiking is often non-Poisson (Pfeiffer & Kiang, 1965; Tuckwell, 1988; Koch & Segev, 1998; Shadlen & Newsome, 1998; Brown, Kass, & Mitra, 2004; Truccolo et al., 2005) and potentially nonuniform in its bandwidth. The cross-validated kernel smoother is clearly not optimal, yet there exist specific circumstances in which the method can be beneficial. Since the estimation of bandwidth in the cross-validated kernel smoother is completely data driven, this method requires only a single trial of spike train data to calculate an estimate of rate. Therefore, no parametric modeling assumptions, prior distributions of initial conditions, or the existence of any external stimuli or correlates are necessary to calculate firing rate. The estimated bandwidth from our method can suggest potential correlated stimuli and thus help motivate the creation of parametric models, which can be applied to the methodologies. An additional benefit to using the cross-validated kernel smoother is that it is extremely efficient and easy to compute. At its core, the cross-validated kernel smoother is simply a convolution of spiking

data with a normalized zero-centered kernel. Thus, the entire algorithm can be programmed as a single iterative loop in relatively few lines of code.

A useful feature of the likelihood framework is that it is orthogonal to the actual smoothing method on which it is applied, as it is primarily a means for model and parameter selection (Cunningham et al., 2009), and thus estimation algorithm independent. In the quest for an optimal rate estimation algorithm, further applications of the general likelihood framework to firing rate estimation can therefore take advantage of more sophisticated rate estimation methods that are able to incorporate aspects of spike timing such as refractory period and history dependence. One possible way to incorporate these timing features into the methodology is to adapt state-space methodologies (Brown et al., 1998; Brown, Nguyen, Frank, Wilson, & Solo, 2001; Eden et al., 2004; Czanner et al., 2008; Kulkarni & Paninski, 2008) to facilitate the cross-validation procedure, which we plan as a future investigation. We can also add the ability to capture a time-dependent bandwidth. A state-space modeling approach may also be helpful in adding a dynamic estimation of bandwidth. Using the current methods, it is possible to get a rudimentary estimate of systems with a changing bandwidth over time by using the cross-validated kernel smoother at several discrete temporal intervals, such as on a trial-by-trial basis. A temporal analysis of bandwidth, even at the trial-by-trial level, can provide a measure of neural plasticity, illustrating change in neural function over time.

In summary, the firing rate is calculated at the outset of the data analysis for the preponderance of electrophysiological research and is used as the basis for all subsequent computations and conclusions. It is therefore vital that these rates be computed in a principled manner to avoid spurious conclusions. By applying a general framework based on the spiking likelihood to any of the preexisting smoothing procedures, we can create data-oriented internally consistent methods and proceed with higher-level rate-based computations with an objective and justifiable foundation for the resultant conclusions.

Appendix: The Time Histogram and the Kernel Smoother

We present an overview of the two most prominently used methods for the temporal smoothing of spike trains: the time histogram and the kernel smoother. For each method, we discuss the relationship of bandwidth selection to the time course of change of the rate estimate produced by the smoothers.

We define a spike train in discrete time with N_T observations as the sequence of spike counts,

$$s_i = \Delta N_{(t_i, t_{i+1}]}, \tag{A.1}$$

where $\Delta N_{(a,b]}$ is the spike count between times a and b, and,

$$t_i = i\,\Delta t, \tag{A.2}$$

given a fixed time resolution, Δt.

A.1 The Time Histogram. To calculate a time histogram, time ranging from 0 to t_{N_T} is partitioned into K equal and nonoverlapping bins of size Δb, where

$$\Delta b = \frac{t_{N_t}}{K}, \tag{A.3}$$

and K is the parameter that governs the time histogram bandwidth.

We define the time histogram rate estimate, $r\,(t)$, as

$$r\,(t) = \frac{\Delta N_{(k\Delta t,(k+1)\Delta t]}}{\Delta b}, \tag{A.4}$$

where k is an integer ranging from 0 to $K-1$, which denotes the bin number into which t falls, such that $k\Delta b < t \leq (k+1)\,\Delta b$. $\Delta N_{(k\Delta b,(k+1)\Delta b]}$ consequently defines the spike count of the bin into which t falls. Hence, the value of time histogram rate in each bin is the mean firing rate for that given bin. When the bounds of the bins are based on spatial bounds as opposed to temporal bounds, as is often useful in hippocampal place cell electrophysiology studies (Wood, Dudchenko, Robitsek, & Eichenbaum, 2000; Eichenbaum, 2004; Lipton et al., 2007), the rate calculation is termed an occupancy normalized histogram.

Time histograms produce discontinuous step function estimates of rate, which jump at each bin boundary. The bandwidth and placement of these nonoverlapping bins greatly change the structure and smoothness of the rate estimate. A small bin size (see Figure 1a) produces a greater fluctuating, small-bandwidth estimate of rate, whereas a large bin size (see Figure 1b) produces a slowly changing, large-bandwidth estimate of rate. When there is a single bin containing all of T, the time histogram converges to the calculation of the mean firing rate of the entire spike train.

While the time histogram is simple to compute, improperly specified fixed bin boundaries or inappropriate bandwidth can lead to misleading conclusions about the underlying spiking data (see Figures 1e and 1f). Though it appears to be possible to construct an appropriate rate estimate for trivial firing examples by fitting parameters by eye, it can quite difficult to determine accurate bin size for complex experimental spike trains in which the trends in firing rate are not obvious.

A.2 The Kernel Smoother. The kernel smoother is essentially a weighted moving average. Kernel smoothers avoid the problems associated with the fixed, nonoverlapping bin boundaries of the time histogram by using a convolution of a function with the spikes to smooth the data. This, in effect, creates a sliding window across the data, as opposed to static bins.

The kernel smoother is defined as the discrete convolution of data, s, with a function $w(t)$,

$$r(t_i) = \sum_{j=1}^{N_T} w\left(t_i - t_j\right) s_j \Delta t. \tag{A.5}$$

In the kernel smoother, the function $w(t)$ integrates to 1 and is termed the kernel or, more colloquially, the window function. The kernel acts as a temporal weighting function for the data, and its structure determines how past and future data are weighed into the estimate of firing rate at a given point. Common choices for $w(t)$are gaussian or Hanning functions, which, due to their symmetry about 0, integrate equally both past and future local information. The choice of a gaussian kernel, for example, would create a smoother that would weigh the information closest to the current data point the highest, then decrease the importance of the spiking data as it became increasingly temporally distant. This is in contrast to the time histogram, which is based on spike count and thus disregards the location of the spikes within any given bin. In addition to symmetric functions, kernels can be causal functions, which are functions that are asymmetric about 0. Causal function kernels are used to integrate only past information into the convolution. Because kernels weigh the data based on the time relative to the current temporal bin, kernel smoothers can create rate estimates with a higher degree of temporal information than time histograms, which is important in neural systems in which information is contained in spike timing.

The choice of kernel bandwidth, as with the bin size of the time histogram, governs how quickly the rate estimation can change over time. The bandwidth of kernel smoothers is the temporal extent over which the kernel incorporates information for a specific temporal bin. For example, if the kernel is gaussian, the smoother bandwidth is proportional to its standard deviation. Large kernels (see Figure 1d) combine information over many contiguous local data points and allow for a slower change in rate over time than do kernels with small bandwidths (see Figure 1c), resulting in potentially large differences in the temporal variations of the rate estimate. As with the histogram, both oversmoothing and undersmoothing based on bandwidth selection can result in the loss or misinterpretation of data.

Acknowledgments

This work was supported by the National Science Foundation under grant number IIS-0643995. We thank Paul Lipton and Howard Eichenbaum for the use of the experimental data.

References

Adrian, E. D., & Zotterman, Y. (1926). The impulses produced by sensory nerve endings: Part 3. Impulses set up by touch and pressure. *Journal of Physiology*, *61*(4), 465–483.

Brown, E. N., Frank, L. M., Tang, D., Quirk, M. C., & Wilson, M. A. (1998). A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *Journal of Neuroscience*, *18*(22), 7411–7425.

Brown, E. N., Kass, R. E., Mitra, P. P. (2004). Multiple neural spike train data analysis: State-of-the-art and future challenges. *Nature Neuroscience*, *7*(5), 456–461.

Brown, E. N., Nguyen, D. P., Frank, L. M., Wilson, M. A., & Solo, V. (2001). An analysis of neural receptive field plasticity by point process adaptive filtering. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(21), 12261–12266.

Cunningham, J. P., Gilja, V., Ryu, S. I., & Shenoy, K. V. (2009). Methods for estimating neural firing rates, and their application to brain-machine interfaces. *Neural Networks, 9*, 1235–1246.

Cunningham, J. P., Yu, B. M., Shenoy, K. V., & Sahani, M. (2008). *Inferring neural firing rates from spike trains using Gaussian processes*. Cambridge, MA: MIT Press.

Czanner, G., Eden, U. T., Wirth, S., Yanike, M., Suzuki, W. A., & Brown, E. N. (2008). Analysis of between-trial and within-trial neural spiking dynamics. *Journal of Neurophsiology*, *99*(5), 2672–2693.

Daley, D. J., & Vere-Jones, D. (2003). *An introduction to the theory of point processes*. New York: Springer.

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge, MA: MIT Press.

DiMatteo, I., Genovese, C. R., & Kass, R. E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika*, *88*(4), 1055–1071.

Duong, T., & Hazelton, M. L. (2005). Cross-validation bandwidth matrices for multivariate kernel density estimation. *Scandinavian Journal of Statistics*, *32*(3), 485–506.

Eden, U. T., Frank, L. M., Barbieri, R., Solo, V., & Brown, E. N. (2004). Dynamic analysis of neural encoding by point process adaptive filtering. *Neural Computation*, *16*(5), 971–998.

Eden, U. T., Truccolo, W., Fellows, M. R., Donoghue, J. P., & Brown, E. N. (2004). Reconstruction of hand movement trajectories from a dynamic ensemble of spiking motor cortical neurons. In *Proceedings of the 26th International IEEE EMBS Conference*. Piscataway, NJ: IEEE.

Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *American Statistician*, *37*(1), 36–48.

Eichenbaum, H. (2004). Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron*, *44*(1), 109–120.

Endres, D., & Oram, M. (2009). Feature extraction from spike trains with Bayesian binning: "Latency is where the signal starts." *Journal of Computational Neuroscience, 29,* 149–169.

Gerstein, G. L., & Kiang, N. Y. (1960). An approach to the quantitative analysis of electrophysiological data from single neurons. *Biophysical Journal, 1*, 15–28.

Hall, P., Racine, J., & Li, Q. (2004). Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association, 99*(468), 1015–1026.

Harris, K. D., Csicsvari, J. , Hirase, H., Dragoi, G., & Buzsáki, G. (2003). Organization of cell assemblies in the hippocampus. *Nature, 424*(6948), 552–556.

Itskov, V., Curto, C., & Harris, K. D. (2008). Valuations for spike train prediction. *Neural Computation, 20*(3), 644–667.

Kass, R. E., Ventura, V., & Cai, C. (2003). Statistical smoothing of neuronal data. *Network—Computation in Neural Systems, 14*(1), 5–15.

Koch, C., & Segev, I. (1998). *Methods in neuronal modeling: From ions to networks*. Cambridge, MA: MIT Press.

Kulkarni, J. E., & Paninski, L. (2008). State-space decoding of goal-directed movements. *IEEE Signal Processing Magazine, 25*(1), 78–86.

Leung, D.H.Y. (2005). Cross-validation in nonparametric regression with outliers. *Annals of Statistics, 33*(5), 2291–2310.

Lipton, P. A., White, J. A., & Eichenbaum, H. (2007). Disambiguation of overlapping experiences by neurons in the medial entorhinal cortex. *Journal of Neuroscience, 27*(21), 5787–5795.

McCullagh, P. (1984). Generalized linear-models. *European Journal of Operational Research, 16*(3), 285–292.

Moran, D. W., & Schwartz, A. B. (1999). Motor cortical representation of speed and direction during reaching. *Journal of Neurophsiology, 82*(5), 2676–2692.

Nawrot, M., Aertsen, A. , & Rotter, S. (1999). Single-trial estimation of neuronal firing rates: From single-neuron spike trains to population activity. *Journal of Neuroscience Methods, 94*(1), 81–92.

Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics, 33*, 1065–1076.

Pfeiffer, R. R., & Kiang, N. Y. (1965). Spike discharge patterns of spontaneous and continuously stimulated activity in the cochlear nucleus of anesthetized cats. *Journal of Biophysics, 5*(3), 301–316.

Picard, R. R., & Cook, R. D. (1984). Cross-validation of regression-models. *Journal of the American Statistical Association, 79*(387), 575–583.

Racine, J. (1993). An efficient cross-validation algorithm for window width selection for nonparametric kernel regression. *Communications in Statistics-Simulation and Computation, 22*(4), 1107–1114.

Richmond, B. J., Optican, L. M., & Spitzer, H. (1990). Temporal encoding of 2-dimensional patterns by single units in primate primary visual-cortex. 1. stimulus-response relations. *Journal of Neurophsiology, 64*(2), 351–369.

Sanderson, A. C. (1980). Adaptive filtering of neuronal spike train data. *IEEE Transactions on Biomedical Engineering, 27*(5), 271–274.

Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *Journal of Neuroscience, 18*(10), 3870–3896.

Shimazaki, H., & Shinomoto, S. (2007a). Kernel width optimization in the spike-rate estimation. *Neural Coding. Abstract*, 120–123 (Montevideo, Uruguay).

Shimazaki, H., & Shinomoto, S. (2007b). A method for selecting the bin size of a time histogram. *Neural Computation, 19*(6), 1503–1527.

Silverman, B. W. (1984). A fast and efficient cross-validation method for smoothing parameter choice in spline regression. *Journal of the American Statistical Association, 79*(387), 584–589.

Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., & Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of Neurophysiology, 93*(2), 1074–1089.

Tuckwell, H. C. (1988). *Introduction to theoretical neurobiology*. Cambridge: Cambridge University Press.

Turlach, B. A. (1993). *Bandwidth selection in kernel density estimation: A review.* Louvain-la-Neuve, Belgium: Institut de Statistique.

Walk, H. (2002). On cross-validation in kernel and partitioning regression estimation. *Statistics and Probability Letters, 59*(2), 113–123.

Wong, W. H. (1983). On the consistency of cross-validation in kernel nonparametric regression. *Annals of Statistics, 11*(4), 1136–1141.

Wood, E. R., Dudchenko, P. A., Robitsek, R. J., & Eichenbaum, H. (2000). Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron, 27*(3), 623–633.

Received December 8, 2009; accepted March 3, 2011.