Likelihood Methods for Point Processes with Refractoriness

Citation: Citi, Luca, Demba Ba, Emery N. Brown, and Riccardo Barbieri. "Likelihood Methods for Point Processes with Refractoriness." Neural Computation 26, no. 2 (February 2014): 237-263. © 2014 Massachusetts Institute of Technology

As Published: http://dx.doi.org/10.1162/NECO_a_00548

Publisher: MIT Press

Persistent URL: http://hdl.handle.net/1721.1/85015

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

ARTICLE **Communicated by Jakob Macke**

Likelihood Methods for Point Processes with Refractoriness

Luca Citi
lciti@neurostat.mit.edu
Demba Ba
demba@neurostat.mit.edu
Emery N. Brown
enb@neurostat.mit.edu
Riccardo Barbieri
barbieri@neurostat.mit.edu
Department of Anesthesia, Massachusetts General Hospital–Harvard Medical School, Boston, MA 02129, and Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02142, U.S.A.

Likelihood-based encoding models founded on point processes have received significant attention in the literature because of their ability to reveal the information encoded by spiking neural populations. We propose an approximation to the likelihood of a point-process model of neurons that holds under assumptions about the continuous time process that are physiologically reasonable for neural spike trains: the presence of a refractory period, the predictability of the conditional intensity function, and its integrability. These are properties that apply to a large class of point processes arising in applications other than neuroscience. The proposed approach has several advantages over conventional ones. In particular, one can use standard fitting procedures for generalized linear models based on iteratively reweighted least squares while improving the accuracy of the approximation to the likelihood and reducing bias in the estimation of the parameters of the underlying continuous-time model. As a result, the proposed approach can use a larger bin size to achieve the same accuracy as conventional approaches would with a smaller bin size. This is particularly important when analyzing neural data with high mean and instantaneous firing rates. We demonstrate these claims on simulated and real neural spiking activity. By allowing a substantive increase in the required bin size, our algorithm has the potential to lower the barrier to the use of point-process methods in an increasing number of applications.

L.C. is currently with the School of Computer Science and Electronic Engineering, University of Essex, Colchester, U.K.

Neural Computation **26, 1–27** (2014)
doi:10.1162/NECO_a_00548

1 Introduction

Technological advances of the past decade have given us an unprecedented glimpse into the inner workings of the brain. A common paradigm in experimental neuroscience is to record the activity of groups of neurons in various behavioral settings or in response to sensory stimuli. In computational neuroscience, this paradigm is exploited by building models of the encoding process from behavior/stimulus to neural responses. In recent years, likelihood-based encoding models based on point processes have received attention in the literature (Paninski, Pillow, & Lewi, 2007; Barbieri, Quirk, Frank, Wilson, & Brown, 2001; Wang & Principe, 2010; Berger et al., 2011; Meng, Kramer, & Eden, 2011; So, Koralek, Ganguly, Gastpar, & Carmena, 2012) because of their ability to significantly improve neural decoding tasks (Brown, Nguyen, Frank, Wilson, & Solo, 2001; Paninski et al., 2010; Srinivasan, Eden, Willsky, & Brown, 2006; Barbieri et al., 2004; Barbieri, Wilson, Frank, & Brown, 2005). The goal of encoding models is to improve our understanding of brain function and, most important, help to design algorithms for inferring behavior/stimulus from previously unseen neural data, a process referred to as decoding and commonly used in the design of brain-machine interfaces (BMIs) (Shoham et al., 2005; Srinivasan, Eden, Mitter, & Brown, 2007; Sanchez et al., 2008).

Likelihood-based models of neural spike trains are based on a continuous-time point-process model of a neuron. A point process is fully characterized by its conditional intensity function (CIF), which generalizes the notion of rate of a Poisson process to include time and history dependence (Andersen, Borgan, Gill, & Keiding, 1995; Daley & Vere-Jones, 2007). Generalized linear models (GLMs) are a class of discrete-time models based on log-linear expansions of a discrete-time version of the CIF of a point process (Berman & Turner, 1992; Paninski, 2004; Truccolo, Eden, Fellows, Donoghue, & Brown, 2005). Berman and Turner (1992) were among the first to suggest the use of discretization for approximate maximum likelihood estimation of point-process models using the GLM framework. In neuroscience, this has resulted in the development of successful frameworks to characterize the dynamics of various neural systems (Paninski, 2004; Truccolo et al., 2005) and to develop algorithms for decoding based on neural data, with applications to BMIs (Chiel & Thomas, 2011).

In this article, we demonstrate that this approximation can be significantly improved by imposing a biophysically motivated structure on the continuous-time CIF of a pointprocess model of a neuron. Coincidentally, this translates into a specific, data-dependent form of the weighting function in Berman's approximation. The proposed approximation holds under assumptions about the continuous time process that are reasonable for neural spike trains: the presence of a refractory period, the predictability and integrability of the CIF. These properties are not exclusive to neural point processes but also apply to point-process models of geysers,

heartbeats, and repeated failures of components in engineering systems (see section 2.4).

The resulting point-process likelihood has several advantages over state-of-the-art approximations (Paninski, 2004; Truccolo et al., 2005). First, one can use standard fitting procedures for GLMs based on iteratively reweighted least-squares (IRWLS) while improving the accuracy of the approximation to the likelihood and reducing bias in the estimation of the parameters of the underlying continuous-time model. Stated otherwise, the proposed approach can use a larger bin size to achieve the same accuracy as the other approaches noted with a smaller bin size. This aspect is particularly important when analyzing neurons with high spiking rates, which often require a sampling frequency of at least 10 kHz. Tested on one example of this type of recordings, our method achieved at 1 kHz results comparable to those of a Poisson GLM (Paninski, 2004; Truccolo et al., 2005) at 10 kHz. Second, we find that our formulation partly obviates the use of the correction techniques introduced in Haslinger, Pipa, and Brown (2010) for goodness-of-fit assessment using the time-rescaling theorem and discrete-time approximations to the CIF. We demonstrate our claims on simulated data, as well as real data, from rat thalamic neurons recorded in response to periodic whisker deflections varying in velocity (Temereanca, Brown, & Simons, 2008). These data are characterized by high mean and instantaneous firing rates—on the order of 20 and 200 Hz, respectively.

The remainder of our treatment proceeds as follows. First, we introduce our model of continuous-time point processes with refractoriness in section 2, along with the resulting discrete-time likelihoods. In this section, we also compare the latter to conventional discrete-time approximations of continuous-time point-process likelihoods. We use simulations to demonstrate the power of our new model, which accounts for refractoriness. Then, in section 3, we propose a class of parametric log-linear models and discuss the connections with conventional GLMs and iteratively reweighted least squares. We apply the log-linear models to analysis of real neural data. Finally, we provide concluding remarks in section 4.

2 A Model of Continuous-Time Point Process with Refractoriness

We consider a continuous-time point process defined in $(0, T]$, with conditional intensity function $\lambda(t \mid \mathcal{H}_t)$, where $\mathcal{H}_t$ represents all the information accumulated up to time t. This information can include the previous firing pattern of the same neuron and of its afferents (e.g., via postspike and coupling filters), the values of other covariates and hidden states, and the observations of deterministic inputs (e.g., an external stimulus). Let $N(t)$ be the counting process associated with the point process (Andersen et al., 1995); for each realization, $N(t)$ is a function counting the number of events observed up to, and including, time t (see Figure 1).

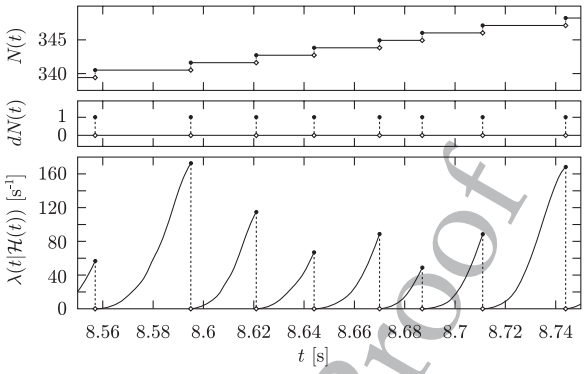

Figure 1: Representative example of a realization of a neural point process. The top plot shows the counting process, $N(t)$, which is a right-continuous function counting the number of events observed up to and including time t. The middle plot shows its differential, $dN(t)$, which is what is conventionally known as a spike train—an indicator function that assumes a value of one if there is an event at time t and zero otherwise. The bottom plot shows the conditional intensity function, $\lambda(t \mid \mathcal{H}_s)$, which jumps to zero after every event because, as a result of the refractory period, an additional event cannot take place arbitrarily close to the previous one.

The likelihood function for the continuous-time point process can be formulated by deriving the joint probability density of a realization of the counting process in the observation interval, $N_{0:T}$ (Daley & Vere-Jones, 2007):

$$\ln f(N_{0:T}) = \int_0^T \ln \lambda(t \mid \mathcal{H}_t)\, \mathrm{d}N(t) - \int_0^T \lambda(t \mid \mathcal{H}_t)\, \mathrm{d}t \ . \tag{2.1}$$

To facilitate the construction of encoding and decoding algorithms, typically a discrete-time representation of the point process is derived and used as an approximation to the underlying continuous-time point process. The observation interval is finely partitioned in I bins of width δ to obtain the discretized spike train $\Delta N_{1:I} = \{\Delta N_i\}_{i=1}^I$ where ΔN_i is a function that takes value zero if no events were observed in the ith bin and one otherwise. The joint probability mass function (PMF) of the discretized spike train is used as a likelihood function (see Truccolo et al., 2005, equation 3):

$$\ln P_{\mathrm{d}}(\Delta N_{1:I}) = \sum_{i=1}^{I} \Delta N_i \ln(\lambda_i \delta) - \lambda_i \delta \ , \tag{2.2}$$

where λ_i is a sampled version of $\lambda(t \mid \mathcal{H}_t)$.

In the rest of this section, we will show that under the assumption that the continuous-time point process is characterized by the presence of a refractory period, a better approximation to the joint probability mass function of the discretized spike train is given by

$$\ln P_{\mathrm{n}'}(\Delta N_{1:I}) = \sum_{i=1}^{I} \Delta N_i \, \ln(\lambda_i \delta) - \left(1 - \frac{\Delta N_i}{2}\right) \lambda_i \delta \, . \tag{2.3}$$

2.1 Key Assumptions. Before delving into the derivation of equation 2.3, we need to make three assumptions about the continuous-time process. The first two require that the CIF be predictable and integrable within a small time bin and free of possible discontinuities due to the incoming flow of information through $\mathcal{H}_t$ (property P1) or due to the intrinsic form of the CIF (property P2). These two properties are a prerequisite for the discretization of the point process and are needed to derive equations 2.2 and 2.3 alike. In particular, they are required for the existence of a well-defined value λ_i in each bin that can serve as an accurate representative value of the behavior of the function $\lambda(t \mid \mathcal{H}_t)$ inside the bin.

The third property, P3, is refractoriness—the presence of a period of time following each event during which another event cannot occur. Excitable cells, like neurons and cardiac cells, have an absolute refractory period, during which a second action potential cannot be initiated and a relative refractory period during which the cell is hyperpolarized and requires a greater stimulus to reach the threshold and initiate another depolarization.

All three properties are needed to derive equation 2.3. Formally, they can be formulated as follows:

P1: *Predictability.* The CIF changes slowly in response to new information coming from covariates and deterministic inputs. For all $\epsilon > 0$, there exist $\delta > 0$ such that given $s \in (0, T]$, $\tau \in (0, \delta]$, and $N(s) = N(s + \tau)$, we have:

$$\sup_{t \in (s, s+\tau)} \left| \lambda(t \mid \mathcal{H}_t) - \lambda(t \mid \mathcal{H}_s) \right| < \epsilon \, \lambda(t \mid \mathcal{H}_s) \, . \tag{2.4}$$

P2: *Integrability.* Given the history at a time s, the function $\lambda(t \mid \mathcal{H}_s)$ is Riemann integrable for $t \in (s, T]$. This implies that for all $\epsilon > 0$, there exists $\delta > 0$ such that for $t_1 \in (s, s + \delta]$,

$$\left| \lambda(t_1 \mid \mathcal{H}_s) \, \delta - \int_s^{s+\delta} \lambda(t \mid \mathcal{H}_s) \, \mathrm{d}t \right| < \epsilon \, . \tag{2.5}$$

Therefore, we can define a representative value $\tilde{\lambda}$ of $\lambda(t \mid \mathcal{H}_s)$ in $(s, s + \delta]$ such that the former approximates the integral mean value of the latter inside the interval:

$$\tilde{\lambda} \simeq \frac{1}{\delta} \int_s^{s+\delta} \lambda(t \mid \mathcal{H}_s) \, \mathrm{d}t \, . \tag{2.6}$$

P3: *Refractoriness.* The CIF, $\lambda(t \mid \mathcal{H}_t)$, is a piecewise Lipschitz continuous function in t that jumps to zero immediately after the occurrence of every event:

$$\lim_{\epsilon \to 0^+} \lambda(t + \epsilon \mid \mathcal{H}_{t+\epsilon}) = 0 \quad \forall t : \mathrm{d}N(t) = 1 \,. \tag{2.7}$$

A graphical representation of this property is given in Figure 1. This assumption is plausible for point processes with a refractory period such as neural spike trains, motor-unit firing trains (Hu, Tong, & Hung, 2004), and others (see section 2.4).

2.2 Implications of Key Assumptions on Discrete-Time Approximation. Given δ small enough so that the key assumptions hold and that the probability of two events in the same bin is negligible, we partition the observation interval in I bins of width δ and attempt to relate the original continuous-time point process with the resulting discrete-time point process.

2.2.1 Probabilities of Key Events. When P1 and P2 apply, the probability of observing no events in $(\bar{\imath}\delta, i\delta]$ ($\bar{\imath}$ is short for $i-1$) can be approximated as

$$\Pr\left(N(i\delta) - N(\bar{\imath}\delta) = 0\right) = e^{-\int_{\bar{\imath}\delta}^{i\delta} \lambda(t \mid \mathcal{H}_t)\,\mathrm{d}t} \simeq e^{-\int_{\bar{\imath}\delta}^{i\delta} \lambda(t \mid \mathcal{H}_{\bar{\imath}\delta})\,\mathrm{d}t} \simeq e^{-\lambda_i \delta}, \tag{2.8}$$

where λ_i is the representative value of the CIF inside the ith bin according to equation 2.6. If P3 also applies and δ is sufficiently small, the probability of more than one event in a given bin is negligible because the first event causes the CIF to be approximately zero for the rest of the bin. In fact, while for a generic point process the probability of two events occurring in the same bin of infinitesimal size δ is $O(\delta^2)$, when P3 holds, one can prove that this probability is $O(\delta^3)$ (see appendix A). Therefore, the probability of observing exactly one event is simply $1 - e^{-\lambda_i \delta}$. As expected, for δ infinitesimal, the first-order Taylor expansions of the probability of no events and of that of one event coincide with $1 - \lambda_i \delta$ and $\lambda_i \delta$, respectively that is, with the results of a Bernoulli approximation of the Poisson process.

2.2.2 Discrete-Time Point-Process PMF. We formally define the discretized spike train as $\Delta N_i = \min\{1, N(i\delta) - N(\bar{\imath}\delta)\}$. Therefore, a given realization of the point process can be represented as a binary sequence $\Delta N_{1:I} = \{\Delta N_i\}_{i=1}^{I}$ with probability mass function (PMF)

$$P_{\mathrm{n}}(\Delta N_{1:I}) = \prod_{i=1}^{I} \left(1 - e^{-\lambda_i \delta}\right)^{\Delta N_i} \left(e^{-\lambda_i \delta}\right)^{1 - \Delta N_i}, \tag{2.9}$$

where we use the subscript n to remind us that this is the PMF of a discrete-time version of the continuous-time point process obtained using the properties of neural spike trains, in particular, the presence of a refractory period, as per P3. The log probability is simply

$$\ln P_{\mathrm{n}}(\Delta N_{1:I}) = \sum_{i=1}^{I} \Delta N_i \ln\left(1 - e^{-\lambda_i \delta}\right) - (1 - \Delta N_i)\lambda_i \delta \ . \tag{2.10}$$

In appendix B we prove that as $\delta \to 0$, equation 2.10 converges to the continuous-time log likelihood, equation 2.1.

2.2.3 Useful Approximation. We show here that equation 2.3 can be obtained as an approximation of equation 2.10. In fact, noting that

$$\ln\left(\frac{1 - e^{-\xi}}{\xi}\right) = -\frac{\xi}{2} + o(\xi) \quad \text{as} \quad \xi \to 0 \ , \tag{2.11}$$

equation 2.10 can be approximated as

$$\begin{aligned} \ln P_{\mathrm{n}}(\Delta N_{1:I}) &= \sum_{i=1}^{I} \Delta N_i \left[\ln(\lambda_i \delta) + \ln\left(\frac{1 - e^{-\lambda_i \delta}}{\lambda_i \delta}\right)\right] - (1 - \Delta N_i)\lambda_i \delta \\ &= \sum_{i=1}^{I} \Delta N_i \ln(\lambda_i \delta) - \left(1 - \frac{\Delta N_i}{2}\right)\lambda_i \delta + o(\delta). \end{aligned} \tag{2.12}$$

We will see in section 2.5 that in most cases, this approximation introduces an acceptable error. In section 3, we will show how equation 2.3 allows an efficient procedure of fitting a log-linear model by means of an IRWLS algorithm. One can interpret equation 2.3 as a data-dependent choice of the weighting proposed by Berman and Turner (1992). We stress that Berman's motivation behind the use of weighting functions is mostly heuristic. However, in our setting, the weighting function is dictated by the conditions imposed on the continuous-time point process—in particular, refractoriness.

2.3 Comparison with Conventional Discrete-Time Approximation. We now compare equation 2.3 with equation 2.2. While both of these approximations converge in the limit to the continuous PDF, equation 2.1, they do so with a different linear coefficient for δ. Specifically, while their first terms are the same, their second terms differ. From an empirical point of view, we can think of this second term of equation 2.3 and 2.2, as the numerical computation of the integral in the second term of equation 2.1 and the two implementations, equations 2.2 and 2.3, differ only in the way

the CIF is integrated inside those bins containing an event. The contribution of one such bin to the second term of equation 2.1 is

$$\int_{\bar{\imath}\delta}^{i\delta} \lambda(t \mid \mathcal{H}_t)\, \mathrm{d}t \Big|_{\Delta N_i = 1} \simeq \begin{cases} \lambda_i \delta/2 & \text{in equation 2.3,} \\ \lambda_i \delta & \text{in equation 2.2.} \end{cases} \tag{2.13}$$

Defining τ as the time when the event occurs relative to the start of the bin and using the properties P3, P1, and P2 in this order, we obtain

$$\int_{\bar{\imath}\delta}^{i\delta} \lambda(t \mid \mathcal{H}_t)\, \mathrm{d}t \Big|_{\Delta N_i = 1} \simeq \int_{\bar{\imath}\delta}^{\bar{\imath}\delta+\tau} \lambda(t \mid \mathcal{H}_t)\, \mathrm{d}t \simeq \int_{\bar{\imath}\delta}^{\bar{\imath}\delta+\tau} \lambda(t \mid \mathcal{H}_{\bar{\imath}\delta})\, \mathrm{d}t \simeq \lambda_i \tau \ . \tag{2.14}$$

We use a truncated exponential distribution to find the expected value of τ conditioned on the fact that an event occurred inside the bin:

$$E[\tau \mid \Delta N_i = 1] = \frac{\int_0^\delta \tau \lambda_i e^{-\lambda_i \tau}\, \mathrm{d}\tau}{\int_0^\delta \lambda_i e^{-\lambda_i \tau}\, \mathrm{d}\tau} = \frac{e^{-\lambda_i \delta}}{\lambda_i} \frac{e^{\lambda_i \delta} - \lambda_i \delta - 1}{1 - e^{-\lambda_i \delta}} = \frac{1}{2}\delta + o(\delta). \tag{2.15}$$

Replacing equation 2.15 in 2.14, we see that for those bins where an event occurs, approximately half of $\lambda_i \delta$ should be considered. Looking at equation 2.13, we see that this corresponds to what our solution does, while equation 2.3 tends to overestimate the left-hand side of equation 2.13.

2.4 Point Processes with Refractoriness Are Pervasive. The assumptions about the continuous time process under which the approximation holds (the presence of a refractory period, the predictability of the CIF, and its integrability) are not exclusive to neural point processes, but also apply to a wide spectrum of point processes, such as models of geysers, heartbeats, and repeated failures of components in engineering systems (e.g., light bulbs).

There are two main approaches to modeling point-process data: model either the CIF or the probability density (assuming it exists) of the interevent intervals (IEIs). In some sense, these two approaches are equivalent because the following one-to-one transformation between CIFs and IEI PDFs holds (Daley & Vere-Jones, 2007):

$$\lambda(t \mid \mathcal{H}_t) = \frac{f(t \mid \mathcal{H}_t)}{\int_t^\infty f(\tau \mid \mathcal{H}_t)\, d\tau} \ . \tag{2.16}$$

In this article, we have adopted the former approach. In light of the above transformation, it is not hard to show that the refractory property of the

CIF in equation 2.7 holds if and only if the IEI distribution has a PDF that vanishes as its argument goes to zero. Therefore, the proposed framework for modeling point processes with refractoriness is also applicable to a much broader range of problems. For example, the log-normal, inverse-gaussian, Weibull, and gamma (with shape parameter $k > 1$) distributions, which are commonly used in survival analysis (Kalbfleisch & Prentice, 2002), models of heartbeats (Barbieri, Matten, Alabi, & Brown, 2005), and geyser and seismic data (Nishenko & Buland, 1987), satisfy the property that their respective PDFs vanish as their argument goes to zero. The corresponding CIFs must therefore satisfy the refractory property of equation 2.7, which is the backbone of the framework developed in this article.

2.5 Simulations. In this section, we simulate three examples of renewal continuous-time point processes to compare, as a function of the bin size δ, the accuracy of estimates of the log-PDF (and consequently the log likelihood) obtained using equations 2.10, 2.3, and 2.2:

Example 1: A homogeneous Poisson (P) process that, by definition, has constant CIF λ_P and exponentially distributed interspike intervals (ISIs), as shown in Figure 2. For such a process, P1 and P2 hold, while P3 does not.

Example 2: A Rayleigh (R) distribution for the ISIs. The corresponding CIF is $\lambda_R(z) = z/\sigma_R^2$, which increases linearly with the time since the last event, z. We chose this example because it has the simplest form of CIF that also complies with P3. The Rayleigh distribution is a special case of the Weibull distribution; it is skewed and bell shaped (see Figure 2) and can be used to approximate the ISIs of neural spike trains (Lansky & Greenwood, 2005; Tyrcha, 2008).

Example 3: A renewal process for which the ISIs follow an inverse gaussian (IG) distribution. The IG distribution is particularly suited to model spike trains because it represents the distribution of the first threshold passage time of an integrate-and-fire neuron driven by a Wiener process with drift (Brown, 2005). Given the time since the last event, the CIF corresponding to IG-distributed ISIs with mean μ and shape parameter k is

$$\lambda_{IG}(z) = \frac{\alpha_3(z)\,\varphi(\alpha_1(z)-\alpha_2(z))}{\phi(\alpha_1(z)-\alpha_2(z)) - e^{2\alpha_1(z)\alpha_2(z)}\phi(\alpha_1(z)+\alpha_2(z))}, \tag{2.17}$$

where the vector $\alpha(z) = [-\sqrt{kz}/\mu, -\sqrt{k/z}, \sqrt{k/z^3}]$ while $\varphi(x)$ and $\phi(x)$ are the probability density function and the cumulative distribution function of a standard gaussian random variable.

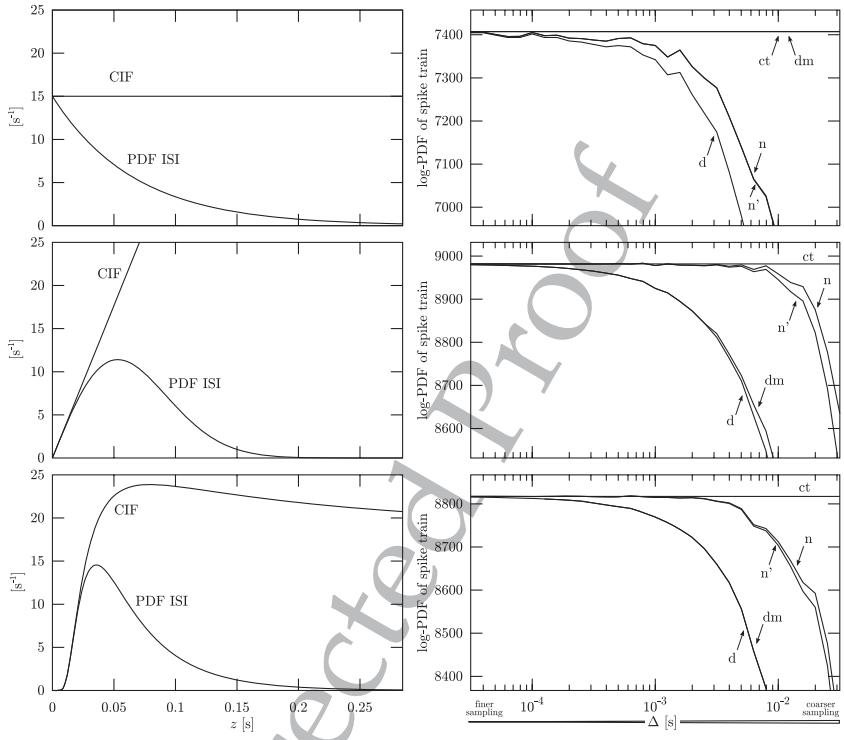

Figure 2: The plots on the left show the CIF of the point process and the PDF of the associated ISI distribution for the three examples: homogeneous Poisson process with exponentially distributed ISIs (top), Rayleigh-distributed ISIs (center), and inverse-gaussian-distributed ISIs (bottom). On the right, for each case, the plots show the corresponding log-PDF of the whole spike train evaluated from the simulated data (see section 2.5) using ct, the continuous-time version, equation 2.18; n, our discrete-time definition, equation 2.10; n, its approximation, equation 2.3; d, the conventional discrete-time formulation, equation 2.2; and dm, a variant of equation 2.2, allowing for more events per bin.

2.5.1 Simulation Parameters. For each of the three distributions, we simulated one realization in $(0, T]$ with $T = 300$ s, assuming that the last event before the observation interval happened at $u_0 = 0$. Given the set of event times, $\{u_k\}_{k=1}^{N(T)}$, the continuous-time log-PDF is simply

$$\ln f_D(N_{0:T}) = \sum_{k=1}^{N(T)} \ln \mathfrak{f}_D(u_k - u_{k-1}) + \ln\left(1 - \mathfrak{F}_D(T - u_{N(T)})\right), \quad (2.18)$$

where $D \in \{\mathrm{P}, \mathrm{R}, \mathrm{IG}\}$ and $\mathfrak{f}_D(z)$ and $\mathfrak{F}_D(z)$ are the PDF and CDF associated with the ISIs.

We binned each process for 31 values of the bin size, δ, logarithmically spaced between $10^{-4.5}$ and $10^{-1.5}$, that is, roughly 0.03 ms and 30 ms. Then we evaluated the log-PMF using equations 2.10, 2.3, and 2.2, in turn, followed by subtraction of the correcting factor $(\ln(\delta) \sum_{i=1}^{I} \Delta N_i)$ to obtain an estimate of the log-PDF from the log-PMF. For large bin sizes, realizations of all three processes (and P in particular) may have more than one event in some bins. To account for this, we also tested a variant of equation 2.2, where ΔN_i is replaced with a noncapped version $\Delta N_i^* = N(\bar{\imath}\delta) - N(i\delta)$.

2.5.2 Results. Results of these simulations are presented in the right panels of Figure 2. For the homogeneous Poisson process, all methods of estimating the log-PDF starting from a discrete-time process fail miserably unless a very small bin size is used. The only exception is the one that allows for more than one event per bin. Because this type of point process does not comply with P3, this result was expected.

For both the Rayleigh and the inverse gaussian distributions, the two formulations presented in this article, equations 2.10 and 2.3, show a remarkable improvement compared with the legacy approach of equation 2.2. For example, for the Rayleigh distribution, the conventional approach requires a bin size of 0.1 ms to achieve a result comparable to the one that our method shows at 5 ms, while for the IG distribution, it requires 0.3 ms to achieve the same result that our method shows at around 4 ms.

3 Log-Linear Models of Discrete-Time Point-Process Data

In this section, we compare our formulation of the point-process PMF with the legacy approach in parametric, likelihood-based modeling of neural data. Parametric, likelihood-based models of neural spike trains use the log-PMF of the data as the likelihood function. It is natural to expect that inference algorithms for these models would benefit from our novel formulation of the discretized point-process likelihood. In the next section, we show that this is indeed the case.

Parametric models of the log likelihood in equation 2.2 have been successfully applied to the identification of neural systems (Paninski, 2004; Truccolo et al., 2005). Their main advantage is the ability to relate neuronal firing to extrinsic covariates such as stimuli or intrinsic ones, which capture the internal dynamics of neurons. In a linear-nonlinear-Poisson (LNP) model, one expresses the intensity function of an inhomogeneous Poisson process as a nonlinear function of the output of a linear filter applied to the set of covariates: $\lambda_i = h_\theta(\beta^{\mathrm{T}} x_i)$, where x_i is the vector of covariates, θ is a vector parameterizing the nonlinear function $h(\cdot)$, and β represents the coefficients of the linear filter. The unknown parameters (θ, β) can be estimated by maximum likelihood estimation by plugging this model in the log-likelihood function. Sufficient conditions guaranteeing that our

likelihood function, equation 2.3, has no nonglobal local maxima in the parameters (θ, β), regardless of the observations, are the same as those for the conventional likelihood—that is, that $h_\theta(w)$ is jointly convex and log concave in θ and w (Paninski, 2004).

A simple example of convex and log-concave function is the exponential function that gives rise to log-linear models (Truccolo et al., 2005). This turns equation 2.2 into a generalized linear model (GLM) with Poisson observations and log link (Fahrmeir & Tutz, 2001). Below, we show that a log-linear model of λ_i in equation 2.3 leads to a parametric model that can be fit as efficiently as conventional GLMs of discrete-time point-process data (Paninski, 2004; Truccolo et al., 2005).

3.1 IRWLS Algorithm for Log-Linear Model Fitting. GLMs are a generalization of linear least squares to observations from the exponential family. They are computationally very attractive because they can be fit by iteratively reweighted least squares (IRWLS), that is, by solving a sequence of weighted least-squares problems (Fahrmeir & Tutz, 2001).

If we assume $\ln(\lambda_i) = \beta^{\mathrm{T}} x_i$, for a d-dimensional vector of parameters β and a vector x_i of covariates of the same dimension, then equation 2.3 becomes

$$\ln P_{\mathrm{n}'}(\Delta N_{1:I}, \beta; X) \propto \sum_{i=1}^{I} \Delta N_i \beta^{\mathrm{T}} x_i - \left(1 - \frac{\Delta N_i}{2}\right) \exp(\beta^{\mathrm{T}} x_i), \tag{3.1}$$

where X is the I-by-d matrix with x_i^{T} in the rows. One can maximize equation 3.1 using Newton's method. In order to draw a parallel between equation 3.1 and conventional GLMs, we consider the following generalization of equation 3.1:

$$\ln P_m(\Delta N_{1:I}, \beta; X, \rho) \propto \sum_{i=1}^{I} \Delta N_i \beta^{\mathrm{T}} x_i - \rho_i \exp(\beta^{\mathrm{T}} x_i), \tag{3.2}$$

where $\rho_i = 1 - \frac{\Delta N_i}{2}$ and $m = \mathrm{n}'$ for equation 3.1, while $\rho_i = 1$ and $m = \mathrm{d}$ for conventional GLMs arising from maximizing P_{d} in equation 2.2 using a log-linear parameterization of λ_i. In appendix C, we derive the IRWLS algorithm for maximizing equation 3.2 showing that up to a choice of a weighting function (ρ), maximizing equation 3.1 and fitting conventional GLMs are equivalent. This implies that fast implementations (Komarek & Moore, 2005) of IRWLS can be modified in a simple way to maximize the log likelihood of equation 3.1.

3.2 Goodness-of-Fit Assessment. Quantitatively measuring the agreement between a proposed model and a spike train is crucial for establishing

the model's validity prior to using it to make inferences about the neural system being studied. The time rescaling theorem can be used to check the goodness of fit of statistical models of neural spike trains (Brown, Barbieri, Ventura, Kass, & Frank, 2002).

Haslinger and colleagues (2010) have drawn attention to the fact that the finite temporal resolution of discrete-time models introduces some bias in the application of the time rescaling theorem, which can lead to misleading results, that is, indicating poor goodness of fit even for accurate models. They found two root causes and proposed analytical corrections for these effects. The first cause is that as an unavoidable consequence of binning, there is a lower bound on the smallest possible ISI (one bin). The second type of bias is introduced when one uses a Bernoulli distribution to fit the model and obtain p_i, the probability of one event in a discrete bin. Then if one naıvely applies the time rescaling theorem (which implicitly assumes a Poisson distribution) with $\lambda_i = p_i/\delta$, a bias is introduced because this relationship holds only for λ_i, $p_i \to 0$.

Our method does not suffer from the second issue because it directly estimates λ_i, a discretized version of the continuous-time function $\lambda(t)$, rather than the probability of one event in a discrete bin (as in the Bernoulli case). As we have presented in section 2.2, when properties P3 and P2 hold, these two variables are related by $p_i = \Pr(\Delta N_i = 1) = 1 - e^{-\lambda_i \delta}$, which is exactly the inverse of the correction that the authors propose in their equation 2.35.

In our proposed approach, the goodness-of-fit assessment can be simply performed using the following procedure, adapted from Haslinger et al. (2010):

1. Use the log-linear model, equation 3.2, based on equations 2.3 or 2.2, to fit a model to the observed spike train and obtain an estimate of the sampled CIF, $\hat{\lambda}_i$.
2. Given the set of bins containing a spike $\{i_k \mid \Delta N_{i_k} = 1\}$, for each ISI, find the integrated CIF as

$$\xi_k = \sum_{i=i_k+1}^{i_{k+1}-1} \hat{\lambda}_i \delta + \hat{\lambda}_{i_{k+1}} \tau_k, \tag{3.3}$$

where τ_k represents the random time of the event relative to the start of the bin and can be obtained by first drawing a random variable r_k uniformly distributed in $[0, 1]$ and then calculating

$$\tau_k = -\frac{\ln[1 - r_k\,(1 - \exp(-\hat{\lambda}_{i_{k+1}} \delta))]}{\hat{\lambda}_{i_{k+1}}}\,. \tag{3.4}$$

In most cases this expression can be approximated as $\tau_k = r_k \delta$.

3. After a further variable transform, $z_k = 1 - e^{-\xi_k}$, the goodness of fit can be assessed by plotting the empirical CDF of the transformed ISIs, z_k, against the CDF of the uniform distribution.

3.3 Simulations. In this section, we present the results of the comparison between the conventional GLM approach and our novel log-linear model, equation 3.1, in the estimation of the regression terms of $\ln \lambda(t)$. We assumed a simple yet plausible form for $\ln \lambda(t)$, allowing for simulation in continuous time, and then assessed the estimation accuracy as a function of bin size using either method.

3.3.1 Generation of Simulated Spike Trains. We modeled $\ln \lambda(t)$ as the convolution of a postspike filter (Pillow et al., 2008) with the comb of previously observed events,

$$\ln \lambda(t) = \ln(\ell_0) + \int_0^t \psi(t-\tau)\, dN(\tau), \tag{3.5}$$

where ℓ_0 is a constant term. We chose a function $\psi(z)$ that, by modulating the conditional intensity function, was able to capture dependencies on spike train history:

$$\psi(z) = \begin{cases} \ln\left(c_1\left(\dfrac{z}{c_3}\right)^3 + c_2\left(\dfrac{z}{c_3}\right)^2\right) & \text{if } 0 < z < c_3 \\ 0 & \text{otherwise} \end{cases}. \tag{3.6}$$

We chose this specific function (with parameters $c = [-8, 9, 0.1]$ and $\ell_0 = 100$) because it allows the simulation of spike trains in continuous time without resorting to discretization and because it resembles a prototypical postspike filter (Pillow et al., 2008) with three distinct regions able to explain refractoriness, burstiness, and adaptation (see Figure 3). Thanks to this choice of $\psi(z)$, the CIF and its integral, $\Lambda(t)$, are piecewise polynomials in t, and from equations 3.5 and 3.6, we obtain

$$\lambda(t) = \ell_0 \prod_{k=N(t-c_3)+1}^{N(t)} \left(c_1\left(\frac{t-u_k}{c_3}\right)^3 + c_2\left(\frac{t-u_k}{c_3}\right)^2\right), \tag{3.7}$$

where $\{u_k\}_{k=1}^{N(T)}$ is the ordered set of event times. To simulate a spike train, the time rescaling theorem (Brown et al., 2002) was used to calculate the next event u_{k+1} from the set of previous events as the solution of $\Lambda(u_{k+1}) - \Lambda(u_k) = -\ln(q_{k+1})$ where q_{k+1} is a random variable uniformly sampled in $(0, 1)$. The solutions of this polynomial equation can be found as the eigenvalues of its Frobenius companion matrix (Edelman & Murakami,

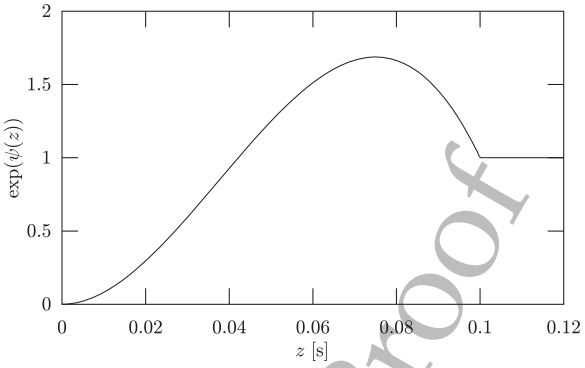

Figure 3: The plot represents the exponential of the postspike filter in equation 3.6, that is, a continuous function in $(0, +\infty)$ that assumes value $c_1(z/c_3)^3 + c_2(z/c_3)^2$ if $0 < z < c_3$ and one otherwise. It can be interpreted as a spike-induced gain adjustment that accounts for the effects of the neuron's spike history. The first region ($z < 50$ ms, $\exp(\psi(z)) < 1$) accounts for the presence of refractoriness, the second region (50 ms $< z <$ 100 ms, $\exp(\psi(z)) > 1$) for bursting activity, and in the third region ($z > 100$ ms, $\exp(\psi(z)) = 1$), the filter is inactive.

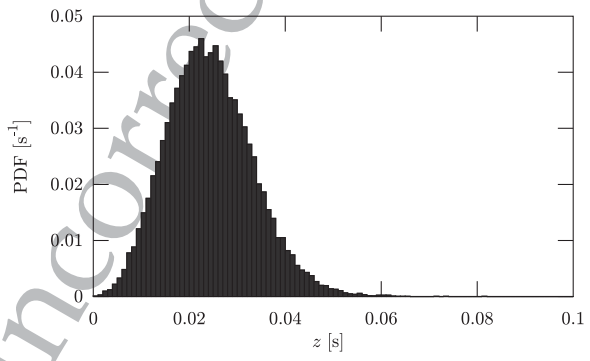

Figure 4: Empirical PDF (histogram) of the ISIs of the simulated neural spike train generated using the procedure described in section 3.3.

1995). We take the smallest real positive eigenvalue as time of the next event, u_{k+1}. Note that the CIF of Figure 1 was generated in this fashion.

Using this procedure, we generated 250 realizations of the continuous-time point process, each one with an observation interval of 600 s and approximately 40 events per second. The resulting distribution of ISIs is reported in Figure 4.

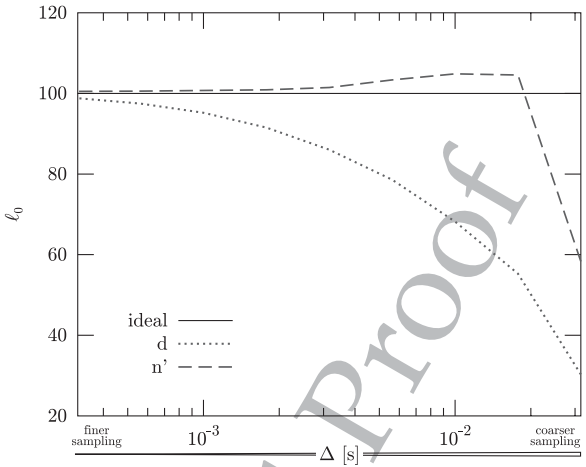

Figure 5: Value of ℓ_0 estimated using the conventional GLM (dotted line) and using our log-linear model (dashed line) as a function of the bin size, δ.

3.3.2 Model Fitting. We considered the following model for the discrete-time CIF,

$$\ln \lambda_i = \beta_0 + \sum_{j=1}^{\lceil c_3/\delta \rceil} \beta_j \, \Delta N_{i-j}, \tag{3.8}$$

which is the discrete-time equivalent of equation 3.5. Then we fitted the conventional GLM and our log-linear model on each realization for nine values of δ, logarithmically spaced between $10^{-3.5}$ and $10^{-1.5}$, that is, between approximately 0.3 ms and 30 ms.

3.3.3 Results. We assessed the difference between the original baseline firing rate $\ell_0 = 100$ and the value $\ln \beta_0$ estimated by the two methods as a function of δ. The results are presented in Figure 5. Compared to the conventional GLM, our method approaches the ideal value of ℓ_0 with higher accuracy and less sensitivity to the bin size. In fact, the bin size can be increased without much detriment up to the value (approximately 0.1 s) where the probability of more spikes in a bin becomes significant, while the accuracy of the traditional GLM progressively decreases for all bin sizes. We also compared the accuracy of the two methods at the task of recovering the postspike filter. Figures 6a to 6c show the true postspike filter $\psi(z)$ and the estimated one β_j for three values of δ, while Figure 6d shows the error $\beta_j - \psi(z)$, with $j = \lfloor z/\delta + 1/2 \rfloor$, for $z = 0.02\,\text{s}$ and $z = 0.07\,\text{s}$ as a function of δ. From these results, we notice that the traditional GLM, which

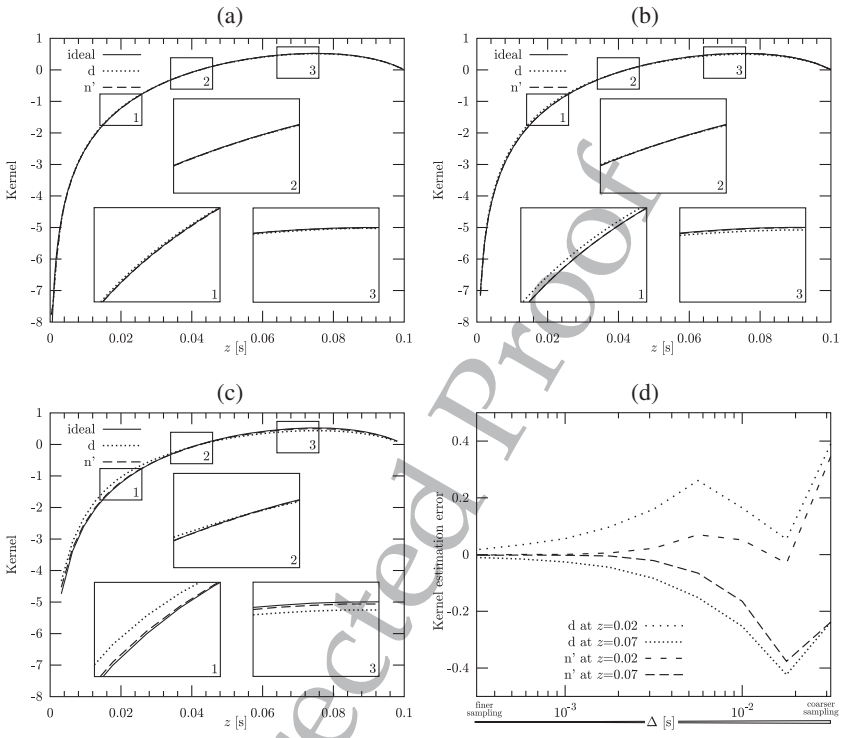

Figure 6: The figure compares the accuracies of the conventional GLM and of our log-linear model in the estimation of the postspike filter $\psi(z)$ (see equation 3.6). Panels a to c show the reconstructed filter (represented by dotted lines for the traditional GLM and dashed lines for our log-linear model) and the target function $\psi(z)$ (solid lines) for three values of bin sizes: $\delta \simeq 0.3$ ms, $\delta \simeq 1$ ms, and $\delta \simeq 3$ ms, respectively. Panel d shows the estimation error (difference between the estimated value and the ideal one) for $z = 0.02$ s and $z = 0.07$ s (i.e., the centers of the zoom insets 1 and 3 in panels a–c) as a function of the bin size, δ.

maximizes the likelihood given by equation 2.2, tends to underestimate the base firing rate ℓ_0, overestimate the filter for small values of z (i.e., briefly after each event), and underestimate it for bigger z. Our log-linear model, which maximizes equation 2.10, instead is able to estimate the postspike filter with much higher accuracy.

3.4 Real Data. In this section, we demonstrate that the improvement that our method shows on simulated data also holds for real neural spike trains. In particular, we tested our approach on single unit recordings from

individual thalamic barreloid neurons in the somatosensory whisker/barrel system of an anesthetized rat.

3.4.1 Experiment. Here, we provide a brief description of the experimental protocol that was reported in detail by Temereanca et al. (2008). The principal whisker was deflected in caudal direction by means of a piezoelectric stimulator using periodic waveforms of different velocity delivered at a repetition rate of eight per second. Each deflection was 1 mm in amplitude and began and ended from the neutral position of the whisker as the trough (phase of $3\pi/2$) of a single sine wave. The fast stimulus was generated by taking one period of a 40 Hz sine wave and delivering it every 125 ms for 2 s. The other two stimuli were generated similarly by taking one cycle of a 25 Hz sine wave and one cycle of an 8 Hz sine wave. The whisker was deflected pseudorandomly using the three stimuli, and pseudorandomized blocks were delivered 50 times, with interstimulus intervals of 1.5 s. Spike waveforms were analyzed off-line, clustered based on waveform shape and interspike interval, and the spike stamps saved at 100 μs resolution.

3.4.2 Log-Linear Model of the CIF. In our analysis, we assumed a log-linear CIF given by

$$\ln \lambda_i = \beta_0 + \gamma_1 s(i\delta) + \gamma_2 s(i\delta - q_1) + \sum_{j=1}^{q_2/\delta} \beta_j \Delta N_{i-j}, \tag{3.9}$$

where $s(t)$ is the stimulus: the whisker deflection at time t. The parameter q_1 represents the delay between the two observations of $s(t)$ ($s(i\delta)$ and $s(i\delta - q_1)$) that allow the CIF to depend on the derivative of the stimulus input in addition to its instantaneous value. The parameter q_2 represents the duration of the postspike filter. Their values ($q_1 = 1$ ms and $q_2 = 6$ ms) were chosen based on prior extensive analyses of the same neurons considered here. We used the conventional GLM and our log-linear model to estimate the vector $\beta = [\beta_0, \ldots, \beta_{q_2/\delta}, \gamma_1, \gamma_2]$ that best explains the observed spike train given the stimulus and the process history. We repeated the analysis for bin sizes $\delta \in \{0.1, 0.2, 0.5, 1\}$ milliseconds.

3.4.3 Results. The top three panels of Figure 7 show the results of the estimation of β_0, γ_1, and γ_2 as a function of the bin size when using conventional GLM and our log-linear model. Although in the case of real data the ground truth is unknown, it is fair to assume that the true values are more likely to be somewhere around those obtained with a finer sampling. It is clear from the figure that when using the traditional GLM based on equation 2.2, there is a strong effect of the bin size on the estimated values, leading to a significant bias when a coarser sampling is used. This effect is much less pronounced when using the log-linear model coming from our

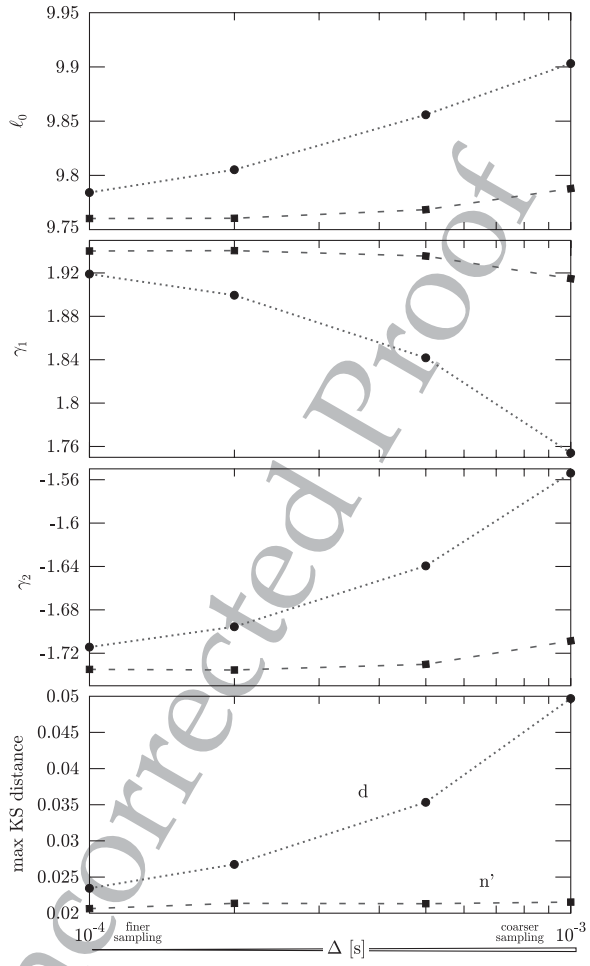

Figure 7: The top three panels show the results of the estimation of the parameters β_0, γ_1, and γ_2 of equation 3.9 as a function of the bin size when using conventional GLM (circles connected by dotted segments) and our log-linear model (squares connected by dashed segments). The bottom panel shows the KS distance achieved by the fitted models for different bin sizes.

new likelihood formulation, equation 2.3. In fact, the values obtained with our model at 1 ms (1 kHz sampling rate) are similar to those obtained by the GLM for 0.1 ms (10 kHz).

Using the time-rescaling procedure described in section 3.2, we performed a goodness-of-fit analysis and assessed the Kolmogorov-Smirnov (KS) distance achieved by the fitted models for different bin sizes. The

results are reported in the bottom panel of Figure 7, which confirms the increasing inaccuracy of the GLM model for increasing bin sizes. Our log-linear model instead achieves approximately the same fit for all the sampling rates tested. For even larger bin sizes (not shown in the figure), the fit degrades rapidly for both methods as soon as the assumption of one spike per bin is violated, similar to what is shown in Figure 2 for the likelihood.

4 Discussion

The traditional discretized version of the Poisson process PDF, equation 2.2, as reported by Truccolo et al. (2005), serves as a building block for many advanced statistical analysis and signal processing techniques (Paninski et al., 2007; Friedman, Hastie, & Tibshirani, 2010; Zhao et al., 2012; Zhao & Iyengar, 2010; Pillow, Ahmadian, & Paninski, 2011; Paninski et al., 2010; Czanner et al., 2008; Lawhern, Wu, Hatsopoulos, & Paninski, 2010) applied to neural signals. These techniques have been used for basic neuroscience research (Kass, Kelly, & Loh, 2011; Okatan, Wilson, & Brown, 2005; Eldawlatly, Jin, & Oweiss, 2009; Berger et al., 2011; Jenison, Rangel, Oya, Kawasaki, & Howard, 2011; So et al., 2012) to improve biophysical neural models (Ahrens, Paninski, & Sahani, 2008; Meng et al., 2011; Mensi et al., 2012) or to design better BMIs (Shoham et al., 2005; Srinivasan et al., 2006, 2007; Truccolo, Friehs, Donoghue, & Hochberg, 2008; Wang & Principe, 2010; Saleh, Takahashi, & Hatsopoulos, 2012).

In this article, we presented a new formulation for the probability mass function of observing discrete-time realizations of continuous-time point processes arising from neural spike trains. This new theory holds under assumptions about the continuous-time point process that are reasonable for neural spike trains: the presence of a refractory period, the predictability of the conditional intensity function, and its integrability within a time bin. These properties are not exclusive to neural point processes, but also apply to a much wider spectrum of point processes, including models of geysers, heartbeats, and repeated failures of components in engineering systems. Our new definition represents a remarkable theoretical improvement over the traditional discretized version of the continuous-time point-process PDF in the presence of refractoriness. As a result, the estimated value approaches the solution of the continuous problem, as the bin size goes to zero, at significantly higher speed of convergence.

Based on this new definition, we also introduced a log-linear model that shows an improvement over a traditional GLM fit. In both simulations and with real data, our novel algorithm converges to the asymptotic value for bin sizes one order of magnitude larger than the traditional GLM. This can be advantageous when analyzing neurons with a high firing rate. For example, we showed that our method achieves, on neural data resampled at 1 KHz, more refined outcomes (e.g., in terms of goodness of fit) than those of a Poisson GLM on the original recordings sampled at 10 kHz.

The improvement shown by our log-linear model over the traditional GLM is likely to extend to most models and algorithms based on the point-process GLM framework and comes at virtually no cost because, as we have shown with our log-linear model, the number of floating-point operations is practically the same. In most cases, instead, the computational cost can decrease drastically because our new definition allows the use of a coarser sampling rate while still providing comparable accuracy.

As stated in section 1, an active field of research is the application of point-process models to the decoding of neural signals, for example, for BMI or sensory processing. Compared to spike-rate-based decoding algorithms, point-process models have a potential advantage in terms of accuracy and temporal resolution. Unfortunately, this comes at the price of a higher computational cost (Sanchez et al., 2008) because they need to process the neural signals at the timescale of the spike times ($\leq$1 ms) instead of that of the modulating signal. We are currently investigating whether the use of our likelihood model, equation 2.3, within a decoding framework results in the possibility of using larger time steps, which would be extremely useful for BMIs and neural prostheses requiring real-time decoding. Preliminary results with simulated data suggest that this is indeed the case and that a recursive decoding filter algorithm (Barbieri et al., 2004) using our novel likelihood allows for a twofold reduction in the sampling rate. An extensive treatment of the extension of our method to the decoding problem and a thorough analysis with simulated and real data will constitute the subject of future work.

In conclusion, our method improves over conventional approaches by taking advantage of an essential feature of spike trains, refractoriness, to reformulate the discrete likelihood in a principled way that can improve the performance of this powerful statistical model. Because it requires only a minor modification to the likelihood term, it can replace the legacy point-process likelihood (Truccolo et al., 2005) in virtually all instances where such a probabilistic framework is used. By allowing a substantive increase in the required bin size, our algorithm has the potential of lowering the barrier to the use of point-process methods in an increasing number of applications.

Appendix A: Asymptotic Probability of More Than One Event in a Bin

A function $f(t)$ is piecewise Lipschitz continuous in an interval $(a, b]$ if there is a positive constant C and a partition $\bigcup_{j=1}^{J} I_j = (a, b]$ such that

$$\left|f(t_1) - f(t_2)\right| \leq C\,|t_1 - t_2| \quad \forall t_1, t_2 \in I_j,\ 1 \leq j \leq J\,. \tag{A.1}$$

In our specific case, using the Lipschitz condition and the fact that the only jumps allowed are decreasing (the jumps to zero after each event), we obtain

that the following condition holds for the CIF:

$$0 \leq \lambda(t_2 | \mathcal{H}_{t_2}) \leq \lambda(t_1 | \mathcal{H}_{t_1}) + C\,(t_2 - t_1) \quad \forall\, t_1, t_2 \in (0, T] : t_1 \leq t_2\ . \tag{A.2}$$

Let us consider the ith bin and attempt to find the asymptotic behavior of the probability that the number of events contained, $n = N(i\delta) - N(\bar{\imath}\delta)$, is at least one. Using condition A.2, we have that for $t \in (\bar{\imath}\delta, i\delta)$,

$$\lambda(t \mid \mathcal{H}_t) \leq \lambda(\bar{\imath}\delta \mid \mathcal{H}_{\bar{\imath}\delta}) + C\,(t - \bar{\imath}\delta) \leq \lambda(\bar{\imath}\delta \mid \mathcal{H}_{\bar{\imath}\delta}) + C\,\delta\ , \tag{A.3}$$

leading to

$$\begin{aligned} \Pr(n \geq 1) &= 1 - \exp\left\{ - \int_{\bar{\imath}\delta}^{i\delta} \lambda(t \mid \mathcal{H}_t)\, \mathrm{d}t \right\} \\ &\leq 1 - \exp\left\{ - \left[\lambda(\bar{\imath}\delta \mid \mathcal{H}_{\bar{\imath}\delta}) + C\,\delta \right] \delta \right\} = O(\delta)\ . \end{aligned} \tag{A.4}$$

Let us now find the probability of more than one event conditioned on the presence of at least one event in the bin: $\Pr(n > 1 \mid n \geq 1)$. Calling $t_1 \in (\bar{\imath}\delta, i\delta]$ the time when the first event occurs and using property P3, we have that $\lim_{\epsilon \to 0^+} \lambda(t_1 + \epsilon \mid \mathcal{H}_{t_1 + \epsilon}) = 0$ and therefore $\lambda(t \mid \mathcal{H}_t) \leq C\,(i\delta - t_1) \leq C\delta$ for $t \in (t_1, i\delta]$. Observing that $n > 1$ implies at least one event in $(t_1, i\delta]$, we can follow a reasoning similar to equation A.4 and obtain

$$\Pr(n > 1 \mid n \geq 1) \leq \left[1 - \exp\left(-C\,\delta^2 \right) \right] = O(\delta^2)\ . \tag{A.5}$$

Finally,

$$\Pr(n > 1) = \Pr(n > 1 \mid n \geq 1)\ \Pr(n \geq 1) = O(\delta^3), \tag{A.6}$$

which is what we wanted to prove.

Appendix B: Convergence for Infinitesimal Bin Size

We can define the logarithm of the normalized PMF (log-nPMF) as

$$\begin{aligned} \ln\left(P_{\mathrm{n}}(\Delta N_{1:I}) \,/\, \delta^{N(T)} \right) &= -N(T) \ln \delta \\ &\quad + \sum_{i=1}^{I} \Delta N_i \ln\left(1 - e^{-\lambda_i \delta} \right) - (1 - \Delta N_i) \lambda_i \delta \end{aligned} \tag{B.1}$$

and show that in the limit $\delta \to 0$, it converges to the logarithm of the continuous-time probability density function (log-PDF) of observing

exactly those $N(T)$ events in $(0, T]$, given in equation 2.1. We start by noting that as $\delta \to 0$,

$$\ln\left(1 - e^{-\lambda_i \delta}\right) = \ln\left(\lambda_i\, \delta\, \frac{1 - e^{-\lambda_i \delta}}{\lambda_i \delta}\right) = \ln \lambda_i + \ln \delta + o(1)\ . \tag{B.2}$$

Then we take the limit of equation B.1, replace equation B.2, expand the terms, and finally obtain

$$\begin{aligned}
\lim_{\delta \to 0} \ln\left(P_{\mathrm{n}}(\Delta N_{1:I}) \,/\, \delta^{N(T)}\right) &= -N(T) \ln \delta \\
&\quad + \lim_{\delta \to 0} \left(\sum_{i=1}^{\lceil T/\delta \rceil} \Delta N_i\, \ln \lambda_i + \sum_{i=1}^{\lceil T/\delta \rceil} \Delta N_i \ln \delta - \sum_{i=1}^{\lceil T/\delta \rceil} \lambda_i \delta + \sum_{i=1}^{\lceil T/\delta \rceil} \Delta N_i\, \lambda_i \delta \right) \\
&= -N(T) \ln \delta + \int_0^T \ln \lambda(t \mid \mathcal{H}_t)\, \mathrm{d}N(t) + N(T) \ln \delta \\
&\quad - \int_0^T \lambda(t \mid \mathcal{H}_t)\, \mathrm{d}t + 0,
\end{aligned} \tag{B.3}$$

which simplifies to the continuous-time log-PDF in equation 2.1.

Appendix C: Log-Linear Model

Differentiating equation 3.2 with respect to β, we obtain its gradient and its Hessian,

$$\begin{aligned}
g(\beta) = \sum_{i=1}^{I} \left(\Delta N_i - \rho_i \lambda_i(\beta)\right) x_i &= \sum_{i=1}^{I} \left(\Delta N_i - \rho_i \exp\{\beta^{\mathrm{T}} x_i\}\right) x_i \\
&= X^{\mathrm{T}}(\Delta N - \lambda_\rho(\beta)),
\end{aligned} \tag{C.1}$$

$$\begin{aligned}
H(\beta) = -\sum_{i=1}^{I} \rho_i \lambda_i(\beta) \cdot x_i x_i^{\mathrm{T}} &= -\sum_{i=1}^{I} \rho_i \exp\{\beta^{\mathrm{T}} x_i\} \cdot x_i x_i^{\mathrm{T}} \\
&= X^{\mathrm{T}} W_\rho(\beta) X,
\end{aligned} \tag{C.2}$$

where X is the I-by-d matrix with x_i^{T} in the rows, ΔN and $\lambda_\rho(\beta)$ are I-length vectors satisfying with entries ΔN_i and $\rho_i \lambda_i(\beta)$, respectively, and $W_\rho(\beta)$ is the I-by-I diagonal matrix with diagonal elements $\rho_i \lambda_i(\beta)$.

One can maximize equation 3.2 by taking Newton steps as follows:

$$\beta_{(k+1)} = \beta_{(k)} - H^{-1}(\beta_{(k)}) g(\beta_{(k)}). \tag{C.3}$$

Substituting equations C.1 and C.2 in C.3 and rearranging, we find that $\beta_{(k+1)}$ is the solution of a quadratic approximation to the objective function, which we refer to as weighted least squares (WLS):

$$\beta_{(k+1)} = \underset{\beta}{\operatorname{argmax}} -\frac{1}{2}(b - A\beta)^{\mathrm{T}} C (b - A\beta), \tag{C.4}$$

where $b = X\beta_{(k)} + W_{\rho}^{-1}(\beta_{(k)})(\Delta N - \lambda_{\rho}(\beta_{(k)}))$, $C = W_{\rho}(\beta_{(k)})$ and $A = X$. One maximizes equation 3.2 by iteratively solving equation C.4, hence the name *iteratively reweighted least squares (IRWLS)*. Assuming X is full rank, equation 3.2 is a concave function of β. Therefore, there exists a unique solution to which the Newton algorithm, implemented by IRWLS, converges. The general formulation of equation 3.2 shows that maximizing equation 3.1 and fitting conventional GLMs of neural data (using equation 2.2) are equivalent up to the choice of ρ_i.

The computational cost of each iteration is dominated by the matrix product in equation C.2, which is $O(2Id^2)$, and by the Newton step, whose cost is $O\left(\frac{1}{3}d^3\right)$ using a Cholesky factorization to solve the linear system. If ρ_i is precomputed, the additional cost per iteration due to our likelihood function is only I multiplications.

Acknowledgments

We thank Simona Temereanca for kindly providing the experimental data used in section 3.4. This work was supported by the National Institutes of Health through grants R01-HL084502 (R.B.) and DP1-OD003646 (E.N.B.).

References

Ahrens, M. B., Paninski, L., & Sahani, M. (2008). Inferring input nonlinearities in neural encoding models. *Network*, *19*(1), 35–67.

Andersen, P., Borgan, O., Gill, R., & Keiding, N. (1995). *Statistical models based on counting processes*. New York: Springer.

Barbieri, R., Frank, L. M., Nguyen, D. P., Quirk, M. C., Solo, V., Wilson, M. A., & Brown, E. N. (2004). Dynamic analyses of information encoding in neural ensembles. *Neural Computation*, *16*, 277–307.

Barbieri, R., Matten, E., Alabi, A., & Brown, E. (2005). A point-process model of human heartbeat intervals: New definitions of heart rate and heart rate variability. *American Journal of Physiology–Heart C*, *288*(1), H424–H435.

Barbieri, R., Quirk, M. C., Frank, L. M., Wilson, M. A., & Brown, E. N. (2001). Construction and analysis of non-Poisson stimulus-response models of neural spiking activity. *Journal of Neuroscience Methods*, *105*(1), 25–37.

Barbieri, R., Wilson, M. A., Frank, L. M., & Brown, E. N. (2005). An analysis of hippocampal spatio-temporal representations using a Bayesian algorithm for

neural spike train decoding. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *13*(2), 131–136.

Berger, T. W., Hampson, R. E., Song, D., Goonawardena, A., Marmarelis, V. Z., & Deadwyler, S. A. (2011). A cortical neural prosthesis for restoring and enhancing memory. *Journal of Neural Engineering*, *8*(4), 046017.

Berman, M., & Turner, T. (1992). Approximating point process likelihoods with GLIM. *Applied Statistics*, *41*, 31–38.

Brown, E. (2005). Theory of point processes for neural systems. In C. C. Chow, B. Gutkin, D. Hansel, C. Meunier, & J. Dalibard (Eds.), *Methods and models in neurophysics* (pp. 691–726). Amsterdam: Elsevier.

Brown, E., Barbieri, R., Ventura, V., Kass, R., & Frank, L. (2002). The time-rescaling theorem and its application to neural spike train data analysis. *Neural Computation*, *14*(2), 325–346.

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.

Chiel, H. J., & Thomas, P. J. (2011). Applied neurodynamics: From neural dynamics to neural engineering. *Journal of Neural Engineering*, *8*(6), 060201.

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 Neurophysiology*, *99*(5), 2672–2693.

Daley, D., & Vere-Jones, D. (2007). *An introduction to the theory of point processes: General theory and structure*, vol. 2. New York: Springer-Verlag.

Edelman, A., & Murakami, H. (1995). Polynomial roots from companion matrix eigenvalues. *Mathematics of Computation*, *64*(210), 763–776.

Eldawlatly, S., Jin, R., & Oweiss, K. G. (2009). Identifying functional connectivity in large-scale neural ensemble recordings: A multiscale data mining approach. *Neural Computation*, *21*(2), 450–477.

Fahrmeir, L., & Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models*. New York: Springer-Verlag.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22.

Haslinger, R., Pipa, G., & Brown, E. (2010). Discrete time rescaling theorem: Determining goodness of fit for discrete time statistical models of neural spiking. *Neural Computation*, *22*(10), 2477–2506.

Hu, X. L., Tong, K. Y., & Hung, L. K. (2004). Oscillations in the power spectra of motor unit signals caused by refractoriness variations. *Journal of Neural Engineering*, *1*(3), 174–185.

Jenison, R. L., Rangel, A., Oya, H., Kawasaki, H., & Howard, M. A. (2011). Value encoding in single neurons in the human amygdala during decision making. *Journal of Neuroscience*, *31*(1), 331–338.

Kalbfleisch, J., & Prentice, R. (2002). *The statistical analysis of failure time data* (2nd ed.). New York: Willey.

Kass, R. E., Kelly, R. C., & Loh, W.-L. (2011). Assessment of synchrony in multiple neural spike trains using loglinear point process models. *Annals of Applied Statistics*, *5*(2B), 1262–1292.

Komarek, P., & Moore, A. (2005). Making logistic regression a core data mining tool with TR-IRLS. In *Proccedings of the Fifth IEEE International Conference on Data Mining* (pp. 685–688). Piscataway, NJ: IEEE.

Lansky, P., & Greenwood, P. E. (2005). Optimal signal estimation in neuronal models. *Neural Computation*, *17*(10), 2240–2257.

Lawhern, V., Wu, W., Hatsopoulos, N., & Paninski, L. (2010). Population decoding of motor cortical activity using a generalized linear model with hidden states. *Journal of Neuroscience Methods*, *189*(2), 267–280.

Meng, L., Kramer, M. A., & Eden, U. T. (2011). A sequential Monte Carlo approach to estimate biophysical neural models from spikes. *Journal of Neural Engineering*, *8*(6), 065006.

Mensi, S., Naud, R., Pozzorini, C., Avermann, M., Petersen, C.C.H., & Gerstner, W. (2012). Parameter extraction and classification of three cortical neuron types reveals two distinct adaptation mechanisms. *Journal of Neurophysiology*, *107*(6), 1756–1775.

Nishenko, S., & Buland, R. (1987). A generic recurrence interval distribution for earthquake forecasting. *Bulletin of the Seismological Society of America*, *77*(4), 1382–1399.

Okatan, M., Wilson, M. A., & Brown, E. N. (2005). Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity. *Neural Computation*, *17*(9), 1927–1961.

Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems*, *15*(4), 243–262.

Paninski, L., Ahmadian, Y., Ferreira, D. G., Koyama, S., Rad, K. R., Vidne, M., et al. (2010). A new look at state-space models for neural data. *Journal of Computational Neuroscience*, *29*(1–2), 107–126.

Paninski, L., Pillow, J., & Lewi, J. (2007). Statistical models for neural encoding, decoding, and optimal stimulus design. *Progress in Brain Research*, *165*, 493–507.

Pillow, J. W., Ahmadian, Y., & Paninski, L. (2011). Model-based decoding, information estimation, and change-point detection techniques for multineuron spike trains. *Neural Computation*, *23*(1), 1–45.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E., et al. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, *454*(7207), 995–999.

Saleh, M., Takahashi, K., & Hatsopoulos, N. G. (2012). Encoding of coordinated reach and grasp trajectories in primary motor cortex. *Journal of Neuroscience*, *32*(4), 1220–1232.

Sanchez, J., Principe, J., Nishida, T., Bashirullah, R., Harris, J., & Fortes, J. (2008). Technology & signal processing for brain-machine interfaces. *IEEE Signal Processing Magazine*, *25*(1), 29–40.

Shoham, S., Paninski, L. M., Fellows, M. R., Hatsopoulos, N. G., Donoghue, J. P., & Normann, R. A. (2005). Statistical encoding model for a primary motor cortical brain-machine interface. *IEEE Transactions on Biomedical Engineering*, *52*(7), 1312–1322.

So, K., Koralek, A. C., Ganguly, K., Gastpar, M. C., & Carmena, J. M. (2012). Assessing functional connectivity of neural ensembles using directed information. *Journal of Neural Engineering*, *9*(2), 026004.

Srinivasan, L., Eden, U. T., Mitter, S. K., & Brown, E. N. (2007). General-purpose filter design for neural prosthetic devices. *Journal of Neurophysiology*, *98*(4), 2456–2475.

Srinivasan, L., Eden, U. T., Willsky, A. S., & Brown, E. N. (2006). A state-space analysis for reconstruction of goal-directed movements using neural signals. *Neural Computation*, *18*(10), 2465–2494.

Temereanca, S., Brown, E. N., & Simons, D. J. (2008). Rapid changes in thalamic firing synchrony during repetitive whisker stimulation. *Journal of Neuroscience*, *28*(44), 11153–11164.

Truccolo, W., Eden, U., Fellows, M., Donoghue, J., & Brown, E. (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.

Truccolo, W., Friehs, G. M., Donoghue, J. P., & Hochberg, L. R. (2008). Primary motor cortex tuning to intended movement kinematics in humans with tetraplegia. *Journal of Neuroscience*, *28*(5), 1163–1178.

Tyrcha, J. (2008). Dynamics of integrate-and-fire models. In A. Deutch, R. Bravo de la Parra, R. de Boer, O. Diekmann, P. Jagers, E. Kisdi et al. (Eds.), *Mathematical modeling of biological systems* (Vol. 2, pp. 225–236). Boston: Birkäuser.

Wang, Y., & Principe, J. C. (2010). Instantaneous estimation of motor cortical neural encoding for online brain-machine interfaces. *Journal of Neural Engineering*, *7*(5), 056010.

Zhao, M., Batista, A., Cunningham, J. P., Chestek, C., Rivera-Alvidrez, Z., Kalmar, R., et al. (2012). An L1-regularized logistic model for detecting short-term neuronal interactions. *Journal of Computational Neuroscience*, *32*, 479–497.

Zhao, M., & Iyengar, S. (2010). Nonconvergence in logistic and poisson models for neural spiking. *Neural Computation*, *22*(5), 1231–1244.

Received March 19, 2013; accepted September 1, 2013.