

Estimating a State-Space Model from Point Process Observations

Anne C. Smith

asmith@neurostat.mgh.harvard.edu

Neuroscience Statistics Research Laboratory, Department of Anesthesia and Critical Care, Massachusetts General Hospital, Boston, MA 02114, U.S.A.

Emery N. Brown

brown@neurostat.mgh.harvard.edu

Neuroscience Statistics Research Laboratory, Department of Anesthesia and Critical Care, Massachusetts General Hospital, Boston, MA 02114, U.S.A., and Division of Health Sciences and Technology, Harvard Medical School/Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

A widely used signal processing paradigm is the state-space model. The state-space model is defined by two equations: an observation equation that describes how the hidden state or latent process is observed and a state equation that defines the evolution of the process through time. Inspired by neurophysiology experiments in which neural spiking activity is induced by an implicit (latent) stimulus, we develop an algorithm to estimate a state-space model observed through point process measurements. We represent the latent process modulating the neural spiking activity as a gaussian autoregressive model driven by an external stimulus. Given the latent process, neural spiking activity is characterized as a general point process defined by its conditional intensity function. We develop an approximate expectation-maximization (EM) algorithm to estimate the unobservable state-space process, its parameters, and the parameters of the point process. The EM algorithm combines a point process recursive nonlinear filter algorithm, the fixed interval smoothing algorithm, and the state-space covariance algorithm to compute the complete data log likelihood efficiently. We use a Kolmogorov-Smirnov test based on the time-rescaling theorem to evaluate agreement between the model and point process data. We illustrate the model with two simulated data examples: an ensemble of Poisson neurons driven by a common stimulus and a single neuron whose conditional intensity function is approximated as a local Bernoulli process.

1 Introduction

A widely used signal processing paradigm in many fields of science and engineering is the state-space model. The state-space model is defined by two equations: an observation equation that defines what is being measured or observed and a state equation that defines the evolution of the process through time. State-space models, also termed *latent process models* or *hidden Markov models*, have been used extensively in the analysis of continuous-valued data. For a linear gaussian observation process and a linear gaussian state equation with known parameters, the state-space estimation problem is solved using the well-known Kalman filter. Many extensions of this algorithm to both nongaussian, nonlinear state equations and nongaussian, nonlinear observation processes have been studied (Ljung & Söderström, 1987; Kay, 1988; Kitagawa & Gersh, 1996; Roweis & Ghahramani, 1999; Ghahramani, 2001). An extension that has received less attention, and the one we study here, is the case in which the observation model is a point process.

This work is motivated by a data analysis problem that arises from a form of the stimulus-response experiments used in neurophysiology. In the stimulus-response experiment, a stimulus under the control of the experimenter is applied, and the response of the neural system, typically the neural spiking activity, is recorded. In many experiments, the stimulus is explicit, such as the position of a rat in its environment for hippocampal place cells (O'Keefe & Dostrovsky, 1971; Wilson & McNaughton, 1993), velocity of a moving object in the visual field of a fly H1 neuron (Bialek, Rieke, de Ruyter van Steveninck, & Warland, 1991), or light stimulation for retinal ganglion cells (Berry, Warland, & Meister, 1997). In other experiments, the stimulus is implicit, such as for a monkey executing a behavioral task in response to visual cues (Riehle, Grün, Diesmann, & Aertsen, 1997) or trace conditioning in the rabbit (McEchron, Weible, & Disterhoft, 2001). The neural spiking activity in implicit stimulus experiments is frequently analyzed by binning the spikes and plotting the peristimulus time histogram (PSTH). When several neurons are recorded in parallel, cross-correlations or unitary events analysis (Riehle et al., 1997; Grün, Diesmann, Grammont, Riehle, & Aertsen, 1999) have been used to analyze synchrony and changes in firing rates. Parametric model-based statistical analysis has been performed for the explicit stimuli of hippocampal place cells using position data (Brown, Frank, Tang, Quirk, & Wilson, 1998). However, specifying a model when the stimulus is latent or implicit is more challenging. State-space models suggest an approach to developing a model-based framework for analyzing stimulus-response experiments when the stimulus is implicit.

We develop an approach to estimating state-space models observed through a point process. We represent the latent (implicit) process modulating the neural spiking activity as a gaussian autoregressive model driven by an external stimulus. Given the latent process, neural spiking activity

is characterized as a general point process defined by its conditional intensity function. We will be concerned here with estimating the unobservable state or latent process, its parameters, and the parameters of the point process model. Several approaches have been taken to the problem of simultaneous state estimation and model parameter estimation, the latter being termed *system identification* (Roweis & Ghahramani, 1999). In this article, we present an approximate expectation-maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977) to solve this simultaneous estimation problem. The approximate EM algorithm combines a point process recursive nonlinear filter algorithm, the fixed interval smoothing algorithm, and the state-space covariance algorithm to compute the complete data log likelihood efficiently. We use a Kolmogorov-Smirnov test based on the time-rescaling theorem to evaluate agreement between the model and point process data. We illustrate the algorithm with two simulated data examples: an ensemble of Poisson neurons driven by a common stimulus and a single neuron whose conditional intensity function is approximated as a local Bernoulli process.

2 Theory

2.1 Notation and the Point Process Conditional Intensity Function.

Let $(0, T]$ be an observation interval during which we record the spiking activity of C neurons. Let $0 < u_{c1} < u_{c2} < \dots < u_{cJ_c} \leq T$ be the set of J_c spike times (point process observations) from neuron c for $c = 1, \dots, C$. For $t \in (0, T]$, let $N_{0,t}^c$ be the sample path of the spike times from neuron c in $(0, t]$. It is defined as the event $N_{0,t}^c = \{0 < u_{c1} < u_{c2}, \dots, u_{cj} \leq t \cap N^c(t) = j\}$, where $N^c(t)$ is the number of spikes in $(0, t]$ and $j \leq J_c$. The sample path is a right continuous function that jumps 1 at the spike times and is constant otherwise (Snyder & Miller, 1991). This function tracks the location and number of spikes in $(0, t]$ and therefore contains all the information in the sequence of spike times. Let $N_{0,t} = \{N_{0,t}^1, \dots, N_{0,t}^C\}$ be the ensemble spiking activity in $(0, t]$.

The spiking activity of each neuron can depend on the history of the ensemble, as well as that of the stimulus. To represent this dependence, we define the set of stimuli applied in $(0, t]$ as $S_{0,t} = \{0 < s_1 < \dots < s_\ell \leq t\}$. Let $H_t = \{N_{0,t}, S_{0,t}\}$ be the history of all C neurons up to and including time t . To define a probability model for the neural spiking activity, we define the conditional intensity function for $t \in (0, T]$ as (Cox & Isham, 1980; Daley & Vere-Jones, 1988):

$$\lambda_c(t | H_t) = \lim_{\Delta \rightarrow 0} \frac{\Pr(N_{0,t+\Delta}^c - N_{0,t}^c = 1 | H_t)}{\Delta}. \quad (2.1)$$

The conditional intensity function is a history-dependent rate function that generalizes the definition of the Poisson rate (Cox & Isham, 1980; Daley &

Vere-Jones, 1988). If the point process is an inhomogeneous Poisson process, the conditional intensity function is $\lambda_c(t | H_t) = \lambda_c(t)$. It follows that $\lambda_c(t | H_t)\Delta$ is the probability of a spike in $[t, t + \Delta)$ when there is history dependence in the spike train. In survival analysis, the conditional intensity is termed the *hazard function* because, in this case, $\lambda_c(t | H_t)\Delta$ measures the probability of a failure or death in $[t, t + \Delta)$ given that the process has survived up to time t (Kalbfleisch & Prentice, 1980).

2.2 Latent Process Model, Sample Path Probability Density, and the Complete Data Likelihood. It is possible to define the latent process in continuous time. However, to simplify the notation for our filtering, smoothing, and EM algorithms, we assume that the latent process is defined on a discrete set of evenly spaced lattice points. To define the lattice, we choose K large, and divide $(0, T]$ into K intervals of equal width $\Delta = T/K$, so that there is at most one spike per interval. The latent process model is evaluated at $k\Delta$ for $k = 1, \dots, K$. We also assume the stimulus inputs can be measured at a resolution of Δ .

We define the latent model as the first-order autoregressive model,

$$x_k = \rho x_{k-1} + \alpha I_k + \varepsilon_k, \quad (2.2)$$

where x_k is the unknown state at time $k\Delta$, ρ is a correlation coefficient, I_k is the indicator function that is 1 if there is a stimulus at $k\Delta$ and zero otherwise, α modulates the effect of the stimulus on the latent process, and ε_k is a gaussian random variable with mean zero and variance σ_ε^2 . While more complex latent process models can certainly be defined, equation 2.2 is adequate to illustrate the essential features of our algorithm.

The joint probability density of the latent process is

$$p(x | \rho, \alpha, \sigma_\varepsilon^2) = \left[\frac{(1 - \rho^2)}{2\pi\sigma_\varepsilon^2} \right]^{\frac{1}{2}} \times \exp \left\{ -\frac{1}{2} \left[\frac{(1 - \rho^2)}{\sigma_\varepsilon^2} x_0^2 + \sum_{k=1}^K \frac{(x_k - \rho x_{k-1} - \alpha I_k)^2}{\sigma_\varepsilon^2} \right] \right\}, \quad (2.3)$$

where $x = (x_0, x_1, \dots, x_K)$.

We assume that the conditional intensity function is $\lambda_c(k\Delta | x_k, H_k^c, \theta_c^*)$, where θ_c^* is an unknown parameter. We can express the joint probability density of the sample path of neuron c conditional on the latent process as (Barbieri, Quirk, Frank, Wilson, & Brown, 2001; Brown, Barbieri, Ventura,

Kass, & Frank, 2002)

$$p(N_{0,T}^c \mid x, H_T^c, \theta_c^*) = \exp \left[\int_0^T \log \lambda_c(u \mid x(u), H_u^c, \theta_c^*) dN^c(u) - \int_0^T \lambda_c(u \mid x(u), H_u^c, \theta_c^*) du \right], \quad (2.4)$$

where $dN^c(u) = 1$ if there is a spike at u from neuron c and 0 otherwise. Under the assumption that the neurons in the ensemble are conditionally independent given the latent process, the joint probability density of the sample paths of the ensemble is

$$p(N_{0,T} \mid x, H_T, \theta^*) = \prod_{c=1}^C p(N_{0,T}^c \mid x, H_T^c, \theta_c^*), \quad (2.5)$$

where $\theta^* = (\theta_1^*, \dots, \theta_C^*)$.

2.3 Parameter Estimation: Expectation-Maximization Algorithm. To illustrate the algorithm, we choose a simple form of the conditional intensity function. That is, we take the conditional intensity function for neuron c as

$$\lambda_c(k\Delta) = \exp(\mu_c + \beta_c x_k), \quad (2.6)$$

where μ_c is the log of the background firing rate and β_c is its gain parameter that governs how much the latent process modulates the firing rate of this neuron. Here we have $\theta_c^* = (\mu_c, \beta_c)$. Equations 2.3 and 2.6 define a doubly stochastic point process (Cox & Isham, 1980). If we condition on the latent process, then equation 2.6 defines an inhomogeneous Poisson process. Under this model, all the history dependence is through the stimulus. We let $\theta = (\rho, \alpha, \sigma_\epsilon^2, \theta^*)$. Because our objective is to estimate the latent process, x , and to compute the maximum likelihood estimate of the model parameter, θ , we develop an EM algorithm (Dempster et al., 1977). In our EM algorithm, we treat the latent process x as the missing or unobserved quantity. The EM algorithm requires us to maximize the expectation of the complete data log likelihood. It follows from equations 2.3 and 2.6 that the complete data likelihood for our model is

$$p(N_{0,T}, x \mid \theta) = p(N_{0,T} \mid x, \theta^*) p(x \mid \rho, \alpha, \sigma_\epsilon^2). \quad (2.7)$$

2.3.1 E-Step. At iteration $\ell+1$ of the algorithm, we compute in the E-step the expectation of the complete data log likelihood given H_K , the ensemble

spiking activity and stimulus activity in $(0, T]$, and $\theta^{(\ell)}$, the parameter estimate from iteration ℓ . By our notation convention in the previous section, since $K\Delta = T$, $H_K = H_T$ and

$$\begin{aligned}
Q(\theta \mid \theta^{(\ell)}) &= \mathbb{E}[\log[p(N_{0,T}, x \mid \theta)] \parallel H_K, \theta^{(\ell)}] \\
&= \mathbb{E} \left[\sum_{k=0}^K \sum_{c=1}^C (dN^c(k\Delta)(\mu_c + \beta_c x_k + \log \Delta) \right. \\
&\quad \left. - \exp(\mu_c + \beta_c x_k) \Delta) \parallel H_K, \theta^{(\ell)} \right] \\
&\quad + \mathbb{E} \left[\sum_{k=1}^K -\frac{1}{2} \frac{(x_k - \rho x_{k-1} - \alpha I_k)^2}{\sigma_\varepsilon^2} \right. \\
&\quad \left. - \frac{K}{2} \log 2\pi - \frac{K}{2} \log \sigma_\varepsilon^2 \parallel H_K, \theta^{(\ell)} \right] \\
&\quad + \mathbb{E} \left[\frac{1}{2} \log(1 - \rho^2) - \frac{1}{2} \frac{x_0^2(1 - \rho^2)}{\sigma_\varepsilon^2} \parallel H_K, \theta^{(\ell)} \right]. \quad (2.8)
\end{aligned}$$

Upon expanding the right side of equation 2.8, we see that calculating the expected value of the complete data log likelihood requires computing the expected value of the latent process $\mathbb{E}[x_k \parallel H_K, \theta^{(\ell)}]$ and the covariances $\mathbb{E}[x_k^2 \parallel H_K, \theta^{(\ell)}]$ and $\mathbb{E}[x_k x_{k+1} \parallel H_K, \theta^{(\ell)}]$. We denote them as

$$x_{k|K} \equiv \mathbb{E}[x_k \parallel H_K, \theta^{(\ell)}] \quad (2.9)$$

$$W_k \equiv \mathbb{E}[x_k^2 \parallel H_K, \theta^{(\ell)}] \quad (2.10)$$

$$W_{k,k+1} \equiv \mathbb{E}[x_k x_{k+1} \parallel H_K, \theta^{(\ell)}], \quad (2.11)$$

for $k = 1, \dots, K$ where the notation $k \mid j$ denotes the expectation of the latent process at $k\Delta$ given the ensemble spiking activity and the stimulus up to time $j\Delta$. To compute these quantities efficiently, we decompose the E-step into three parts: a forward nonlinear recursive filter to compute $x_{k|K}$; a backward, fixed interval smoothing (FIS) algorithm to estimate $x_{k|K}$; and a state-space covariance algorithm to estimate W_k and $W_{k,k+1}$. This approach for evaluating the complete data log likelihood was suggested first by Shumway and Stoffer (1982). They used the FIS but a more complicated form of the state-covariance algorithm. An alternative covariance algorithm was given in Brown (1987). The logic of this approach is to compute the forward mean and covariance estimates and combine them with the backward mean and covariance estimates to obtain equations 2.10 and 2.11. This approach is exact for linear gaussian latent process models and linear gaussian

observation processes. For our model, it will be approximate because our observations form a point process.

E-Step I: Nonlinear Recursive Filter. The following equations comprise a recursive nonlinear filtering algorithm to estimate $x_{k|k}$ and $\sigma_{k|k}^2$ using equation 2.6 as the conditional intensity. The algorithm is based on the maximum a posteriori derivation of the Kalman filter algorithm (Mendel, 1995; Brown et al., 1998). It recursively computes a gaussian approximation to the posterior probability density $p(x_k | H_k, \theta^{(\ell)})$. The approximation is based on recursively computing the posterior mode $x_{k|k}$ and computing its variance $\sigma_{k|k}^2$ as the negative inverse of the second derivative of the log posterior probability density (Tanner, 1996). The nonlinear recursive algorithm is:

(Observation Equation)

$$p(dN(k\Delta) | x_k) = \prod_{c=1}^C [\exp(\mu_c + \beta_c x_k) \Delta]^{dN^c(k\Delta)} \times \exp(-\exp(\mu_c + \beta_c x_k) \Delta) \quad (2.12)$$

(One-Step Prediction)

$$x_{k|k-1} = \rho x_{k-1|k-1} + \alpha I_k \quad (2.13)$$

(One-Step Prediction Variance)

$$\sigma_{k|k-1}^2 = \rho^2 \sigma_{k-1|k-1}^2 + \sigma_\varepsilon^2 \quad (2.14)$$

(Posterior Mode)

$$x_{k|k} = x_{k|k-1} + \sigma_{k|k-1}^2 \sum_{c=1}^C \beta_c [dN^c(k\Delta) - \exp(\mu_c + \beta_c x_{k|k}) \Delta] \quad (2.15)$$

(Posterior Variance)

$$\sigma_{k|k}^2 = - \left[-(\sigma_{k|k-1}^2)^{-1} - \sum_{i=1}^C \beta_i^2 \exp(\mu_i + \beta_i x_{k|k}) \Delta \right]^{-1} \quad (2.16)$$

for $k = 1, \dots, K$. The initial condition is x_0 and $\sigma_{0|0}^2 = \sigma_\varepsilon^2(1 - \rho^2)^{-1}$. The algorithm is nonlinear because $x_{k|k}$ appears on the left and right of equation 2.15. The derivation of this algorithm for an arbitrary point process model is given in the appendix.

E-Step II: Fixed Interval Smoothing (FIS) Algorithm. Given the sequence of posterior mode estimates $x_{k|k}$ (see equation 2.15) and the variance $\sigma_{k|k}^2$ (see

equation 2.16) we use the fixed interval smoothing algorithm to compute $x_{k|K}$ and $\sigma_{k|K}^2$. The algorithm is (Mendel, 1995; Brown et al., 1998)

$$x_{k|K} = x_{k|k} + A_k(x_{k+1|K} - x_{k+1|k}), \quad (2.17)$$

where

$$A_k = \rho \sigma_{k|k}^2 (\sigma_{k+1|k}^2)^{-1} \quad (2.18)$$

and

$$\sigma_{k|K}^2 = \sigma_{k|k}^2 + A_k^2 (\sigma_{k+1|K}^2 - \sigma_{k+1|k}^2) \quad (2.19)$$

for $k = K - 1, \dots, 1$ and initial conditions $x_{K|K}$ and $\sigma_{K|K}^2$.

E-Step III: State-Space Covariance Algorithm. The covariance estimate, $\sigma_{k,u|K}$, can be computed from the state-space covariance algorithm (de Jong & MacKinnon, 1988) and is given as

$$\sigma_{k,u|K} = A_k \sigma_{k+1,u|K} \quad (2.20)$$

for $1 \leq k \leq u \leq K$. It follows that the covariance terms required for the E-step are

$$W_{k,k+1} = \sigma_{k,k+1|K} + x_{k|K} x_{k+1|K} \quad (2.21)$$

and

$$W_k = \sigma_{k|K}^2 + x_{k|K}^2. \quad (2.22)$$

2.3.2 M-Step. In the M-step, we maximize the expected value of the complete data log likelihood in equation 2.8 with respect to $\theta^{(\ell+1)}$. In so doing, we obtain the following closed-form solutions for $\rho^{(\ell+1)}$, $\alpha^{(\ell+1)}$ and $\sigma_\varepsilon^{2(\ell+1)}$,

$$\begin{bmatrix} \rho^{(\ell+1)} \\ \alpha^{(\ell+1)} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^K W_{k-1} & \sum_{k=1}^K x_{k-1|k} I_k \\ \sum_{k=1}^K x_{k-1|k} I_k & \sum_{k=1}^K I_k \end{bmatrix}^{-1} \begin{bmatrix} \sum_{k=1}^K W_{k,k-1} \\ \sum_{k=1}^K x_{k|k} I_k \end{bmatrix} \quad (2.23)$$

$$\sigma_\varepsilon^{2(\ell+1)} = K^{-1} \left[\sum_{k=1}^K W_k + \rho^{2(\ell+1)} \sum_{k=1}^K W_{k-1} + \alpha^{2(\ell+1)} \sum_{k=1}^K I_k \right]$$

$$\begin{aligned}
& -2\rho^{(\ell+1)} \sum_{k=1}^K W_{k,k-1} - 2\alpha^{(\ell+1)} \sum_{k=1}^K x_{k|K} I_k + 2\rho^{(\ell+1)} \alpha^{(\ell+1)} \\
& \times \sum_{k=1}^K x_{k-1|K} I_k + W_0 \left(1 - \rho^{2(\ell+1)}\right) \Big], \quad (2.24)
\end{aligned}$$

where initial conditions for the latent process are estimated from $x_0^{(\ell+1)} = \rho^{(\ell+1)} x_{1|K}$ and $\sigma_{0|0}^{2(\ell+1)} = \sigma_\varepsilon^{2(\ell+1)} (1 - \rho^{2(\ell+1)})^{-1}$. The closed-form solution for $\rho^{(\ell+1)}$ in equation 2.23 is obtained by neglecting the last two terms in the expectation of the complete data log likelihood (see equation 2.8). This approximation means that we estimate $\rho^{(\ell+1)}$ from the probability density of x_1, \dots, x_K given x_0 and the point process measurements instead of the probability density of x_0, \dots, x_K given the point process measurements. Inclusion of the last two terms results in a cubic equation for computing $\rho^{(\ell+1)}$, which is avoided by using the closed-form approximation. We report only the results of the closed-form solution in section 3 because we found that for our algorithms, the absolute value of the fractional difference between the two solutions was less than 10^{-6} (i.e., $|\text{cubic solution} - \text{closed form solution}| / \text{cubic solution} < 10^{-6}$).

The parameter $\mu_c^{(\ell+1)}$ is estimated as

$$\begin{aligned}
\mu_c^{(\ell+1)} &= \log N^c(T) \\
& - \log \left(\sum_{k=1}^K \exp \left(\beta_c^{(\ell+1)} x_{k|K} + \frac{1}{2} \beta_c^{2(\ell+1)} \sigma_{k|K}^2 \right) \Delta \right), \quad (2.25)
\end{aligned}$$

whereas $\beta_c^{(\ell+1)}$ is the solution to the nonlinear equation,

$$\begin{aligned}
& \sum_{k=1}^K dN^c(k\Delta) x_{k|K} \\
& = \exp \mu_c^{(\ell+1)} \left\{ \sum_{k=1}^K \exp \left(\beta_c^{(\ell+1)} x_{k|K} + \frac{1}{2} \beta_c^{2(\ell+1)} \sigma_{k|K}^2 \right) \right. \\
& \quad \times (x_{k|K} + \beta_c^{(\ell+1)} \sigma_{k|K}^2) \Delta \Big\}, \quad (2.26)
\end{aligned}$$

which is solved by Newton's method after substituting $\mu_c^{(\ell+1)}$ from equation 2.26. The expectations needed to derive equations 2.25 and 2.26 were computed using the lognormal probability density and the approximation of $p(x_k | H_K, \theta^{(\ell)})$ as a gaussian probability density with mean $x_{k|K}$ and variance $\sigma_{k|K}^2$.

2.4 Assessing Model Goodness-of-Fit by the Time-Rescaling Theorem. The latent process and point process models, along with the EM algorithm, provide a model and an estimation procedure for computing the latent process and the parameter vector θ . It is important to evaluate model goodness-of-fit, that is, determine how well the model describes the neural spike train data series data. Because the spike train models are defined in terms of an explicit point process model, we can use the time-rescaling theorem to evaluate model goodness-of-fit. To do this, we compute for each neuron the time-rescaled or transformed interspike intervals

$$\tau_j = \int_{u_{j-1}}^{u_j} \lambda(u \mid \hat{\theta}) du, \quad (2.27)$$

where the u_j s are the spike times from the neuron and $\lambda(t \mid \theta)$ is the conditional intensity function in equation 2.6 evaluated at the maximum likelihood estimate $\hat{\theta}$ for $j = 1, \dots, J$, where we have dropped the subscript c to simplify notation. The u_j s are a point process with a well-defined conditional intensity function and, hence, by the time-rescaling theorem, the τ_j s are independent, exponential random variables with a unit rate (Barbieri et al., 2001; Brown et al., 2002). Under the further transformation $z_j = 1 - \exp(-\tau_j)$, the z_j s are independent, uniform random variables on the interval (0,1). Therefore, we can construct a Kolmogorov-Smirnov (K-S) test to measure agreement between the z_j s and the uniform probability density (Barbieri et al., 2001; Brown et al., 2002). First, we order the z_j s from the smallest to the largest value. Then we plot values of the cumulative distribution function of the uniform density defined as $b_j = \frac{j-1}{J}$ for $j = 1, \dots, J$ against the ordered z_j s. The points should lie on the 45 degree line. Because the transformation from the u_j s to the z_j s is one-to-one, a close agreement between the probability density of the z_j s and the uniform probability density on (0,1) indicates close agreement between the (latent process-point process) model and the point process measurements. Hence, the time-rescaling theorem provides a direct means of measuring agreement between a point process or neural spike train time series and a probability model intended to describe its stochastic structure.

3 Applications

3.1 Example 1. Multiple Neurons Driven by a Common Latent Process.

To illustrate our analysis paradigm, we simulate $C = 20$ simultaneously recorded neurons from the model described by equations 2.2 and 2.6. The time interval for the simulation was $T = 10$ seconds, and the latent process model parameters were $\rho = 0.99$, $\alpha = 3$, and $\sigma_\varepsilon^2 = 10^{-3}$ with the implicit stimulus I_k applied at 1 second intervals. The parameters for the conditional intensity function defining the observation process were the log of the background firing rate $\mu = -4.9$ for all neurons, whereas the gain coefficients

Table 1: True Parameter Values and EM Algorithm Parameter Estimates.

Parameter	True	Estimate	Parameter	True	Estimate
ρ	0.990	0.993	β_{10}	1.079	1.190
α	3.000	2.625	β_{11}	1.008	1.039
μ	-4.900	-5.105	β_{12}	1.078	1.247
β_1	1.075	1.261	β_{13}	1.009	1.031
β_2	0.989	1.060	β_{14}	1.078	1.129
β_3	1.035	1.039	β_{15}	0.980	1.121
β_4	0.973	1.101	β_{16}	0.976	1.164
β_5	1.003	1.190	β_{17}	0.990	1.234
β_6	0.992	1.132	β_{18}	0.968	1.154
β_7	0.977	1.118	β_{19}	0.956	0.979
β_8	1.027	1.176	β_{20}	1.098	1.350
β_9	1.066	1.089			

β_c for the 20 neurons were chosen randomly on the interval [0.9 1.1]. These parameter values for the latent and observation processes were chosen to give an approximate average firing rate of 10 Hz for each neuron. The neural spike trains were simulated using the time-rescaling theorem algorithm described in Brown et al. (2002), and the state equations in the EM algorithm were updated at $\Delta = 1$ msec.

Using the EM algorithm described in the previous section, we fit the model in equations 2.2 and 2.6 simultaneously to the 20 simulated neural spike trains, assuming a fixed noise variance, $\sigma_\varepsilon^2 = 10^{-3}$. The convergence criteria for the algorithm were absolute changes of less than 10^{-2} in consecutive iterates of the parameters and relative changes in the parameter iterates of less than 10^{-3} , that is, $|old - new|/old < 10^{-3}$. The parameter estimates from the EM algorithm were in good agreement with the true values used to simulate the spike train data (see Table 1). In this case, the overestimates of the gain coefficients, β_c , are offset by the underestimates of α and μ .

Approximating the probability density of the state at $k\Delta$ as the gaussian density with mean \hat{x}_k and variance, $\hat{\sigma}_k^2$, it follows from equation 2.6 and the standard change of variables formula from probability theory that the probability density of the rate for neuron c at time $k\Delta$ is the lognormal probability density defined as

$$p(\lambda_k^c | \hat{\mu}_c, \hat{\beta}_c, \hat{x}_k) = (2\pi\hat{\sigma}_k^2)^{-\frac{1}{2}} \hat{\beta}_c(\lambda_k^c)^{-1} \times \exp\left(-\frac{1}{2} \left[\frac{\hat{\beta}_c^{-1}(\log \lambda_k^c - \hat{\mu}_c) - \hat{x}_k}{\hat{\sigma}_k} \right]^2\right), \quad (3.1)$$

where $\hat{\mu}_c$ and $\hat{\beta}_c$ are the EM algorithm estimates of μ_c and β_c for $c = 1, \dots, 20$. The $1 - \xi$ confidence limits for the rate are computed simply by using the

relation between the lognormal and standard gaussian probabilities to find the $\xi/2$ and $1 - \xi/2$ quantiles of the probability density in equation 3.1 for $\xi \in (0, 1)$. In our analyses, we take $\xi = 0.05$ and construct 95% confidence intervals.

To compare our model-based analysis with current practices for analyzing neural spike train data using empirical smoothing methods, we also estimated the rate function for each of the 20 neurons by dividing the number of spikes in a 100 msec window by 100 msec. The window was then shifted 1 msec to give the same temporal resolution as in our updating algorithms. Because the latent process drives all the neurons, we also estimated the population rate by averaging the rates across all 20 neurons. This is a commonly used empirical temporal smoothing algorithm for computing spike rate that does not make use of stimulus information in the estimation (Riehle et al., 1997; Grün et al., 1999; Wood, Dudchenko, & Eichenbaum, 1999).

The confidence limits of the model-based rate function give a good estimate of the true firing rate used to generate the spikes (see Figure 1). In particular, the estimates reproduce the magnitude and duration of the effect of the implicit stimulus on the spike firing rate. The population firing rate estimated using temporal smoothing across all neurons is misleading (see Figure 1, dot-dashed line) in that around the time of the stimuli, it has diminished amplitude and is spread out in time. Furthermore, if we smooth a single spike train without averaging across neurons, spurious peaks in the firing rate can be produced due to noise (see Figure 1, solid gray line). By using information about the timing of the stimulus, the model firing rate estimate follows the true rate function more closely.

The 95% confidence bounds for the state process estimated from the EM algorithm cover almost completely the time course of the true state process (see Figure 2). The true state lies sometimes outside the confidence limits in regions where there are very few spikes and, hence, little information about the latent process.

To assess how well the model fits the data, we apply the K-S goodness-of-fit tests based on the time-rescaling theorem as described in section 2.4 (see Figure 3). In Figures 3A, 3B, and 3C, the solid black line represents exact agreement between the model and spike data, and dotted lines represent 95% confidence limits. For 18 of 20 neurons, the model lies within the confidence limits, indicating a good fit to the data. In contrast, for 14 of the 20 neurons, the empirical estimate lies outside the confidence limits. For 6 of 20 neurons, both the model and empirical estimates lie within the confidence limits (see Figure 3B). In two cases, the model lies outside the confidence limits (see Figure 3C). In both cases where the model does not fit the data, the empirical estimate does not either. Thus, the model appears to give a considerably more accurate description of the spike train than the empirical estimate.

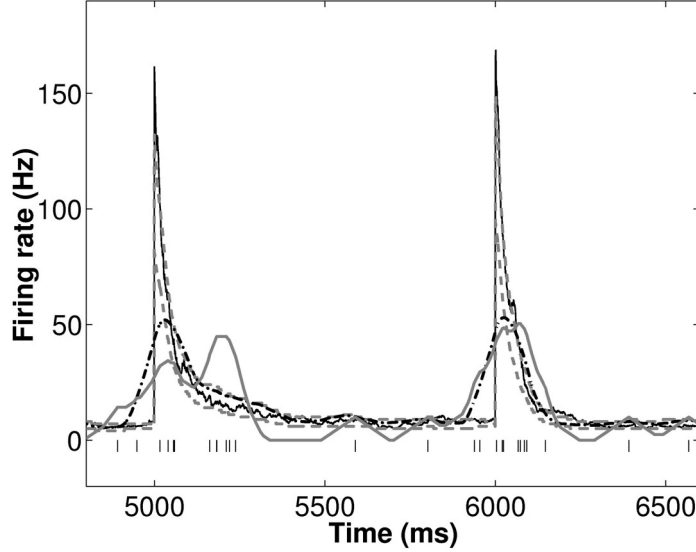


Figure 1: A 2000 millisecond segment of data from neuron 8 in the 20-neuron simulation study of latent process estimation by the EM algorithm from example 1. During this period, 26 spikes (vertical bars above abscissa) are concentrated around two stimulus events at 5000 msec and 6000 msec. The 95% confidence bands for the firing rate estimated by the latent model (dashed lines) cover the true rate (solid black line) used to generate the spikes most of the time. In contrast, when the firing rate for neuron 8 is estimated empirically (solid gray line) using a 100 msec smoothing window (i.e., by dividing the number of spikes in a 100 msec window by 100 msec), the stimuli appear diminished in amplitude and spread out in time. Furthermore, a spurious third peak in firing rate is produced approximately 200 ms after the true first peak. This third peak is avoided by using a smoothed estimate that is averaged over all cells (dash-dotted black line), but the EM algorithm estimate of the firing rate based on the model more accurately reproduces the features of the stimuli.

3.2 Example 2. Single Neuron Latent Stimulus-Response Model. A common practice is to use the binomial probability mass function as a local model for analyzing neural spike trains. We demonstrate how the parameters of a local Bernoulli model may be estimated using our EM algorithm. We generate a single spike train subject to multiple repeated stimuli over a period of approximately 1 minute. In this example, the conditional intensity is given by

$$\lambda(k\Delta)\Delta = \exp(\mu + x_k)\Delta \approx \frac{\exp(\mu + x_k)\Delta}{1 + \exp(\mu + x_k)\Delta} = p_k, \quad (3.2)$$

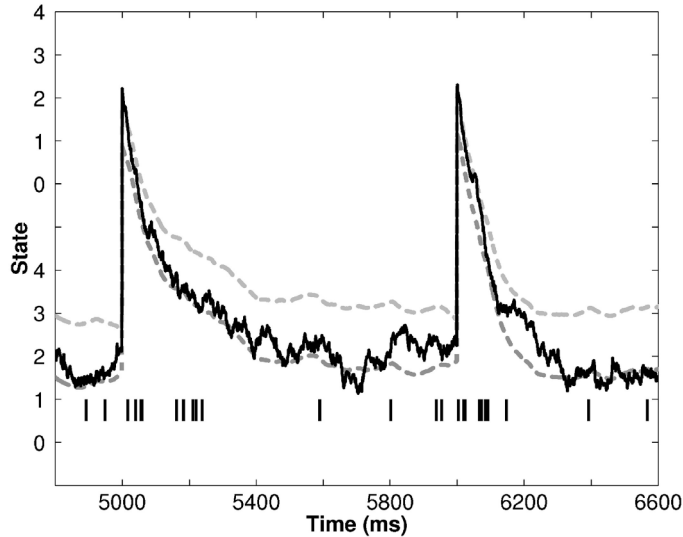
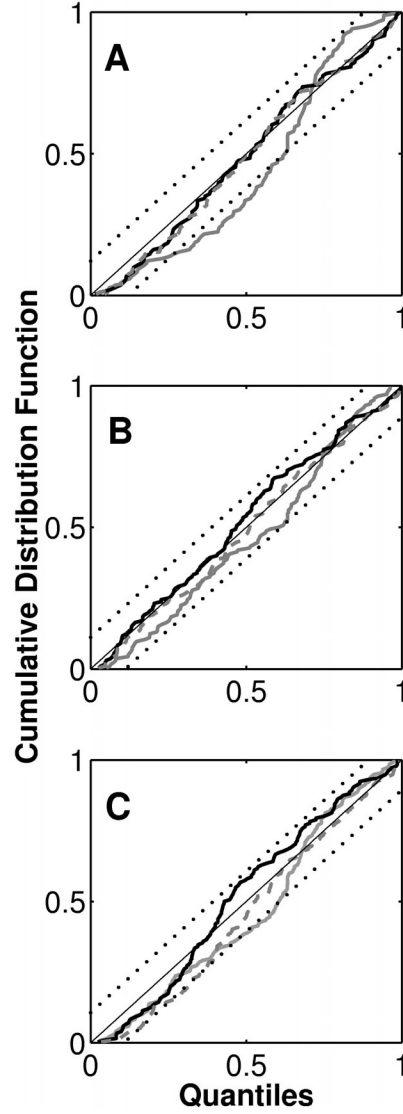


Figure 2: True state (solid black line) and estimated 95% confidence limits (dashed lines) for the true state computed from the spike train observed in Figure 1 using the EM algorithm. The estimated confidence limits computed as $x_{k|k} \pm 1.96\sigma_{k|k}$ fail to cover the true state when few spikes are discharged (e.g., near 5700 ms).

Figure 3: *Facing page.* Kolmogorov-Smirnov goodness-of-fit analyses based on the time-rescaling theorem for three representative neurons. Each panel is the K-S plots comparing the model rate estimate (dashed line), the empirical rate estimate (solid gray line), and the true rate (solid black line). In these figures, the 45 degree line in black represents an exact agreement between the model and the spike data. Dotted lines in each panel are the 95% confidence limits (see Brown et al., 2002, for details). Since the true rate was used to generate spikes, the true rate KS plot always lies within the confidence limits. (A) An example of the KS plot from 1 of the 18 out of the 20 neurons for which the model-based estimate of the KS plot was entirely within the confidence limits, indicating close agreement between the overall model fit and the simulated data. (B) An example of 1 of the 6 out of the 20 neurons for which the K-S plot based on the empirical rate estimate completely lies within the 95% confidence limits. (C) An example of 1 of the 2 out of the 20 neurons for which the K-S plot based on the model estimate of the rate function fails to fall within the 95% confidence limits. For both of these neurons, as this panel suggests, the KS plot based on the empirical rate model did not remain in the 95% confidence bounds.



and spikes are generated with the local Bernoulli model,

$$p(dN(k\Delta) | H_k) = p_k^{dN(k\Delta)} (1 - p_k)^{1-dN(k\Delta)}, \quad (3.3)$$

with parameters $\mu = -4.6$, $\alpha = 4$, and $\rho = 0.8$. The implicit stimulus is applied at 40 time points. We assume the noise σ_ε^2 in the latent process is 0.2. In this example, we set $\Delta = 5$ msec. This generates a spike train with

Table 2: True Parameter Values and the EM Algorithm Parameter Estimates.

Parameter	ρ	α	σ_ε^2	μ
True	0.800	4.000	0.200	-4.600
Estimate	0.804	3.573	0.125	-4.404

an average firing rate over the minute period of approximately 10 Hz. The parameters, including the noise variance σ_ε^2 , are again estimated using the EM algorithm and are in good agreement with their true values (see Table 2).

As in the previous example, confidence intervals were computed for the true rate function by deriving the probability density function for λ_k using the standard change of variable formula. Firing rates estimated by the model are compared with the empirical temporally smoothed rate computed by counting spikes in a 100 msec window and then dividing by 100 msec.

The firing rate (95% confidence limits) computed by the model compares favorably with the original rate used to generate the spikes (see Figure 4). The 95% confidence limits for the model estimate the magnitude and shape of the stimulus effect on the rate even when the stimulus has little obvious effect on the spike train. This is because the estimates for each stimulus's effect are made based on the whole time series. In contrast, the temporally smoothed estimate of rate does not clearly identify stimulus effects on the spike train (see Figure 4, solid gray line). The choice of a 100 msec window for these simulated data appears too large as increases in firing rate at the stimuli are smoothed out. We can also make an overall estimate of the accuracy of the firing rates by comparing the model and the empirical firing rates with the true firing at the stimulus times, namely, at 40 points in time where we applied the stimulus during the 1 minute epoch of simulated data. The mean differences (standard error) between the empirical and model rates compared to the true rate at these times are -113.9 (4.5) Hz and -8.5 (4.4) Hz, respectively. Thus, the model provides a significantly better approximation to the true rate than the empirical method (Student's t -test, $p < 10^{-6}$). The estimate of the state variable is found to compare well with the true state used to generate the spike train (see Figure 5).

Again we assess model goodness-of-fit using the K-S plot (see Figure 6). The fit of model lies within the 95% confidence limits except at very small quantiles and follows closely the curve for the true firing rate. In contrast, the fit for the empirical 100 msec temporal smoothing method lies outside the confidence limits for a large portion of the total time, indicating poor fit to the distribution over low and high percentiles.

3.3 An Extension to More General Point Process and State-Space Models. In the two simulated data examples we considered, the log conditional intensity function $\log \lambda_c(t | H_t)$ was a linear function of both the state vari-

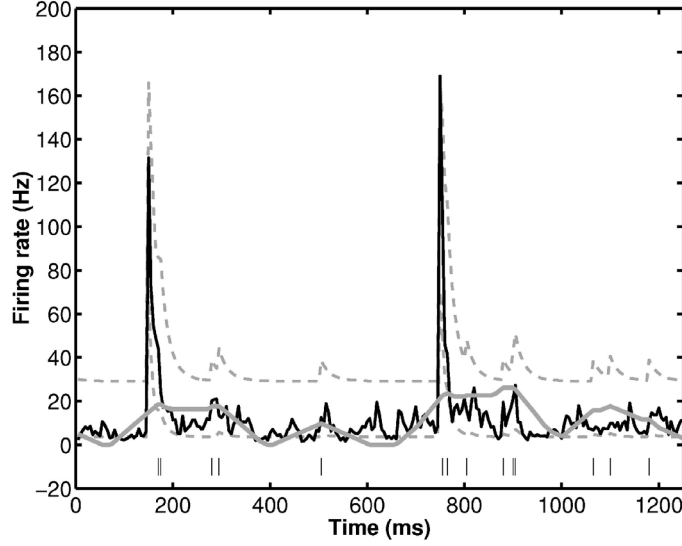


Figure 4: Simulated spike train (vertical bars above the abscissa), true firing rate (solid black line) from the local Bernoulli model (see equation 3.2) in example 2, 95% confidence limits for the rate from the model-based EM algorithm (dashed lines), and empirical firing rate estimate (solid gray line) computed by temporal smoothing over the 100 msec window. In this time segment, two external stimuli are applied at 150 msec and 750 msec. The 95% confidence limits cover nearly everywhere the true rate function. Although the spike count does not obviously increase at these times, the algorithm estimates effectively the amplitude and duration of the stimulus because it uses information from the entire spike train. The 100 msec window for the empirical rate function appears too large as increases in firing rate at the stimuli are smoothed out.

able x_k and certain components of the parameter vector θ . The EM algorithm is straightforward to modify when this is not the case. For an arbitrary $\lambda_c(t | H_t)$, the E-step in equation 2.8 becomes

$$\begin{aligned}
 Q(\theta | \theta^{(\ell)}) &= E[\log[p(N_{0,T}, x | \theta)] | H_K, \theta^{(\ell)}] \\
 &\approx E \left[\sum_{k=0}^K \sum_{c=1}^C dN^c(k\Delta) \log \lambda_c(x_k | H_k, \theta) \right. \\
 &\quad \left. - \lambda_c(x_k | H_k, \theta) \Delta | H_K, \theta^{(\ell)} \right] \\
 &\quad + E[\log p(x | \theta) | H_K, \theta^{(\ell)}],
 \end{aligned} \tag{3.4}$$

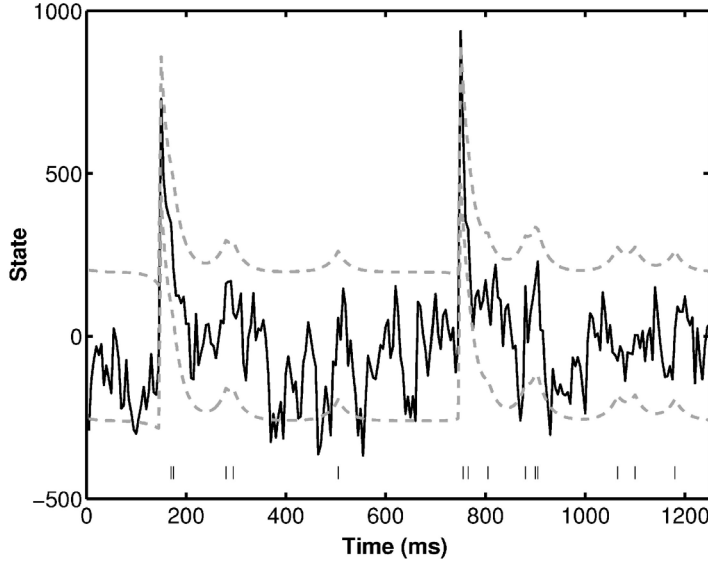


Figure 5: The true state (black line) and the model-based estimates of the 95% confidence intervals (dashed lines) computed using the EM algorithm for the local Bernoulli probability model corresponding to the spike train and rate function in Figure 4. As in example 1, Figure 5 shows that the 95% confidence limits cover the true state completely except when the neural spiking activity is low (around 700 msec).

where the last term in equation 3.4 is the sum of the last two terms on the right-hand side of equation 2.8. We assume $\lambda_c(x_k | H_k, \theta)$ and $\log \lambda_c(x_k | H_k, \theta)$ are twice differentiable functions that we denote generically as $g(x_k)$. To evaluate the first term on the right side of equation 3.4, it suffices to compute $E[g(x_k | H_k, \theta) | H_K, \theta^{(\ell)}]$. This can be accomplished by expanding $g(x_k)$ in a Taylor series about $\hat{x}_{k|K}$ and taking the expected value to obtain the approximation

$$E[g(x_k | H_k, \theta) | H_K, \theta^{(\ell)}] \doteq g(\hat{x}_{k|K}) + \frac{1}{2} \sigma_{k|K}^2 g''(\hat{x}_{k|K}), \quad (3.5)$$

where $g''(\hat{x}_{k|K})$ is the second derivative of $g(x_k)$ evaluated at $\hat{x}_{k|K}$. The right-hand side of equation 3.5 is substituted into equation 3.4 to evaluate the E-step. The evaluation of the second term on the right of equation 3.4 proceeds as in the evaluation of the second and third terms on the right of equation 2.8.

If the log conditional intensity function is no longer a linear or approximately linear function of the parameters, the M-step takes the more

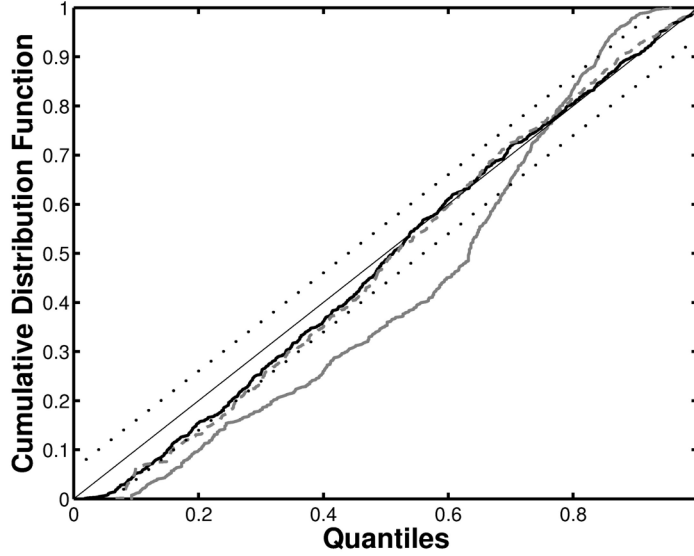


Figure 6: Kolmogorov-Smirnov goodness-of-fit analysis for local Bernoulli probability model. The K-S plot for the EM algorithm model-based estimate (dashed line) lies almost entirely within the 95% confidence bounds (dotted lines) and follows closely the KS plot computed using the true firing rate (solid black line), which lies almost entirely on the 45 degree line of exact agreement. In contrast, the empirical rate estimate (solid gray line) lies mostly outside the confidence limits, suggesting that this model does not agree closely with the data.

general form,

$$\nabla Q(\theta \mid \theta^{(\ell)}) = 0, \quad (3.6)$$

where the gradient in $\nabla Q(\theta \mid \theta^{(\ell)})$ is with respect to θ . The zero of $\nabla Q(\theta \mid \theta^{(\ell)})$ and, hence $\theta^{(\ell+1)}$, has to be found using Newton's method or another appropriate numerical optimization procedure. Similarly, the recursive filtering and smoothing algorithms generalize in a straightforward way when the dimension of the state-space model is greater than one (Brown et al., 1998). While these modifications increase the computational requirements of our algorithm, they illustrate how it can be applied to a wider range of point process and state-space models.

4 Discussion

We have presented a method for estimating from point process (spike train) data a conditional intensity (rate) function modulated by an unobservable

or latent continuous-valued state variable. The latent variable relates the effect of the external stimuli applied at specified times by the experimenter to the spike train rate function. We compute maximum likelihood estimates of the model parameters by the EM algorithm in which the E-step combines forward and backward point process filtering algorithms. The model performs better than smoothed histogram estimates of rate because it makes explicit use of the timing of the stimulus to analyze changes in background firing. Also, the model gives a more accurate description of the neural spike train as evaluated by the goodness-of-fit K-S test.

Several authors have discussed the analyses of state-space models in which the observation process is a point process. Diggle, Liang, and Zeger (1995) briefly mention state estimation from point process observations but no specific algorithms are given. MacDonald and Zucchini (1997) discuss state estimation for point processes without using the smoothing and filtering approach suggested here. West and Harrison (1997) define the state-space model implicitly and use the discount concept to construct an approximate forward filter. This approach is difficult to generalize (Fahrmeir, 1992). Kitagawa and Gersch (1996) described numerical algorithms to carry out state-space updating with forward recursion algorithms for binomial and Poisson observation processes. For the Poisson model in equation 2.6, Chan and Ledolter (1995) provided a computationally intensive Markov chain Monte Carlo EM algorithm to conduct the state updating and parameter estimation. The forward updating algorithms of Fahrmeir (1992), Fahrmeir and Tutz (1994), and Durbin and Koopman (2000) resemble most closely the ones we present, particularly, in the special case where the observation process is a point process from an exponential family, and the natural parameter is modeled as a linear function of the latent process. Both the examples we present follow these two special cases. The Fahrmeir forward recursion algorithm for example 1 with a single neuron is

$$x_{k|k} = x_{k|k-1} + \frac{\sigma_{k|k-1}^2 \beta}{[\lambda(k\Delta) \beta^2 \sigma_{k|k-1}^2 + 1]} [dN(k\Delta) - \lambda(k\Delta)] \quad (4.1)$$

$$\sigma_{k|k}^2 = [(\sigma_{k|k-1}^2)^{-1} + \lambda(k\Delta) \beta^2]^{-1}, \quad (4.2)$$

whereas the Durbin and Koopman (2000) update is

$$x_{k|k} = x_{k|k-1} + \frac{\sigma_{k|k-1}^2 \beta}{(\lambda(k\Delta) + \sigma_{k|k-1}^2 \beta^2) \lambda(k\Delta)} [dN(k\Delta) - \lambda(k\Delta)] \quad (4.3)$$

$$\sigma_{k|k}^2 = [(\sigma_{k|k-1}^2)^{-1} + \beta^2 \lambda(k\Delta)]^{-1}. \quad (4.4)$$

The variance updating algorithm in the Fahrmeir algorithm agrees because the observed and expected Fisher information are the same for the Poisson model in our example. The state updating equation differs from our

updating formula in equation 2.15 because their update is computed from the Kalman filter and not directly by finding the root of the log posterior probability density. The state and variance update formulae in the Durbin and Koopman algorithm differ from ours because theirs use a Taylor series approximation of the score function, first derivative of the log likelihood, instead of the exact score function. Fahrmeir (1992) and Fahrmeir and Tutz (1994) suggest using the EM algorithm for estimating the unknown parameters, but details are not given.

In an example applied to spike train data, Sahini (1999) describes the use of a latent model for neural firing where spikes are generated as an inhomogeneous Polya process. In his model, parameters are computed by optimizing the marginalized posterior by gradient ascent, and Monte Carlo goodness-of-fit is used to compare the model fit with measured spike train stochastic process.

The fact that our state-space models fit the simulated data better than the empirical method is expected given that the spikes were generated with the model. In applications to real data, it will be possible to use the same approach, testing reasonable and ideally parsimonious forms of the state-space and point process models for a given neurophysiological experiment. In any case, we may use the time-rescaling theorem to assess goodness-of-fit of any candidate models.

To study the problem of estimating a latent process simultaneously with its model parameters and the parameters of the observation process, we discretized time and assumed that the observation process and the latent process occur on a lattice of points spaced Δ time units apart. Using the EM algorithm, we computed the maximum likelihood estimate of θ and empirical Bayes' estimates of the latent process conditional on the maximum likelihood estimate of θ . A Bayesian alternative would be to specify a prior distribution for θ and compute a joint posterior probability density for the latent process and θ . Liu and Chen (1998) developed sequential Monte Carlo algorithms that may be adapted to this approach.

As another alternative, it is useful to point out how the latent process and parameter estimation may be carried out if both the point process and the latent process are assumed to be measured in continuous time. Nonlinear continuous time filtering and smoothing algorithms for point process observations have been studied extensively in the control theory literature (Snyder, 1975; Segall, Davis, & Kailath, 1975; Boel & Beneš, 1980; Snyder & Miller, 1991; Twum-Danso, 1997; Solo, 2000; Twum-Danso & Brockett, 2001). If the normalized conditional probability density, $p(x(t) \mid N_{0,t})$, is to be evaluated in continuous time, then a nonlinear stochastic partial differential equation in this probability density must be solved at each step (Snyder, 1975; Snyder and Miller, 1991). Here we let $x(t)$ be the continuous time value of x_k . If the updating is performed with respect to the unnormalized conditional probability density, $p(x(t), N_{0,t})$, then a linear stochastic partial differential equation must be solved in this probability density at

each step (Boel & Beneš, 1980; Twum-Danso, 1997; Solo, 2000). For either the normalized or unnormalized probability density updating algorithms, the essential steps in their derivations use the posterior prediction equation in equation A.1 and the one-step prediction equation in equation A.2 to derive Fokker-Planck equations (Snyder, 1975; Snyder & Miller, 1991; Twum-Danso, 1997; Solo, 2000; Twum-Danso & Brockett, 2001). If the parameters of the continuous time system are nondynamic and unknown, then, as in the discretized time case we present here, either the normalized or unnormalized partial differential equation updating of the conditional probability density may be embedded in an EM algorithm to compute maximum likelihood estimates of the parameters and empirical Bayes estimates of the latent process. Similarly, a Bayesian procedure can be derived if a prior probability density for θ can be specified. While the normalized and unnormalized conditional probability density updating equations have been known for several years, the computational requirements of these algorithms may be the reason they have not been more widely used (Manton, Krishnamurthy, & Elliott, 1999). Discretized approximations (Snyder & Miller, 1991; Twum-Danso, 1997; Twum-Danso & Brockett, 2001) and sequential Monte Carlo algorithms (Solo, 2000; Doucet, de Freitas, & Gordon, 2001; Shoham, 2001) have been suggested as more plausible alternatives. The sequential Monte Carlo methods use simulations to compute recursively the solutions to equations A.1 and A.2 on a discrete lattice of time points, whereas our nonlinear recursive algorithm, equations 2.12 through 2.16, uses sequential gaussian approximations to perform the same computations.

A potential application of this analysis paradigm would be to estimate the effect of external cue on a spike train in the delayed-response hand-pointing task described in Riehle et al. (1997). In this experiment, parallel spike data, measured in primary motor cortex of the monkey, are analyzed to estimate differences between spike rate increases corresponding to actual motion and those caused by expectation of a stimulus. The structure of our model enables us to estimate above-random firing propensity in a single cell while incorporating the history of cell firing. For these reasons, this approach may have advantages over the unitary events analysis methods (Grün et al., 1999), which may be difficult to apply to neurons with low firing rates (Roy, Steinmetz, & Niebur, 2000). A second potential application would be in the analysis of cell firing in the trace conditioning paradigm (McEchron, Weible, & Disterhof, 2001). In this case, a conditioned stimulus is followed after a trace interval by an unconditioned stimulus. After many such trials, an association between the two stimuli develops, as evidenced by changes in the firing rate of the neuron. These data are conventionally analyzed using PSTH techniques. However, because these studies involve an implicit relation between the stimulus and the neural spiking activity, this relation may be more clearly delineated by using the paradigm presented here. Finally, we are currently investigating the application of these methods to learning and memory experiments during

recordings from the medial temporal lobe of the monkey (Wirth et al., 2002; Yanike et al., 2002).

The state-space approach suggests several advantages. First, the approach uses a latent variable to relate explicitly the timing of the stimulus input and history of the experiment to the observed spiking activity. Second, use of explicit probability models makes it possible to compute probability density functions and confidence intervals for model quantities of interest. Finally, formulation of the analysis in terms of a general point process model allows us to assess model goodness-of-fit using the K-S tests based on the time-rescaling theorem. In our opinion, this latter step is the most critical as it forces us to assess how sensitive our inferences may be to lack of agreement between the model and the experimental data.

In summary, we have presented a computationally tractable method for state-space and parameter estimation from point process observations and suggested that these algorithms may be useful for analyzing neurophysiologic experiments involving implicit stimuli. In a future publication, we will apply these methods to actual experimental studies.

Appendix: Derivation of the Recursive Nonlinear Filter Algorithm

We derive a form of the recursive filter equations appropriate for an arbitrary point process model. The algorithm in equations 2.12 through 2.16 is obtained by taking the special case of the Poisson. To derive the nonlinear recursive filter, we require the posterior prediction equation,

$$p(x_k | H_k) = \frac{p(x_k | H_{k-1})p(dN(k\Delta) | x_k, H_k)}{p(dN(k\Delta) | H_{k-1})}, \quad (\text{A.1})$$

and the one-step prediction or Chapman-Kolmogorov equation,

$$p(x_k | H_{k-1}) = \int p(x_k | x_{k-1})p(x_{k-1} | H_{k-1})dx_{k-1}. \quad (\text{A.2})$$

The derivation of the algorithm proceeds as follows. Assume that at time $(k-1)\Delta$, $x_{k-1|k-1}$ and $\sigma_{k-1|k-1}^2$ are given. Under a gaussian continuity assumption on x_k , the distribution of x_k given $x_{k-1|k-1}$ is $N(\rho x_{k-1|k-1} + \alpha I_k, \sigma_{k|k-1}^2)$, where $\sigma_{k|k-1}^2 = \sigma_\varepsilon^2 + \rho^2 \sigma_{k-1|k-1}^2$. By equations 2.4, A.1, and A.2 and the gaussian continuity assumption, the posterior probability density $p(x_k | H_k)$ and the log posterior probability density $\log p(x_k | H_k)$ are, respectively,

$$\begin{aligned} p(x_k | H_k) &\propto \exp \left\{ -\frac{1}{2} \frac{(x_k - \rho x_{k-1|k-1} - \alpha I_k)^2}{\sigma_{k|k-1}^2} \right\} \\ &\times \prod_{c=1}^C \exp \{ \log \lambda_c(k\Delta | H_k^c) dN^c(k\Delta) - \lambda_c(k\Delta | H_k^c) \Delta \} \quad (\text{A.3}) \end{aligned}$$

$$\begin{aligned} \log p(x_k | H_k) \propto & -\frac{1}{2} \frac{(x_k - \rho x_{k-1|k-1} - \alpha I_k)^2}{\sigma_{k|k-1}^2} \\ & + \sum_{c=1}^C [\log \lambda_c(k\Delta | H_k^c) dN^c(k\Delta) - \lambda_c(k\Delta | H_k^c) \Delta]. \quad (\text{A.4}) \end{aligned}$$

To find the optimal estimate of x_k , we apply a gaussian approximation to equation A.3. This gaussian approximation is distinct from the gaussian continuity assumption and means that we can use the mode and variance of the probability density in equation A.3 to approximate it as a gaussian probability density. As a result, we differentiate with respect to x_k to find the mode, and we compute the second derivative to obtain the approximate variance (Tanner, 1996). Differentiating equation A.4 with respect to x_k gives

$$\begin{aligned} \frac{\partial \log p(x_k | H_k)}{\partial x_k} = & -\frac{(x_k - \rho x_{k-1|k-1} - \alpha I_k)}{\sigma_{k|k-1}^2} \\ & + \sum_{c=1}^C \frac{1}{\lambda_c(k\Delta)} \frac{\partial \lambda_c}{\partial x_k} [dN^c(k\Delta) - \lambda_c(k\Delta) \Delta], \quad (\text{A.5}) \end{aligned}$$

and solving for x_k yields

$$\begin{aligned} x_k = & \rho x_{k-1|k-1} + \alpha I_k \\ & + \sum_{c=1}^C \sigma_{k|k-1}^2 \lambda_c(k\Delta | H_k^c)^{-1} \\ & \times \frac{\partial \lambda_c(k\Delta | H_k^c)}{\partial x_k} [dN^c(k\Delta) - \lambda_c(k\Delta | H_k^c) \Delta]. \quad (\text{A.6}) \end{aligned}$$

Equation A.6 is in general nonlinear in x_k and can be solved using Newton's method. The second derivative of equation A.5 is

$$\begin{aligned} \frac{\partial^2 \log p(x_k | H_k)}{\partial x_k^2} = & -\frac{1}{\sigma_{k|k-1}^2} \\ & + \sum_{c=1}^C \left[\left(\frac{\partial^2 \lambda_c(k\Delta)}{\partial x_k^2} \frac{1}{\lambda_c(k\Delta)} - \left(\frac{\partial \lambda_c(k\Delta)}{\partial x_k} \right)^2 \frac{1}{\lambda_c(k\Delta)^2} \right) \right. \\ & \quad \times [dN^c(k\Delta) - \lambda_c(k\Delta) \Delta] \\ & \quad \left. - \left(\frac{\partial \lambda_c(k\Delta)}{\partial x_k} \right)^2 \frac{1}{\lambda_c(k\Delta)} \Delta \right], \quad (\text{A.7}) \end{aligned}$$

and the variance of x_k , under the gaussian approximation to equation A.3, is

$$\sigma_{k|k}^2 = - \left[-\frac{1}{\sigma_{k|k-1}^2} + \sum_{c=1}^C \left[\left(\frac{\partial^2 \lambda_c(k\Delta)}{\partial x_k^2} \frac{1}{\lambda_c(k\Delta)} - \left(\frac{\partial \lambda_c(k\Delta)}{\partial x_k} \right)^2 \frac{1}{\lambda_c(k\Delta)^2} \right) \times [dN^c(k\Delta) - \lambda_c(k\Delta)\Delta] - \left(\frac{\partial \lambda_c(k\Delta)}{\partial x_k} \right)^2 \frac{1}{\lambda_c(k\Delta)} \Delta \right] \right]^{-1}. \quad (\text{A.8})$$

Equations A.6 and A.8 constitute the basis for the general form of the filter equations 2.12 through 2.16.

Acknowledgments

Support was provided in part by NIMH grants MH59733, MH61637, NIDA grant DA015644, NHLBI grant HL07901, and NSF grant IBN-0081458. Part of this research was performed while E.N.B. was on sabbatical at the Laboratory for Information and Decision Systems in the Department of Electrical Engineering and Computer Science at MIT.

References

- 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 spike train activity. *J. Neurosci. Methods*, 105, 25–37.
- Berry, M. J., Warland, D. K., & Meister, M. (1997). The structure and precision of retinal spike trains. *Proc. Nat. Acad. Sci. USA*, 94, 5411–5416.
- Bialek, W., Rieke, F., de Ruyter van Steveninck, R. R., & Warland, D. (1991). Reading a neural code. *Science*, 252, 1854–1857.
- Boel, R. K., & Beneš, V. E. (1980). Recursive nonlinear estimation of a diffusion acting as the rate of an observed Poisson process. *IEEE Trans. Inf. Theory*, 26(5), 561–574.
- Brown, E. N. (1987). *Identification and estimation of differential equation models for circadian data*. Unpublished doctoral dissertation, Harvard University, Cambridge, MA.
- Brown, E. N., Barbieri, R., Ventura, V., Kass, R. E., & Frank, L. M. (2002). The time-rescaling theorem and its application to neural spike train data analysis. *Neural Comp.*, 14(2), 325–346.
- 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. *J. Neurosci.*, 18, 7411–7425.
- Chan, K. S., & Ledolter, J. (1995). Monte Carlo estimation for time series models involving counts. *J. Am. Stat. Assoc.*, 90, 242–252.

- Cox, D. R., & Isham, V. (1980). *Point processes*. New York: Chapman and Hall.
- Daley, D. J., & Vere-Jones, D. (1988). *An introduction to the theory of point processes*. New York: Springer-Verlag.
- de Jong, P., & Mackinnon, M. J. (1988). Covariances for smoothed estimates in state space models. *Biometrika*, 75, 601–602.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. B*, 39, 1–38.
- Diggle, P. J., Liang, K-Y., & Zeger, S. L. (1995). *Analysis of longitudinal data*. Oxford: Clarendon.
- Doucet, A., de Freitas, N., & Gordon, N. (2001). *Sequential Monte Carlo methods in practice*. New York: Springer-Verlag.
- Durbin, J., & Koopman, S. J. (2000). Time series analysis of non-gaussian observations based on state space models from both classical and Bayesian perspectives. *J. Roy. Statist. Soc. B*, 62, 3–56.
- Fahrmeir, L. (1992). Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models. *J. Am. Stat. Assoc.*, 87, 501–509.
- Fahrmeir, L., & Tutz, D. (1994). Dynamic-stochastic models for time-dependent ordered paired-comparison systems. *J. Am. Stat. Assoc.*, 89, 1438–1449.
- Gharamani, Z. (2001). An introduction to hidden Markov models and Bayesian networks. *Int. J. Pattern Recognition*, 15(1), 9–42.
- Grün, S., Diesmann, M., Grammont, F., Riehle A., & Aertsen, A. (1999). Detecting unitary events without discretization of time. *J. Neurosci. Meth.*, 93, 67–79.
- Kalbfleisch, J. D., & Prentice, R. L. (1980). *The statistical analysis of failure time data*. New York: Wiley.
- Kay, S. M. (1988). *Modern spectral estimation: Theory and applications*. Upper Saddle River, NJ: Prentice Hall.
- Kitagawa, G., & Gersh, W. (1996). *Smoothness priors analysis of time series*. New York: Springer-Verlag.
- Liu, J. S., & Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems. *J. Am. Stat. Assoc.*, 93(443), 567–576.
- Ljung, L., & Söderström, S. (1987). *Theory and practice of recursive identification*. Cambridge, MA: MIT Press.
- MacDonald, I. L., & Zucchini, W. (1997). *Hidden Markov and other models for discrete-valued time series*. New York: Chapman and Hall.
- Manton, J. H., Krishnamurthy, V., & Elliott, R. J. (1999). Discrete time filters for doubly stochastic Poisson processes and other exponential noise models. *Int. J. Adapt. Control Sig. Proc.*, 13, 393–416.
- McEchron, M. D., Weible, A. P., & Disterhoft, J. F. (2001). Aging and learning-specific changes in single-neuron activity in CA1 hippocampus during rabbit trace eyeblink conditioning. *J. Neurophysiol.*, 86, 1839–1857.
- Mendel, J. M. (1995). *Lessons in estimation theory for signal processing, communication, and control*. Upper Saddle River, NJ: Prentice Hall.
- O’Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34, 171–175.

- Riehle, A., Grün, S., Diesmann, M., & Aertsen, A. (1997). Spike synchronization and rate modulation differentially involved in motor cortical function. *Science*, 278, 1950–1953.
- Roweis, S., & Ghahramani, Z. (1999). A unifying review of linear gaussian models. *Neural Comp.*, 11, 305–345.
- Roy, A., Steinmetz, P. N., & Niebur, E. (2000). Rate limitations of unitary event analysis. *Neural Comp.*, 12(9), 2063–2082.
- Sahini, M. (1999). *Latent variable models for neural data analysis*. Unpublished doctoral dissertation, California Institute of Technology, Pasadena.
- Segall, A., Davis, M. H. A., & Kailath, T. (1975). Nonlinear filtering with counting observations. *IEEE Trans. Inf. Theory*, 21(2), 143–149.
- Shoham, S. (2001). *Advances towards an implantable motor cortical interface*. Unpublished doctoral dissertation, University of Utah, Salt Lake City.
- Shumway, R. H., & Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Series Analysis*, 3, 253–264.
- Snyder, D. (1975). *Random point processes*. New York: Wiley.
- Snyder, D. L., & Miller, M. I. (1991). *Random point processes in time and space*. New York: Springer-Verlag.
- Solo, V. (2000). Unobserved Monte-Carlo method for identification of partially-observed nonlinear state-space systems, part II: counting process observations. In *Proc. IEEE Conference on Decision and Control*. New York: Institute of Electrical and Electronic Engineers.
- Tanner, M. A. (1996). *Tools for statistical inference*. New York: Springer-Verlag.
- Twum-Danso, N. T. (1997). *Estimation, information and neural signals*. Unpublished doctoral dissertation, Harvard University, Cambridge, MA.
- Twum-Danso, N. T., & Brockett, R. (2001). Trajectory information from place cell data. *Neural Networks*, 14(6–7), 835–844.
- West, M., & Harrison, J. (1997). *Bayesian forecasting and dynamic models*. New York: Springer-Verlag.
- Wilson, M. A., & McNaughton, B. L. (1993). Dynamics of the hippocampal ensemble code for space. *Science*, 261, 1055–1058.
- Wirth, S., Chiu, C., Sharma, V., Frank, L. M., Smith, A. C., Brown, E. N., & Suzuki, W. A. (2002). *Medial temporal lobe activity during the acquisition of new object-place associations*. Program 676.1, 2002 Abstract Viewer/Itinerary Planner, Society for Neuroscience. Available on-line: <http://sfn.scholarone.com>.
- Wood, E. R., Dudchenko, P. A., & Eichenbaum, H. (1999). The global record of memory in hippocampal neuronal activity. *Nature*, 397(6720), 613–616.
- Yanike, M., Wirth, S., Frank, L. M., Smith, A. C., Brown, E. N., & Suzuki, W. A. (2002). *Categories of learning-related cells in the primate medial temporal lobe*. Program 676.2, 2002 Abstract Viewer/Itinerary Planner, Society for Neuroscience. Available on-line: <http://sfn.scholarone.com>.

A Point Process Framework for Relating Neural Spiking Activity to Spiking History, Neural Ensemble, and Extrinsic Covariate Effects

Wilson Truccolo,¹ Uri T. Eden,^{2,3} Matthew R. Fellows,¹ John P. Donoghue,¹ and Emery N. Brown^{2,3}

¹Neuroscience Department, Brown University, Providence, Rhode Island; ²Neuroscience Statistics Research Laboratory, Department of Anesthesia and Critical Care, Massachusetts General Hospital, Boston; and ³Division of Health Sciences and Technology, Harvard Medical School/Massachusetts Institute of Technology, Cambridge, Massachusetts

Submitted 8 July 2004; accepted in final form 23 August 2004

Truccolo, Wilson, Uri T. Eden, Matthew R. Fellows, John P. Donoghue, and Emery N. Brown. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J Neurophysiol* 93: 1074–1089, 2005. First published September 8, 2004; doi:10.1152/jn.00697.2004. Multiple factors simultaneously affect the spiking activity of individual neurons. Determining the effects and relative importance of these factors is a challenging problem in neurophysiology. We propose a statistical framework based on the point process likelihood function to relate a neuron's spiking probability to three typical covariates: the neuron's own spiking history, concurrent ensemble activity, and extrinsic covariates such as stimuli or behavior. The framework uses parametric models of the conditional intensity function to define a neuron's spiking probability in terms of the covariates. The discrete time likelihood function for point processes is used to carry out model fitting and model analysis. We show that, by modeling the logarithm of the conditional intensity function as a linear combination of functions of the covariates, the discrete time point process likelihood function is readily analyzed in the generalized linear model (GLM) framework. We illustrate our approach for both GLM and non-GLM likelihood functions using simulated data and multivariate single-unit activity data simultaneously recorded from the motor cortex of a monkey performing a visuomotor pursuit-tracking task. The point process framework provides a flexible, computationally efficient approach for maximum likelihood estimation, goodness-of-fit assessment, residual analysis, model selection, and neural decoding. The framework thus allows for the formulation and analysis of point process models of neural spiking activity that readily capture the simultaneous effects of multiple covariates and enables the assessment of their relative importance.

INTRODUCTION

Understanding what makes a neuron spike is a challenging problem, whose solution is critical for deciphering the nature of computation in single cells and neural ensembles. Multiple factors simultaneously affect spiking activity of single neurons and thus assessing the effects and relative importance of each factor creates the challenge. Neural activity is often studied in relation to 3 types of covariates. First, spiking activity is associated with extrinsic covariates such as sensory stimuli and behavior. For example, the spiking activity of neurons in the rat hippocampus is associated with the animal's position in its environment, the theta rhythm, theta phase precession, and the animal's running velocity (Frank et al. 2002; Mehta et al. 1997, 2000; O'Keefe and Dostrovsky 1971; O'Keefe and Recce

1993). Retinal neurons respond to light intensity and light contrast, and V1 neurons are influenced by the spatiotemporal structure outside their classic receptive fields (Knierim and Vanessen 1992; Sillito et al. 1995; Vinje and Gallant 2000). The spiking activity of neurons in the arm region of the primary motor cortex (MI) is strongly associated with several covariates of motor behavior such as hand position, velocity, acceleration, and generated forces (Ashe and Georgopoulos 1994; Fu et al. 1995; Scott 2003). Second, the current spiking activity of a neuron is also related to its past activity, reflecting biophysical properties such as refractoriness and rebound excitation or inhibition (Hille 2001; Keat et al. 2001; Wilson 1999).

Third, current capabilities to record the simultaneous activity of multiple single neurons (Csicsvari et al. 2003; Donoghue 2002; Nicolelis et al. 2003; Wilson and McNaughton 1993) make it possible to study the extent to which spiking activity in a given neuron is related to concurrent ensemble spiking activity (Grammont and Riehle 1999, 2003; Hatsopoulos et al. 1998, 2003; Jackson et al. 2003; Maynard et al. 1999; Sanes and Truccolo 2003). Therefore, a statistical modeling framework that allows the analysis of the simultaneous effects of extrinsic covariates, spiking history, and concurrent neural ensemble activity would be highly desirable.

Current studies investigating the relation between spiking activity and these 3 covariate types have used primarily linear (reverse correlation) or nonlinear regression methods (e.g., Ashe and Georgopoulos 1994; Fu et al. 1995; Luczak et al. 2004). Although these methods have played an important role in characterizing the spiking properties in many neural systems, 3 important shortcomings have not been fully addressed. First, neural spike trains form a sequence of discrete events or point process time series (Brillinger 1988). Standard linear or nonlinear regression methods are designed for the analysis of continuous-valued data and not point process observations. To model spike trains with conventional regression methods the data are frequently smoothed or binned, a preprocessing step that can alter their stochastic structure and, as a consequence, the inferences made from their analysis. Second, although it is accepted that extrinsic covariates, spiking history, and neural ensemble activity affect neural spiking, current approaches make separate assessments of these effects, thereby making it difficult to establish their relative importance. Third, model goodness-of-fit assessments as well as the analysis of neural

Address for reprint requests and other correspondence: W. Truccolo, Neuroscience Department, Brown University, 190 Thayer St., Providence, RI (E-mail: Wilson_Truccolo@Brown.edu).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

ensemble representation based on decoding should be carried out using methods appropriate for the point process nature of neural spike trains.

To address these issues, we present a point process likelihood framework to analyze the simultaneous effects and relative importance of spiking history, neural ensemble, and extrinsic covariates. We show that this likelihood analysis can be efficiently conducted by representing the logarithm of the point process conditional intensity function in terms of linear combinations of general functions of the covariates and then using the discrete time point process likelihood function to fit the model to spike train data in the generalized linear model (GLM) framework. Because the discrete time point process likelihood function is general, we also show how it may be used to relate covariates to neural spike trains in a non-GLM model. We illustrate the methods in the analysis of a simulated data example and an example in which multiple single neurons are recorded from MI in a monkey performing a visuomotor pursuit-tracking task.

METHODS

In this section we present the statistical theory underlying our approach. First, we define the conditional intensity function for a point process. Second, we present a discrete time approximation to the continuous time point process likelihood function, expressed in terms of the conditional intensity function. Third, we show that when the logarithm of the conditional intensity is a linear combination of functions of the covariates, the discrete time point process likelihood function is equivalent to the likelihood of a GLM under a Poisson distribution and log link function. Alternatively, if the point process is represented as a conditionally independent Bernoulli process and the probability of the events is modeled by a logistic function, then the likelihood function is equivalent to the likelihood of a GLM under a Bernoulli distribution and a logistic link function. Fourth, we present several forms of conditional intensity models for representing spiking history, neural ensemble, and extrinsic covariate effects. Finally, we define our approach to maximum likelihood estimation, goodness-of-fit assessment, model comparison, residuals analysis, and decoding from point process observations by combining the GLM framework with analysis methods for point processes.

A *point process* is a set of discrete events that occur in continuous time. For a neural spike train this would be the set of individual spike times. Given an observation interval $(0, T]$, a sequence of J spike times $0 < u_1 < \dots < u_j < \dots < u_J \leq T$ constitutes a point process. Let $N(t)$ denote the number of spikes counted in the time interval $(0, t]$ for $t \in (0, T]$. We define a single realization of the point process during the time interval $(0, t]$ as $N_{0:t} = \{0 < u_1 < u_2 < \dots < u_j \leq t \cap N(t) = j\}$ for $j \leq J$.

Conditional intensity function

A stochastic neural point process can be completely characterized by its conditional intensity function $\lambda(t | H(t))$ (Daley and Vere-Jones 2003), defined as

$$\lambda(t | H(t)) = \lim_{\Delta \rightarrow 0} \frac{P[N(t + \Delta) - N(t) = 1 | H(t)]}{\Delta} \quad (1)$$

where $P[\cdot | \cdot]$ is a conditional probability and $H(t)$ includes the neuron's spiking history up to time t and other relevant covariates. The conditional intensity is a strictly positive function that gives a history-dependent generalization of the rate function of a Poisson process. From Eq. 1 we have that, for small Δ , $\lambda(t | H(t))\Delta$ gives approximately the neuron's spiking probability in the time interval $(t,$

$t + \Delta]$. Because defining the conditional intensity function completely defines the point process, to model the neural spike train in terms of a point process it suffices to define its conditional intensity function. We use parametric models to express the conditional intensity as a function of covariates of interest, therefore relating the neuron's spiking probability to the covariates. We use $\lambda(t | \theta, H(t))$ to denote the parametric representation of the conditional intensity function in Eq. 1, where θ denotes an unknown parameter to be estimated. The dimension of θ depends on the form of the model used to define the conditional intensity function.

A discrete time representation of the point process will facilitate the definition of the point process likelihood function and the construction of our estimation algorithms. To obtain this representation, we choose a large integer K and partition the observation interval $(0, T]$ into K subintervals $(t_{k-1}, t_k]_{k=1}^K$ each of length $\Delta = TK^{-1}$. We choose large K so that there is at most one spike per subinterval. The discrete time versions of the continuous time variables are now denoted as $N_k = N(t_k)$, $N_{1:k} = N_{0:t_k}$, and $H_k = H(t_k)$. Because we chose large K , the differences $\Delta N_k = N_k - N_{k-1}$ define the spike train as a binary time series of zeros and ones. In discrete time, the parametric form of the conditional intensity function becomes $\lambda(t_k | \theta, H_k)$.

Point process likelihood and GLM framework

Because of its several optimality properties, we choose a likelihood approach (Pawitan 2001) for fitting and analyzing the parametric models of the conditional intensity function. As in all likelihood analyses, the likelihood function for a continuous time point process is formulated by deriving the joint probability density of the spike train, which is the joint probability density of the J spike times $0 < u_1 < u_2 < \dots < u_J \leq T$ in $(0, T]$. For any point process model satisfying Eq. 1, this probability density can be expressed in terms of the conditional intensity function (Daley and Vere-Jones 2003). Similarly, in the discrete time representation, this joint probability density can be expressed in terms of the joint probability mass function of the discretized spike train (see APPENDIX Eqs. A1 and A2) and is expressed here as a product of conditionally independent Bernoulli events (Andersen et al. 1992; Berman and Turner 1992; Brillinger 1988; Brown et al. 2003)

$$P(N_{1:K} | \theta) = \prod_{k=1}^K [\lambda(t_k | \theta, H_k)\Delta]^{N_k} [1 - \lambda(t_k | \theta, H_k)\Delta]^{1-N_k} + o(\Delta^J) \quad (2)$$

where the term $o(\Delta^J)$ relates to the probability of observing a spike train with 2 or more spikes in any subinterval $(t_{k-1}, t_k]$. From Eqs. A3–A5 in the APPENDIX, it follows that Eq. 2 can be reexpressed as

$$P(N_{1:K} | \theta) = \exp \left\{ \sum_{k=1}^K \log [\lambda(t_k | \theta, H_k)\Delta] N_k - \sum_{k=1}^K \lambda(t_k | \theta, H_k)\Delta \right\} + o(\Delta^J) \quad (3)$$

If we view Eq. 3 as a function of θ , given the spike train observations $N_{1:K}$, then this probability mass function defines our discrete time point process likelihood function and we denote it as $L(\theta | H_K) = P(N_{1:K} | \theta)$. From Eq. A6 in the APPENDIX, it can be seen that Eq. 3 is a discrete time approximation to the joint probability density of a continuous time point process.

To develop a computationally tractable and efficient approach to estimating θ we note that for any subinterval $(t_{k-1}, t_k]$, the conditional intensity function is approximately constant so that, by Eq. 3, $P(\Delta N_k) = \exp\{\log [\lambda(t_k | \theta, H_k)\Delta] N_k - \lambda(t_k | \theta, H_k)\Delta\}$ is given by the Poisson probability mass function. Because Δ is small, this is equivalent to the Bernoulli probability $P(\Delta N_k) = [\lambda(t_k | \theta, H_k)\Delta]^{N_k} [1 - \lambda(t_k | \theta, H_k)\Delta]^{1-N_k}$ in Eq. 2. If we now express the logarithm of the conditional intensity function as a linear combination of general functions of the covariates

$$\log \lambda(t_k | \theta, H_k) = \sum_{i=1}^q \theta_{g_i} [v_i(t_k + \tau)] \quad (4)$$

where g_i is a general function of a covariate $v_i(t_k)$ at different time lags τ , and q is the dimension of the estimated parameter θ , then Eq. 3 has the same form as the likelihood function for a GLM under a Poisson probability model and a log link function (see APPENDIX, Eqs. A7–A8). Thus, maximum likelihood estimation of model parameters and likelihood analyses can be carried out using the Poisson–GLM framework. Alternatively, if we extend the results in Brillinger (1988), we obtain

$$\log \{[1 - \lambda(t_k | \theta, H_k)\Delta]^{-1} [\lambda(t_k | \theta, H_k)\Delta]\} = \sum_{i=1}^q \theta_{g_i} [v_i(t_k + \tau)] \quad (5)$$

then Eq. 2 has the same form as the likelihood function for a GLM under a Bernoulli probability distribution and a logistic link function (Eqs. A9 and A10). Thus, maximum likelihood estimation of model parameters and likelihood analyses can also be carried out using the Bernoulli–GLM framework (see also Kass and Ventura 2001). In other words, for Δ sufficiently small (i.e., at most one spike per time subinterval), likelihood analyses performed with either the Bernoulli or the Poisson model are equivalent. However, because we are interested in modeling the conditional intensity function directly, instead of the probability of events in our discrete time likelihoods, we used the Poisson–GLM framework in our analyses.

Therefore, we can take advantage of the computational efficiency and robustness of the GLM framework together with all of the analysis tools from the point process theory: goodness-of-fit based on the time rescaling theorem, residual analysis, model selection, and stochastic decoding based on point process observations. We refer to this combined framework as the *point process–GLM framework*. This framework covers a very large class of models because Eq. 4 allows for general functions of the covariates and of interaction terms consisting of combinations of the covariates. An application of GLM analysis to spike train data, without the support of the derived relations between the point process and GLM likelihood functions, would remain purely heuristic in nature.

Finally, Eqs. 2 and 3 are generally applicable discrete time approximations for the point process likelihood function. Thus, when a parametric model of the conditional intensity function cannot be expressed in terms of either Eq. 4 or Eq. 5, the GLM framework may be replaced with standard algorithms for computing maximum likelihood estimates (Pawitan 2001).

Models for the conditional intensity function

We formulate specific models for the conditional intensity function that incorporate the effects of spiking history, ensemble, and extrinsic covariates. For the exposition in the remainder of this section, we extend our notation to include the neural ensemble activity. Consider an observation time interval $t \in (0, T]$ with corresponding sequences of J^c spike times $0 < u_1^c < \dots < u_{J^c}^c < \dots < u_{J^c}^c \leq T$, for $c = 1, \dots, C$ recorded neurons. Let $N_{1:K}^{1:C} = \bigcup_{c=1}^C N_{1:K}^c$ denote the sample path for the entire ensemble.

CONDITIONAL INTENSITY MODELS IN THE POINT PROCESS–GLM FRAMEWORK. The general form for the conditional intensity function we use to model a single cell's spiking activity is

$$\lambda(t_k | N_{1:K}^{1:C}, \mathbf{x}_{k+\tau}, \theta) = \lambda_f(t_k | N_{1:K}, \theta_f) \lambda_E(t_k | N_{1:K}^{1:C}, \theta_E) \lambda_X(t_k | \mathbf{x}_{k+\tau}, \theta_X) \quad (6)$$

where $\theta = \{\theta_X, \theta_E, \theta_f\}$, $\lambda_f(t_k | N_{1:K}, \theta_f)$ is the component of the intensity function conditioned on the spiking history $N_{1:K}$ of the neuron whose intensity is being modeled, $\lambda_E(t_k | N_{1:K}^{1:C}, \theta_E)$ is the component related to the ensemble history contribution, and $\lambda_X(t_k |$

$\mathbf{x}_{k+\tau}, \theta_X)$ is the component related to an extrinsic covariate $\mathbf{x}_{k+\tau}$, where τ is an integer time shift. Note that the term H_k , used in the previous section, is now replaced by more specific information according to the model.

We consider the following specific models for each of these 3 covariate types. We begin with a model incorporating the spiking history component.

The spiking history component is modeled as

$$\lambda_f(t_k | N_{1:K}, \theta_f) = \exp \left\{ \gamma_0 + \sum_{n=1}^Q \gamma_n \Delta N_{k-n} \right\} \quad (7)$$

where Q is the order of the autoregressive process, γ_n represents the autoregressive coefficients, and γ_0 relates to a background level of activity. This model is henceforth referred to as the autoregressive spiking history model. We apply Akaike's standard information criterion (AIC, see Eq. 16 below) to estimate the parameter Q . We expect this autoregressive spiking history model to capture mostly refractory effects, recovery periods, and oscillatory properties of the neuron.

The contributions from the ensemble are expressed in terms of a regression model of order R

$$\lambda_E(t_k | N_{1:K}^{1:C}, \theta_E) = \exp \left\{ \beta_0 + \sum_{c=1}^R \sum_{r=1}^R \beta_r^c \Delta N_{k-r}^c \right\} \quad (8)$$

where the first summation is over the ensemble of cells with the exception of the cell whose conditional intensity function is being modeled. Thus the above model contains $R \times (C - 1)$ parameters plus one additional parameter for the background level. Note that the coefficients in the ensemble model capture spike effects at 1-ms time resolution and in this way they may reflect lagged-synchrony between spikes of the modeled cell and other cells in the ensemble. Alternatively, to investigate ensemble effects at lower time precision, we consider the ensemble rates model

$$\lambda_E(t_k | N_{1:K}^{1:C}, \theta_E) = \exp \left\{ \beta_0 + \sum_{c=1}^R \sum_{r=1}^R \beta_r^c (N_{k-(r-1)W}^c - N_{k-rW}^c) \right\} \quad (9)$$

where the term $N_{k-(r-1)W}^c - N_{k-rW}^c$ is the spike count in a time window of length W covering the time interval $(t_{k-rW}, t_{k-(r-1)W}]$. The coefficients in this model may reflect spike covariances on slow time scales.

In our application to MI data, the extrinsic covariate $\mathbf{x}_{k+\tau}$ will specify the hand velocity. To model this component we employ a variation of the Moran and Schwartz (1999) model, henceforth referred to as the velocity model

$$\lambda_X(t_k | \mathbf{x}_{k+\tau}, \theta_X) = \exp \{ \alpha_0 + |V_{k+\tau}| [\alpha_1 \cos(\phi_{k+\tau}) + \alpha_2 \sin(\phi_{k+\tau})] \} \quad (10)$$

where $|V_{k+\tau}|$ and $\phi_{k+\tau}$ are, respectively, the magnitude and angle of the 2-D hand velocity vector in polar coordinates at time $t_{k+\tau}$. In this model $\mathbf{x}_{k+\tau} = [|V_{k+\tau}|, \phi_{k+\tau}]'$. For illustration purposes, we have considered only a single, fixed-time shift τ in the above model. Based on previous results (Paninski et al. 2004) we set $\tau = 150$ ms. A much more generic model form including linear or nonlinear functions of covariates at many different time lags could be easily formulated.

The most complex conditional intensity function models we investigate are the autoregressive spiking history plus velocity and ensemble activity, and the autoregressive spiking history plus velocity and ensemble rates models. For the former, the full conditional intensity function model is given by

$$\lambda(t_k | N_{1:k}^{1:C}, \mathbf{x}_{k+\tau}, \theta) = \exp\left\{\mu + \sum_{n=1}^Q \gamma_n \Delta N_{k-n} + \sum_c \sum_{r=1}^R \beta_r^c \Delta N_{k-r}^c + |V_{k+\tau}| [\alpha_1 \cos(\phi_{k+\tau}) + \alpha_2 \sin(\phi_{k+\tau})]\right\} \quad (11)$$

where μ relates to the background activity.

It should be noticed that although these models are in the “generalized linear” model class, the relation between the conditional intensity function and spiking history, ensemble, and extrinsic covariates can be highly nonlinear. These models are linear only after the transformation of the natural parameter (here the conditional intensity function) by the log link function and only with respect to the model parameters being estimated. As seen in Eq. 4, general functions (e.g., quadratic, cubic, etc.) of the actual measured covariates can be used.

NON-GLM CONDITIONAL INTENSITY FUNCTION MODEL. To illustrate the generality of the proposed point process framework, we construct and analyze a non-GLM conditional intensity function model that also incorporates effects of spiking history, neural ensemble, and extrinsic covariates. Additionally, this example demonstrates a procedure for obtaining a conditional intensity function by first modeling the interspike interval (ISI) conditional probability density function. The conditional intensity is obtained from the ISI probability density model using the relation (Brown et al. 2003)

$$\lambda(t_k | \theta, H_k) = \frac{p(t_e | \theta, H_k)}{1 - \int_{u_{N_{k-1}}}^{t_k} p[t | \theta, H(t)] dt} \quad (12)$$

where $t_e = t_k - u_{N_{k-1}}$ is the elapsed time since the most recent spike of the modeled cell before time t_k and $p(t_e | \theta, H_k)$ is the ISI probability density, specified here by the inhomogeneous inverse Gaussian (Barbieri et al. 2001). This probability density is given in Eq. A11 in the APPENDIX. This density is specified by a time-varying scaling parameter $s(t_k | \cdot)$ that, in our application to MI spiking data, captures the velocity and ensemble rates covariate effects

$$s(t_k | \mathbf{x}_{k+\tau}, N_{1:k}^{1:C}, \theta_X, \theta_E) = \exp\left\{\mu + \sum_c \sum_{r=1}^R \beta_r^c (N_{k-(r-1)W}^c - N_{k-rW}^c) + |V_{k+\tau}| [\alpha_1 \cos(\phi_{k+\tau}) + \alpha_2 \sin(\phi_{k+\tau})]\right\} \quad (13)$$

and a location parameter ψ . The set of parameters defining the inhomogeneous inverse Gaussian density and therefore the conditional intensity function is denoted $\theta = \{\theta_X, \theta_E, \psi\}$. This model (Eqs. 12, 13, and A11) is henceforth referred to as the inhomogeneous inverse Gaussian plus velocity and ensemble rates model. The history dependence in this model extends back to the time of the previous, most recent spike.

Maximum likelihood parameter estimation

Maximum likelihood parameter estimates for the models in the point process–GLM framework were efficiently computed using the iterative reweighted least squares (IRLS) algorithm. This method is the standard choice for the maximum likelihood estimation of GLMs because of its computational simplicity, efficiency, and robustness. IRLS applies the Newton–Raphson method to the reweighted least squares problem (McCullagh and Nelder 1989). Given the conditional intensity model in Eq. 4, the log-likelihood function is strictly concave. Therefore, if the maximum log-likelihood exists, it is unique (Santner and Duffy 1989). Confidence intervals and p -values were obtained following standard

computations based on the observed Fisher information matrix (Pawitan 2001). Statistically nonsignificant parameters (e.g. $P \geq 0.001$) were set to zero for all of the models. In the non-GLM case, the inhomogeneous inverse Gaussian model was fit by direct maximization of the likelihood function using a quasi-Newton method (IMSL, C function *min_uncon_multivar*, from Visual Numerics, 2001). For the data sets used here, the most intensive computations involved operations on large matrices of size about $10^6 \times 200$. Algorithms were coded in C and run on dual-processor 3.9-GHz IBM machines, 2 GB of RAM memory. Standard GLM estimation using IRLS is also available in several statistical packages (S-Plus, SPSS, and Matlab Statistics toolbox).

Goodness-of-fit, point process residual analyses and model comparison

KOLMOGOROV–SMIRNOV (K-S) TEST ON TIME RESCALED ISIS. Before making an inference from a statistical model, it is crucial to assess the extent to which the model describes the data. Measuring quantitatively the agreement between a proposed model and a spike train data series is a more challenging problem than for models of continuous-valued processes. Standard distance measures applied in continuous data analyses, such as average sum of squared errors, are not designed for point process data. One alternative solution to this problem is to apply the time-rescaling theorem (Brown et al. 2002; Ogata 1988; Papangelou 1972) to transform point processes like spike trains into continuous measures appropriate for goodness-of-fit assessment. Once a conditional intensity function model has been fit to a spike train data series, we can compute rescaled times z_j from the estimated conditional intensity and from the spike times as

$$z_j = 1 - \exp\left\{-\int_{u_j}^{u_{j+1}} \lambda(t | H(t), \hat{\theta}) dt\right\} \quad (14)$$

for $j = 1, \dots, J-1$, where $\hat{\theta}$ is the maximum likelihood estimate of the parameters. The z_j values will be independent uniformly distributed random variables on the interval $[0, 1]$ if and only if the conditional intensity function model corresponds to the true conditional intensity of the process. Because the transformation in Eq. 14 is one to one, any statistical assessment that measures the agreement between the z_j values and a uniform distribution directly evaluates how well the original model agrees with the spike train data. To construct the K-S test, we order the z_j values from smallest to largest, denoting the ordered values as $z_{(j)}$, and then plot the values of the cumulative distribution function of the uniform density function defined as $b_j = (j - 1/2)/J$ for $j = 1, \dots, J$ against the $z_{(j)}$. We term these plots *K-S plots*. If the model is correct, then the points should lie on a 45° line. Confidence bounds for the degree of agreement between a model and the data may be constructed using the distribution of the Kolmogorov–Smirnov statistic (Johnson and Kotz 1970). For moderate to large sample sizes the 95% confidence bounds are well approximated by $b_j \pm 1.36 \cdot J^{-1/2}$ (Johnson and Kotz 1970).

To assess how well a model performs in terms of the original ISIs ($ISI_j = u_j - u_{j-1}$), we relate the ISIs to the computed z_j values in the following manner. First, the empirical probability density of the z_j values is computed, and the ratio of the empirical to the expected (uniform) density is calculated for each bin in the density. Second, the ISI values in the data are rounded to integer milliseconds and collected into bins. For these ISIs, all the corresponding z_j values as well as the ratios of empirical to expected densities in the related bins are obtained. This correspondence between ISIs and z_j values is easily available from Eq. 14. Third, we compute the mean ratio R (i.e., the mean of all the ratios for this particular ISI value). A mean ratio $R > 1$ ($R < 1$) implies that there are more (less) rescaled ISIs of length z_j than expected and that the intensity is being underestimated (overestimated), on average, at this particular ISI value.

If the model is correct, the z_j values should be not only uniformly distributed, but also independent. Thus, even when the K-S statistic is small, we still need to show that the rescaled times are independent. Here we assess independence up to 2nd-order temporal correlations by computing the autocorrelation function of the transformed rescaled times. As a visualization aid, we plot z_{j+1} against z_j .

POINT PROCESS RESIDUAL ANALYSIS. A standard approach in goodness-of-fit analysis is to examine structure in the data that is not described by the model. For continuous valued data, this is done by analyzing the residual error (i.e., the difference between the true and predicted values). For point process data, a different definition of residuals is needed to relate the conditional intensity function to the observed spike train data. The point process residual (Andersen et al. 1992) over nonoverlapping moving time windows is defined as

$$M(t_k) = \sum_{i=k-B}^k \Delta N_i - \int_{t_{k-B}}^{t_k} \lambda(t | H(t), \hat{\theta}) dt \quad (15)$$

for $k - B \geq 1$. In the application to MI data, we will look for relations between the point process residual and motor covariates (e.g., speed or direction) by computing their cross-correlation functions. Existence of correlations would imply that there is some structure left in the residuals that is not captured by the conditional intensity function model.

MODEL SELECTION. An additional tool for comparing models comes from the statistical theory of model selection (Burnham and Anderson 2002). The idea consists of choosing the best models to approximate an underlying process generating the observed data, a process whose complexity can be potentially infinite dimensional. To achieve this goal, we adopt Akaike's standard information criterion (AIC) (Akaike 1973). This criterion also provides a way to rank different candidate models. The AIC was originally derived as an estimate of the expected relative Kullback–Leibler distance (Cover and Thomas 1991) between a distribution given by an approximating model and the distribution of the true underlying process generating the data. This criterion is formulated as

$$AIC(q) = -2 \log L(\hat{\theta} | H_K) + 2q \quad (16)$$

where $L(\hat{\theta} | H_K)$ is the likelihood function, $L(\hat{\theta} | H_K) = P(N_{1:K} | \hat{\theta}, H_K)$; $\hat{\theta}$ is the maximum likelihood estimate of the model parameters θ ; and q is the total number of parameters in the model. By this criterion, the best model is the one with the smallest AIC, implying that the approximate distance between this model and the “true process” generating the data is smallest. The AIC is frequently interpreted as a measure of the trade-off between how well the model fits the data and the number of parameters required to achieve that fit, or of the desired trade-off between bias and variance (Burnham and Anderson 2002). An equivalence between AIC and cross-validation for the purpose of model selection has been established (Stone 1977). AIC can be applied to both nested and nonnested models, and to models with different distributions in their stochastic component. We compute AIC values for different models to guide our model comparison. Specifically, we provide the difference between the AIC of all of the models with respect to the AIC of the best model. We also use the AIC to estimate the order of the autoregressive spiking history component in Eq. 7.

Neural decoding analysis by state estimation with point process observations

Beyond assessing the goodness-of-fit of a single cell model with respect to its individual spike train data, we also analyze the ability of the model, over the entire cell population, to decode an m -dimensional extrinsic covariate $\mathbf{x}_{k+\tau}$. Such decoding will use the spike times of the entire ensemble of cells and the corresponding conditional intensity

function for each of these cells. We thus perform a state estimation of \mathbf{x}_k based on point process observations and thereby assess the ensemble coding properties of the cell population. The estimated extrinsic covariate will be given by the posterior mode after a Gaussian approximation to the Bayes–Chapman–Kolmogorov system (Eden et al. 2004).

For the particular type of hand kinematics data described above, we model \mathbf{x}_k as a Gaussian autoregressive process of order 1, henceforth AR(1), given by

$$\mathbf{x}_{k+\tau} = \mu_{\mathbf{x}} + \mathbf{F}\mathbf{x}_{k+\tau-1} + \varepsilon_{k+\tau} \quad (17)$$

where $\mu_{\mathbf{x}}$ is an m -dimensional vector of mean parameters, \mathbf{F} is an $m \times m$ state matrix, and ε_k is the noise term given by a zero mean m -dimensional white Gaussian vector with $m \times m$ covariance matrix \mathbf{W}_{ε} . The matrices \mathbf{F} and \mathbf{W}_{ε} are fitted by maximum likelihood.

The point process observation equation is expressed in terms of the modeled conditional intensity functions $\lambda^c(t_k | \cdot)$ for each of the C cells entering the decoding. As an example, for intensity functions conditioned on a motor covariate $\mathbf{x}_{k+\tau}$ and intrinsic spiking history $N_{1:k}^C$, we have the following recursive point process filter.

One step prediction

$$\mathbf{x}_{k+\tau|k+\tau-1} = \mu_{\mathbf{x}} + \mathbf{F}\mathbf{x}_{k+\tau-1|k+\tau-1} \quad (18)$$

One-step prediction covariance

$$\mathbf{W}_{k+\tau|k+\tau-1} = \mathbf{F}\mathbf{W}_{k+\tau-1|k+\tau-1}\mathbf{F} + \mathbf{W}_{\varepsilon} \quad (19)$$

Posterior covariance

$$\begin{aligned} \mathbf{W}_{k+\tau|k+\tau} = & \left[\mathbf{W}_{k+\tau|k+\tau-1}^{-1} + \sum_{c=1}^C [\nabla \log \lambda^c(t_k | N_{1:k}^c, \mathbf{x}_{k+\tau|k+\tau-1}, \hat{\theta}^c)] \right. \\ & \lambda^c(t_k | N_{1:k}^c, \mathbf{x}_{k+\tau|k+\tau-1}, \hat{\theta}^c) \Delta [\nabla \log \lambda^c(t_k | N_{1:k}^c, \mathbf{x}_{k+\tau|k+\tau-1}, \hat{\theta}^c)]' \\ & \left. - \sum_{c=1}^C \nabla^2 \log \lambda^c(t_k | N_{1:k}^c, \mathbf{x}_{k+\tau|k+\tau-1}, \hat{\theta}^c) \right. \\ & \left. [\Delta N_{1:k}^c - \lambda^c(t_k | N_{1:k}^c, \mathbf{x}_{k+\tau|k+\tau-1}, \hat{\theta}^c) \Delta] \right]^{-1} \quad (20) \end{aligned}$$

Posterior mode

$$\begin{aligned} \mathbf{x}_{k+\tau|k+\tau} = & \mathbf{x}_{k+\tau|k+\tau-1} + \mathbf{W}_{k+\tau|k+\tau} \\ & \times \sum_{c=1}^C \nabla \log \lambda^c(t_k | N_{1:k}^c, \mathbf{x}_{k+\tau|k+\tau-1}, \hat{\theta}^c) [\Delta N_{1:k}^c - \lambda^c(t_k | N_{1:k}^c, \mathbf{x}_{k+\tau|k+\tau-1}, \hat{\theta}^c) \Delta] \quad (21) \end{aligned}$$

The term $\nabla(\nabla^2)$ denotes the m -dimensional vector ($m \times m$ matrix) of first (second) partial derivatives with respect to $\mathbf{x}_{k+\tau}$, and $\mathbf{W}_{k+\tau|k+\tau}$ is the posterior covariance matrix of $\mathbf{x}_{k+\tau}$. Similarly, decoding equations based on other models of the conditional intensity function can be obtained. The derivation of the recursive point process filter is based on the well-established (Mendel 1995; Kitagawa and Gersh 1996) relation between the posterior probability density and the Chapman–Kolmogorov (one-step prediction) probability density, and on a Gaussian approximation of the posterior density (for details see Eden et al. 2004). The Gaussian approximation results from a 2nd-order Taylor expansion of the density and it is a standard first approach for approximating probability densities (Tanner 1996; Pawitan 2001). Nonetheless, the spiking activity enters into the computations in a very non-Gaussian way through the point process model instantiated by the conditional intensity function.

The amount of uncertainty in the algorithm about the true state of the decoded parameter is related to the matrix $\mathbf{W}_{k+\tau|k+\tau}$. Confidence

regions and coverage probability for the decoding can thus be obtained as follows. At time $k\Delta t$ an approximate 0.95 confidence region for the true covariate $\mathbf{x}_{k+\tau}$ may be constructed as

$$(\mathbf{x}_{k+\tau} - \mathbf{x}_{k+\tau|k+\tau})' \mathbf{W}_{k+\tau|k+\tau}^{-1} (\mathbf{x}_{k+\tau} - \mathbf{x}_{k+\tau|k+\tau}) \leq \chi_{0.95}^2(m) \quad (22)$$

for $k = 1, 2, \dots, K$, where $\chi_{0.95}^2(m)$ gives the 0.95 quantile of the χ^2 distribution with degrees of freedom equal to the dimension m of the covariate. The coverage probability up to time t_k is given by s_k/k where s_k is the number of times the true covariate is within the confidence regions during the time interval $(0, k\Delta]$. In the decoding analysis we compute the mean of the coverage probability over the entire decoding period. A Monte Carlo simulation is employed to obtain the confidence intervals and coverage probability for the covariate in polar coordinates. We first use the estimated posterior covariance matrix to generate 10^4 Gaussian-distributed samples centered at the current covariate estimates in Cartesian coordinates. Second, these random samples are converted to polar coordinates. Finally, the confidence intervals are then computed from the distribution of the random samples in polar coordinates.

RESULTS

The proposed point process framework is illustrated with 2 examples. The first one is applied to simulated neural spike data and the second to multiple single units simultaneously recorded from monkey primary motor cortex. For the discrete time representation of the neural point process we set $\Delta = 1$ ms.

Simulation study

The goal of the simulation study is 2-fold. First, we illustrate the main properties of the model in Eq. 11 containing the autoregressive history, neural ensemble history, and motor covariate effects. Second, we demonstrate that the parameters of the simulated model are accurately recovered from relatively small spike data sets by maximum likelihood estimation implemented with the IRLS algorithm.

The conditional intensity functions of 6 neurons (A, B, C, D, E, F) were simulated using methods as described in Ogata (1981). The intensity of 5 of them (B–F) was given by the velocity model (Eq. 10); that is, the neurons were modeled as inhomogeneous Poisson processes with mean background spiking rates of 17, 16, 9, 8, and 7 Hz, respectively, and inhomogeneity introduced by the modulating hand velocity signal. Different velocity tuning functions were used for the set of cells. The hand velocity signal was sampled from actual hand trajectories performed by a monkey (see *Application to MI spiking data*, below). The conditional intensity function for neuron A was given by the autoregressive spiking history plus ensemble and velocity model (Eq. 11). The background mean firing rate of this neuron was set to 10 Hz. The autoregressive spiking history component contained 120 coefficients covering 120 ms of spiking history (see Fig. 2B). The autoregressive coefficients mimicked the effects of refractory–recovery periods and rebound excitation. From the ensemble of 5 neurons, only 2 contributed excitatory (neuron B) and inhibitory (neuron C) effects at 3 time lags (-1 , -2 , and -3 ms).

The simulation scheme worked as follows. Starting with the initial simulation time step, first the conditional intensity functions were updated and then, at the same time step, the spiking activities for all of the cells were simulated. The simulation then moved to the next time step. The conditional intensity

functions were updated based on the past intrinsic and ensemble spiking history (neuron A only) and on the current hand velocity state (all neurons).

The main features of the conditional intensity function model in Eq. 11 can be observed in Fig. 1, where the simulated conditional intensity function of neuron A and its own spiking activity are plotted together with the activity of the other 5 neurons and the contribution of velocity signal. The simulated conditional intensity function clearly shows the dependence on spike history: after a spike, the intensity drops to almost zero and slowly recovers, reaching a period of higher than background spiking probability at about 20 ms after the spike. Fast excitatory and inhibitory effects follow the spikes of neurons B and C. Spiking history, neural ensemble, and velocity modulate each other's contributions in a multiplicative fashion.

From the simulated ensemble spike trains and from the velocity signal, we then estimated the conditional intensity function generating the spiking activity of neuron A. The data set entering the estimation algorithm thus consisted of 6 simulated spike trains, each 200 s long, and of the hand velocity time series in polar coordinates. The spike train for the modeled neuron A contained 2,867 spikes. The parametric model for the estimated conditional intensity function consisted of a background mean, 120 autoregressive coefficients, and 5 regressive coefficients for each of the other 5 neurons (i.e., 25 coefficients in total). Each set of 5 coefficients related to spiking activity at lags -1 , -2 , \dots , -5 ms. The IRLS

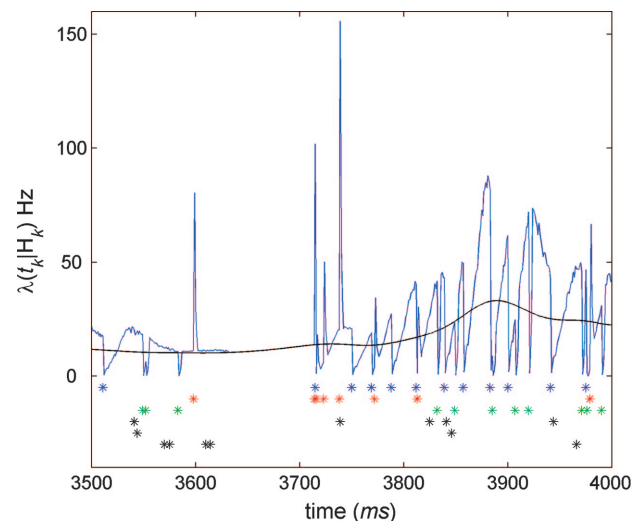


FIG. 1. Simulated conditional intensity function model. Conditional intensity function, modeled as in Eq. 1, was simulated and used to generate a spike train (neuron A, blue asterisks mark the times of the spike events). In this model, the intensity (blue curve) was conditioned on the past spiking history, the spikes of 2 other neurons (neuron B, excitatory, red asterisks; neuron C, inhibitory, green asterisks), and on hand velocity. Past spiking history effect was modeled by a 120-order autoregressive process carrying a refractory period, recovery, and rebound excitation. Coefficient values were based on parameters estimated from a primary motor cortex (MI) cell (see Fig. 5B). The conditional intensity function resulting from the contribution of only hand velocity is shown by the black line. Three other cells were also simulated (neurons D, E, and F; black asterisks). Neurons B–F were modeled as inhomogeneous Poisson processes modulated according to the velocity model (Eq. 10). All cells had different preferred movement directions. Spiking history, ensemble, and velocity modulated each other in a multiplicative fashion. Simulated ensemble spike trains together with hand velocity were used to estimate the parameters for the conditional intensity function model of neuron A (see Fig. 2).

algorithm converged in 12 iterations (tolerance was set to 10^{-6}). Statistically nonsignificant parameters ($P > 0.001$) were set to zero (see METHODS section). The true model parameters used in the simulation of neuron A were accurately recovered (Fig. 2, *B* and *C*), with the estimated model passing the K-S goodness-of-fit test (Fig. 2*D*). Parameter estimation on smaller data sets (about 50 s of data) led to similar successful fittings.

Application to MI spiking data

Experimental data were obtained from the MI area of a behaving monkey. Details of the basic recording hardware and protocols are available elsewhere (Donoghue et al. 1998; Maynard et al. 1999). After task training, a Bionic Technologies LLC (BTL, Salt Lake City, UT) 100-electrode silicon array was implanted in the area of MI corresponding to the arm representation. One monkey (*M. mulatta*) was operantly conditioned to track a smoothly and randomly moving visual target. The monkey viewed a computer monitor and gripped a 2-link, low-friction manipulandum that constrained hand movement to a horizontal plane. The hand (x, y) position signal was digitized and resampled to 1 kHz. Low-pass-filtered finite

differences of position data were used to obtain hand velocities. Some 130 trials (8–9 s each) were recorded. More details about the statistical properties of the distributions for hand position and velocity, spiking sorting methods, and other task details can be found in Paninski et al. (2004).

Models including 1, 2, or all of the 3 types of covariates were analyzed. To start, we focus on the analysis of the velocity and the autoregressive spiking history plus velocity models. Later, we also compare these 2 models using neural decoding based on the observation of the entire ensemble of cells. For this reason, we analyzed these 2 models for each of the 20 cells in the ensemble. More detailed analysis involving K-S plots, point process residuals, and AIC model comparison will be illustrated for one typical cell.

K-S GOODNESS-OF-FIT ANALYSIS FOR THE VELOCITY AND THE AUTOREGRESSIVE SPIKING HISTORY PLUS VELOCITY MODELS. The tuning functions obtained from the velocity model (Eq. 10) are shown in Fig. 3. This model was statistically significant for all of the cells. Preferred direction was diverse across cells, covering the range of possible directions. The corresponding K-S plots are shown in Fig. 4. The quantiles refer to the $z_{(j)}$ s (Eq. 14) and the cumulative distribution function (CDF) refers to the expected uniform distribution for the case when the estimated conditional intensity model was equivalent to the true one. The velocity model tends to overestimate (underestimate) the conditional intensity at lower (middle) quantiles. Introduction of the autoregressive spiking history component (Eq. 7) in the velocity model greatly improved the explanation of the spiking process, almost completely eliminating both the over- and underestimation of the intensity. The maximum order of the autoregressive component was about 120 (i.e., the component incorporated history effects spanning over 120 ms in the past). The most significant history effects extended to 60 ms in the past. For the majority of the cells, this component seemed to capture mostly 3 main history effects: refractory and recovery periods followed by an increase in the firing probability around 20 ms after a spike (see Fig. 5*B*). It should be noticed that the autoregressive coefficients could have also reflected dynamical network properties of nonmeasured neural ensembles such as networks of excitatory and inhibitory neurons where the modeled cell is embedded, or nonmeasured fast extrinsic covariates. No significant differences in the K-S plots were observed between a pure autoregressive history model and the autoregressive history plus velocity models (not shown).

Figure 5, *C* and *D* summarize the above observations for a typical cell (cell 75a, 29,971 spikes over 130 trials, i.e., ~1,040 s) in this example set and relate the fitting problems of the velocity model to the original nontime rescaled ISIs. In the velocity model, the intensity is overestimated (mean ratio $R < 1$) for ISIs below 10 ms and underestimated (mean ratio $R > 1$) for ISIs in the interval 10 to about 40 ms (Fig. 5*D*). The overestimation is likely a reflection of a refractory–recovery period (up to ~10 ms) after the cell has spiked, which is not captured by the velocity model. The underestimation reflects a period of increased firing probability that follows the recovery period. These 2 different regimes are reasonably well captured by the coefficients of the autoregressive component (see Fig. 5*B*), thus resulting in the improved fit observed for the autoregressive spiking history plus velocity model. Introduction of

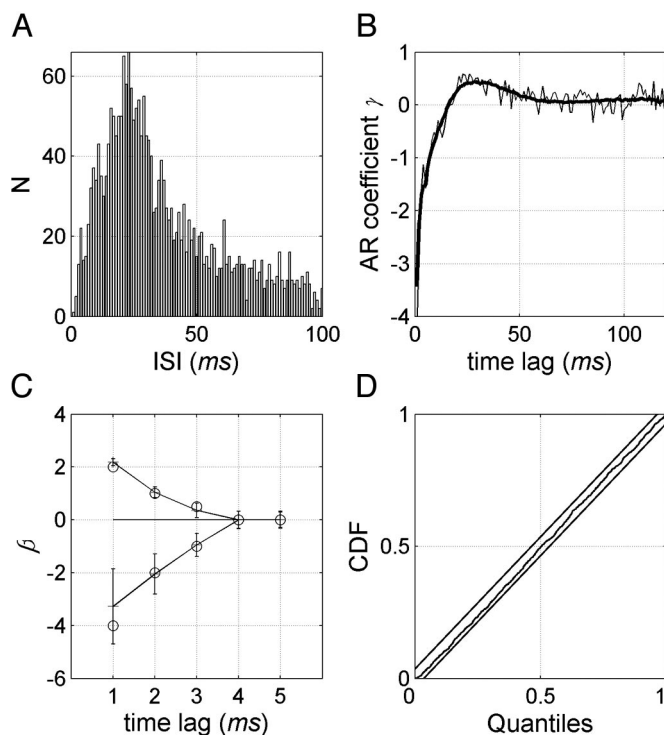


FIG. 2. Model parameter estimation by iteratively reweighted least squares (IRLS) in the generalized linear model (GLM) framework. Spike trains of the 6 simulated cells (each lasting 200 s; see Fig. 1) together with hand velocity produced the data set for the estimation of the parameters of the conditional intensity function model for neuron A. Spike train of neuron A consisted of 2,867 spikes. *A*: ISI distribution for neuron A. *B*: true autoregressive coefficients (thick curve) and the estimated ones. *C*: ensemble covariate in the estimated model contained 5 coefficients per cell (covering 5 ms of the past). Only the 3 coefficients for the excitatory and inhibitory cells were significantly different from zero ($P < 0.001$). Bars indicate the 95% confidence intervals. Small circles show the location of the true coefficients. *D*: Kolmogorov–Smirnov (K-S) goodness-of-fit test shows that the estimated conditional intensity model passed the test (the 95% confidence region is given by the parallel lines). Coefficients for the velocity covariate $\alpha_1 = 0.1$ and $\alpha_2 = -0.05$ were estimated as $\hat{\alpha}_1 = 0.105$ and $\hat{\alpha}_2 = -0.045$.

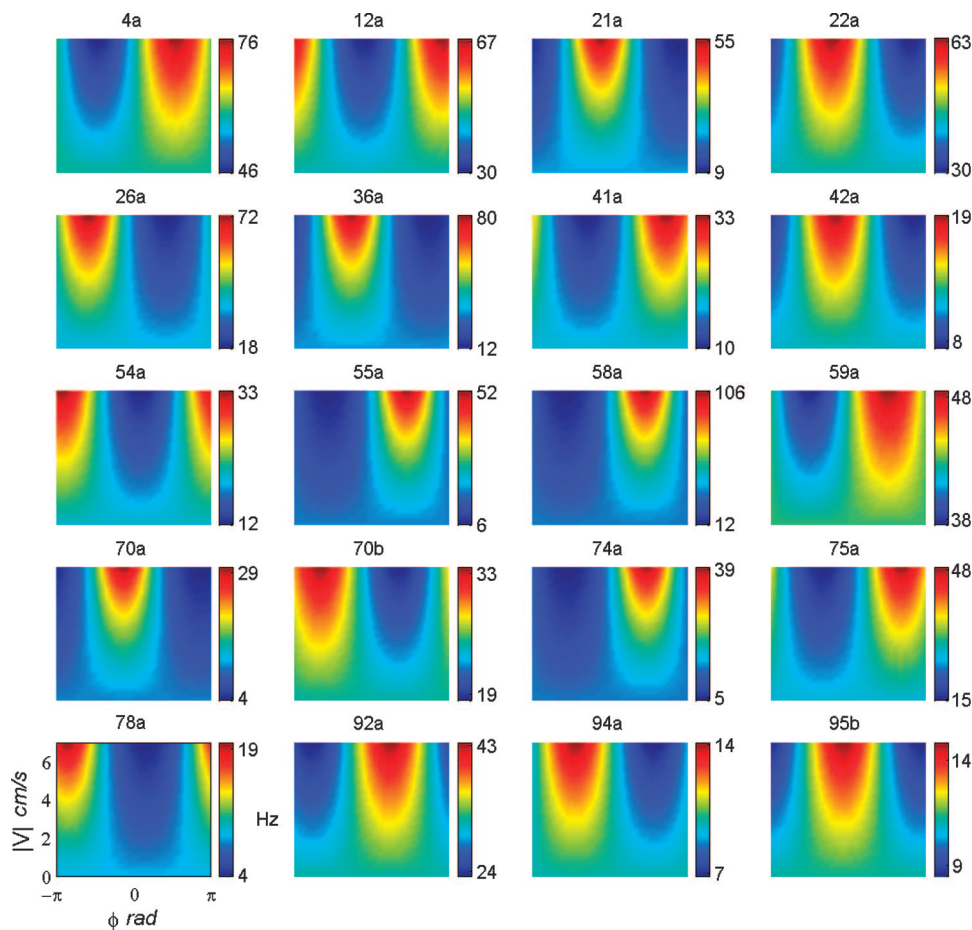


FIG. 3. Velocity tuning functions. Conditional intensity function values, based on the velocity model, are expressed by pseudo-color maps. Velocity is given in polar coordinates, with ϕ representing the movement direction. Each subplot relates to a particular cell, with cells' labels given at the top.

the autoregressive component makes the observed density for the z_j values much closer to the expected uniform density (Fig. 5C).

The K-S statistic measures how close rescaled times are to being uniformly distributed on $[0, 1]$. In addition, a good model should also generate independent and identically distributed rescaled times. To illustrate this point, we checked for temporal correlations at lag 1 in the time-rescaled ISIs (Fig. 6). As expected, some temporal structure remains in the case of the velocity model ($r^2 = 0.25$, $P = 10^{-6}$), whereas this structure is effectively insignificant for the velocity plus autoregressive spiking history ($r^2 = 0.002$, $P = 10^{-6}$). The cross-correlation function computed over a broad range of lags was consistent with this result.

POINT PROCESS RESIDUAL ANALYSIS. Even though the parameters for the velocity model were statistically significant, the K-S plot analysis showed that the velocity model fell short of explaining the entire statistical structure in the observed single-cell spiking activity. It thus remains to be seen how well this model captured the relationship between hand velocity and spiking activity. Besides neural decoding, another approach to address this problem is to measure the correlations among the point process residuals as defined in Eq. 15 and the movement velocity. Existence of correlations would imply that there is some structure left in the residuals that is not captured by the velocity model. On the other hand, a decrease in the correlation level with respect to some other model would imply that the

velocity model does capture some of the structure in the spiking activity related to hand velocity.

We computed the correlations for the residuals from the autoregressive spiking history model (Eq. 7) and compared them to the correlations for the residuals from the velocity and from the autoregressive spiking history plus velocity model. Residuals were computed for nonoverlapping 200-ms moving windows (Fig. 7). Cross-correlation functions were computed between the residuals and the mean of the kinematic variables. Mean (x, y) velocities were computed for each time window and were used to obtain, in polar coordinates, the respective mean movement speed and direction. In the autoregressive model case, peak cross-correlation values between the residuals and direction, speed, and velocities in x and y coordinates were 0.29, 0.10, -0.17 , and 0.50, respectively. For the autoregressive spiking history and velocity model, the peak cross-correlation values for the same variables were 0.08, 0.06, -0.12 , and 0.28. This suggests that, for this particular neuron, the velocity model captures a significant amount of information about hand velocity available in the spiking activity. Nonetheless, it is also clear that there is a residual structure in the spiking activity that is statistically related to the hand velocity in Cartesian coordinates and that is not captured by the autoregressive spiking history plus velocity model. Furthermore, the cross-correlation functions for both the velocity and the autoregressive spiking history plus velocity model show no significant differences, which suggests that the autoregressive

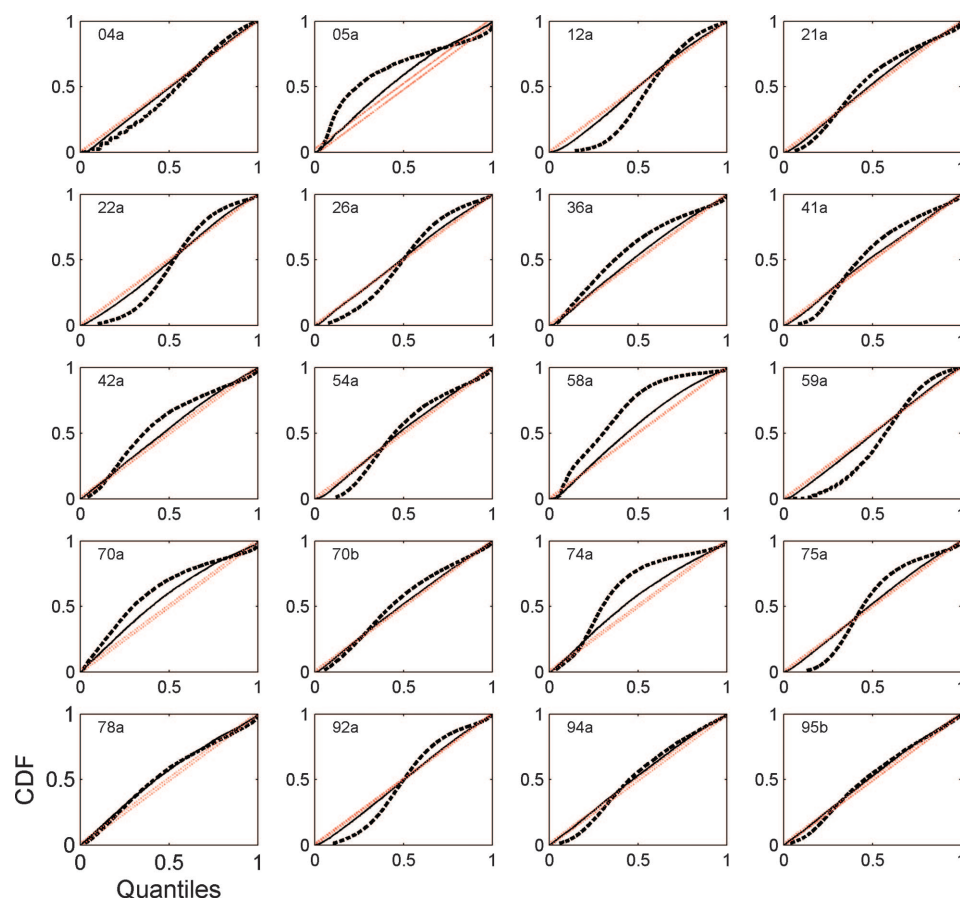


FIG. 4. K-S plots for the velocity model and the autoregressive spiking history plus velocity model. Because the K-S plots are constructed from a uniform distribution on the interval $[0, 1]$, the 50th and 100th percentiles correspond, respectively, to quantiles 0.5 and 1.0 on both the horizontal and vertical axes. Two-sided 95% confidence error bounds of the K-S statistics are displayed for each cell (45° red lines). Visual inspection alone already reveals that, for most of the cells, the autoregressive spiking history plus velocity model (solid curve) improves the fit considerably.

component does not carry additional statistical information about hand velocity.

MODEL COMPARISON. We compared partial and complete (i.e., 1, 2, or 3 covariate types) conditional intensity function models for cell 75a. The ensemble model included spiking activity of each cell at 4 time lags (-1 , -2 , -3 , and -4 ms). The ensemble rates included the spike counts of each cell at 3 nonoverlapping and lagged time windows. The length of each of the time windows, specified by the parameter W in Eq. 9, was 50 ms. The K-S plots in Fig. 8 reveal that the autoregressive spiking history plus velocity and the autoregressive spiking history plus velocity and ensemble rates models provided the best fits among all of the models for this specific cell. The inhomogeneous inverse Gaussian plus velocity and ensemble rates model performed better than the velocity, ensemble, and ensemble rates models. Inspection of the coefficients for the ensemble and ensemble rates models showed that the dependencies were statistically significant for many of the cells in the ensemble. Individual cells contributed either positive or negative effects to the conditional intensity function and the effective ensemble contribution to the modulation of the conditional intensity function could reach tens of hertz.

In the above K-S plot comparisons, some of the models had K-S statistics far from the 95% confidence intervals or nearly identical to those from other models, making a clear comparison difficult. The AIC analysis was then used to provide a more detailed comparison, as well as to take the complexity of the model (i.e., number of parameters) into consideration in the model comparison. Figure 9 shows the ranked models in terms

of their difference with respect to the AIC of the best model. In this context, models with lower AIC difference values are considered better models.

Overall, this criterion provided a fine model ranking and suggested that models containing the autoregressive spiking history component performed better in each instance. Among the alternative models for spiking history, the autoregressive spiking history model performed better than the conditional intensity model based on the inhomogeneous inverse Gaussian ISI distribution model (Eqs. 12, 13, and A11), both in the AIC and K-S goodness-of-fit analyses. Also, the ensemble rates model did better than models containing only the velocity covariate or the ensemble covariate at fine temporal precision.

VELOCITY AND MOVEMENT DIRECTION DECODING ANALYSIS. The velocity (Eq. 10) and the autoregressive spiking history plus velocity models were used in the neural decoding of hand velocity. Models were fit to a training data set (120 trials, about 8–9 s each) and applied to decoding on a different test data set (10 trials, again about 8–9 s each). The state matrix \mathbf{F} for the AR(1) state process (Eq. 17) was estimated to be diagonal with nonzero terms approximately equal to 0.99, and the noise covariance matrix \mathbf{W}_ε to be diagonal with nonzero entries equal to 0.01. Figure 10 shows the resulting decoding of movement direction and, in Cartesian coordinates, the estimated (x, y) velocities for a single test trial based on the velocity model. Overall, decoding of movement direction was remarkably good. Decoded (x, y) velocities captured mostly slower fluctuations. To compare the decoding performance of the 2 models, we computed the coverage probability and the

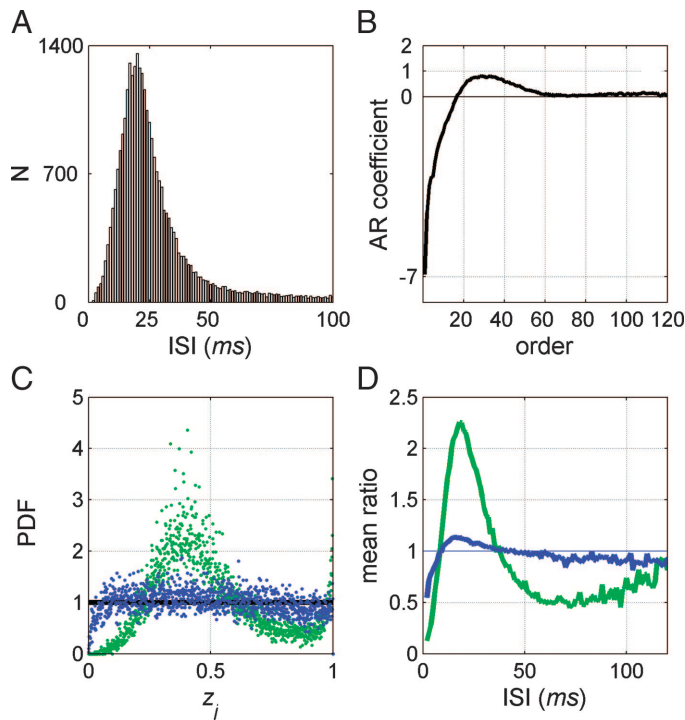


FIG. 5. Contribution of the autoregressive spiking history component. Cell 75a is chosen to illustrate how the addition of the autoregressive component improves the model's fit. *A*: ISI histogram. *B*: estimated coefficients of the autoregressive component. Autoregressive component incorporates a recovery period after the cell spikes, which lasts for about 18 ms (negative coefficients). Cell's firing probability then starts to increase, peaking at about 25 ms after a spike. *Order* refers to the order of the AR coefficient representing increasing times since the last spike. *C*: histogram for the transformed times z_i for both models (green: velocity model; blue: autoregressive spiking history plus velocity model). Black line shows the expected uniform distribution for the case where the estimated intensity function is close enough to the true intensity function underlying the neural point process. *D*: mean ratio of observed to expected z_i values indicates that the velocity model overestimates, on average, the intensity function for periods up to about 10 ms after a spike, while it tends to underestimate the intensity for periods between 10 and 40 ms. Introduction of the negative (positive) autoregressive coefficients almost completely eliminates the over (under) estimation of the conditional intensity function based on the velocity model alone.

decoding error. Table 1 gives the mean values (across time and trials) for the coverage probabilities of the bivariate estimate (velocity magnitude and movement direction) and the coverage probabilities of the univariate estimate (velocity magnitude or movement direction). Mean coverage probability for the movement direction estimate was 0.94 for the velocity model. For the same model, coverage probabilities for the bivariate estimate and velocity magnitude were much smaller, consistent with the observation that the estimated velocities in Cartesian coordinates captured mostly slow fluctuations. Mean coverage probability, mean and median decoding errors, and confidence intervals for the decoding errors were not significantly different between the 2 models.

DISCUSSION

An important problem in neurophysiology is determining the factors that affect a neuron's spiking behavior. To address this question we have presented a point process statistical framework that allowed us to characterize simultaneously the effects of several covariates on the spiking activity of an individual

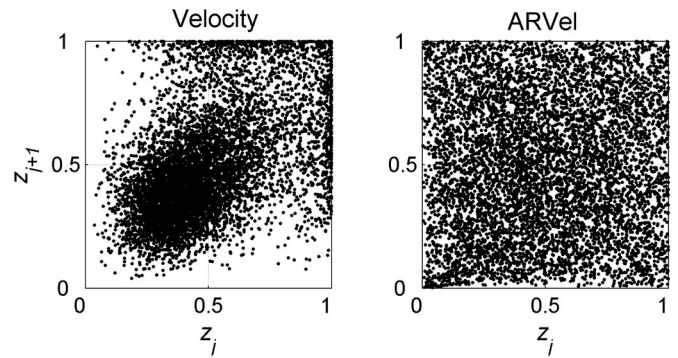


FIG. 6. Temporal correlations in the time-rescaled ISIs. Scatter plots are shown for consecutive z_i values from the velocity and autoregressive spiking history plus velocity (ARVel) models applied to cell 75a. Clearly, the autoregressive spiking history plus velocity model presents a more independent rescaled distribution. Corresponding correlation coefficients are 0.25 ($P = 10^{-6}$) for the velocity model and 0.002 ($P = 10^{-6}$) for the autoregressive spiking history plus velocity model. Cross-correlation functions computed over a broad range of lags led to similar results. Thus, in addition to improving the fit in the K-S plots, the introduction of the autoregressive component also eliminates temporal correlations among the rescaled times observed for the velocity model.

neuron. The 3 types of covariates we considered were the neuron's spiking history, past neural ensemble activity, and extrinsic covariates such as stimuli or behavior. Because de-

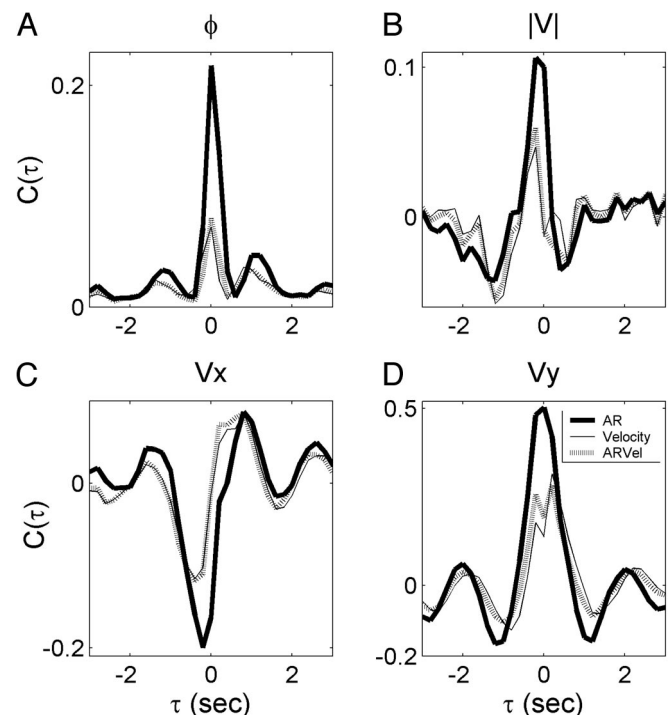


FIG. 7. Point process residual analysis. *A*: cross-correlation function $C(\tau)$ between the hand-movement direction and the residuals from the autoregressive spiking history (AR, thick curve), the velocity (thin curve), and the autoregressive spiking history plus velocity (ARVel, dashed curve) models applied to cell 75a. *B–D*: cross-correlations functions between the residuals and speed, and velocity in Cartesian coordinates (V_x and V_y). Correlations are significantly reduced for the velocity model in comparison to the autoregressive spiking history model. Nonetheless, there remains some structure in the point process residual that is related to the hand velocity but was not captured by the velocity model. Correlations for the velocity model were practically identical to the autoregressive spiking history plus velocity model, suggesting that the autoregressive component does not provide additional information about velocity (see text for details).

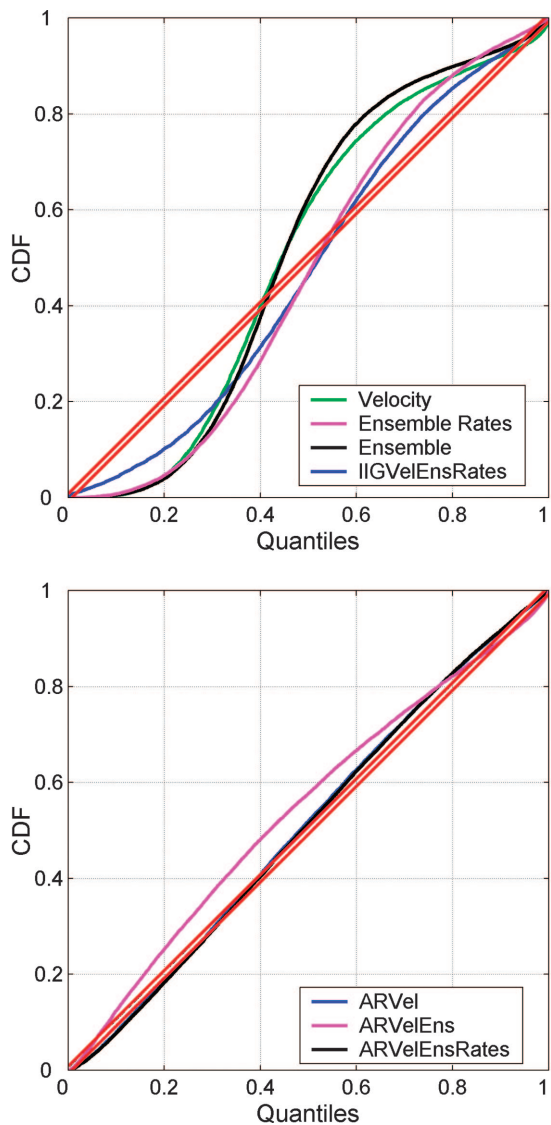


FIG. 8. Goodness-of-fit assessment (K-S plots) of alternative models. For comparison purposes, the models are shown in 2 groups. *Top*: the velocity, ensemble, ensemble rates, and the inhomogeneous inverse Gaussian plus velocity and ensemble rates models (IIGVelEnsRates) are compared. *Bottom*: the autoregressive spiking history plus velocity model (ARVel) is compared to 2 other models that add the ensemble (ARVelEns) or the ensemble rates component (ARVelEnsRates). K-S plots for the ARVel and the ARVelEnsRates partially overlap. (Cell 75a).

fining the conditional intensity function defines a point process model, the explanatory contributions of the covariates were assessed by constructing conditional intensity function models in terms of these covariates and by using a likelihood-based estimation approach to fit these models to data and to assess goodness-of-fit.

Analyses that measure the simultaneous effects of different covariates are crucial because the covariates modulate the neural spiking activity at the same time. Currently used analysis methods do not allow for the modeling of the simultaneous effects of spiking history, neural ensemble, and extrinsic covariates on the spike train treated as a point process. To evaluate the relation between spiking activity and covariates, the spike train data are frequently transformed to a rate function and the relation between the rate function and the covariate

is then assessed using regression methods (e.g., Ashe and Georgopoulos 1994; Luczak et al. 2004). The relation between a neuron's spiking activity, spiking history, and concurrent neural activity are usually assessed using autocorrelation and pairwise cross-correlation analyses performed directly on the spike train (e.g., Hatsopoulos et al. 1998). The use of different methods to assess individually the importance of these covariates precludes an analysis of the neural point process in which the relative importance of all covariates is assessed and may also lead to a misleading estimate of the covariate effects. For example, spiking history effects can interfere with the accurate estimation of extrinsic covariate effects in spike triggered averages and reverse correlation methods (Aguera y Arcas and Fairhall 2003; Aguera y Arcas et al. 2003). Additionally, current analysis techniques also assess the contribution of ensemble covariates separately. For instance, pairwise cross-correlation analyses measure the statistical association of a single neuron to each member of the ensemble separately but not the association between the single neuron's activity and the entire observed ensemble.

The key to our likelihood approach is representing the conditional intensity function of a single neuron in terms of the covariates. In this way, the covariates are directly related to the probability that the neuron spikes. Although this formulation can be used generally to analyze the relation between covariates and neural spiking activity, it usually requires writing a new algorithm or function to carry out the maximum likelihood

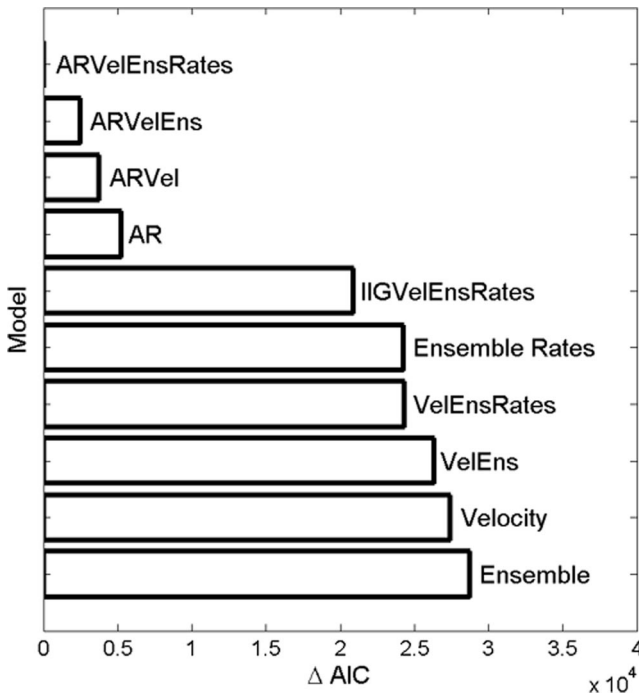


FIG. 9. Akaike's standard information criterion (AIC) model comparison. For convenience, we plot the differences of the AICs, denoted by ΔAIC of all of the models with respect to the AIC of the best model. Following this criterion and convention, better models have smaller AIC differences. Model labels: autoregressive spiking history (AR), autoregressive spiking history plus velocity (ARVel), autoregressive spiking history plus velocity and ensemble (ARVelEns), autoregressive spiking history plus velocity and ensemble rates (ARVelEnsRates), inhomogeneous inverse Gaussian plus velocity and ensemble rates models (IIGVelEnsRates), velocity plus ensemble (VelEns), and velocity plus ensemble rates (VelEnsRates). See METHODS section for model details. (Cell 75a).

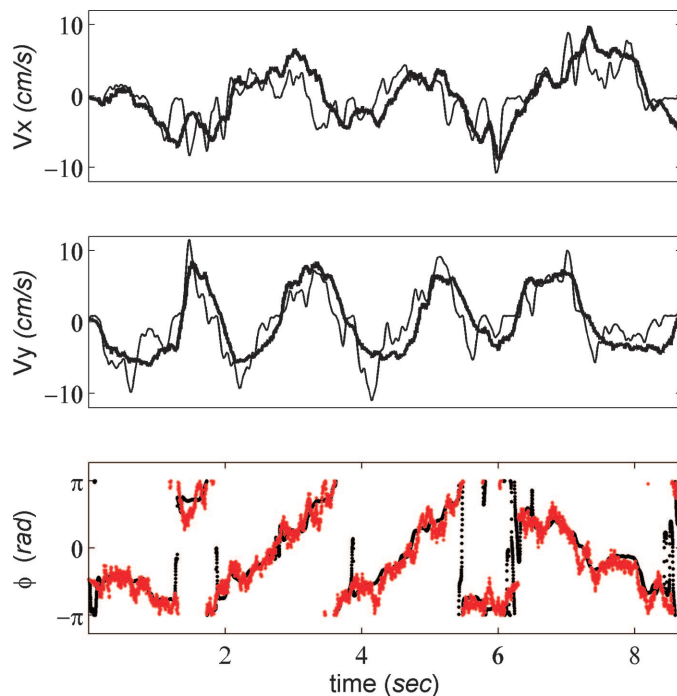


FIG. 10. Neural decoding of (x, y) velocities and movement direction by the point process filter. Estimated velocities (thick curve) and direction (red dots), together with true velocities and direction, are shown for a single decoded test trial. Time was discretized at a 1-ms resolution. At every millisecond, an estimate of the velocity parameters was obtained based on the state of the cells in the ensemble. All 20 recorded cells were used in the decoding. Conditional intensity function for each cell was given by the velocity model. Original decoding was done in polar coordinates. From a total of 130 trials, 120 trials were used for model fitting and 10 test trials for neural decoding. See Table 1 for summary statistics over the entire ensemble of test trials.

estimation with each new model formulation. To make our point process likelihood approach more broadly applicable, we showed that by representing the logarithm of the conditional intensity function as a linear combination of general functions of the covariates, the conditional intensity function model could be efficiently fit to neural spike train data using the GLM framework under a Poisson distribution and a log link function (Eqs. 3, A6, and A8). We also showed that, equivalently, if the point process is represented as a conditionally dependent Bernoulli process (Eqs. 2 and A2) and the probability of the events is modeled by a logistic function, then the point process model could also be fit using the GLM framework under a Bernoulli distribution and a logistic link function (Eqs. 5 and A10). Regarding the time discretization, we note that the 1-ms discretization interval we chose should not be interpreted as the absolute discretization interval for all applications of the point process–GLM framework. How fine the discretization interval should be will depend on the particular problem under study.

Our use of the Poisson distribution to fit point process models does not mean that we are assuming that our original data are Poisson. We are rather exploiting for computational purposes the fact that all point process likelihoods that admit a conditional intensity function, including those that are history dependent, have the same mathematical form given by Eq. A6. Our analysis makes explicit the relation between likelihood methods for point processes, conditional Bernoulli processes, and GLM model fitting with Poisson or Bernoulli distributions.

This relation provides a justification for using this GLM framework to analyze spike trains as point process data.

The point process–GLM framework for analyzing neural spike train data has several important advantages. First, this framework allows us to formulate complex models to relate the spiking activity to covariates. The fact that Eq. 4 is written in terms of general functions of the covariates provides the framework with a very large class of possible models. Second, the GLM framework is part of several standard mathematical and statistical packages (e.g., Splus, Matlab, SPSS), so that the approach is readily accessible to experimentalists analyzing their data. Even though likelihood-based methods are highly desirable because of several optimality properties, the biggest impediment to their widespread use is the lack of readily available software in which a flexible class of neural spike train models can be easily applied to data. The point process–GLM framework offers a practical, broadly applicable solution to the computational problem of fitting potentially complex point process models for neural spike trains by maximum likelihood.

Third, the point process–GLM framework makes it possible to apply a set of goodness-of-fit tools for point processes not available in the GLM. These are the point process residuals analysis, goodness-of-fit tests based on the time-rescaling theorem, and decoding from point process observations. Fourth, the model selection and goodness-of-fit methods available in GLM are extended to spike train data. Thus we have a set of complementary methods to assess the extent to which proposed models explain the structure in neural spike trains. Although we used the point process–GLM framework to carry out most of the analysis, we also illustrated with the conditional intensity function based on the inhomogeneous inverse Gaussian model how non-GLM point process likelihood models may be used to analyze neural spike train data. Finally, by analogy with the way in which linear regression methods are used to analyze the relation between a continuous dependent variable and a set of candidate explanatory variables, the ready availability of software to implement GLMs also makes it possible for neurophysiologists to quickly assess the relevance of a wide range of covariates before proceeding to construct more specific models that may require non-GLM algorithms to carry out the model fitting.

A key objective of the proposed framework is to provide tools for the assessment of the relative importance of the covariates on neural spiking activity. We showed how this objective is accomplished by analyzing the goodness-of-fit of

TABLE 1. Mean coverage probabilities, mean, and median errors for the velocity, and the autoregressive spiking history plus velocity models

	Velocity	(AR) History + Velocity
$\langle V , \phi \rangle$: Mean cov. prob.	0.30 ± 0.10	0.24 ± 0.08
$ V $: Mean cov. prob.	0.32 ± 0.10	0.27 ± 0.08
ϕ : Mean cov. prob.	0.94 ± 0.02	0.90 ± 0.04
(V_x, V_y) : Mean error	3.60 ± 0.3	3.60 ± 0.4
(V_x, V_y) : Median error	3.43	3.38
(V_x, V_y) : Error (95% CI)	(0, 7.8]	(0, 7.8]
ϕ : Mean error	$0.15\pi \pm 0.1$	$0.14\pi \pm 0.1$
ϕ : Median error	0.11π	0.11π
ϕ : Error (95% CI)	(0, 0.65π]	(0, 0.63π]

CI denotes the confidence interval. Velocity is given in cm/s.

each model we proposed. Because no method provides a complete assessment of goodness-of-fit, we used the standard statistical approach of using multiple complementary measures that evaluate different aspects of the model's agreement with the data. We applied 4 complementary types of goodness-of-fit techniques in our example data set: Kolmogorov–Smirnov tests based on the time-rescaling theorem, model selection by AIC, point process residual analysis, and neural spike train decoding based on point process observations.

First, the K-S analyses performed on the example data problems were useful in providing a sense of how close our models were to capturing the stochastic structure in the example data sets. The fact that some of our models captured most of the statistical structure in the spiking activity (Figs. 4 and 8) suggests that developing a parsimonious statistical model of MI activity is a realistic goal. The K-S plots also highlighted the importance of the spiking history. The fact that the coefficients had a structure that seemed to reflect mostly refractory and recovery periods together with rebound excitation in short time scales (effectively shorter than 60 ms) suggests that the autoregressive component successfully captured important history effects.

Second, the AIC analysis provided additional information for model comparison when the K-S plots did not distinguish between different models. By finely ranking the different models, the AIC analysis allowed for the assessment of the distinct effects and relative importance of the 3 types of covariates. Our choice of the AIC for model comparison was motivated by the fact that AIC was derived to be an estimate of the expected relative Kullback–Leibler distance between the distributions generated by the model and the distribution of the underlying stochastic process generating the data. The “true” underlying model does not need to be in the set of tested models for the AIC analysis to suggest the most appropriate model, a situation that we believe is more often the rule than the exception in biology. Because the AIC is penalized with increasing numbers of model parameters, its use is more appropriate than the use of the data likelihood itself in preventing overfitting of the data by the model. Additionally, an equivalence between AIC and cross-validation in model comparison problems has been established (Stone 1977), and model-based penalty methods (e.g., AIC) outperform cross-validation in important aspects when assessing prediction performance (Efron 2004). We also computed the Bayesian information criterion (BIC) (Schwarz 1978), an alternate, more conservative criterion function for each of our models and found that it yielded the same model rankings as did the AIC. When comparing complex models, another protection against overfitting comes from having large quantities of data as compared to the number of model parameters. In our example the $>10^6$ data observations far outnumbered the maximum 200 or so parameters in our most complex models. Apparent inconsistencies between AIC and K-S analyses could occur in cases where, for example, the conditional intensity model is more likely to produce the observed spike train as a whole, but is less accurate in describing a specific aspect of the data structure, such as the regime of small ISIs. This might have been the case when comparing the AIC and K-S plots results for the autoregressive spiking history plus velocity and ensemble model and a simpler autoregressive spiking history plus velocity model (Figs. 8 and 9).

Third, we illustrated how the point process residual analysis can be used to assess the contribution of an extrinsic covariate to a single neuron's spiking activity. In the illustration example, the residual analysis showed that the introduction of the velocity covariate captured a significant amount of the statistical structure related to hand velocity available in the spiking activity of a single neuron. Yet, for the particular cell chosen in this study, the analysis was also able to reveal that there still was a significant amount of structure in the residuals that was correlated to hand velocity but that was not captured by this specific form of the velocity model. Cross-correlation analysis of the point process residuals and extrinsic covariates is thus an important tool for assessing whether a particular model has captured well the effects of the covariate on the spiking activity. The ideal model should produce a residual with no significant correlations to the modeled covariate. It should also be noted that, unlike the decoding analysis, the point process residual analysis is not dependent on the properties of a decoding algorithm.

Fourth, complementing the above 3 goodness-of-fit analysis tools, the spike train decoding allowed for the goodness-of-fit assessment at the neural ensemble level. In conjunction with understanding what makes a neuron spike, we are also interested in assessing how well a model captures the representation of an extrinsic covariate at the ensemble level. At present, decoding is the only technique we have for assessing goodness-of-fit at this level. The key elements for assessing goodness-of-fit in the decoding analysis were the predicted signal and its confidence intervals and coverage probability, and especially the estimation error and its confidence intervals. The confidence intervals and the coverage probability based on the estimated posterior covariance matrix provided an estimate of the amount of uncertainty in the decoding algorithm, whereas the decoding error and its distribution provided a measure of the algorithm's actual performance. For this reason, the mean coverage probability should be interpreted in conjunction with the mean decoding error. As suggested by the narrow distribution of the decoding error and approximately 0.95 mean coverage probability (Table 1), hand movement direction was remarkably well decoded. Velocity estimates in Cartesian coordinates captured reasonably well slow fluctuations in the measured hand velocity. We also illustrated how to assess the contribution of the autoregressive spiking history component to neural decoding. The autoregressive spiking history plus velocity and the velocity models performed similarly well. This preliminary result suggests that short time dynamics captured by the autoregressive component did not play a crucial role in decoding hand velocity or movement direction in this data set. Given that the models used in the decoding analysis did not include the ensemble or ensemble rates covariate, the ensemble decoding assumed independent encoder cells. Nonetheless, the framework also allows the assessment of the contribution of interaction patterns in the neural ensemble to decoding. That could be easily achieved by extending the conditional intensity models to include the ensemble covariates. This analysis is beyond the scope of this paper and will be addressed elsewhere.

In summary, the above 4 complementary goodness-of-fit and model selection analyses are an essential step for achieving our primary objective of assessing the effects and relative importance of the modeled covariates. The proposed point process

framework provides a starting point for building and analyzing complex models of spiking activity including the 3 types of covariates discussed in this paper. In particular, the point process–GLM framework provides a systematic approach to neural spike train data analysis analogous to that for continuous-valued variables under the linear Gaussian regression framework.

We foresee several potential improvements and extensions of this framework. Although we were able to fit models with hundreds of parameters, much larger models will require the development of efficient algorithms for both GLM and non-GLM computations. Also, to analyze the relation between the simultaneous firing of an entire ensemble relative to its spiking history and a set of extrinsic covariates, we can extend the framework by using multivariate point process likelihoods (Chornoboy et al. 1988). A multivariate likelihood model will facilitate the study of the independence, redundancy, and synergy in the ensemble representation. Finally, multielectrode devices (Csicsvari et al. 2003; Donoghue 2002; Nicolelis et al. 2003) now make possible the simultaneous recordings of multiple single cells from many different brain areas. We foresee the proposed framework as a valuable tool for investigating how interacting brain regions represent, compute, and process information. We are currently applying this framework to the analysis of parietal and MI spiking activity in monkeys performing visuomotor tracking tasks and hippocampus activity in rats performing a range of learning tasks.

APPENDIX

Continuous and discrete time point process likelihood function

The likelihood of a neural spike train, like that of any statistical model, is defined by finding the joint probability density of the data. We show below that the joint probability of any point process is easy to derive from the conditional intensity function. We show that the point process likelihood function in Eqs. 2 and 3 gives a discrete time approximation of the likelihood function for a continuous time point process (Eq. A6 below).

Let $0 < u_1 < u_2, \dots, u_J < T$ be a set of neural spike train measurements. Using the discrete time representation given in the METHODS section, define the events

$$\begin{aligned} A_k &= \{\text{spike in } (t_{k-1}, t_k] \mid H_k\} \\ E_k &= \{A_k\}^{\Delta N_k} \{A_k^c\}^{1-\Delta N_k} \\ H_k &= \left\{ \bigcap_{j=1}^{k-1} E_j \right\} \end{aligned} \quad (A1)$$

for $k = 1, \dots, K$ and where A_k^c is the complement of A_k . For simplicity, H_k includes only the intrinsic history of the process. It can be easily extended to incorporate neural ensemble activity and other extrinsic covariates. By construction of the partition of the interval $(0, T]$, introduced in the METHODS section, we must have $u_j \in (t_{k_j-1}, t_{k_j}]$, $j = 1, \dots, J$, for a subset of the intervals satisfying $k_1 < k_2 < \dots < k_J$. The remaining $K - J$ intervals have no spikes.

The probability of exactly J events occurring within the intervals $(t_{k_j-1}, t_{k_j}]_{j=1}^J$ in $(0, T]$, may then be computed as

$$\begin{aligned} P(N_{1:K}) &= P(u_j \in (t_{k_j-1}, t_{k_j}], j = 1, \dots, J, \cap N(T) = J) \\ &= \prod_{k=1}^K P(A_k)^{\Delta N_k} P(A_k^c)^{1-\Delta N_k} \end{aligned} \quad (A2)$$

by the definition of A_k and E_k in Eq. A1.

The spike train thus forms a sequence of conditionally independent Bernoulli trials, with the probability of a spike in the k^{th} time interval given by $P(A_k)$. In any interval $(t_{k-1}, t_k]$ we have

$$\begin{aligned} P(A_k) &= \lambda(t_k \mid H_k) \Delta + o(\Delta) \\ P(A_k^c) &= 1 - \lambda(t_k \mid H_k) \Delta + o(\Delta) \end{aligned} \quad (A3)$$

Substituting Eq. A3 into Eq. A2 yields

$$P(N_{1:K}) = \prod_{k=1}^K [\lambda(t_k \mid H_k) \Delta]^{\Delta N_k} [1 - \lambda(t_k \mid H_k) \Delta]^{1-\Delta N_k} + o(\Delta') \quad (A4)$$

which is Eq. 2. For small Δ , $[1 - \lambda(t_k) \Delta] \approx \exp\{-\lambda(t_k) \Delta\}$ and $\log [\lambda(t_k) \Delta [1 - \lambda(t_k) \Delta]^{-1}] \approx \log (\lambda(t_k) \Delta)$, therefore we obtain

$$\begin{aligned} P(N_{1:K}) &= \prod_{k=1}^K [\lambda(t_k \mid H_k) \Delta]^{\Delta N_k} [1 - \lambda(t_k \mid H_k) \Delta]^{-\Delta N_k} \prod_{k=1}^K [1 - \lambda(t_k \mid H_k) \Delta] + o(\Delta') \\ &= \prod_{k=1}^K \left[\frac{\lambda(t_k \mid H_k) \Delta}{1 - \lambda(t_k \mid H_k) \Delta} \right]^{\Delta N_k} \prod_{k=1}^K \exp\{-\lambda(t_k \mid H_k) \Delta\} + o(\Delta') \\ &= \exp \left\{ \sum_{k=1}^K \log [\lambda(t_k \mid H_k) \Delta] \Delta N_k - \sum_{k=1}^K \lambda(t_k \mid H_k) \Delta \right\} + o(\Delta') \end{aligned} \quad (A5)$$

The probability density of these J exact spikes in $(0, T]$, given by $p(N_{0:T}) = \lim_{\Delta \rightarrow 0} P(N_{1:K})/\Delta^J$, is then obtained as

$$\begin{aligned} P(N_{0:T}) &= \lim_{\Delta \rightarrow 0} \frac{\exp \left\{ \sum_{k=1}^K \log [\lambda(t_k \mid H_k) \Delta] \Delta N_k - \sum_{k=1}^K \lambda(t_k \mid H_k) \Delta \right\} + o(\Delta')}{\Delta^J} \\ &= \lim_{\Delta \rightarrow 0} \frac{\exp \left\{ \sum_{k=1}^K \log \lambda(t_k \mid H_k) \Delta N_k - \sum_{k=1}^K \lambda(t_k \mid H_k) \Delta \right\} \Delta^J + o(\Delta')}{\Delta^J} \\ &= \exp \left\{ \int_0^T \log \lambda(t \mid H(t)) dN(t) - \int_0^T \lambda(t \mid H(t)) dt \right\} \end{aligned} \quad (A6)$$

which is the joint probability density of the point process spike train in continuous time (Brown et al. 2003; Daley and Vere-Jones 2003). Note that we could have derived the likelihood for the continuous time point process (and therefore also Eq. 3) by a generalization of the continuous time Poisson process (Daley and Vere-Jones 2003), without resorting to representing the neural point process as a conditional Bernoulli process. We formulated the spike train joint probability in terms of Eq. A2 only to show (see below) the equivalence between Poisson and Bernoulli–GLMs when Δ is sufficiently small.

The Poisson and Bernoulli–GLMs

We briefly define a generalized linear model and show that for small enough Δ , the Bernoulli and Poisson–GLMs are equivalent in the modeling of spiking train data.

Two main aspects characterize a generalized linear model of a random variable y (McCullagh and Nelder 1989). First, the modeled random variable y has a distribution in the exponential family. Among several members of this family are the Gaussian, the Poisson, and the Bernoulli distribution. The exponential family has the general form

$$f(y | \Theta, \phi) = \exp\{[y\Theta - b(\Theta)]/a(\phi) + c(y, \phi)\} \quad (A7)$$

where $a(\cdot)$, $b(\cdot)$, and $c(\cdot)$ are some specific functions. If ϕ is known, this is an exponential-family model with canonical parameter Θ .

For the particular case of the Poisson distribution, $\Theta = \log \lambda$, $b(\Theta) = e^\Theta$, $a(\phi) = \phi = 1$, $c(y, \phi) = -\log(y!)$. The location and scale parameters are λ and ϕ , respectively. Thus the distribution in Eq. A7 can be expressed as $f(y | \lambda) = \exp\{y \log \lambda - \lambda - \log(y!)\} = \lambda^y e^{-\lambda}/y!$. Note that the canonical parameter Θ has, in the Poisson case, a natural representation in terms of the logarithm of the parameter λ . The joint probability distribution for an independently distributed data set $\mathbf{y} = \{y_k\}_{k=1}^K$ becomes

$$f(\mathbf{y} | \Theta, \phi) \propto \exp\left\{\sum_{k=1}^K y_k \log \lambda_k - \sum_{k=1}^K \lambda_k\right\} \quad (A8)$$

If the rate function λ_k of this Poisson process is generalized by the conditional intensity function (Eq. 1); and $y_k = \Delta N_k$, $\Delta = 1$, then Eq. A8 has the same form as the typical general likelihood function for any discrete time point process (Eqs. 3 and A5).

For the Bernoulli case, we let $y = \{0, 1\}$ with the probability of success denoted by P , and set $\Theta = \log([1 - P]^{-1}P)$, $b(\Theta) = \log(1 + e^\Theta)$, $a(\phi) = \phi$, $c(\cdot) = 1$, and $\phi = 1$. Thus, for single realizations we have $p(y | P) = P^y(1 - P)^{1-y}$. Given an independently distributed data set $\mathbf{y} = \{y_k\}_{k=1}^K$, the likelihood function under the Bernoulli distribution becomes

$$f(\mathbf{y} | \Theta, \phi) = \prod_{k=1}^K [P_k]^{y_k} [1 - P_k]^{1-y_k} \quad (A9)$$

By letting $[P_k]^{y_k} = P(A_k)^{\Delta N_k}$ we obtain Eq. A2.

Second, the defining feature of a generalized linear model follows. The canonical parameter of the exponential family is expressed in a linear form by a transformation given by a monotonic differentiable function. In the Poisson case, if the canonical parameter is modeled as a linear combination of general functions of covariates \mathbf{v} of interest (that is, $\Theta = \log \lambda(\theta, \mathbf{v}) = \sum_{i=1}^q \theta_{ig_i}(\mathbf{v}_i)$ as in Eq. 4 or equivalently as $\lambda(\theta, \mathbf{v}) = \exp\{\sum_{i=1}^q \theta_{ig_i}(\mathbf{v}_i)\}$ as in Eqs. 5–10), then $f(y | \Theta, \phi) \propto \exp\{y \log \lambda(\theta, \mathbf{v}) - \lambda(\theta, \mathbf{v})\}$ gives the distribution for a GLM under a Poisson distribution and a log link function. In the Bernoulli case, if $\Theta = \log([1 - P(A_k | \theta, H_k)]^{-1}P(A_k | \theta, H_k))$ is modeled as linear combination of general functions of the covariates, then Eqs. 2 and A2 give the likelihood function for a GLM under a Bernoulli distribution and a logistic link function.

Finally, we establish the relation between the Poisson and Bernoulli–GLMs in the context of neural point process models. After making explicit the parametric model of the conditional intensity function, we have the probability of a spike event in the time interval $(t_{k-1}, t_k]$ given by $P(A_k | \theta, H_k) = \lambda(t_k | \theta, H_k)\Delta + o(\Delta)$. For small Δ

$$\log \left[\frac{P(A_k | \theta, H_k)}{1 - P(A_k | \theta, H_k)} \right] \approx \log [\lambda(t_k | \theta, H_k)\Delta] \quad (A10)$$

Therefore, for small enough Δ , the Bernoulli and Poisson–GLMs are equivalent when applied to the modeling of spiking train data.

The IIG model for the ISI probability density

For a particular cell, let $t_e = t_k - u_{N_{k-1}}$ denote the time elapsed since the last spike $u_{N_{k-1}}$. The inhomogeneous inverse Gaussian ISI probability density function conditioned on the motor covariate and neural ensemble activity is defined as

$$p(t_e | G) = \frac{s(t_e | \cdot)}{\left[2\pi \int_{u_{N_{k-1}}}^{t_k} s(t | \cdot) dt\right]^{3/2}} \exp \left\{ -\frac{1}{2} \frac{\left(\int_{u_{N_{k-1}}}^{t_k} s(t | \cdot) dt - \psi \right)^2}{\psi^2 \int_{u_{N_{k-1}}}^{t_k} s(t | \cdot) dt} \right\} \quad (A11)$$

where $G = \{u_{N_{k-1}}, \mathbf{x}_{k+\tau}, N_{1:k}^{1:C}, \theta\}$, $\psi > 0$ is the location parameter and $s(t_k | \cdot) > 0$ is the scaling parameter at time t_k , conditioned on the extrinsic and ensemble rates covariates as given in Eq. 13.

GRANTS

This work was supported in part by Burroughs Wellcome Fund, Keck Foundation, National Institute of Neurological Disorders and Stroke Neural Prosthesis Program, National Institutes of Health Grants NS-25074, MH-59733, MH-61637, MH-65018, and DA-015644, National Science Foundation Grant 0081548, and grants from Defense Advanced Research Projects Agency and Office of Naval Research.

DISCLOSURE

J. P. Donoghue is a cofounder and shareholder in Cybernetics, Inc., a neurotechnology company that is developing neural prosthetic devices.

REFERENCES

- Aguera y Arcas B, and Fairhall AL. What causes a neuron to spike? *Neural Comput* 15: 1789–1807, 2003.
- Aguera y Arcas B, Fairhall AL, and Bialek W. Computation in a single neuron: Hodgkin and Huxley revisited. *Neural Comput* 15: 1715–1749, 2003.
- Akaike H. Information theory as an extension of the maximum likelihood principle. In: *Second International Symposium on Information Theory*, edited by Petrov BN and Csaki F. Budapest: Akademiai Kiado, 1973, p. 267–281.
- Andersen PK, Borgan O, Gill RD, and Keiding N. *Statistical Models Based on Counting Processes*. New York: Springer-Verlag, 1992.
- Ashe J and Georgopoulos AP. Movement parameters and neural activity in motor cortex and area-5. *Cereb Cortex* 4: 590–600, 1994.
- Barbieri R, Quirk MC, Frank LM, Wilson MA, and Brown EN. Construction and analysis on non-Poisson stimulus-response models of neural spiking activity. *J Neurosci Methods* 105: 25–37, 2001.
- Berman M and Turner TR. Approximating point process likelihoods with GLIM. *Appl Stat* 41: 31–38, 1992.
- Brillinger DR. Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol Cybern* 59: 189–200, 1988.
- Brown EN, Barbieri R, Eden UT, and Frank LM. Likelihood methods for neural data analysis. In: *Computational Neuroscience: A Comprehensive Approach*, edited by Feng J. London: CRC, 2003.
- Brown EN, Barbieri R, Ventura V, Kass RE, and Frank LM. The time-rescaling theorem and its application to neural spike train data analysis. *Neural Comput* 14: 325–2346, 2002.
- Burnham KP and Anderson DR. *Model Selection and Multimodel Inference: A Practical Information Theoretic Approach* (2nd ed.). New York: Springer-Verlag, 2002.
- Chornoboy ES, Schramm LP, and Karr AF. Maximum likelihood identification of neural point process systems. *Biol Cybern* 59: 265–275, 1988.
- Cover TM and Thomas JA. *Elements of Information Theory*. New York: Wiley, 1991.
- Csicsvari J, Henze DA, Jamieson B, Harris KD, Sirota A, Bartho P, Wise KD, and Buzsaki G. Massively parallel recording of unit and local field potentials with silicon-based electrodes. *J Neurophysiol* 90: 1314–1323, 2003.
- Daley D and Vere-Jones D. *An Introduction to the Theory of Point Processes*. New York: Springer-Verlag, 2003.
- Donoghue JP. Connecting cortex to machines: recent advances in brain interfaces. *Nat Neurosci* 5: 1085–1088, 2002.
- Donoghue JP, Sanes JN, Hatsopoulos NG, and Gaal G. Neural discharge and local field potential oscillations in primate motor cortex during voluntary movements. *J Neurophysiol* 79: 159–173, 1998.
- Eden UT, Frank LM, Barbieri R, Solo V, and Brown EN. Dynamic analysis of neural encoding by point process adaptive filtering. *Neural Comput* 16: 971–998, 2004.

- Efron B.** The estimation of prediction error: Covariance penalties and cross-validation. *J Am Stat Assoc* 99: 619–632, 2004.
- Frank LM, Eden UT, Solo V, Wilson MA, and Brown EN.** Contrasting patterns of receptive field plasticity in the hippocampus and the entorhinal cortex: an adaptive filtering approach. *J Neurosci* 22: 3817–3830, 2002.
- Fu QG, Flament D, Coltz JD, and Ebner TJ.** Temporal encoding of movement kinematics in the discharge of primate primary motor and premotor neurons. *J Neurophysiol* 73: 836–854, 1995.
- Grammont F and Riehle A.** Precise spike synchronization in monkey motor cortex involved in preparation for movement. *Exp Brain Res* 128: 118–122, 1999.
- Grammont F and Riehle A.** Spike synchronization and firing rate in a population of motor cortical neurons in relation to movement direction and reaction time. *Biol Cybern* 88: 360–373, 2003.
- Hatsopoulos NG, Ojakangas CL, Paninski L, and Donoghue JP.** Information about movement direction obtained from synchronous activity of motor cortical neurons. *Proc Natl Acad Sci USA* 95: 15706–15711, 1998.
- Hatsopoulos NG, Paninski L, and Donoghue JP.** Sequential movement representations based on correlated neuronal activity. *Exp Brain Res* 149: 478–486, 2003.
- Hille B.** *Ion Channels of Excitable Membrane* (3rd ed.). Sunderland, MA: Sinauer Associates, 2001.
- Jackson A, Gee VJ, Baker SN, and Lemon RN.** Synchrony between neurons with similar muscle fields in monkey motor cortex. *Neuron* 38: 115–125, 2003.
- Johnson A and Kotz S.** *Distributions in Statistics: Continuous Univariate Distributions*. New York: Wiley, 1970.
- Kass RE and Ventura V.** A spike-train probability model. *Neural Comput* 13: 1713–1720, 2001.
- Keat J, Reinagel P, Reid RC, and Meister M.** Predicting every spike: a model for the responses of visual neurons. *Neuron* 30: 803–817, 2001.
- Kitagawa G and Gersh W.** *Smoothness Priors Analysis of Time Series*. New York: Springer-Verlag, 1996.
- Knierim JJ and Vanessen DC.** Neuronal responses to static texture patterns in area-V1 of the alert macaque monkey. *J Neurophysiol* 67: 961–980, 1992.
- Luczak A, Hackett TA, Kajikawa Y, and Laubach M.** Multivariate receptive field mapping in marmoset auditory cortex. *J Neurosci Methods* 136: 77–85, 2004.
- Maynard EM, Hatsopoulos NG, Ojakangas CL, Acuna BD, Sanes JN, Normann RA, and Donoghue JP.** Neuronal interactions improve cortical population coding of movement direction. *J Neurosci* 19: 8083–8093, 1999.
- McCullagh P and Nelder JA.** *Generalized Linear Models* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC, 1989.
- Mehta MR, Barnes CA, and McNaughton BL.** Experience-dependent, asymmetric expansion of hippocampal place fields. *Proc Natl Acad Sci USA* 94: 8918–8921, 1997.
- Mehta MR, Quirk MC, and Wilson MA.** Experience-dependent asymmetric shape of hippocampal receptive fields. *Neuron* 25: 707–715, 2000.
- Mendel JM.** *Lessons in Estimation Theory for Signal Processing, Communications, and Control*. Englewood Cliffs, NJ: Prentice Hall, 1995.
- Moran DW and Schwartz AB.** Motor cortical representation of speed and direction during reaching. *J Neurophysiol* 82: 2676–2692, 1999.
- Nicolelis MA, Dimitrov D, Carmena JM, Crist R, Lehew G, Kralik JD, and Wise SP.** Chronic, multisite, multielectrode recordings in macaque monkeys. *Proc Natl Acad Sci USA* 100: 11041–11046, 2003.
- Ogata Y.** On Lewis' simulation method for point processes. *IEEE Trans Inform Theory* IT-27: 23–31, 1981.
- Ogata Y.** Statistical models for earthquake occurrences and residual analysis for point processes. *J Am Stat Assoc* 83: 9–27, 1988.
- O'Keefe J and Dostrovsky J.** The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res* 34: 171–175, 1971.
- O'Keefe J and Recce ML.** Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* 3: 317–330, 1993.
- Paninski L, Fellows MR, Hatsopoulos NG, and Donoghue JP.** Spatiotemporal tuning of motor neurons for hand position and velocity. *J Neurophysiol* 91: 515–532, 2004.
- Papangelou F.** Integrability of expected increments of point processes and a related change of scale. *Trans Am Math Soc* 165: 483–506, 1972.
- Pawitan Y.** *In All Likelihood: Statistical Modeling and Inference Using Likelihood*. London: Oxford Univ. Press, 2001.
- Sanes JN and Truccolo W.** Motor “binding”: do functional assemblies in primary motor cortex have a role? *Neuron* 38: 3–5, 2003.
- Santner TJ and Duffy DE.** *The statistical analysis of discrete data*. New York: Springer-Verlag, 1989.
- Schwarz G.** Estimating the dimension of a model. *Ann Stat* 6: 461–464, 1978.
- Scott SH.** The role of primary motor cortex in goal-direction movements: insights from neurophysiological studies on non-human primates. *Curr Opin Neurobiol* 13: 671–677, 2003.
- Sillito AM, Grieve KL, Jones HE, Cudeiro J, and Davis J.** Visual cortical mechanisms detecting focal orientation discontinuities. *Nature* 378: 492–496, 1995.
- Stone M.** An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *J R Stat Soc B* 39: 44–47, 1977.
- Tanner MA.** *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. New York: Springer-Verlag, 1996.
- Vinje WE and Gallant JL.** Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287: 1273–1276, 2000.
- Wilson HR.** Simplified dynamics of human and mammalian neocortical neurons. *J Theor Biol* 200: 375–388, 1999.
- Wilson MA and McNaughton BL.** Dynamics of the hippocampal ensemble code for space. *Science* 261: 1055–1058, 1993.

Multiple neural spike train data analysis: state-of-the-art and future challenges

Emery N Brown, Robert E Kass & Partha P Mitra

Multiple electrodes are now a standard tool in neuroscience research that make it possible to study the simultaneous activity of several neurons in a given brain region or across different regions. The data from multi-electrode studies present important analysis challenges that must be resolved for optimal use of these neurophysiological measurements to answer questions about how the brain works. Here we review statistical methods for the analysis of multiple neural spike-train data and discuss future challenges for methodology research.

Neurophysiologists often administer a stimulus and simultaneously record neural activity from a brain region believed to respond to that stimulus. The stimulus can be physical in nature, such as light used to stimulate retinal or lateral geniculate neurons, or sound used to stimulate neural activity in the auditory cortex. It can also be abstract or cognitive, such as in a working memory task, which elicits neural activity in the hippocampus or pre-frontal cortex. The experimental question can be addressed by characterizing the relation between the stimulus and the individual or ensemble neural responses and/or the relation among the spiking activity of the neurons in the ensemble. In contrast to studying the spiking activity from a single neuron, the recent advent of multiple-electrode recording¹ makes it possible to study the simultaneous spiking activity of many neurons (more than 20). This allows us to understand how groups of neurons act in concert to define the function of a given brain region. Simultaneous recording of multiple neurons offers new promise for investigating fundamental questions, provided the challenging problem of analyzing multiple simultaneously recorded spike trains can be properly addressed.

In probability theory and statistics, a time series of discrete events, such as a spike train, is called a point process². Hence, ensembles of spike trains from simultaneously recorded neurons are multi-dimensional point-process time series. These time series are both dynamic and stochastic. That is, their properties change through time in a manner that can often be characterized by a probability model describing the likelihood of spikes at a given time. These data present

new analysis challenges because most standard signal processing techniques are designed primarily for continuous-valued data and not point processes. Thus, standard methods are of limited use in analyzing multiple neural spike train data. Moreover, because brain regions represent relevant biological signals in the spiking patterns of their constituent neurons, proper analysis of these data requires accurately characterizing the neural interactions.

Spike sorting: identification and classification of spike events

In neurophysiological experiments, individual spikes are not directly recorded. This is because when multiple electrodes are implanted, the extracellular voltage potentials recorded on any electrode represent the simultaneous electrical activity of an unknown number of neurons. From these voltage traces, the spike events or action potentials must be identified, the number of neurons being recorded must be determined, and each spike must be assigned to the neuron that produced it^{3–5}. This three-stage process, termed 'spike sorting' (Fig. 1a,b) is the mandatory first step in all multiple spike train data analyses. The accuracy of the spike sorting critically affects the accuracy of all subsequent analyses.

Many algorithms are used for spike sorting and at present, there is no consensus as to which are best. Different algorithms applied to the same data set can yield different results, illustrating the many complexities of the spike-sorting problem. First, clusters of voltage traces that summarize the spike events often violate the frequently made assumption of stable, Gaussian errors in model-based parametric algorithms. Because neuronal properties and experimental conditions evolve, these clusters change over time. Second, identifying the number of neurons is a challenging problem. One strategy is to assume a number of neurons well in excess of the number believed to be in the data, and then combine clusters that are sufficiently close using a stopping criterion⁵. An alternative Monte Carlo-based strategy has been recently proposed, but has yet to be widely tested⁶. Third, dual intracellular-extracellular recording studies have shown that spike sorting, particularly for large numbers of neurons, has a non-zero error rate because the probability distribution of spike shapes from different neurons share some degree of overlap⁴. Finally, multiple electrodes with different geometries and numbers of electrodes usually require different sorting algorithms.

Cross-correlogram and cross-intensity function

Most current methods for neural spike train data analysis assess only associations between pairs of neurons. As is true for continuous-valued data, techniques to measure the association between neural spike trains can be divided into time-domain and frequency-

Emery N. Brown is in the Neuroscience Statistics Research Laboratory, Department of Anesthesia and Critical Care, Massachusetts General Hospital, and the Division of Health Sciences and Technology, Harvard Medical School, Massachusetts Institute of Technology, Boston, Massachusetts 02114, USA. Robert E. Kass is in the Department of Statistics, Carnegie Mellon University and the Center for the Neural Basis of Cognition, Pittsburgh, Pennsylvania 15208, USA. Partha P. Mitra is at the Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA.
e-mail: brown@neurostat.mgh.harvard.edu

Published online 27 April 2004; doi:10.1038/nn1228

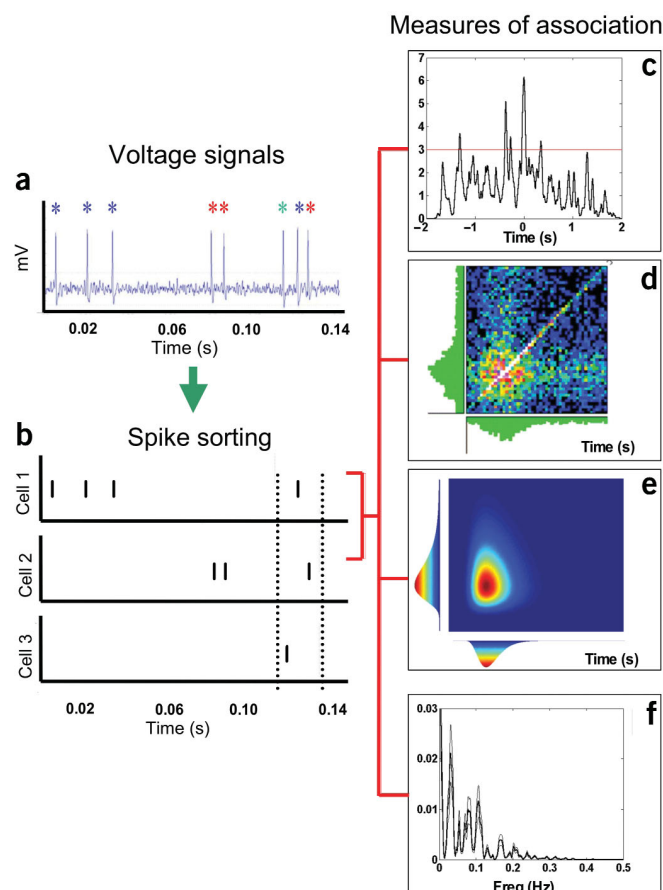


Figure 1 Transition from voltage signal recordings to measures of association for three neural spike trains. (a) Voltage trace containing the spike events of three different neurons recorded on the same electrode. Each colored star indicates a different neuron. (b) Application of a spike sorting algorithm that identifies the spike events, determines the number of neurons and assigns each spike event to a particular neuron. The dotted vertical lines show a spike triplet identified by a spike pattern classification method. (c–f) Measures of association between the spike trains from Cell 1 and Cell 2 computed using an unnormalized cross-correlogram (c), a JPSTH (d), a parametric model fit by maximum likelihood (e) and a cross-coherence function (f, solid black line) and confidence bounds (f, thin black line). The horizontal line in c is the upper 95% confidence bound. Correlations above this line are significantly different from zero.

domain methods. The most commonly used time-domain method for measuring association between neurons is the unnormalized cross-correlogram (Fig. 1c)⁷. Given a pair of neural spike trains and a specified bin width, the un-normalized cross-correlogram is the cross-covariance between the two binned spike trains computed at a series of lags. The method assumes that the two spike trains are stationary. That is, it is assumed that the stochastic properties of the neurons do not change in time. In many cases, this ‘stationarity’ assumption can be hard to justify, given that the neural responses are elicited by time-varying stimuli and frequently adapt with time in response to the same stimulus. Non-stationarity has been addressed by performing the covariance analyses in moving windows; however, this requires a substantial amount of data.

A related measure of association between two stationary point processes (spike trains) is the cross-intensity function⁸. This function estimates the spike rate of one neuron at different lags relative to the

spiking activity of a second neuron. Despite being designed expressly to measure association between two point processes, being simple to compute and having associated confidence interval estimates, this method has received only limited use in neural data analysis. Both the cross-correlogram and the cross-intensity function are histogram-based, and provide only measures of paired associations of neural activity.

Joint peri-stimulus time histogram

The joint peri-stimulus time histogram (JPSTH)⁹ (Fig. 1d) is for a pair of neurons a logical extension of the single-neuron PSTH^{9–11}. Whereas the PSTH displays the spike count per unit time t at each time t , the JPSTH is a two-dimensional histogram that displays the joint spike count per unit time at each time u for neuron 1 and time v for neuron 2. The main diagonal of the JPSTH (the ‘PST coincidence histogram’) displays for each time t the observed rate at which both neurons fire simultaneously (to within the accuracy of the binwidth of the histogram). A modification of the JPSTH, termed the normalized JPSTH, is also used¹². The normalized JPSTH subtracts from the joint firing rate the firing rate expected under independence, and then divides by the product of the two standard deviations (of the two neuronal firing rates) to correct for the possibility that two independent neurons with jointly elevated firing rates can appear to be strongly associated. The normalized JPSTH at the time pair (u, v) is the Pearson correlation (computed across trials) of the firing of neuron 1 at time u with that of neuron 2 at time v . Summing the diagonals of the normalized JPSTH produces the normalized cross-correlogram.

Although the normalized JPSTH and the normalized cross-correlogram (Fig. 1) are useful, both have limitations. First, the Pearson correlation is only one of many possible measures of association, and different measures can produce different results, the accuracy of which depends on the underlying mechanism that produces the joint spiking activity¹³. Second, statistical significance testing can be performed in several ways with these methods and again, the results can differ depending on the assumptions and the methods. A new approach to significance testing using recently developed smoothing procedures and a bootstrap significance test can yield greater statistical power¹¹. The bootstrap is a broadly applicable simulation method for estimating uncertainty in a statistical analysis. Third, the normalized JPSTH and cross-correlogram assume that all trials are statistically indistinguishable⁷. If, instead, there is detectable trial-to-trial variation in the neural firing rates, then this variation can appear artifactually as synchrony or time-lagged joint firing^{7,14}. A fourth, crucial consideration is that whereas all spike train analysis is predicated on good spike sorting, the accuracy of spike time information is particularly important when searching for synchrony or time-lagged joint firing. The effects of spike overlap, which are problematic for most spike-sorting algorithms, can produce spurious correlations between pairs of neurons¹⁵.

Spike pattern classification methods

Algorithms to detect precise patterns of spike timing are another method of measuring associations among neural spike trains^{10,16,17}. The appeal of these methods is that they provide a way of evaluating higher-order neural interactions, that is, greater than pairwise, in ensemble spiking activity¹⁸. For example, these methods can be used to assess the statistical significance of spike triplet occurrences separated by precise interspike intervals or the occurrence of similar patterns among two or more neurons (Fig. 1b)¹⁹. Methods for identifying statistically conspicuous spike coincidences have also been developed. Such coincidences have been labeled ‘unitary events’ when they occur more frequently than would be predicted by chance under the null

hypothesis that spike times are independent^{20–22}. Their occurrence is then studied in relation to behavioral events. The delicate statistical issue involved in applying spike pattern classification methods is choosing the complexity or size of the pattern, and formulation of the null hypothesis and test statistic so that the procedure has the correct significance level under reasonably general assumptions. For this reason, some findings from these analyses have at times been criticized as suffering from statistical artifacts²³. An alternative approach to test for synchrony is to build a distribution of spike trains under a null hypothesis by ‘jittering’ the observed spike trains randomly within a small time window. This intuitive idea has recently been formalized and extended to cover several practical data analysis scenarios²⁴.

Likelihood methods

Likelihood methods are central tools for modeling and analysis in statistical research²⁵. Most likelihood methods assume a specific parametric probability model for a process under study (Fig. 1e). The likelihood is the joint probability density of the experimental data arising from this process viewed as a function of the model’s unknown or free parameters. These free parameters may be estimated from the experimental data by formal estimation procedures such as method of moments or maximum likelihood. If the probability model is a good approximation to the process being studied, then use of the likelihood is an optimal way of analyzing the data being generated by the process²⁵. Likelihood methods for point processes have been used to analyze single neural spike train data^{8,26–29}, and in a few instances to model two or more simultaneously recorded neurons^{8,30}. Likelihood methods hold important promise for this and other neuroscience data analysis problems because they provide in a coherent framework a wide range of well-developed statistical methods for data analysis, including assessing model goodness-of-fit, constructing confidence intervals and testing hypotheses^{8,29}. The challenge in using likelihood point process methods to analyze multiple neural spike trains is defining multivariate point process models that accurately represent joint neural spiking activity and devising efficient algorithms for model fitting³⁰.

Frequency-domain methods

Under the assumption of stationarity, as in the case of continuous-valued data, a frequency domain analysis of ensemble neural spiking activity can be conducted by taking the Fourier transform of the spike trains, and using these to compute the spectrum of the individual trains and the cross-spectrum or coherence between each spike train pair^{8,31,32}. The coherence is a simple frequency-dependent correlation measure of association between two processes (Fig. 1f). It has two important advantages over the time domain counterpart: the normalization is not bin-size dependent, and it can be pooled across neuron pairs. It also allows for analysis of point processes, continuous-valued processes, and hybrid point and continuous-valued pairs using the same measure. Error estimates and confidence intervals can be computed for spectra and coherence estimates from theoretical formulae

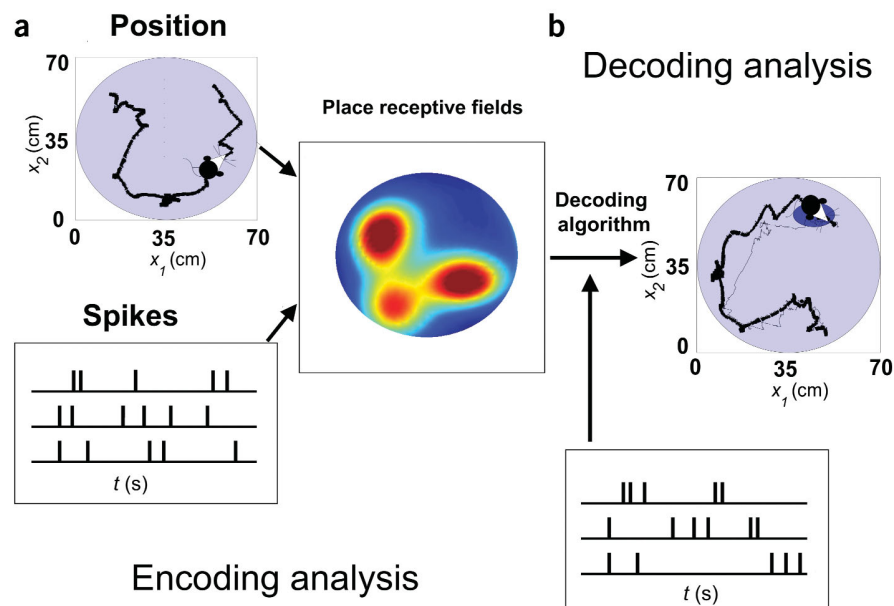


Figure 2 Decoding of position from ensemble rat neural spiking activity^{40,50}. (a) Encoding analysis in which the relation between the biological stimulus (trajectory of the rat in the environment, solid black line in the Position panel) and spiking activity (Spikes panel) is estimated as place receptive fields for three neurons. (b) Decoding analysis in which the estimated place receptive fields are used in a Bayesian decoding algorithm to compute the predicted position (thin black line) of the rat in the environment from new spiking activity of the neural ensemble recorded during the decoding stage. The predicted position is compared with the observed position (thick black line) during the decoding stage. The blue oval defines a 95% confidence region centered at that location.

that are valid when the numbers of spikes in the spike trains are large, or from bootstrap/jackknife procedures (Fig. 1f)³³.

Stimulus-driven non-stationarity is an important feature of neural spike train data, and may be analyzed using moving window estimates of spectra (spectrograms) and coherences (coherograms)³⁴. A key technical yet practical point for use of time-frequency spectral estimates, including moving window and wavelet-based estimates³⁵, is that they must obey the uncertainty principle, which puts a lower bound on the area of the point spread functions of these estimates at all points in the time-frequency plane ($\Delta f \Delta t \geq 1$). Moving window estimates computed in the frequency domain are often less biased than the corresponding time-domain estimates. Thus, even time domain functions, such as the cross-correlogram and the PSTH, may be optimally estimated by inverse Fourier-transforming the corresponding frequency-domain quantities. One principled approach to estimating the frequency-domain quantities is by using multitaper techniques³⁶. These methods have also proved useful in estimating coherence between spike trains and local field potentials³⁷ and are well-suited for error analyses using bootstrap/jackknife procedures.

Neural spike train decoding

Decoding algorithms are the mathematical techniques used in neuroscience to study how spike train firing patterns from a single neuron^{38,39} or an ensemble of neurons⁴⁰ represent external stimuli and biological signals. The decoding analysis proceeds typically in two stages: the encoding stage (Fig. 2a) and the decoding stage (Fig. 2b). In the encoding stage, neural spiking activity is characterized as a function of the biological signal. In the decoding stage, the relation is inverted, and the signal is estimated from the spiking activity of the neurons. Developed initially to study how movements are represented

by neurons in the motor cortex, the population vector is one of the earliest decoding techniques^{41,42}. To decode the neural spiking activity at any time t , one computes the population vector (the normalized dot product of the observed spiking activity in a time window with the firing functions for each neuron) for different values of the signal. The value of the signal for which the dot product is the largest is taken to be the decoded estimate of the signal.

Reverse correlation is a linear regression-based decoding algorithm that has been used to study how groups of neurons represent information in the visual and motor systems and to control neural prosthetic devices^{38,39,43–46}. The appeal of this widely used approach is that binning the spike trains to create continuous-valued regressors avoids use of an explicit encoding model and makes it possible to use standard linear regression theory to fit the model and assess the accuracy of the decoding. Moreover, computation of the regression coefficients implicitly takes account of pairwise correlations in the neural activity. The linear construction of the population vector makes it a special case of the reverse correlation methods.

Decoding algorithms that are based on Bayes' theorem, the elementary probability rule for computing the probability of one event given another event, offer a general approach to estimating the representation of a biological signal in ensemble spiking activity^{40,47–51}. They have been used successfully to study how neural ensembles in the hippocampus represent an animal's position in an environment^{40,47,50} (Fig. 2) and to characterize how motor commands are represented by ensembles of neurons in primary motor cortex^{49,51}. The encoding stage can use likelihood methods to compute the probability of the spiking activity given the signal, and the decoding stage computes the probability of the signal given the spiking activity. The appeal of the Bayesian approach is that it uses probability models to represent different sources of information in the problem, and it formulates decoding in the theoretical framework of other filtering and smoothing methods in statistics and signal processing. When the proposed model is a reasonable approximation to the data, the Bayesian approach, like the likelihood methods, has many optimality properties, including efficiency, which, in the decoding problem, means that its signal estimates have the smallest possible uncertainty²⁵. An important conceptual difference between the Bayesian and reverse correlation decoding methods is that under the standard assumptions of regression theory, the neural firing rates used in a reverse correlation analysis are assumed to be non-random, known constants. In contrast, the Bayesian approach models the spike trains as a stochastic point process and the biological signal as a stochastic process based on its known properties.

Information theory

Information theory measures are used widely in analyses of neural spike train data^{39,52–54}. These include the entropy to quantify spike train variability, and mutual information to measure the association between two processes, such as between two spike trains or between a spike train and a stimulus. These measures have been applied extensively to study how much information a single spike train conveys about a biological signal by using histogram-based methods to estimate empirically the relevant probability densities. Use of the information measures is grounded in thinking about those parts of the nervous system, such as visual pathways, that may be modeled as communication channels³⁹ with a rationale that analyses may be conducted free of assumptions about detailed system properties⁵⁵. There are limitations to this approach. For any neural system, the optimal word length (histogram binwidth) is an unknown that must be estimated taking account that the data requirements for histogram esti-

mation increase exponentially with the word length. The data requirements are far greater for extending this approach to estimating mutual information between multiple spike trains and a biological signal⁵⁴. Information theory methods summarize complex functional relationships between the spike train and the signal as single numbers. Moreover, whether sensory pathways can be treated using information theory as in conventional communications analyses has recently been questioned⁵⁶.

One approach to extending the use of information theory to analyze multiple spike trains may arise from developing probability models of joint spiking activity and likelihood methods to estimate these models. An advantage of modeling explicitly the joint probability density between the ensemble spiking activity and the biological signal is that the mutual information and any other relevant functions of this probability density can be computed directly once this probability density has been estimated. Parametric models may then offer insight into how to construct their more flexible, model-free counterparts.

Future challenges for multiple spike train data analysis

Simultaneous recording of multiple spike trains from several neurons offers a window into how neurons work in concert to generate specific brain functions. Without substantial methodology research in the future, our ability to understand this function will be significantly hampered because current methods fall short of what is ultimately required for the analysis of multiple spike train data. With the exception of spike pattern classification methods, decoding algorithms, partial coherence estimation⁸ and certain graphical methods⁵⁷, current techniques for spike train analysis are designed to analyze—at most—pairs of neurons. Therefore, the future challenge is to design methods that truly allow neuroscientists to perform multivariate analyses of multiple spike train data. This development must be done taking explicit account of the questions being studied and the experimental protocols being used.

Because the accuracy of the spike sorting significantly affects the accuracy of the experimental data, development of the best possible spike-sorting algorithms must be an important goal. The complexity of the spike-sorting problem increases with number of electrodes in the recording systems. There should be systemic study of spike-sorting algorithms taking account of different electrode numbers and configurations, recording conditions and brain regions. A harder, yet no less important, challenge is to devise accurate, real-time spike-sorting algorithms to enable multiple spike trains to be inputs to neural prosthetic devices or brain-machine interfaces^{42,45,46}. Real-time spike sorting could also lead to real-time data analysis, and possibly to real-time changes in experimental protocols.

Graphical methods⁵⁷ for multivariate point process data are important for screening data for errors and inconsistencies prior to analysis, postulating preliminary models and formulating meaningful displays to report findings.

Multiple spike trains are multivariate point processes, yet research in statistics and signal processing on multivariate point process models has not been nearly as extensive as research on models of multivariate continuous-valued processes. Therefore, developing multivariate point process models should be a primary focus of methodology research for multiple neural spike trains. Because there is a canonical representation of univariate and multivariate point processes in terms of the conditional intensity function^{2,29}, developing strategies to construct parametric models of conditional intensity functions and likelihood-based estimation methods seems a good way to proceed^{2,8,27–30}. Other avenues of investigation could include lattice or spin models from statistical physics⁵⁸ and multivariate binary data models from

statistics⁵⁹. Whatever the approach, the objective must be to develop tractable methods for estimating high-dimensional interactions among groups of neurons from their spike train recordings. Furthermore, because plasticity in neural dynamics makes non-stationarity in neural data a rule rather than an exception, developing explicit adaptive estimation algorithms to track these dynamics for multivariate point processes is another important research problem⁶⁰.

Dynamical systems and neural network models have long been central in providing quantitative characterizations of neural processes^{61,62}. Research on data analysis methodology should be conducted in concert with this modeling research⁶³. Multiple-electrode recordings combined with statistical methods explicitly designed to analyze multiple spike train data will offer a better opportunity to explicitly link experimentation and computational modeling by using the information from experiments to quantify better predictions from more complex models, refine model formulation and, as a consequence, design better experiments. Similarly, the computational models can suggest formulations of the statistical methods that may enhance their success at extracting salient features in experimental data.

Although the objective of most current neurophysiological experiments is to relate relevant biological stimuli to ensemble spiking activity, experiments that record simultaneous multimodality data such as neurophysiological, functional imaging and behavioral data are becoming more common⁶⁴. Developing appropriate statistical methods to analyze simultaneous multimodality recordings will require innovative approaches to integrate information properly across the different temporal and spatial scales of various data sources.

There are many benefits of developing multivariate methods for multiple spike train data analysis. First, methods specifically tailored to analyze multiple spike train data will allow neuroscientists to make precise statements of how reliably findings from a given experiment can be stated in terms of standard statistical summaries. Practically speaking, this means that even for this complex, high-dimensional modeling problem, the analysis reports standard errors for firing rates and time constant parameters, provides confidence intervals for measures of neural interactions and associations, and gives quantitative assessments of how well a given model describes the experimental data²⁹. Second, more accurate quantitative summaries will allow neuroscientists to make more reliable statements about how strongly experimental findings support hypotheses or proposed mechanisms. For example, would an analysis measuring time-varying interactions among three or more neurons rather than pairwise correlations offer new insight into the mechanism of persistent activity seen in the oculomotor system⁶⁵?

Third, more accurate multivariate quantitative summaries will make it easier to relate ensemble neural dynamics (within and between specific brain regions) to behavior and to relevant biological stimuli. As an illustration, applying these methods to the study of simultaneously recorded neural activity in the parietal and primary motor cortices could help reveal how these two brain regions communicate during formulation and execution of motor commands. Fourth, as the number of neurons whose interactions can be accurately measured increases, neuroscientists will be able to increase the complexity of the experiments they design, and as a consequence, the questions they investigate. Fifth, more reliable data analyses will provide more refined quantitative constraints and perhaps parameter values for dynamical models of neural systems. Finally, improved multiple spike train data analysis methods, particularly spike-sorting and decoding algorithms, will have immediate, significant implications for improving the design and implementation of neural prosthetic devices and brain-computer interfaces^{42,45,46}.

Multiple spike train recordings are an important component of the data explosion that is currently occurring in neuroscience. Therefore, devising systematic research programs for neuroscience data analysis akin to those currently being undertaken in genomics and bioinformatics is a must. Several initiatives to support such research have already been proposed by the US National Institutes of Health (<http://grants1.nih.gov/grants/guide/pa-files/PA-04-006.html>) and National Science Foundation (www.nsf.gov/bio/progdes/biocrcn.htm). Specific initiatives to encourage quantitative scientists (statisticians, physicists, engineers, computer scientists and mathematicians) to undertake data analysis research in neuroscience should be part of these current and future programs. Neuroscience training for statisticians and incentives to involve them more directly in neuroscience data analysis research should be a priority. Finally, courses on the analysis of neuroscientific data (www.mbl.edu/education/courses/special_topics/neufo.html) should be part of the curriculum in neuroscience programs, as are courses on computational modeling. This will ensure that instruction in the most contemporary data analysis principles and methods are an integral part of undergraduate, graduate and postdoctoral training in neuroscience, and in the disciplines that support computational research in this field.

ACKNOWLEDGMENTS

Support for this work was provided in part by NIH grants MH66410 to P.M. and E.N.B., MH62528 to P.M., MH64537 to R.E.K., and MH59733, MH61637 and DA015664 to E.N.B. We thank S. Grün for comments on an earlier draft of this manuscript, G. Gerstein for permission to use Fig. 1d and R. Barbieri for help preparing the figures.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at <http://www.nature.com/natureneuroscience/>

1. Wilson, M.A. & McNaughton, B.L. Dynamics of the hippocampal ensemble code for space. *Science* **261**, 1055–1058 (1993).
2. Daley, D. & Vere-Jones, D. *An Introduction to the Theory of Point Process*, 2nd ed. (Springer-Verlag, New York, 2003).
3. Lewicki, M.S. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network Comput. Neural Syst.* **9**, R53–R78 (1998).
4. Harris, K.D., Henze, D.A., Csicsvari, J., Hirase, H. & Buzsáki, G. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *J. Neurophysiol.* **84**, 401–414 (2000).
5. Fee, M.S., Mitra, P.P. & Kleinfeld, D. Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability. *J. Neurosci. Methods* **69**, 175–188 (1996).
6. Nguyen, D., Frank, L.M. & Brown, E.N. An application of reversible-jump MCMC to spike classification of multiunit extracellular recordings. *Network Comput. Neural Syst.* **14**, 61–82 (2003).
7. Brody, C.D. Correlations without synchrony. *Neural Comput.* **11**, 1537–1551 (1999).
8. Brillinger, D.R. Nerve cell spike train data analysis: a progression of techniques. *J. Amer. Stat. Assoc.* **87**, 260–271 (1992).
9. Gerstein, G.L. & Perkel, D.H. Simultaneously recorded trains of action potentials: analysis and functional interpretation. *Science* **164**, 828–830 (1969).
10. Abeles, M. Quantification, smoothing, and confidence limits for single-unit histograms. *J. Neurosci. Methods* **5**, 317–325 (1982).
11. Kass, R.E., Ventura, V. & Cai, C. Statistical smoothing of neuronal data. *Network Comput. Neural Syst.* **14**, 5–15 (2003).
12. Aertsen, A., Gerstein, G.L., Habib, M.K. & Palm, G. Dynamics of neural firing correlation: modulation of “effective connectivity.” *J. Neurophysiol.* **61**, 900–917 (1989).
13. Ito, H. & Tsuji, S. Model dependence in quantification of spike interdependency by joint peri-stimulus time histogram. *Neural Comput.* **12**, 195–217 (2000).
14. Baker, S.N. & Gerstein, G.L. Determination of response latency and its application to normalization of cross-correlation measures. *Neural Comput.* **13**, 1351–1377 (2001).
15. Bar-Gad, I., Ritov, Y., Vaadia, E. & Bergmann, H. Failure in identification of overlapping spikes from multiple neuron activity causes artificial correlations. *J. Neurosci. Methods* **107**, 1–13 (2001).
16. Softky, W.R. & Koch, C. The highly irregular firing of cortical cells is consistent with temporal integration of random EPSPs. *J. Neurosci.* **13**, 334–350 (1993).
17. Shadlen, M.T. & Newsome, W.N. The variable discharge of cortical neurons: impli-

- cations for connectivity, computation, and information coding. *J. Neurosci.* **18**, 3870–3896 (1998).
18. Martignou, L. *et al.* Neural coding: higher-order temporal patterns in the neurostatistics of cell assemblies. *Neural Comput.* **12**, 2621–2653 (2000).
 19. Abeles, M. & Gerstein, G.L. Detecting spatiotemporal firing patterns among simultaneously recorded single neurons. *J. Neurophysiol.* **60**, 909–924 (1988).
 20. Grün, S., Diesmann, M. & Aertsen, A. Unitary events in multiple single-neuron spiking activity: II. Nonstationary data. *Neural Comput.* **14**, 81–119 (2002).
 21. Gütig, R., Aertsen, A. & Rotter, S. Statistical significance of coincident spikes: count-based versus rate-based statistics. *Neural Comput.* **14**, 121–153 (2002).
 22. Pipa, G. & Grün, S. Non-parametric significance estimation of joint-spike events by shuffling and resampling. *Neurocomputing* **52–54**, 31–37 (2003).
 23. Oram, M.W., Wiener, M.C., Lestienne, R. & Richmond, B.J. Stochastic nature of precisely timed spike patterns in visual system neuronal responses. *J. Neurophysiol.* **81**, 3021–3033 (1999).
 24. Amarasingham, A. *Temporal Structure in Nervous System Activity* dissertation, Brown Univ. (2002).
 25. Pawitan, Y. *In All Likelihood: Statistical Modelling and Inference Using Likelihood* (Oxford Univ. Press, New York, 2001).
 26. Brillinger, D.R. Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol. Cybern.* **59**, 189–200 (1988).
 27. Barbieri, R., Quirk, M.C., Frank, L.M., Wilson, M.A. & Brown, E.N. Construction and analysis of non-Poisson stimulus response models of neural spike train activity. *J. Neurosci. Methods* **105**, 25–37 (2001).
 28. Kass, R.E. & Ventura, V. A spike train probability model. *Neural Comput.* **13**, 1713–1720 (2001).
 29. Brown, E.N., Barbieri, R., Eden, U.T. & Frank, L.M. Likelihood methods for neural data analysis. in *Computational Neuroscience: A Comprehensive Approach* (ed. Feng, J.) 253–286 (CRC Press, Boca Raton, 2003).
 30. Chornoboy, E.S., Schramm, L.P. & Karr, A.F. Maximum likelihood identification of neural point process systems. *Biol. Cybern.* **59**, 265–275 (1988).
 31. Brillinger, D.R. Comparative aspects of the study of ordinary time series and of point processes. in *Developments in Statistics* Vol. 1, 33–129 (Academic Press, Orlando, 1978).
 32. Jarvis, M.R. & Mitra, P.P. Sampling properties of the spectrum and coherency in sequences of action potentials. *Neural Comput.* **13**, 717–749 (2001).
 33. Thomson, D.J. & Chave, A.D. Jackknifed error estimates for spectra, coherences, and transfer functions. in *Advances in Spectrum Analysis and Array Processing* (ed. Haykin, S.) 58–113 (Prentice Hall, Englewood Cliffs, NJ, 1991).
 34. Brillinger, D.R. *Time Series* (Holt, Rinehart and Winston, New York, 1981).
 35. Percival, D.B. & Walden, A.T. *Wavelet Methods for Time Series Analysis* (Cambridge Univ. Press, Cambridge, UK, 2000).
 36. Percival, D.B. & Walden, A.T. *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques* (Cambridge Univ. Press, Cambridge, UK, 2002).
 37. Pesaran, B., Pezaris, J.S., Sahani, M., Mitra, P.P. & Andersen, R.A. Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat. Neurosci.* **5**, 805–811 (2002).
 38. Bialek, W., Rieke, F., de Ruyter van Steveninck, R.R. & Warland, D. Reading a neural code. *Science* **252**, 1854–1857 (1991).
 39. Rieke, F., Warland, D., de Ruyter van Steveninck, R. & Bialek, W. *Spikes: Exploring the Neural Code* (MIT Press, Cambridge, 1997).
 40. Brown, E.N., Frank, L.M., Tang, D., Quirk, M.C. & Wilson, M.A. A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *J. Neurosci.* **18**, 7411–7425 (1998).
 41. Georgopoulos, A.P., Kettner, R.E. & Schwartz, A.B. Neuronal population coding of movement direction. *Science* **233**, 1416–1419 (1986).
 42. Taylor, D.M., Tillery, S.I. & Schwartz, A.B. Direct cortical control of 3D neuroprosthetic devices. *Science* **296**, 1829–1832 (2002).
 43. Warland, D.K., Reinagel, P. & Meister, M. Decoding visual information from a population of retinal ganglion cells. *J. Neurophysiol.* **78**, 2336–2350 (1997).
 44. Stanley, G.B., Li, F.F. & Dan, Y. Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. *J. Neurosci.* **19**, 8036–8042 (1999).
 45. Wessberg, J. *et al.* Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature* **408**, 361–365 (2000).
 46. Serruya, M.D., Hatsopoulos, N.G., Paninski, L., Fellows, M.R. & Donoghue, J.P. Instant neural control of a movement signal. *Nature* **416**, 141–142 (2002).
 47. Zhang, K., Ginzburg, I., McNaughton, B.L. & Sejnowski, T.J. Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. *J. Neurophysiol.* **79**, 1017–1044 (1998).
 48. Wiener, M.C. & Richmond, B.J. Decoding spike trains instant by instant using order statistics and the mixture-of-Poissons model. *J. Neurosci.* **23**, 2394–2406 (2003).
 49. Gao, Y., Black, M.J., Bienenstock, E., Wu, W. & Donoghue, J.P. A quantitative comparison of linear and non-linear models of motor cortical activity for the encoding and decoding of arm motions. *First Intl. IEEE/EMBS Conf. on Neural Eng.* 189–192 (2003).
 50. Barbieri, R. *et al.* Dynamic analyses of information encoding by neural ensembles. *Neural Comput.* **16**, 277–307 (2004).
 51. Brockwell, A.E., Rojas, A.L. & Kass, R.E. Recursive Bayesian decoding of motor cortical signals by particle filtering. *J. Neurophysiol.* **91**, 1899–1907 (2004).
 52. Borst, A. & Theunissen, F.E. Information theory and neural coding. *Nat. Neurosci.* **2**, 947–957 (1999).
 53. Reich, D.S., Melcher, F. & Victor, J.D. Independent and redundant information in nearby cortical neurons. *Science* **294**, 2566–2568 (2001).
 54. Nirenberg, S., Carcieri, S.M., Jacobs, A.L. & Latham, P.E. Retinal ganglion cells act largely as independent encoders. *Nature* **411**, 698–701 (2001).
 55. Strong, S.P., Koberle, R., de Ruyter van Steveninck, R.R. & Bialek, W. Entropy and information in neural spike trains. *Phys. Rev. Lett.* **80**, 197–200 (1998).
 56. Berger, T. Living information theory. *IEEE Information Theory Society Newsletter* **53**, 1–20 (2003).
 57. Stuart, L., Walter, M. & Borisyuk, R. Visualization of synchronous firing in multidimensional spike trains. *BioSystems* **67**, 265–279 (2002).
 58. Abdollahi, L.M., La Rota, C., Beguin, M. & François, O. Parameter estimation in a model for multidimensional recording of neuronal data: a Gibbsian approximation approach. *Biol. Cybern.* **89**, 170–178 (2003).
 59. Quakish, B.F. A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means. *Biometrika* **90**, 455–464 (2003).
 60. Brown, E.N., Nguyen, D.P., Frank, L.M., Wilson, M.A. & Solo, V. An analysis of neural receptive field plasticity by point process adaptive filtering. *Proc. Natl. Acad. Sci. USA* **98**, 12261–12266 (2001).
 61. Dayan, P. & Abbott, L.F. *Theoretical Neuroscience* (MIT Press, Cambridge, 2001).
 62. Shikri, O., Hansel, D. & Sompolinsky, H. Rate models for conductance-based cortical neuronal networks. *Neural Comput.* **5**, 1809–1841 (2003).
 63. Victor, J.D. & Brown, E.N. Information and statistical structure in spike trains. *Network Comput. Neural Syst.* **14**, 1–4 (2003).
 64. Logothetis, N.K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412**, 150–157 (2001).
 65. Aksay, E., Baker, R., Seung, H.S. & Tank, D.W. Correlated discharge among pairs within the oculomotor horizontal velocity-to-position integrator. *J. Neurosci.* **23**, 10852–10858 (2003).

Measuring the signal-to-noise ratio of a neuron

Gabriela Czanner^{a,1}, Sridevi V. Sarma^b, Demba Ba^c, Uri T. Eden^d, Wei Wu^e, Emad Eskandar^f, Hubert H. Lim^g, Simona Temereanca^h, Wendy A. Suzukiⁱ, and Emery N. Brown^{c,j,k,1}

^aDepartment of Biostatistics and Department of Eye and Vision Science, Faculty of Health and Life Sciences, University of Liverpool, Liverpool L69 3GA, United Kingdom; ^bDepartment of Biomedical Engineering, Institute for Computational Medicine, Johns Hopkins University, Baltimore, MD 21218; ^cDepartment of Brain and Cognitive Sciences and ^dInstitute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139; ^eDepartment of Mathematics and Statistics, Boston University, Boston, MA 02215; ^fSchool of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China; Departments of ^gNeurosurgery and ^hAnesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114; ⁱDepartment of Biomedical Engineering, Institute for Translational Neuroscience, University of Minnesota, Minneapolis, MN 55455; ^jMartinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129; and ^kCenter for Neural Science, New York University, New York, NY 10003

Contributed by Emery N. Brown, March 27, 2015 (sent for review December 5, 2014; reviewed by Mingzhou Ding, Satish Iyengar, and Hualou Liang)

The signal-to-noise ratio (SNR), a commonly used measure of fidelity in physical systems, is defined as the ratio of the squared amplitude or variance of a signal relative to the variance of the noise. This definition is not appropriate for neural systems in which spiking activity is more accurately represented as point processes. We show that the SNR estimates a ratio of expected prediction errors and extend the standard definition to one appropriate for single neurons by representing neural spiking activity using point process generalized linear models (PP-GLM). We estimate the prediction errors using the residual deviances from the PP-GLM fits. Because the deviance is an approximate χ^2 random variable, we compute a bias-corrected SNR estimate appropriate for single-neuron analysis and use the bootstrap to assess its uncertainty. In the analyses of four systems neuroscience experiments, we show that the SNRs are -10 dB to -3 dB for guinea pig auditory cortex neurons, -18 dB to -7 dB for rat thalamic neurons, -28 dB to -14 dB for monkey hippocampal neurons, and -29 dB to -20 dB for human subthalamic neurons. The new SNR definition makes explicit in the measure commonly used for physical systems the often-quoted observation that single neurons have low SNRs. The neuron's spiking history is frequently a more informative covariate for predicting spiking propensity than the applied stimulus. Our new SNR definition extends to any GLM system in which the factors modulating the response can be expressed as separate components of a likelihood function.

SNR | signal-to-noise ratio | neuron | simulation | point processes

The signal-to-noise ratio (SNR), defined as the amplitude squared of a signal or the signal variance divided by the variance of the system noise, is a widely applied measure for quantifying system fidelity and for comparing performance among different systems (1–4). Commonly expressed in decibels as $10\log_{10}(\text{SNR})$, the higher the SNR, the stronger the signal or information in the signal relative to the noise or distortion. Use of the SNR is most appropriate for systems defined as deterministic or stochastic signals plus Gaussian noise (2, 4). For the latter, the SNR can be computed in the time or frequency domain.

Use of the SNR to characterize the fidelity of neural systems is appealing because information transmission by neurons is a noisy stochastic process. However, the standard concept of SNR cannot be applied in neuronal analyses because neurons transmit both signal and noise primarily in their action potentials, which are binary electrical discharges also known as spikes (5–8). Defining what is the signal and what is the noise in neural spiking activity is a challenge because the putative signals or stimuli for neurons differ appreciably among brain regions and experiments. For example, neurons in the visual cortex and in the auditory cortex respond respectively to features of light (9) and sound stimuli (10) while neurons in the somatosensory thalamus respond to tactile stimuli (11). In contrast, neurons in the rodent hippocampus respond robustly to the animal's position in its environment (11, 12), whereas monkey hippocampal neurons respond to the process of task learning (13). As part of responding to a putative stimulus,

a neuron's spiking activity is also modulated by biophysical factors such as its absolute and relative refractory periods, its bursting propensity, and local network and rhythm dynamics (14, 15). Hence, the definition of SNR must account for the extent to which a neuron's spiking responses are due to the applied stimulus or signal and to these intrinsic biophysical properties.

Formulations of the SNR for neural systems have been studied. Rieke et al. (16) adapted information theory measures to define Gaussian upper bounds on the SNR for individual neurons. Coefficients of variation and Fano factors based on spike counts (17–19) have been used as measures of SNR. Similarly, Gaussian approximations have been used to derive upper bounds on neuronal SNR (16). These approaches do not consider the point process nature of neural spiking activity. Moreover, these measures and the Gaussian approximations are less accurate for neurons with low spike rates or when information is contained in precise spike times.

Lyamzin et al. (20) developed an SNR measure for neural systems using time-dependent Bernoulli processes to model the neural spiking activity. Their SNR estimates, based on variance formulae, do not consider the biophysical properties of the neuron and are more appropriate for Gaussian systems (16, 21, 22). The Poisson regression model used widely in statistics to relate count observations to covariates provides a framework for studying the SNR for non-Gaussian systems because it provides an analog of the square of the multiple correlation coefficient (R^2) used to measure goodness of fit in linear regression analyses

Significance

Neurons represent both signal and noise in binary electrical discharges termed action potentials. Hence, the standard signal-to-noise ratio (SNR) definition of signal amplitude squared and divided by the noise variance does not apply. We show that the SNR estimates a ratio of expected prediction errors. Using point process generalized linear models, we extend the standard definition to one appropriate for single neurons. In analyses of four neural systems, we show that single neuron SNRs range from -29 dB to -3 dB and that spiking history is often a more informative predictor of spiking propensity than the signal or stimulus activating the neuron. By generalizing the standard SNR metric, we make explicit the well-known fact that individual neurons are highly noisy information transmitters.

Author contributions: G.C. and E.N.B. designed research; G.C., S.V.S., D.B., U.T.E., W.W., E.E., H.H.L., S.T., and W.A.S. performed research; and G.C. and E.N.B. wrote the paper. Reviewers: M.D., University of Florida; S.I., University of Pittsburgh; and H.L., Drexel University.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

¹To whom correspondence may be addressed. Email: czanner@liverpool.ac.uk or enb@neurostat.mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1505545112/-DCSupplemental.

(23). The SNR can be expressed in terms of the R^2 for linear and Poisson regression models. However, this relationship has not been exploited to construct an SNR estimate for neural systems or point process models. Finally, the SNR is a commonly computed statistic in science and engineering. Extending this concept to non-Gaussian systems would be greatly aided by a precise statement of the theoretical quantity that this statistic estimates (24, 25).

We show that the SNR estimates a ratio of expected prediction errors (EPEs). Using point process generalized linear models (PP-GLM), we extend the standard definition to one appropriate for single neurons recorded in stimulus–response experiments. In analyses of four neural systems, we show that single-neuron SNRs range from –29 dB to –3 dB and that spiking history is often a more informative predictor of spiking propensity than the signal being represented. Our new SNR definition generalizes to any problem in which the modulatory components of a system’s output can be expressed as separate components of a GLM.

Theory

A standard way to define the SNR is as the ratio

$$\text{SNR} = \frac{\sigma_{\text{signal}}^2}{\sigma_{\text{noise}}^2}, \quad [1]$$

where σ_{signal}^2 is structure in the data induced by the signal and σ_{noise}^2 is the variability due to the noise. To adapt this definition to the analysis of neural spike train recordings from a single neuron, we have: to (i) define precisely what the SNR estimates; (ii) extend the definition and its estimate to account for covariates that, along with the applied stimulus or signal input, also affect the neural response; and (iii) extend the SNR definition and its estimate so that it applies to point process models of neural spiking activity.

By analyzing the linear Gaussian signal plus noise model ([Supporting Information](#)), we show that standard SNR computations (Eq. S5) provide an estimator of a ratio of EPEs (Eq. S4). For the linear Gaussian model with covariates, this ratio of EPEs is also well defined (Eq. S6) and can be estimated as a ratio of sum of squares of residuals (Eq. S7). The SNR definition further extends to the GLM with covariates (Eq. S8). To estimate the SNR for the GLM, we replace the sums of squares by the residual deviances, their extensions in the GLM framework Eqs. S9 and S10. The residual deviance is a constant multiple of the Kullback–Leibler (KL) divergence between the data and the model. Due to the Pythagorean property of the KL divergence of GLM models with canonical link functions (26–28) evaluated at the maximum likelihood estimates, the SNR estimator can be conveniently interpreted as the ratio of the explained KL divergence of the signal relative to the noise. We propose an approximate bias correction for the GLM SNR estimate with covariates (Eq. S11), which gives the estimator better performance in low signal-to-noise problems such as single-neuron recordings. The GLM framework formulated with point process models has been used to analyze neural spiking activity (5–7, 29). Therefore, we derive a point process GLM (PP-GLM) SNR estimate for single-neuron spiking activity recorded in stimulus–response experiments.

A Volterra Series Expansion of the Conditional Intensity Function of a Spiking Neuron. Volterra series are widely used to model biological systems (30), including neural spiking activity (16). We develop a Volterra series expansion of the log of the conditional intensity function to define the PP-GLM for single-neuron spiking activity (31). We then apply the GLM framework outlined in [Supporting Information](#) to derive the SNR estimate.

We assume that on an observation interval $(0, T]$, we record spikes at times $0 < u_1 < u_2 < \dots < u_J < T$. If we model the spike events as a point process, then the conditional intensity function of the spike train is defined by (5)

$$\lim_{\Delta \rightarrow 0} \frac{\Pr(N(t+\Delta) - N(t) | H_t)}{\Delta} = \lambda(t | H_t), \quad [2]$$

where $N(t)$ is the number of spikes in the interval $(0, t]$ for $t \in (0, T]$ and H_t is the relevant history at t . It follows that for Δ small,

$$\Pr(\text{spike in } (t, t + \Delta] | H_t) \approx \lambda(t | H_t) \Delta. \quad [3]$$

We assume that the neuron receives a stimulus or signal input and that its spiking activity depends on this input and its biophysical properties. The biophysical properties may include absolute and relative refractory periods, bursting propensity, and network dynamics. We assume that we can express $\log \lambda(t | H_t)$ in a Volterra series expansion as a function of the signal and the biophysical properties (31). The first-order and second-order terms in the expansion are

$$\begin{aligned} \log \lambda(t | H_t) = & \int_0^t s(t-u) \beta_S(u) du + \int_0^t \beta_H(u) dN(t-u) \\ & + \int_0^t \int_0^t s(t-u) s(t-v) h_1(u, v) dudv \\ & + \int_0^t \int_0^t h_2(u, v) dN(t-u) dN(t-v) \\ & + \int_0^t \int_0^t h_3(u, v) s(t-u) dN(t-v) + \dots, \end{aligned} \quad [4]$$

where $s(t)$ is the signal at time t , $dN(t)$ is the increment in the counting process, $\beta_S(u)$ is the one-dimensional signal kernel, $\beta_H(t)$ is the one-dimensional temporal or spike history kernel, $h_1(u, v)$ is the 2D signal kernel, $h_2(u, v)$ is the 2D temporal kernel, and $h_3(u, v)$ is the 2D signal–temporal kernel.

Eq. 4 shows that up to first order, the stimulus effect on the spiking activity and the effect of the biophysical properties of the neuron, defined in terms of the neuron’s spiking history, can be expressed as separate components of the conditional intensity function. Assuming that the second-order effects are not strong, then the approximate separation of these two components makes it possible to define the SNR for the signal, also taking account of the effect of the biophysical properties as an additional covariate and vice versa. We expand the log of the conditional intensity function in the Volterra series instead of the conditional intensity function itself in the Volterra series to ensure that the conditional intensity function is positive. In addition, using the log of the conditional intensity function simplifies the GLM formulation by using the canonical link function for the local Poisson model.

Likelihood Analysis Using a PP-GLM. We define the likelihood model for the spike train using the PP-GLM framework (5). We assume the stimulus–response experiment consists of R independent trials, which we index as $r = 1, \dots, R$. We discretize time within a trial by choosing L large and defining the L subintervals $\Delta = T^{-1}L$. We choose L large so that each subinterval contains at most one spike. We index the subintervals $\ell = 1, \dots, L$ and define $n_{r,\ell}$ to be 1 if, on trial r , there is a spike in the subinterval $((\ell-1)\Delta, \ell\Delta)$ and it is 0 otherwise. We let $n_r = (n_{r,1}, \dots, n_{r,L})$ be the set of spikes recorded on trial r in $(0, T]$. Let $H_{r,\ell} = \{n_{r,\ell-j}, \dots, n_{r,\ell-1}\}$ be the relevant history of the spiking activity at time $\ell\Delta$. We define the discrete form of the Volterra expansion by using the first two terms of Eq. 4 to obtain

$$\log \lambda_r(\ell\Delta | H_{r,\ell}, \beta) \approx \beta_0 + \sum_{k=0}^K \beta_{S,k} s_{\ell-k} + \sum_{j=1}^J \beta_{H,j} n_{r,\ell-j}, \quad [5]$$

where $\beta = (\beta_0, \beta_S, \beta_H)'$, $\beta_S = (\beta_{S,0}, \dots, \beta_{S,K})'$, and $\beta_H = (\beta_{H,1}, \dots, \beta_{H,J})'$, and hence the dependence on the stimulus goes back a period of $K\Delta$, whereas the dependence on spiking history goes back a period of $J\Delta$. Exponentiating both sides of Eq. 5 yields

$$\lambda_r(\ell\Delta|H_\ell, \beta) \approx \exp \left\{ \beta_0 + \sum_{k=0}^K \beta_{S,k} s_{\ell-k} + \sum_{j=1}^J \beta_{H,j} n_{r,\ell-j} \right\}. \quad [6]$$

The first and third terms on the right side of Eq. 6 measure the intrinsic spiking propensity of the neuron, whereas the second term measures the effect of the stimulus or signal on the neuron's spiking propensity.

The likelihood function for β given the recorded spike train is (5)

$$L(n, \beta) = \exp \left\{ \sum_{r=1}^R \left[\sum_{\ell=1}^L n_{r,\ell} \log \lambda(\ell\Delta) | \beta, H_{r,\ell} \Delta - \sum_{\ell=1}^L \lambda(\ell\Delta | \beta, H_{r,\ell} \Delta) \right] \right\}. \quad [7]$$

Likelihood formulations with between-trial dependence (32) are also possible but are not considered here.

The maximum likelihood estimate of β can be computed by maximizing Eq. 7 or, equivalently, by minimizing the residual deviance defined as

$$Dev(n, \hat{\beta}) = -2[\log L(n, \hat{\beta}) - \log L(n, n)], \quad [8]$$

where $n = (n_1, \dots, n_R)$ and $L(n, n)$ is the saturated model or the highest possible value of the maximized log likelihood (26). Maximizing $\log L(n, \beta)$ to compute the maximum likelihood estimate of β is equivalent to minimizing the deviance, because $L(n, n)$ is a constant. The deviance is the generalization to the GLM of the sum of squares from the linear Gaussian model (33).

As in the standard GLM framework, these computations are carried out efficiently using iteratively reweighted least squares. In our PP-GLM likelihood analyses, we use Akaike's Information Criterion (AIC) to help choose the order of the discrete kernels β_H and β_S (34). We use the time-rescaling theorem and Kolmogorov–Smirnov (KS) plots (35) along with analyses of the Gaussian transformed interspike intervals to assess model goodness of fit (36). We perform the AIC and time-rescaling goodness-of-fit analyses using cross-validation to fit the model to half of the trials in the experiments (training data set) and then evaluating AIC, the KS plots on the second half the trials (test data set). The model selection and goodness-of-fit assessments are crucial parts of the SNR analyses. They allow us to evaluate whether our key assumption is valid, that is, that the conditional intensity function can be represented as a finite-order Volterra series whose second-order terms can be neglected. Significant lack of fit could suggest that this assumption did not hold and would thereby weaken, if not invalidate, any subsequent inferences and analyses.

SNR Estimates for a Single Neuron. Applying Eq. S11, we have that for a single neuron, the SNR estimate for the signal given the spike history (biophysical properties) with the approximate bias corrections is

$$\hat{SNR}_S = \frac{Dev(n, \hat{\beta}_0, \hat{\beta}_H) - Dev(n, \hat{\beta}) - \dim(\hat{\beta}_0) - \dim(\hat{\beta}_H) + \dim(\hat{\beta})}{Dev(n, \hat{\beta}) + \dim(\hat{\beta})}, \quad [9]$$

and that for a single neuron, the SNR estimates of the spiking propensity given the signal is

$$\hat{SNR}_H = \frac{Dev(n, \hat{\beta}_S) - Dev(n, \hat{\beta}) - \dim(\hat{\beta}_S) + \dim(\hat{\beta})}{Dev(n, \hat{\beta}) + \dim(\hat{\beta})}, \quad [10]$$

where $\dim(\hat{\beta})$ is the dimension or the number of parameters in $\hat{\beta}$. Application of the stimulus activates the biophysical properties of the neuron. Therefore, to measure the effect of the stimulus, we fit the GLM with and without the stimulus and use the difference between the deviances to estimate the \hat{SNR}_S (Eq. 9). Similarly, to measure the effect of the spiking history, we fit the GLM with and without the spike history and use the difference between the deviances to estimate the \hat{SNR}_H (Eq. 10).

Expressed in decibels, the SNR estimates become

$$\hat{SNR}_S^{dB} = 10 \log_{10}(\hat{SNR}_S) \quad [11]$$

$$\hat{SNR}_H^{dB} = 10 \log_{10}(\hat{SNR}_H). \quad [12]$$

Applications

Stimulus–Response Neurophysiological Experiments. To illustrate our method, we analyzed neural spiking activity data from stimulus–response experiments in four neural systems. The stimulus applied in each experiment is a standard one for the neural system being studied. The animal protocols executed in experiments 1–3 were approved by the Institutional Animal Care and Use Committees at the University of Michigan for the guinea pig studies, the University of Pittsburgh for the rat studies, and New York University for the monkey studies. The human studies in experiment 4 were approved by the Human Research Committee at Massachusetts General Hospital.

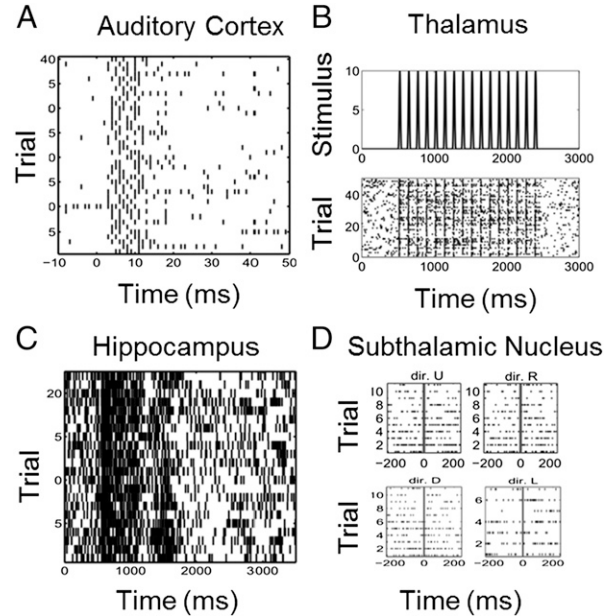


Fig. 1. Raster plots of neural spiking activity. (A) Forty trials of spiking activity recorded from a neuron in the primary auditory cortex of an anesthetized guinea pig in response to a 200 μ s/phase biphasic electrical pulse applied in the inferior colliculus at time 0. (B) Fifty trials of spiking activity from a rat thalamic neuron recorded in response to a 50 mm/s whisker deflection repeated eight times per second. (C) Twenty-five trials of spiking activity from a monkey hippocampal neuron recorded while executing a location scene association task. (D) Forty trials of spiking activity recorded from a subthalamic nucleus neuron in a Parkinson's disease patient before and after a hand movement in each of four directions (dir.): up (dir. U), right (dir. R), down (dir. D), and left (dir. L).

In experiment 1 (Fig. 1A), neural spike trains were recorded from 12 neurons in the primary auditory cortex of anesthetized guinea pigs in response to a 200 μ s/phase biphasic electrical pulse at 44.7- μ A applied in the inferior colliculus (10). Note that the neural recordings were generally multi-unit responses recorded on 12 sites but we refer to them as neurons in this paper. The stimulus was applied at time 0, and spiking activity was recorded from 10 ms before the stimulus to 50 ms after the stimulus during 40 trials. In experiment 2, neural spiking activity was recorded in 12 neurons from the ventral posteromedial (VPM) nucleus of the thalamus (VPM thalamus) in rats in response to whisker stimulation (Fig. 1B) (11). The stimulus was deflection of the whisker at a velocity of 50 mm/s at a repetition rate of eight deflections per second. Each deflection was 1 mm in amplitude and began from the whiskers' neutral position as the trough of a single sine wave and ended smoothly at the same neutral position. Neural spiking activity was recorded for 3,000 ms across 51 trials.

In experiment 3 (Fig. 1C), neural spiking activity was recorded in 13 neurons in the hippocampus of a monkey executing a location scene association task (13). During the experiment, two to four novel scenes were presented along with two to four well-learned scenes in an interleaved random order. Each scene was presented for between 25 and 60 trials. In experiment 4, the data were recorded from 10 neurons in the subthalamic nucleus of human Parkinson's disease patients (Fig. 1D) executing a directed movement task (15). The four movement directions were up, down, left, and right. The neural spike trains were recorded in 10 trials per direction beginning 200 ms before the movement cue and continuing to 200 ms after the cue.

The PP-GLM was fit to the spike trains of each neuron using likelihood analyses as described above. Examples of the model goodness of fit for a neuron from each system is shown in [Supporting Information](#). Examples of the model estimates of the stimulus and history effects for a neuron from each system are shown in Fig. 2.

SNR of Single Neurons. We found that the $\hat{\text{SNR}}_S^{dB}$ estimates (Eq. 11) of the stimulus controlling for the effect of the biophysical model

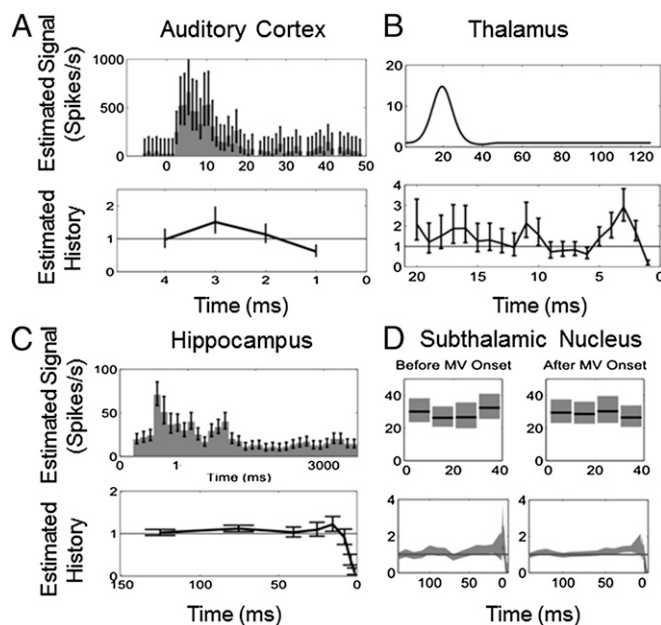


Fig. 2. Stimulus and history component estimates from the PP-GLM analyses of the spiking activity in Fig. 1. (A) Guinea pig primary auditory cortex neuron. (B) Rat thalamic neuron. (C) Monkey hippocampal neuron. (D) Human subthalamic nucleus neuron. The stimulus component (Upper) is the estimated stimulus-induced effect on the spike rate in A, C, and D and the impulse response function of the stimulus in B. The history components (Lower) show the modulation constant of the spike firing rate.

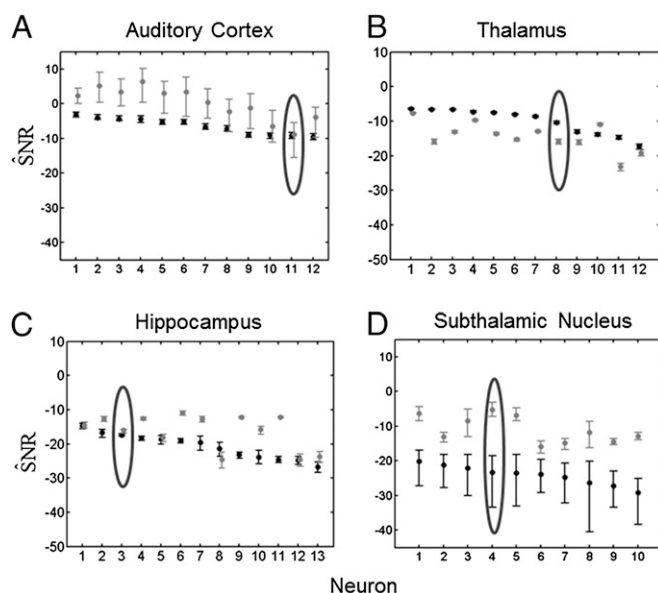


Fig. 3. KL-based SNR for (A) 12 guinea pig auditory cortex neurons, (B) 12 rat thalamus neurons, (C) 13 monkey hippocampal neurons, and (D) 10 subthalamic nucleus neurons from a Parkinson's disease patient. The black dots are $\hat{\text{SNR}}_S^{dB}$, the SNR estimates due to the stimulus correcting for the spiking history. The black bars are the 95% bootstrap confidence intervals for $\hat{\text{SNR}}_S^{dB}$. The gray dots are $\hat{\text{SNR}}_H^{dB}$, the SNR estimates due to the intrinsic biophysics of the neuron correcting for the stimulus. The gray bars are the 95% bootstrap confidence intervals for $\hat{\text{SNR}}_H^{dB}$. The encircled points are the SNR and 95% confidence intervals for the neural spike train raster plots in Fig. 1.

properties were (median [minimum, maximum]): -6 dB [-10 dB, -3 dB] for guinea pig auditory cortex neurons; -9 dB [-18 dB, -7 dB] for rat thalamic neurons; -20 dB [-28 dB, -14 dB] for the monkey hippocampus; and -23 dB [-29 dB, -20 dB] for human subthalamic neurons (Fig. 3, black bars). The higher SNRs (from Eq. 11) in experiments 1 and 2 (Fig. 3A and B) are consistent with the fact that the stimuli are explicit, i.e., an electrical current and mechanical displacement of the whisker, respectively, and that the recording sites are only two synapses away from the stimulus. It is also understandable that SNRs are smaller for the hippocampus and thalamic systems in which the stimuli are implicit, i.e., behavioral tasks (Fig. 3C and D).

We found that $\hat{\text{SNR}}_H^{dB}$ estimates (from Eq. 12) of the biophysical properties controlling for the stimulus effect were: 2 dB [-9 dB, 7 dB] for guinea pig auditory cortex; -13 dB [-22 dB, -8 dB] for rat thalamic neurons; -15 dB [-24 dB, -11 dB] for the monkey hippocampal neurons; and -12 dB [-16 dB, -5 dB] for human subthalamic neurons (Fig. 3, gray bars). They were greater than $\hat{\text{SNR}}_S^{dB}$ for the guinea pig auditory cortex (Fig. 3A), the monkey hippocampus (Fig. 3C), and the human subthalamic experiments (Fig. 3D), suggesting that the intrinsic spiking propensities of neurons are often greater than the spiking propensity induced by applying a putatively relevant stimulus.

A Simulation Study of Single-Neuron SNR Estimation. To analyze the performance of our SNR estimation paradigm, we studied simulated spiking responses of monkey hippocampal neurons with specified stimulus and history dynamics. We assumed four known SNRs of -8.3 dB, -17.4 dB, -28.7 dB, and $-\infty$ dB corresponding, respectively, to stimulus effects on spike rates ranges of 500, 60, 10, and 0 spikes per second (Fig. 4, row 1). For each of the stimulus SNRs, we assumed spike history dependence (Fig. 4, row 2) to be similar to that of the neuron in Fig. 1C. For each of four stimulus effects, we simulated 300 experiments, each consisting of 25 trials (Fig. 4, row 3). To each of the 300 simulated data sets at each SNR level, we applied our SNR estimation paradigm: model

performance of our SNR estimate is attributable to the fact that it is based on the likelihood, whereas the variance-based SNR estimate uses only the first two sample moments of the data. This improvement is no surprise, as it is well known that likelihood-based estimates offer the best information summary in a sample given an accurate or approximately statistical model (34). We showed that for each of the four neural systems, the PP-GLM accurately described the spike train data in terms of goodness-of-fit assessments.

A General Paradigm for SNR Estimation. Our SNR estimation paradigm generalizes the approach commonly used to analyze SNRs in linear Gaussian systems. We derived the generalization by showing that the commonly computed SNR statistic estimates a ratio of EPEs (*Supporting Information*): the expected prediction of the error of the signal representing the data corrected for the nonsignal covariates relative to the EPE of the system noise. With this insight, we used the work of ref. 26 to extend the SNR definition to systems that can be modeled using the GLM framework in which the signal and relevant covariates can be expressed as separate components of the likelihood function. The linear Gaussian model is a special case of a GLM. In the GLM paradigm, the sum of squares from the standard linear Gaussian model is replaced by the residual deviance (Eq. S10). The residual deviance may be viewed as an estimated KL divergence between data and model (26). To improve the accuracy of our SNR estimator, particularly given the low SNRs of single neurons, we devised an approximate bias correction, which adjusts separately the numerator and the denominator (Eqs. 9 and 10). The bias-corrected estimator performed well in the limited simulation study we reported (Figs. 4 and 5). In future work, we will replace the separate bias corrections

for the numerator and denominator with a single bias correction for the ratio, and extend our paradigm to characterize the SNR of neuronal ensembles and those of other non-Gaussian systems.

In *Supporting Information*, we describe the relationship between our SNR estimate and several commonly used quantities in statistics, namely the R^2 , coefficient of determination, the F statistic, the likelihood ratio (LR) test statistic and f^2 , Cohen's effect size. Our SNR analysis offers an interpretation of the F statistic that is not, to our knowledge, commonly stated. The F statistic may be viewed as a scaled estimate of the SNR for the linear Gaussian model, where the scale factor is the ratio of the degrees of freedom (Eq. S21). The numerator of our GLM SNR estimate (Eq. S9) is a LR test statistics for assessing the strength of the association between data Y and covariates X_2 . The generalized SNR estimator can be seen as generalized effect size. This observation is especially important because it can be further developed for planning neurophysiological experiments, and thus may offer a way to enhance experimental reproducibility in systems neuroscience research (39).

In summary, our analysis provides a straightforward way of assessing the SNR of single neurons. By generalizing the standard SNR metric, we make explicit the well-known fact that individual neurons are noisy transmitters of information.

ACKNOWLEDGMENTS. This research was supported in part by the Clinical Eye Research Centre, St. Paul's Eye Unit, Royal Liverpool and Broadgreen University Hospitals National Health Service Trust, United Kingdom (G.C.); the US National Institutes of Health (NIH) Biomedical Research Engineering Partnership Award R01-DA015644 (to E.N.B. and W.A.S.), Pioneer Award DP1 OD003646 (to E.N.B.), and Transformative Research Award GM 104948 (to E.N.B.); and NIH grants that supported the guinea pig experiments P41 EB2030 and T32 DC00011 (to H.H.L.).

- Chen Y, Beaulieu N (2007) Maximum likelihood estimation of SNR using digitally modulated signals. *IEEE Trans Wirel Comm* 6(1):210–219.
- Kay S (1993) *Fundamentals of Statistical Signal Processing* (Prentice-Hall, Upper Saddle River, NJ).
- Welvaert M, Rosseel Y (2013) On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data. *PLoS ONE* 8(11):e77089.
- Mendel J (1995) *Lessons in Estimation Theory for Signal Processing, Communications, and Control* (Prentice-Hall, Upper Saddle River, NJ), 2nd Ed.
- Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN (2005) A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J Neurophysiol* 93(2):1074–1089.
- Brillinger DR (1988) Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol Cybern* 59(3):189–200.
- Brown EN, Barbieri R, Eden UT, Frank LM (2003) Likelihood methods for neural data analysis. *Computational Neuroscience: A Comprehensive Approach*, ed Feng J (CRC, London), pp 253–286.
- Brown EN (2005) Theory of point processes for neural systems. *Methods and Models in Neurophysiology*, eds Chow CC, Gutkin B, Hansel D, Meunier C, Dalibard J (Elsevier, Paris), pp 691–726.
- MacEvoy SP, Hanks TD, Paradiso MA (2008) Macaque V1 activity during natural vision: Effects of natural scenes and saccades. *J Neurophysiol* 99(2):460–472.
- Lim HH, Anderson DJ (2006) Auditory cortical responses to electrical stimulation of the inferior colliculus: Implications for an auditory midbrain implant. *J Neurophysiol* 96(3):975–988.
- Temereanca S, Brown EN, Simons DJ (2008) Rapid changes in thalamic firing synchrony during repetitive whisker stimulation. *J Neurosci* 28(44):11153–11164.
- Wilson MA, McNaughton BL (1993) Dynamics of the hippocampal ensemble code for space. *Science* 261(5124):1055–1058.
- Wirth S, et al. (2003) Single neurons in the monkey hippocampus and learning of new associations. *Science* 300(5625):1578–1581.
- Dayan P, Abbott LF (2001) *Theoretical Neuroscience* (Oxford Univ Press, London).
- Sarma SV, et al. (2010) Using point process models to compare neural spiking activity in the subthalamic nucleus of Parkinson's patients and a healthy primate. *IEEE Trans Biomed Eng* 57(6):1297–1305.
- Rieke F, Warland D, de Ruyter van Steveninck RR, Bialek W (1997) *Spikes: Exploring the Neural Code* (MIT Press, Cambridge, MA).
- Optican LM, Richmond BJ (1987) Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. *J Neurophysiol* 57(1):162–178.
- Softky WR, Koch C (1993) The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J Neurosci* 13(1):334–350.
- Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *J Neurosci* 18(10):3870–3896.
- Lyamzin DR, Macke JH, Lesica NA (2010) Modeling population spike trains with specified time-varying spike rates, trial-to-trial variability, and pairwise signal and noise correlations. *Front Comput Neurosci* 4(144):144.
- Soofi ES (2000) Principal information theoretic approaches. *Am Stat* 95(452):1349–1353.
- Erdogmus D, Larsson EG, Yan R, Principe JC, Fitzsimmons JR (2004) Measuring the signal-to-noise ratio in magnetic resonance imaging: A caveat. *Signal Process* 84(6):1035–1040.
- Mittlboeck M, Waldh r T (2000) Adjustments for R^2 -measures for Poisson regression models. *Comput Stat Data Anal* 34(4):461–472.
- Kent JT (1983) Information gain and a general measure of correlation. *Biometrika* 70(1):163–173.
- Alonso A, et al. (2004) Prentice's approach and the meta-analytic paradigm: A reflection on the role of statistics in the evaluation of surrogate endpoints. *Biometrics* 60(3):724–728.
- Hastie T (1987) A closer look at the deviance. *Am Stat* 41(1):16–20.
- Simon G (1973) Additivity of information in exponential family probability laws. *J Am Stat Assoc* 68(342):478–482.
- Cameron AC, Windmeijer FAG (1997) An R -squared measure of goodness of fit for some common nonlinear regression models. *J Econom* 77(2):329–342.
- Paninski L (2004) Maximum likelihood estimation of cascade point-process neural encoding models. *Network* 15(4):243–262.
- Marmarelis VZ (2004) *Nonlinear Dynamic Modeling of Physiological Systems* (John Wiley, Hoboken, NJ).
- Plourde E, Delgutte B, Brown EN (2011) A point process model for auditory neurons considering both their intrinsic dynamics and the spectrotemporal properties of an extrinsic signal. *IEEE Trans Biomed Eng* 58(6):1507–1510.
- Czanner G, Sarma SV, Eden UT, Brown EN (2008) A signal-to-noise ratio estimator for generalized linear model systems. *Lect Notes Eng Comput Sci* 2171(1):1063–1069.
- McCullagh P, Nelder JA (1989) *Generalized Linear Models* (Chapman and Hall, New York).
- Pawitan Y (2013) *In All Likelihood. Statistical Modelling and Inference Using Likelihood* (Oxford Univ Press, London).
- Brown EN, Barbieri R, Ventura V, Kass RE, Frank LM (2002) The time-rescaling theorem and its application to neural spike train data analysis. *Neural Comput* 14(2):325–346.
- Czanner G, et al. (2008) Analysis of between-trial and within-trial neural spiking dynamics. *J Neurophysiol* 99(5):2672–2693.
- Sekhar SC, Sreenivas TV (2006) Signal-to-noise ratio estimation using higher-order moments. *Signal Process* 86(4):716–732.
- Cajigas I, Malik WQ, Brown EN (2012) nSTAT: Open-source neural spike train analysis toolbox for Matlab. *J Neurosci Methods* 211(2):245–264.
- Collins FS, Tabak LA (2014) Policy: NIH plans to enhance reproducibility. *Nature* 505(7485):612–613.

Supporting Information

Czanner et al. 10.1073/pnas.1505545112

What Does the SNR Estimate?

The SNR has been most studied for linear model systems $Y = X\beta + \varepsilon$ in which one has observations $y = (y_1, \dots, y_n)'$ of a random vector $Y = (Y_1, \dots, Y_n)'$, $X\beta$ is the signal, $X = (x_1, \dots, x_p)$ is the $n \times p$ design matrix, x_k are fixed known vectors of covariates ($k = 1, \dots, p$), β is a $p \times 1$ vector of unknown coefficients, and ε is an $n \times 1$ vector of independent, identically distributed Gaussian random errors with zero mean and variance σ_ε^2 . The first column in X is a $n \times 1$ vector of 1s denoted as 1_n . The unconditional mean of the random vector Y can be defined as $EY = 1_n\beta_0$, where scalar parameter β_0 is typically unknown.

A standard way to define the SNR is as a ratio of variances as

$$\text{SNR}_X = \frac{\sigma_{\text{signal}}^2}{\sigma_{\text{noise}}^2}, \quad [\text{S1}]$$

where σ_{signal}^2 is the variance of the signal representing the expected variability in the data induced by the signal, where

$$\sigma_{\text{signal}}^2 = (X\beta - 1_n\beta_0)'(X\beta - 1_n\beta_0),$$

and $\sigma_{\text{noise}}^2 = n\sigma_\varepsilon^2$ is the variance of noise. In other words, SNR is the true expected proportion of variance in the data due the signal divided by the variance due to the noise.

We can obtain an alternative interpretation of SNR (Eq. S1) if we view the two variances in terms of EPEs in the squared error sense. The variance of the noise can be viewed as the expected error of predicting Y when using covariates X (1). That is,

$$\sigma_{\text{noise}}^2 = \text{EPE}(Y, X\beta) = E[(Y - X\beta)'(Y - X\beta)]. \quad [\text{S2}]$$

Analogously, the expected error of predicting Y when using $1_n\beta_0$ is

$$\text{EPE}(Y, 1_n\beta_0) = E[(Y - 1_n\beta_0)'(Y - 1_n\beta_0)]. \quad [\text{S3}]$$

Due to the Pythagorean property of EPE in linear Gaussian system (1–3), at the parameter values β_0 and β that minimize the EPE, the variance of the signal can be expressed as

$$\sigma_{\text{signal}}^2 = \text{EPE}(Y, 1_n\beta_0) - \text{EPE}(Y, X\beta),$$

the EPE of predicting the values of Y with overall mean, $1_n\beta_0$ minus the EPE of predicting Y with the approximating model, $X\beta$ (4). Hence, σ_{signal}^2 is the reduction in the EPE achieved by using the covariates X . This leads to an alternative definition of the SNR as

$$\text{SNR}_X = \frac{\text{EPE}(Y, 1_n\beta_0) - \text{EPE}(Y, X\beta)}{\text{EPE}(Y, X\beta)}, \quad [\text{S4}]$$

which is the reduction EPE due to the signal, divided by the EPE due to noise, in the squared-error sense. For this reason, we will refer to Eqs. S1 and S4 as a variance-based or a squared-error-based SNR.

The variance-based SNR_X is the true expected SNR obtained if the parameters β and β_0 that give the minimum EPEs are known. In practice, however, the SNR_X is estimated by replacing the parameters β and β_0 by their least-squares estimates $\hat{\beta}$ and \bar{y} , respectively. This leads to the estimate of SNR_X (Eq. S4)

$$\hat{\text{SNR}}_X = \frac{SS\text{Residual}(y, 1_n\bar{y}) - SS\text{Residual}(y, X\hat{\beta})}{SS\text{Residual}(y, X\hat{\beta})} \quad [\text{S5}]$$

where

$$SS\text{Residual}(y, 1_n\bar{y}) = (y - 1_n\bar{y})'(y - 1_n\bar{y})$$

$$SS\text{Residual}(y, X\hat{\beta}) = (y - X\hat{\beta})'(y - X\hat{\beta}).$$

In linear model analyses, $SS\text{Residual}(y, 1_n\bar{y})$ is the variance of the data around their estimated overall mean and $SS\text{Residual}(y, X\hat{\beta})$ is the estimated variability in the data around the estimated signal $X\hat{\beta}$, i.e., the variability that is not explained by the covariate X .

Defining the SNR for a Linear Gaussian Signal Plus Covariates Plus Noise System

If the system is driven by a signal and a nonsignal component and if the two components can be separated by an approximate linear additive model, then the SNR definition and estimate must be modified. We assume the covariate component of the linear model $Y = X\beta + \varepsilon$ can be partitioned as $X\beta = X_1\beta_1 + X_2\beta_2$, where the first component, $X_1\beta_1$, is a covariate not related to the signal, and the second component, $X_2\beta_2$, is the signal. There exist values of vectors β , β_1 , and β_2 that give the minimum EPEs for describing Y in terms of minimizing $\text{EPE}(Y, X\beta)$, $\text{EPE}(Y, X_1\beta_1)$, and $\text{EPE}(Y, X_2\beta_2)$, respectively. For this case, we can define SNR in which only a part of the variability in random vector Y is attributed to the signal to extend the SNR definition in Eq. S1 by replacing $\text{EPE}(Y, 1_n\beta_0)$ with $\text{EPE}(Y, X_1\beta_1)$ to obtain

$$\text{SNR}_{X_2} = \frac{\text{EPE}(Y, X_1\beta_1) - \text{EPE}(Y, X\beta)}{\text{EPE}(Y, X\beta)} \quad [\text{S6}]$$

where the first column of X_1 and of X is the vector 1_n . Eq. S6 gives the expected SNR in Y about the signal, $X_2\beta_2$, while controlling for the effect of nonsignal component, $X_1\beta_1$. The numerator in Eq. S6 is the reduction in the EPE due to the signal, $X_2\beta_2$, when controlling for $X_1\beta_1$, the systematic changes in random vector Y unrelated to the signal whereas the denominator is the EPE due to the noise. By analogy with Eq. S5, we can estimate the squared-error based SNR_{X_2} (Eq. S6) as

$$\hat{\text{SNR}}_{X_2} = \frac{SS\text{Residual}(y, X_1\hat{\beta}_1) - SS\text{Residual}(y, X\hat{\beta})}{SS\text{Residual}(y, X\hat{\beta})}, \quad [\text{S7}]$$

where we replace $SS\text{Residual}(y, 1_n\bar{y})$ with $SS\text{Residual}(y, X_1\hat{\beta}_1)$ in Eq. S5.

Defining the SNR for GLM Systems

The SNR definition and estimate in Eqs. S6 and S7 extend to the GLM framework, the established statistical paradigm for conducting regression analyses when data from the exponential family are observed with covariates (5). We extend SNR to GLM systems in which the covariates may be partitioned into signal and nonsignal components by replacing the squared-error EPE in Eq. S6 with the KL EPE of Y from the approximating model and by replacing the residual sums of squares in Eq. S7 by the residual deviances (1, 3, 5). This leads to the following KL generalization of the true SNR in Y about the signal X_2 , while taking

into account the nonsignal effects X_1 , for the system approximated by the GLM

$$\text{SNR}_{X_2} = \frac{\text{EPE}(Y, X_1\beta_1) - \text{EPE}(Y, X\beta)}{\text{EPE}(Y, X\beta)}, \quad [\text{S8}]$$

and its deviance-based estimate

$$\hat{\text{SNR}}_{X_2} = \frac{\text{Dev}(y, X_1\hat{\beta}_1) - \text{Dev}(y, X\hat{\beta})}{\text{Dev}(y, X\hat{\beta})} \quad [\text{S9}]$$

$$\text{EPE}(Y, X_1\beta_1) = E[-2 \log f(Y|X_1\beta_1)]$$

where the expectation is taken with respect to true generating probability distribution of random vector Y and the “2” in the definition makes the log-likelihood loss for the Gaussian distribution match squared-error loss. For this reason we refer to Eq. S8 as a KL-based SNR and Eq. S9 as its KL- or deviance-based SNR estimator.

The deviance is

$$\text{Dev}(y, X\hat{\beta}) = -2 \log \frac{L(y, X\hat{\beta})}{L(y, y)} \quad [\text{S10}]$$

where $L(y, X\hat{\beta})$ is the likelihood evaluated at the maximum likelihood estimate $\hat{\beta}$ of the model parameter β . $L(y, y)$ is the saturated likelihood defined as the highest value of the likelihood (5).

By the Pythagorean property of the KL divergence estimate in a GLM with canonical link (1–3), the numerator in Eq. S8 is the reduction in KL EPE due to the signal, $X_2\beta_2$, while controlling for the effect of the nonsignal component, $X_1\beta_1$. The KL-based SNR_{X_2} has squared error-based SNR_{X_2} as a special case in which the exponential family model has the Gaussian distribution. The numerator of the SNR estimate (Eq. S9) gives the reduction in deviance due to signal, $X_2\hat{\beta}_2$, while controlling for the nonsignal component, $X_1\hat{\beta}_1$. The estimates $\hat{\beta}$ and $\hat{\beta}_1$ are computed from two separate maximum-likelihood fits of the two models to data y (6).

We define a bias correction for the SNR estimator (Eq. S9), as this problem is especially prevalent in data with a weak signal (4, 7). By definition, the SNR estimate is always positive. Under regularity conditions, the asymptotic biases of the numerator and denominator in Eq. S9 are respectively $\dim(\beta_1) - \dim(\beta)$ and $\dim(\beta)$, suggesting the approximate bias-corrected SNR estimate

$$\hat{\text{SNR}}_{X_2} = \frac{\text{Dev}(y, X_1\hat{\beta}_1) - \text{Dev}(y, X\hat{\beta}) + \dim(\beta_1) - \dim(\beta)}{\text{Dev}(y, X\hat{\beta}) + \dim(\beta)}. \quad [\text{S11}]$$

This SNR estimate remains biased because a ratio of unbiased estimators is not necessarily an unbiased estimator of the ratio. Our simulation studies in Fig. 4 (rows 4 and 5) suggest that the bias is small for neural spike trains (4).

Variance-Based and KL-Based SNR Are the Same in Linear Systems with Independent and Additive Gaussian Noise

We assume that y_1, \dots, y_n is a realization of independent random variables Y_1, \dots, Y_n , from a linear regression model, with means $E[Y_i|X_i] = X_i\beta$, zero covariances, and a common random error variance, σ_e^2 . We also assume overall (unconditional) mean $E[Y_i] = \beta_0$. Furthermore, we assume a reduced model β_1 , i.e., $\beta_1 \subset \beta$. An example of reduced model is a model with the overall mean, β_0 , representing the background firing constant (see Eq. 6), or a model with parameter vector β_1 for background firing constant and for nonsignal covariates. The full model is always the generating model or a good approximating model.

Then, under the above assumptions, the divergence between data y_1, \dots, y_n and the model $X\beta_1$ is

$$\text{KL}(y_1..y_n, X\beta_1) = (y - X\beta_1)^T (y - X\beta_1), \quad [\text{S12}]$$

and, assuming the vector value β_1 that minimizes EPE, the mean is equal to

$$\begin{aligned} \text{EPE}_{\text{KL}}(Y_1..Y_n, X_1\beta_1) &= E[\text{KL}(Y_1..Y_n, X\beta_1)] = \sum E(Y_i - X_i\beta_1)^2 \\ &= \text{EPE}_{\text{SE}}(Y_1..Y_n, X_1\beta_1) \end{aligned} \quad [\text{S13}]$$

i.e., KL-based EPE reduces to squared-error-based EPE for the Gaussian linear system with independent noise. Furthermore,

$$\begin{aligned} \sum E(Y_i - X_i\beta_1)^2 &= \sum E(Y_i - X_i\beta_1)^2 + \sum (X_i\beta - X_i\beta_1)^2 \\ &= n\sigma_e^2 + (X\beta - X\beta_1)^T (X\beta - X\beta_1) \end{aligned} \quad [\text{S14}]$$

where Y_i and X_i are i th component of Y and i th row of X , respectively. Hence, for a linear Gaussian system, we have $\text{EPE}_{\text{KL}}(Y_1..Y_n, X_1\beta_1) = n\sigma_e^2 + \sum (X_i\beta - X_i\beta_1)^2$, with a special case being $\beta_1 = \beta$ that gives $\text{EPE}_{\text{KL}}(Y_1..Y_n, X\beta) = n\sigma_e^2 = \sigma_{\text{noise}}^2$. If we substitute this into Eq. S6, we obtain

$$\begin{aligned} \text{SNR}_{X_1} &= \frac{\text{EPE}_{\text{KL}}(Y, X_1\beta_1) - \text{EPE}_{\text{KL}}(Y, X\beta)}{\text{EPE}_{\text{KL}}(Y, X\beta)} \\ &= \frac{n\sigma_e^2 + \sum (X_i\beta - X_i\beta_1)^2 - n\sigma_e^2}{n\sigma_e^2} \\ &= \frac{(X\beta - X\beta_1)^T (X\beta - X\beta_1)}{\sigma_{\text{noise}}^2}. \end{aligned}$$

That is,

$$\text{SNR}_{\text{KL}, X_1} = \text{SNR}_{\text{SE}, X_1} \quad [\text{S15}]$$

in systems that are linear with additive, independent, and Gaussian noise. Lastly, for completeness, we note here that the scale parameter of a linear Gaussian system is $\phi = \sigma_e^2$.

Variance-Based and KL-Based SNR Are Not the Same for Independent Binomial Observations

We assume that data y_1, \dots, y_L are recorded at 1-ms resolution and that they are realizations of independent random variables Y_1, \dots, Y_L , from a Bernoulli distribution with parameters K and p_l , $l = 1, \dots, L$ i.e., their means are $K \times p_l$ and the variances are $K \times p_l \times (1 - p_l)$. Then the overall expected probability of an event (such as a spike) is $p = L^{-1} \sum K \times p_l$ and the total variance is $\sum \text{Var}(y_l) = \sum K \times p_l \times (1 - p_l)$, and hence the squared-error-based SNR (Eqs. S1 and S4) can be shown to be

$$\text{SNR} = \frac{\sum (K \times p_l - K \times p)^2}{\sum K \times p_l \times (1 - p_l)}. \quad [\text{S16}]$$

The formula in Eq. S16 can be used to calculate SNR for spike trains when spikes trains are independent across trials, and when times of spikes are independent within each trial, such as when there is no spike history dependence. Then, one can summarize the data into a 1-ms peristimulus time histogram, which can be seen as a realization of independent Binomial random variables. Then the SNR numerator in Eq. S16 is the variance of signal, and the denominator contains the sum of variances of Binomial random variables across L bins. This idea was used in ref. 8, and

it was extended to incorporate the spike history, which was estimated in a sequential manner rather than in one single analysis. Nevertheless, our simulations in Fig. 5 indicate that expected variance-based SNR (Eq. S16) is smaller than KL-based expected SNR (Eq. S8).

Variance-Based SNR and the Coefficient-of-Determination

In linear models with Gaussian noise, the coefficient of determination, R^2 , is a commonly used measure of the fit of the model to the data. The coefficient of determination ranges from 0 to 1, with 1 indicating perfect fit. Specifically,

$$R^2 = \frac{(1_n \bar{y} - X \hat{\beta})^T (1_n \bar{y} - X \hat{\beta})}{(y - 1_n \bar{y})^T (y - 1_n \bar{y})} = \frac{SS_{Residual}(y, 1_n \bar{y}) - SS_{Residual}(y, X \hat{\beta})}{SS_{Residual}(y, 1_n \bar{y})}, \quad [\text{S17}]$$

i.e., the numerators of SNR estimator (see Eq. S5) and of R^2 are the same, and it is the sum-of-squares explained by the model (i.e., the signal), and it is often referred to as SS_{Model} or $SS_{Regression}$ in statistical software output. The denominators of R^2 and SNR are different. The denominator of R^2 is the sum of squares around the grand mean, $SS_{Residual}(y, 1_n \bar{y})$, representing the total variability in the data and hence often referred to as SS_{Total} . On the other hand, the variability of the data around the estimated linear function is summarized in the term $SS_{Residual}(y, X \hat{\beta})$, which is often referred to in the statistical software as $SS_{Residual}$. In summary, the R^2 can be written as

$$R^2 = \frac{SS_{Model}}{SS_{Total}} = \frac{SS_{Model}}{SS_{Model} + SS_{Residual}}, \quad [\text{S18}]$$

and we have that

$$\hat{\text{SNR}} = \frac{SS_{Model}}{SS_{Residual}}.$$

It follows that

$$1/R^2 = \frac{SS_{Model} + SS_{Residual}}{SS_{Model}} = 1 + \frac{SS_{Residual}}{SS_{Model}} = 1 + 1/\hat{\text{SNR}}$$

and that

$$\hat{\text{SNR}} = \begin{cases} \frac{R^2}{1 - R^2} & \text{if } R^2 \neq 1 \\ \text{Inf} & \text{if } R^2 = 1 \\ 0 & \text{if } R^2 = 0 \end{cases} \quad [\text{S19}]$$

$$\hat{\text{SNR}}^{dB} = \begin{cases} 10 \log_{10} \left(\frac{R^2}{1 - R^2} \right) & \text{if } R^2 \neq 1 \\ \text{Inf} & \text{if } R^2 = 1 \\ -\text{Inf} & \text{if } R^2 = 0 \\ 0 & \text{if } R^2 = 0.5 \end{cases}. \quad [\text{S20}]$$

Hence, by Eq. S20, we have that squared-error-based $\hat{\text{SNR}}$ is an increasing function of R^2 (Fig. S1). Furthermore, both quantities R^2 and SNR decrease with increasing level of noise (Fig. S2).

A well-known problem with R^2 is that it always increases, even if unimportant covariates are added to the model. Hence an adjusted R^2 was proposed (6, 9) that adjusts for the number of explanatory terms in a model. Unlike R^2 , the adjusted R^2 in-

creases only if the new term improves the model more than would be expected by chance. The adjusted R^2 can be negative—just like bias-adjusted SNR—and will always be less than or equal to the R^2 . While R^2 is a measure of fit, the adjusted R^2 is used for comparison of nested models and for feature (i.e., variable) selection in model building and machine learning. By analogy, the adjusted SNR can also be used for feature selection in biological systems to quantify the amount of information in features.

There are many generalizations of R^2 for GLM models (called pseudo- R^2). Some generalizations are based on likelihoods (9, 10). Their bias-adjusted versions for independent data are known and implemented in statistical software (e.g., statistical software R). These bias-adjusted pseudo- R^2 measures can be directly used to obtain the bias-adjusted SNR via Eq. S20. However, even if an unbiased R^2 estimate is used in Eq. S20 under the assumption that the data are independent, then the SNR estimate can still be biased because the ratio of unbiased estimates is not necessarily an unbiased estimate.

Variance-Based SNR and F-Test Statistic

In linear regression models with independent Gaussian errors, the F test is a commonly used test to evaluate the importance of a set of covariates, X , in explaining the variability of dependent variable, Y . The F -test statistic has the form

$$F = \frac{SS_{Model}/df(\text{Model})}{SS_{Residual}/df(\text{Residual})}$$

where $df(\text{Model}) = k - 1$, $df(\text{Residual}) = n - k - 1$ are degrees of freedom of the model and residuals, and k is the number of covariates (i.e., the number of columns of X). Hence, using Eq. S5,

$$F = \hat{\text{SNR}} \times \frac{df(\text{Residual})}{df(\text{Model})} \quad [\text{S21}]$$

i.e., the bias-unadjusted SNR estimate Eq. S5 is a multiple of the F statistic.

If there is no signal, then $(\sigma_{\text{signal}}^2 / \sigma_{\text{noise}}^2) = 0$, i.e., $\text{SNR} = 0$ (in Eq. S1). In this case, none of the covariates in matrix X is related to Y . In other words, the true generating model is a model with a constant only. In this case, the F statistic has a central Fisher distribution with degrees of freedom $df(\text{Model})$ and $df(\text{Residual})$. It is easy to see that the mean of the F statistic (if $df(\text{Residual}) > 2$) is

$$E(F) = \frac{df(\text{Residual})}{df(\text{Residual}) - 2}$$

and hence, when there is no signal, it follows from Eq. S21 and properties of the central F distribution that the mean of the variance-based $\hat{\text{SNR}}$ is

$$E(\hat{\text{SNR}}) = \frac{df(\text{Residual})}{df(\text{Residual}) - 2} \times \frac{df(\text{Model})}{df(\text{Residual})} = \frac{df(\text{Model})}{df(\text{Residual}) - 2},$$

while the true $\text{SNR} = 0$; hence the bias of $\hat{\text{SNR}}$ is $df(\text{Model}) / [df(\text{Residual}) - 2]$, which converges to zero when the ratio of data size to number of parameters becomes large.

In the general case, when the true variance-based $\text{SNR} \neq 0$, then the associated F statistic (Eq. S21) has a noncentral Fisher distribution with degrees of freedom $df(\text{Model})$ and $df(\text{Residual})$ and with a noncentrality parameter equal to $\sigma_{\text{signal}}^2 / \sigma_{\text{e}}^2 = n \times \text{SNR}$. In such a case, it can be shown that

$$E[\hat{\text{SNR}}] = E\left[F \frac{df(\text{Model})}{n - df(\text{Model}) - 1}\right]$$

$$= \frac{df(\text{Model})}{df(\text{Residual}) - 2} + \text{SNR} \frac{n}{df(\text{Residual}) - 2}$$

and the confidence intervals for SNR can be constructed using quantiles of noncentral Fisher distribution (11).

The equivalent theory for the bias correction and confidence intervals of SNR is not available in GLM models with history dependence. Therefore, here we offered a simple bias correction Eq. S11 that removes some bias, and we showed that it can work well in simulations. However, it can be proved that our bias correction is asymptotically equivalent to the bias correction above for independent data from linear Gaussian model.

SNR and LR Test

The concept of SNR is also related to the concept of the LR test (5). Specifically, the scaled numerator of the generalized SNR estimate (Eq. S9) is an LR test statistic for testing the association

between covariates X and variable Y in GLMs. Under independence of the observations, the LR test statistics have asymptotically χ^2 distributions with degrees of freedom equal to the number of estimated parameters associated with the covariates. Hence, low levels of LR lead to the conclusion that there is not enough evidence for the association, which corresponds to low values of SNR estimate (Eq. S9).

Variance-Based SNR and Effect Size for Linear Regression

Another related measure is effect size. Cohen's effect size for linear regression models (6, 12), defined as $f^2 = R^2 / (1 - R^2)$, is the same as the squared-error-based SNR in Eq. S5. Cohen's f^2 is not typically reported in studies, but it is often used for sample size calculations in linear regression. For linear regressions, effect sizes of 0.02, 0.15, and 0.35 are considered small, medium, and large, respectively. These three effect sizes correspond to an R^2 of 0.02, 0.13, and 0.26 SNR of -17 dB, -8.2 dB, and -4.6 dB, which are consistent with the SNR values that we reported for some of the neurons.

- Hastie T (1987) A closer look at the deviance. *Am Stat* 41(1):16–20.
- Simon G (1973) Additivity of information in exponential family probability laws. *J Am Stat Assoc* 68(342):478–482.
- Cameron AC, Windmeijer FAG (1997) An R -squared measure of goodness of fit for some common nonlinear regression models. *J Econom* 77(2):329–342.
- Czanner G, Sarma SV, Eden UT, Brown EN (2008) A signal-to-noise ratio estimator for generalized linear model systems. *Lect Notes Eng Comput Sci* 2171(1):1063–1069.
- McCullagh P, Nelder JA (1989) *Generalized Linear Models* (Chapman and Hall, New York).
- Cohen J (1992) A power primer. *Psychol Bull* 112(1):155–159.
- Chen Y, Beaulieu N (2007) Maximum likelihood estimation of SNR using digitally modulated signals. *IEEE Trans Wirel Comm* 6(1):210–219.
- Lyamzin DR, Macke JH, Lesica NA (2010) Modeling population spike trains with specified time-varying spike rates, trial-to-trial variability, and pairwise signal and noise correlations. *Front Comput Neurosci* 4(144):144.
- Agresti A, Caffo B (2002) Measures of relative model fit. *Comput Stat Data Anal* 39:127–136.
- Nagelkerke NJD (1991) A note on a general definition of the coefficient of determination. *Biometrika* 78(3):691–692.
- Kelley K (2007) Confidence intervals for standardized effect sizes: Theory, application, and implementation. *J Stat Softw* 20(8):1–24.
- Fritz CO, Morris PE, Richler JJ (2012) Effect size estimates: current use, calculations, and interpretation. *J Exp Psychol Gen* 141(1):2–18.

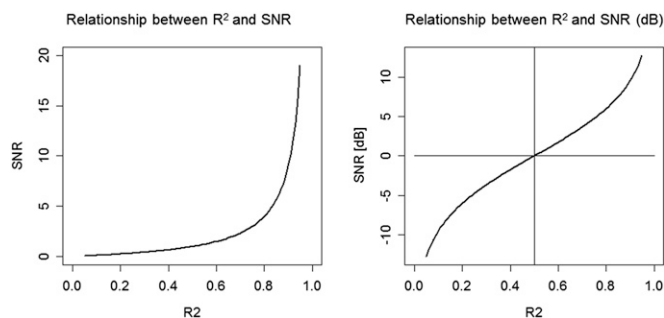
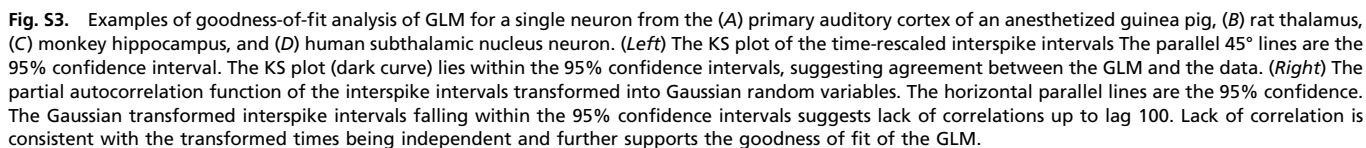


Fig. S1. Relationship between SNR and R^2 . Both plots are created for R^2 values between 0.05 and 0.95. (Left) Eq. S19 and (Right) Eq. S20.



State-space multitaper time-frequency analysis

Seong-Eun Kim^{a,b,1}, Michael K. Behr^{c,1}, Demba Ba^{d,1}, and Emery N. Brown^{a,c,e,f,2}

^aPicower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, MA 02139; ^bDepartment of Electronics and Control Engineering, Hanbat National University, Daejeon 34158, Korea; ^cDepartment of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; ^dSchool of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138; ^eDepartment of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, MA 02114; and ^fInstitute of Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by David L. Donoho, Stanford University, Stanford, CA, and approved November 10, 2017 (received for review February 19, 2017)

Time series are an important data class that includes recordings ranging from radio emissions, seismic activity, global positioning data, and stock prices to EEG measurements, vital signs, and voice recordings. Rapid growth in sensor and recording technologies is increasing the production of time series data and the importance of rapid, accurate analyses. Time series data are commonly analyzed using time-varying spectral methods to characterize their nonstationary and often oscillatory structure. Current methods provide local estimates of data features. However, they do not offer a statistical inference framework that applies to the entire time series. The important advances that we report are state-space multitaper (SS-MT) methods, which provide a statistical inference framework for time-varying spectral analysis of nonstationary time series. We model nonstationary time series as a sequence of second-order stationary Gaussian processes defined on nonoverlapping intervals. We use a frequency-domain random-walk model to relate the spectral representations of the Gaussian processes across intervals. The SS-MT algorithm efficiently computes spectral updates using parallel 1D complex Kalman filters. An expectation-maximization algorithm computes static and dynamic model parameter estimates. We test the framework in time-varying spectral analyses of simulated time series and EEG recordings from patients receiving general anesthesia. Relative to standard multitaper (MT), SS-MT gave enhanced spectral resolution and noise reduction (>10 dB) and allowed statistical comparisons of spectral properties among arbitrary time series segments. SS-MT also extracts time-domain estimates of signal components. The SS-MT paradigm is a broadly applicable, empirical Bayes' framework for statistical inference that can help ensure accurate, reproducible findings from nonstationary time series analyses.

nonparametric spectral analysis | spectral representation theorem | complex Kalman filter | statistical inference | big data

The importance of developing principled methods to solve big data problems is now broadly appreciated ([sites.nationalacademies.org/DEPS/BMSA/DEPS.171738](https://www.nationalacademies.org/DEPS/BMSA/DEPS.171738)). Time series are an important big data class that includes signals ranging from gravitational waves (1), solar variations (2), radar emissions (3), seismic activity (4), global positioning data (5), and stock prices (6) to neural spike train measurements (7), EEG recordings (8), vital signs (9), and voice recordings (10). Rapid growth in sensor and recording technologies in science, engineering, and economics is increasing time series data production and with it, the importance of conducting rapid, accurate analyses. Such analyses require extracting specific data features and characterizing their uncertainty in a way that makes possible formal statistical inferences the same way as they are conducted in simpler problems.

A range of time-frequency methods is used to characterize the nonstationary and often oscillatory features in time series data. Standard nonparametric spectral methods estimate the spectrum (i.e., the frequency content of the time series in a small time interval on which the data are presumed to be stationary) (8, 11, 12). Fourier-based spectral methods estimate only signal power

as a function of frequency and therefore, cannot provide time-domain signal estimates. Spectrogram estimation (time-varying spectral analysis), which entails estimating the frequency content as function of time for nonstationary data, is carried out by simply repeating spectrum estimation on overlapping or nonoverlapping time intervals. Spectrum estimates on adjacent intervals (8, 10–14) are not formally related. While recently developed time-frequency methods address the general problem of minimizing the resolution tradeoff between time and frequency, these techniques are computationally intensive, give their best performance in high signal-to-noise problems, and to date, have had limited application in actual time series analyses (15, 16). Despite their use to study important problems, current time-frequency methods have a critical shortcoming. None of these methods provides a statistical inference framework applicable to the entire time series. (1–10).

State-space (SS) modeling is a flexible, established inference framework for analyzing systems with properties that evolve with time (17–21). This paradigm has been used for time-frequency analysis of nonstationary time series with parametric time series models (22–24), harmonic regression models (25), and nonparametric time series models using batch processing (26). Given stationary data recorded on a finite interval, multitaper (MT) methods are optimal for balancing the bias-variance tradeoff in spectrum estimation conducted by combining Fourier-based

Significance

Rapid growth in sensor and recording technologies is spurring rapid growth in time series data. Nonstationary and oscillatory structure in time series is commonly analyzed using time-varying spectral methods. These widely used techniques lack a statistical inference framework applicable to the entire time series. We develop a state-space multitaper (SS-MT) framework for time-varying spectral analysis of nonstationary time series. We efficiently implement the SS-MT spectrogram estimation algorithm in the frequency domain as parallel 1D complex Kalman filters. In analyses of human EEGs recorded under general anesthesia, the SS-MT paradigm provides enhanced denoising (>10 dB) and spectral resolution relative to standard multitaper methods, a flexible time-domain decomposition of the time series, and a broadly applicable, empirical Bayes' framework for statistical inference.

Author contributions: S.K., D.B., and E.N.B. designed research; S.K., M.K.B., D.B., and E.N.B. performed the research; S.K., M.K.B., D.B., and E.N.B. contributed new analytic tools; S.K., M.K.B., D.B., and E.N.B. analyzed the data; and S.K., and E.N.B. wrote the paper.

Conflict of interest statement: The authors have applied for patents on the use of these methods to track brain states of patients receiving general anesthesia and sedation.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹S.K., M.K.B., and D.B. contributed equally to this work.

²To whom correspondence should be addressed. Email: enb@neurostat.mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1702877115/-DCSupplemental.

methods with tapering (8, 11, 12). Therefore, a plausible approach to analyzing nonstationary and oscillatory time series is to combine SS modeling with MT methods to optimize locally the bias–variance tradeoff, relate spectral estimates across local intervals, and conduct formal statistical inference.

The important advances that we report are state-space multitaper (SS-MT) methods, which provide a statistical inference framework for time-varying spectral analysis of nonstationary time series. The balance of the paper is organized as follows. In *Theory*, we define the SS-MT time-frequency model, SS-MT spectrogram estimation algorithm, time-domain signal extraction algorithm, and empirical Bayes' inference framework. In *Applications*, we illustrate use of the algorithms in the analysis of a simulated nonstationary time series and of EEG time series recorded from patients under general anesthesia. *Discussion* summarizes the properties of the SS-MT paradigm and highlights the implications of this future research for solving large data inference problems.

Theory

An SS-MT Time-Frequency Model. Assume that we observe a nonstationary time series of the form

$$y_t = x_t + \varepsilon_t, \quad [1]$$

where x_t is a zero mean, second-order, locally stationary Gaussian process and ε_t is independent, zero mean Gaussian noise with common variance σ_ε^2 for $t = 1, 2, \dots, T$. A common practice in analyses of nonstationary time series is to assume a minimal interval length and that data are stationary on intervals having this minimal length (*SI Appendix, Table S1*). We define the local stationarity of x_t by assuming that we can write $T = KJ$, where K defines the number of distinct, nonoverlapping stationary intervals in x_t and J is the number of observations per stationary interval. We index the stationary intervals as $k = 1, \dots, K$ and the points per interval as $j = 1, \dots, J$. For example, if we have 1,440 s of a time series that is stationary on 1-s intervals and recorded at 250 Hz, then $K = 1,440$, $J = 250$, and $T = 3,60,000$.

We present the data on stationary interval k as the vector Y_k of length J , with the j th element that is $Y_{k,j} = y_{J(k-1)+j}$, $X_{k,j} = x_{J(k-1)+j}$, and $\varepsilon_{k,j} = \varepsilon_{J(k-1)+j}$ for $k = 1, \dots, K$ and $j = 1, \dots, J$. By the spectral representation theorem (27), we can express each Y_k as

$$Y_k = X_k + \varepsilon_k \\ = W \Delta Z_k + \varepsilon_k, \quad [2]$$

where W is a $J \times J$ matrix with the (l, j) th element that is $(W)_{l,j} = J^{-1/2} \exp(i2\pi(l-1)j/J)$. $\Delta Z_k = (\Delta Z_k(\omega_1), \dots, \Delta Z_k(\omega_J))'$ are differences of orthogonal Gaussian increments, and we define $\omega_j = 2\pi(j-1)/J$.

To relate the data on adjacent intervals, we assume that the Gaussian increment differences are linked by the random walk model

$$\Delta Z_k = \Delta Z_{k-1} + v_k, \quad [3]$$

where we assume that v_k is a zero mean, independent complex Gaussian process with $J \times J$ diagonal covariance matrix $I(\sigma_{v,j}^2)$ for $j = 1, \dots, J$. Eq. 3 defines a stochastic continuity constraint on the nonstationary time series in the frequency domain.

To represent the observation model [2] in the frequency domain, we let F be the Fourier transform operator defined as the $J \times J$ matrix with the (j, l) th element that is $(F)_{j,l} = J^{-1/2} \exp(-i2\pi(l-1)j/J)$. Taking the Fourier transform of Eq. 2 yields

$$Y_k^F = \Delta Z_k + \varepsilon_k^F, \quad [4]$$

where $Y_k^F = FY_k$, $FW = I$, and $\varepsilon_k^F = F\varepsilon_k$ is a zero mean, complex Gaussian vector with $J \times J$ diagonal covariance matrix

$I(\sigma_\varepsilon^2)$. Eqs. 2 and 3 define a frequency-domain SS model for the nonstationary time series.

To combine the SS and MT paradigms, we note that, in the absence of Eq. 3, MT methods with Slepian functions selected as tapers would be used to estimate the spectrum on each stationary interval (11, 12). Therefore, as in the application of MT methods, given J , the number of data points in the stationary interval and Δ , the data sampling rate, we also assume that ω_r , the desired frequency resolution for the spectral analysis, has been specified for the problem. Next, M , the number of tapers, is chosen to minimize the local bias–variance tradeoff for spectrum estimation on each stationary interval using the standard MT formula $M \leq 2[J\Delta^{-1}\omega_r] - 1$ (28). We index the tapers as $m = 1, \dots, M$.

We let $S^{(m)}$ denote the operator for applying the m th Slepian taper to the data, $Y_k^{(m)} = S^{(m)}Y_k$ denote the tapered data, and $Y_k^{(m),F} = FY_k^{(m)}$ denote the Fourier transform of the tapered data. If we take the Slepian tapers to be orthonormal, then by theorem 4.4.2 in ref. 29, the Fourier transform of each tapered series has the same probability distribution and thus, the same spectral representation as Y_k^F in Eq. 4. Therefore, we write

$$Y_k^{(m),F} = \Delta Z_k^{(m)} + \varepsilon_k^{(m),F}, \quad [5]$$

and we view $\Delta Z_k^{(m)}$ and $\varepsilon_k^{(m),F}$ as a realization of ΔZ_k and of ε_k^F , respectively, observable through the m th tapered series. It follows that the random walk model in Eq. 3 has the realization

$$\Delta Z_k^{(m)} = \Delta Z_{k-1}^{(m)} + v_k^{(m)}, \quad [6]$$

where we assume that $v_k^{(m)}$ is a zero mean, independent complex Gaussian vector with a $J \times J$ diagonal covariance matrix $I(\sigma_{v,j}^{2,(m)})$ for $j = 1, \dots, J$ and $m = 1, \dots, M$. Eq. 6 induces stochastic continuity constraints on the tapered nonstationary time series. Eqs. 5 and 6 define an SS-MT time-frequency model. Use of the Slepian tapers to achieve the desired frequency resolution given the assumed length of the local stationary intervals transforms the original time series [2] and its state model [3] into M independent time series [5] and their respective M independent state models [6].

SS-MT Spectrogram Estimation Algorithm. The linear complex Gaussian form of Eqs. 5 and 6 suggests that a Kalman filter algorithm can be used to compute the sequence of increment differences (23) and thus, the sequence of spectrum estimates. For this problem, the Kalman filter has a special structure. Because the $M \Delta Z_k^{(m)}$ are independent, there are M separate, independent J -dimensional Kalman filters. In addition, because $\Delta Z_k^{(m)}(\omega_j)$ is orthogonal across frequencies, there are, for each tapered series, J parallel 1D complex Kalman filter algorithms, one for each frequency ω_j . Hence, the Gaussian increment differences can be recursively estimated by applying $M \cdot J$ 1D complex Kalman filter algorithms to the M tapered time series. Assuming that the increment difference estimates have been computed on interval $k-1$, then for tapered series m , the 1D complex Kalman filter algorithm for estimating $\Delta Z_k^{(m)}(\omega_j)$ on interval k is

$$\Delta Z_{k|k-1}^{(m)}(\omega_j) = \Delta Z_{k-1|k-1}^{(m)}(\omega_j) \quad [7a]$$

$$\sigma_{k|k-1,j}^{2,(m)} = \sigma_{k-1|k-1,j}^{2,(m)} + \sigma_{v,j}^{2,(m)} \quad [7b]$$

$$\Delta Z_{k|k}^{(m)}(\omega_j) = \Delta Z_{k|k-1}^{(m)}(\omega_j) \\ + C_{k,j}^{(m)}(Y_{k,j}^{(m),F} - \Delta Z_{k|k-1}^{(m)}(\omega_j)) \quad [7c]$$

$$\sigma_{k|k,j}^{2,(m)} = (1 - C_{k,j}^{(m)})\sigma_{k-1|k-1,j}^{2,(m)}, \quad [7d]$$

where the Kalman gain for $m=1, \dots, M$, $k=1, \dots, K$, and $j=1, \dots, J$ is

$$C_{k,j}^{(m)} = (\sigma_{\varepsilon}^{2,(m)} + \sigma_{k|k-1,j}^{2,(m)})^{-1} \sigma_{k|k-1,j}^{2,(m)}. \quad [8]$$

The notation $k|s$ denotes the estimate on stationary interval k given all of the data observed through stationary interval s . The derivation of the Kalman filter algorithm is in [SI Appendix](#). We assume that the algorithm has initial conditions $\Delta Z_0^{(m)}(\omega_j)$ and $\sigma_{0,j}^{2,(m)}$. We carry out their estimation along with the model parameters using an expectation–maximization (EM) algorithm, which we describe in [SI Appendix](#). Given the Kalman filter estimate of the increment differences on interval k , the SS-MT spectrogram estimate at frequency ω_j on interval k is

$$f_{k|k}^{SS-MT}(\omega_j) = M^{-1} \sum_{m=1}^M \|\Delta Z_{k|k}^{(m)}(\omega_j)\|^2, \quad [9]$$

where $\|\Delta Z_{k|k}^{(m)}(\omega_j)\|^2$ is the m th SS eigenspectrogram at frequency ω_j (11). Each SS eigenspectrogram is a spectrogram estimate computed by weighting the data with a different Slepian taper. Like the MT spectrogram defined below [10], SS-MT spectrogram estimate [9] is the average of the M approximately independent SS eigenspectrograms.

Eqs. 7–9 define the SS-MT algorithm for spectrogram estimation for nonstationary time series. For each tapered series, the increment difference estimate on interval k is a weighted average between the increment difference estimate on interval $k-1$ and the difference between the Fourier transform of the tapered series and the increment difference estimate on interval $k-1$. The weighting depends on the Kalman gain, which is between zero and one by construction. If the Kalman gain is close to zero, then the one-step prediction variance $\sigma_{k|k-1,j}^{2,(m)}$ is small relative to the observation variance $\sigma_{\varepsilon}^{2,(m)}$, and hence, the increment difference estimate on interval k is close to the estimate on interval $k-1$. If the Kalman gain is close to one, then the one-step prediction variance is large relative to the observation variance, meaning that the uncertainty in the prediction of the increment difference on interval k based on the data up through interval $k-1$ is large. In this case, the increment difference estimate on interval k is close to the Fourier transform of the tapered series observed on interval k .

In the absence of the state models [3 and 6], Eq. 9 becomes the MT spectrogram estimate

$$f_k^{MT}(\omega_j) = M^{-1} \sum_{m=1}^M \|Y_{k,j}^{(m),F}\|^2, \quad [10]$$

where $Y_k^{(m),F} = (Y_{k,1}^{(m),F}, \dots, Y_{k,J}^{(m),F})'$ and $\|Y_{k,j}^{(m),F}\|^2$ is the m th MT eigenspectrogram at frequency ω_j (12). In the absence of tapering, Eq. 9 becomes the SS periodogram estimate

$$f_{k|k}^{SS-P}(\omega_j) = \|\Delta Z_{k|k}^{SS-P}(\omega_j)\|^2, \quad [11]$$

which is computed by applying J parallel 1D complex Kalman filters to the Fourier transformed data Y_k^F . In the absence of tapering and the SS model, Eq. 11 becomes the periodogram estimate

$$f_k^P(\omega_j) = \|Y_{k,j}^F\|^2, \quad [12]$$

where $Y_k^F = (Y_{k,1}^F, \dots, Y_{k,J}^F)'$. By comparing the SS-MT algorithm [7–9] with the standard MT [10], the periodogram [12], and the SS periodogram [11] algorithms, it is possible to understand the effects on spectrogram estimation of combining the MT approach with SS modeling. In addition, the SS-MT paradigm can be applied to compute cross-spectrograms between two or more time series that are described in [SI Appendix \(SI Appendix, Fig. S12\)](#).

Time-Domain Signal Extraction. Given the $\Delta Z_{k|k}^{(m)}$, we can estimate the denoised time-domain signal as

$$X_{k|k} = W \Delta Z_{k|k}, \quad [13]$$

where $\Delta Z_{k|k} = M^{-1} \sum_{m=1}^M \Delta Z_{k|k}^{(m)}$. The extracted signal is a linear combination of the estimated increment differences across all of the frequencies. Frequency components on different stationary intervals are related, because all are estimated by the complex Kalman filter algorithm in Eqs. 7a–7d. Hence, selective filtering, such as high-pass, low-pass, and band-pass filtering, can be performed by simply choosing the components of $\Delta Z_{k|k}$ in the desired frequency range. Given a set of L , not necessarily sequential frequencies, ω_j for $j = s_1, \dots, s_L$, we can obtain the filtered time-domain signal as

$$X_{k|k}^L = W \Delta Z_{k|k}^L, \quad [14]$$

where the components of $\Delta Z_{k|k}^L$, outside the L frequencies and their conjugate symmetric frequencies, are all zero. Eq. 14 provides a highly flexible alternative to an empirical mode decomposition that allows extraction of a time-domain signal composed of any specified frequency components. The analytic version of the filtered time-domain signal can be computed as

$$R_{k|k,t}^L + iI_{k|k,t}^L = 2J^{-\frac{1}{2}} \sum_{j=s_1}^{s_L} \Delta Z_{k|k}^L(\omega_j) e^{i\omega_j t} \quad [15]$$

for $t = J(k-1) + l$ and $l = 1, \dots, J$. Here, $[(R_{k|k,t}^L)^2 + (I_{k|k,t}^L)^2]^{1/2}$ and $\tan(-I_{k|k,t}^L/R_{k|k,t}^L)$ are the instantaneous amplitude and phase of the time-domain signal in the specified frequency range, respectively ([SI Appendix, Figs. S1 and S2](#)). This computation obviates the need to apply a Hilbert–Huang transform to either filtered data or data processed by an empirical mode decomposition to estimate instantaneous amplitude and phase.

Inferences for Functions of the Increment Differences. To make inferences for functions of the increment differences at any time points, we compute the joint distribution of the increment differences conditional on all of the data in the time series using the fixed interval smoothing algorithm (20, 21), which is

$$\begin{aligned} \Delta Z_{k|K}^{(m)}(\omega_j) &= \Delta Z_{k|k}^{(m)}(\omega_j) \\ &\quad + A_{k,j}(\Delta Z_{k+1|K}^{(m)}(\omega_j) - \Delta Z_{k+1|k}^{(m)}(\omega_j)) \\ \sigma_{k|K,j}^{2,(m)} &= \sigma_{k|k,j}^{2,(m)} + A_{k,j}^2(\sigma_{k+1|K,j}^{2,(m)} - \sigma_{k+1|k,j}^{2,(m)}) \\ A_{k,j} &= \sigma_{k|k,j}^{2,(m)}(\sigma_{k+1|k,j}^{2,(m)})^{-1}, \end{aligned} \quad [16]$$

where the initial conditions are $\Delta Z_{K|K}^{(m)}(\omega_j)$ and $\sigma_{K|K,j}^{2,(m)}$ for $k = K-1, K-2, \dots, 1$ and $j = 1, 2, \dots, J$. To compute the covariances between any two states, we use the covariance smoothing algorithm defined as (20)

$$\sigma_{k,u|K,j}^{(m)} = A_{k,j} \sigma_{k+1,u|K,j}^{(m)} \quad [17]$$

for $1 \leq k \leq u \leq K$. Eqs. 16 and 17 allow us to compute the joint distribution of the increment differences conditional on all of the data. Therefore, we can compute the distribution of any function of the increment differences by Monte Carlo methods (30). For each Monte Carlo sample, we draw from the joint distribution and compute the function of interest. The histogram of the function of interest provides a Monte Carlo estimate of its posterior probability density. The estimate is empirical Bayes, because it is computed conditional on the maximum likelihood parameter estimates (31).

the three peaks (Fig. 1 *D*, *Right* and *F*, *Right* and *SI Appendix*, Fig. S6).

A key difference between the MT and the SS-MT spectrograms appears as the power increases. As the heights of the spectral peaks at 9.5 and 11.5 Hz increase, the depth of the “valley” in the spectrogram between them increases also (Fig. 1*B* and *SI Appendix*, Fig. S6 *C* and *D*). The valley is at 5 dB between minutes 24 and 27 (Fig. 1*B*, *Right* and *SI Appendix*, Fig. S6*D*). The MT spectrogram estimates the valley to be at 10 dB (Fig. 1*D*, *Right* and *SI Appendix*, Fig. S6*D*). In contrast, the SS-MT spectrograms estimates the valley to be at 4 dB (Fig. 1*D*, *Right* and *SI Appendix*, Fig. S6*D*). In addition, the mean squared error was lower at all frequencies for the SS-MT algorithm compared with the other three methods (*SI Appendix*, Table S2). We explain the enhanced denoising and enhanced spectral resolution of the SS-MT algorithm relative to the MT algorithm after we analyze the real EEG recordings in the next examples. In *SI Appendix*, we assess the effect of stationary interval length on spectrogram estimation.

Spectrogram Analysis of the EEG During General Anesthesia. Anesthetic drugs act in the brain to create the altered states of general anesthesia by producing highly structured oscillations that disrupt normal information flow between brain regions (34, 35). Because these oscillations are readily visible in the EEG, EEG and EEG-derived measures are commonly used to track

in real time the brain states of patients receiving general anesthesia and sedation (36). We illustrate the SS-MT methods by comparing them with the other three spectrogram methods in the analysis of EEG recordings from patients during general anesthesia.

The EEG recordings, in this example and the subsequent examples, are deidentified data collected as part of protocols at the Massachusetts General Hospital (MGH) that have been approved by the MGH Human Research Committee. For EEGs recorded from patients, informed consent was not required, whereas for EEGs recorded from volunteer subjects, informed consent was required and was obtained. For each patient, the EEG was continuously recorded during general anesthesia using the Sedline monitor (Masimo) with the standard six-electrode frontal montage. The Sedline array records from electrodes located approximately at positions Fp1, Fp2, F7, and F8. On each channel, the electrode impedance was less than 5 kohms. We used the EEG data recorded at Fp1 for the spectral analyses and the EEG data recorded at Fp1 and Fp2 for the coherence analyses (*SI Appendix*, Fig. S12). We began the EEG recordings approximately 3–5 min before induction of general anesthesia and continued the recordings for ~3–5 min after extubation.

The data analyzed in Fig. 2 consist of 190 min of EEG recorded at 250 Hz during maintenance of general anesthesia using the ether anesthetic sevoflurane with oxygen. Hence, we take $T = 2,850,000$. We set $J = 500$ based on our several years of

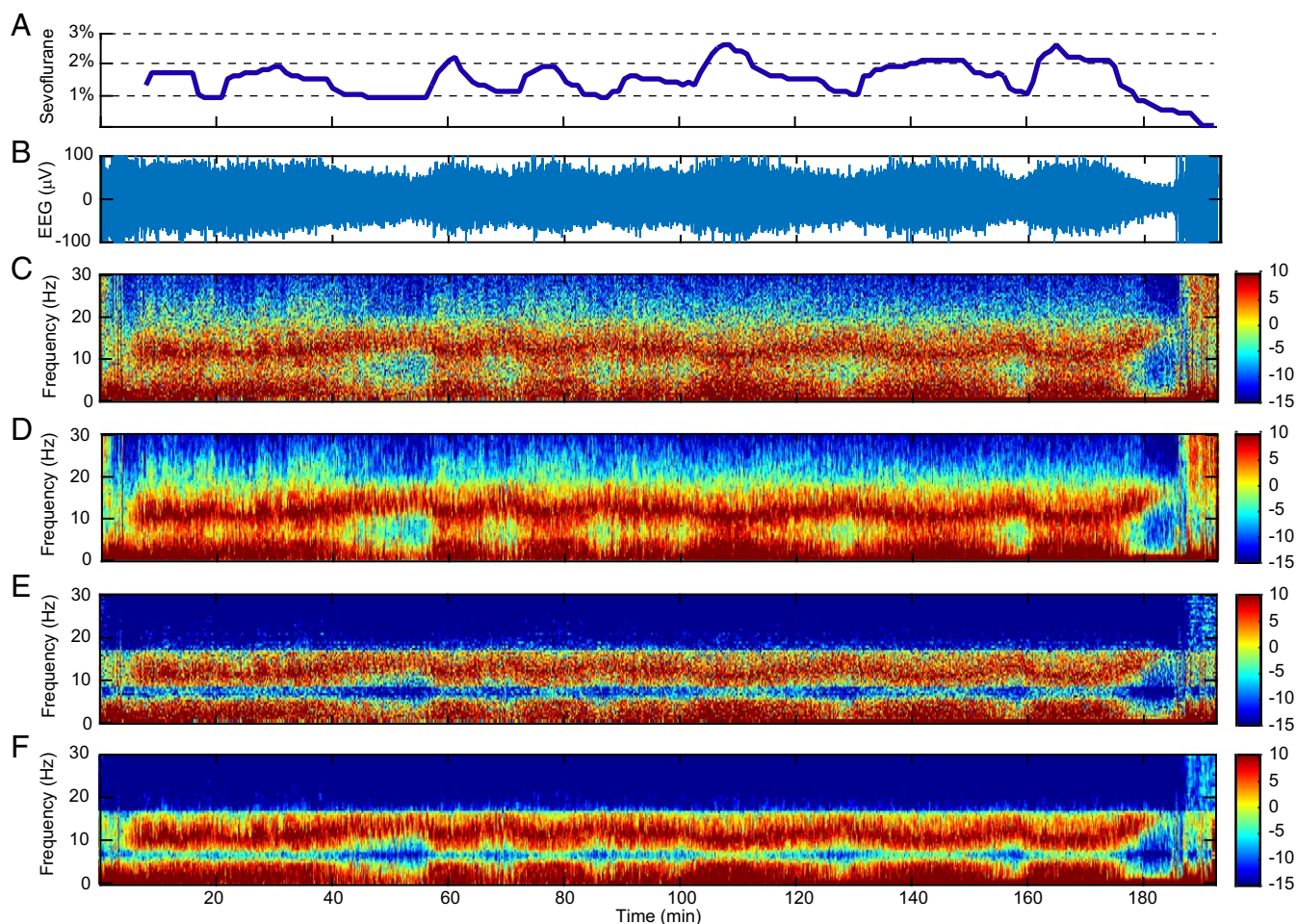


Fig. 2. Spectrogram analysis of EEG time series recorded from a patient under general anesthesia maintained with sevoflurane and oxygen. (A) The expired concentration of sevoflurane. (B) Raw EEG signals. (C) Periodogram. (D) MT method spectrogram. (E) SS periodogram. (F) SS-MT spectrogram. The color scale is in decibels.

experience analyzing the EEG of anesthetized patients. Hence, we have $K = 5,750$. We chose M , the number of tapers, to be three. This corresponds to a spectral resolution of 2 Hz for the MT method. We estimated the model parameters and initial conditions from the first 5 min of data using the EM algorithm. To estimate the observation noise variance in the EM algorithm, we restricted the analysis to frequencies in the physiologically relevant range from 0.1 to 30 Hz. The raw EEG signal (Fig. 2B) shows strong modulation with changes in the sevoflurane concentration (Fig. 2A).

All four spectrograms for these data show the well-known alpha-beta oscillations (8–17 Hz) and slow-delta oscillations (0.1–4 Hz) that are characteristic of general anesthesia maintained by sevoflurane (36). When the sevoflurane concentration increases, the power in the alpha-beta band shifts to lower frequencies, while the power in the slow-delta band power shifts to higher frequencies. The opposite changes occur when the sevoflurane concentration decreases. The spectral changes associated with increases in the sevoflurane concentration appear as increases in theta oscillation power (4–8 Hz) (36). The periodogram (Fig. 2C) shows diffuse, grainy power between 10 and 17 Hz and in the slow-delta range. As expected, the MT spectrogram (Fig. 2D) has higher spectral resolution relative to the periodogram. Both the periodogram and the MT spectrogram show diffuse power ranging from 7 to -2 dB in the theta range and from -5 to -15 dB in the beta-gamma range (>17 Hz). Relative to the periodogram and the MT spectrogram, the SS periodogram (Fig. 2E) and the SS-MT spectrogram (Fig. 2F) show substantially greater denoising defined as a reduction in power in the frequency bands with low power. For the latter

two spectrograms, the power in the beta-gamma range is uniformly at -15 dB, which is a 5–15 dB power reduction relative to the MT spectrogram. Both the SS periodogram and the SS-MT spectrogram estimate the power in the theta band to be 10–15 dB less than that for either the periodogram or the MT spectrogram. Like the periodogram, the prominent alpha-beta and the slow-delta power in the SS periodogram is grainy and diffuse.

Spectrogram Analysis of the EEG During Transitions Among Anesthetic States. To illustrate the full potential of the SS-MT algorithm, we reanalyze 155.4 min of EEG data recorded at 250 Hz from a frontal lead in a human volunteer subject receiving i.v. propofol administered by a computer-controlled infusion at an increasing and then, a decreasing infusion rate (35). Gradually increasing the propofol infusion rate allows the subject to transition gradually from being awake to unconsciousness (UNC). Gradually decreasing the propofol infusion rate from the rate required to achieve the maximum target effect site concentration (model-derived brain concentration) allows the subject to transition from UNC to the awake state. Baseline EEG was recorded for 20 min while the subject lay supine with eyes closed and received no propofol. After the baseline period, propofol was administered through a computer-controlled infusion to achieve five different effect site concentrations in increasing levels (Fig. 3A) (35). After completing the fifth level, the propofol infusion rate was systematically decreased to achieve a similar sequence of target effect site concentrations in decreasing order until the infusion was stopped. Each target effect site concentration was maintained for 14 min.

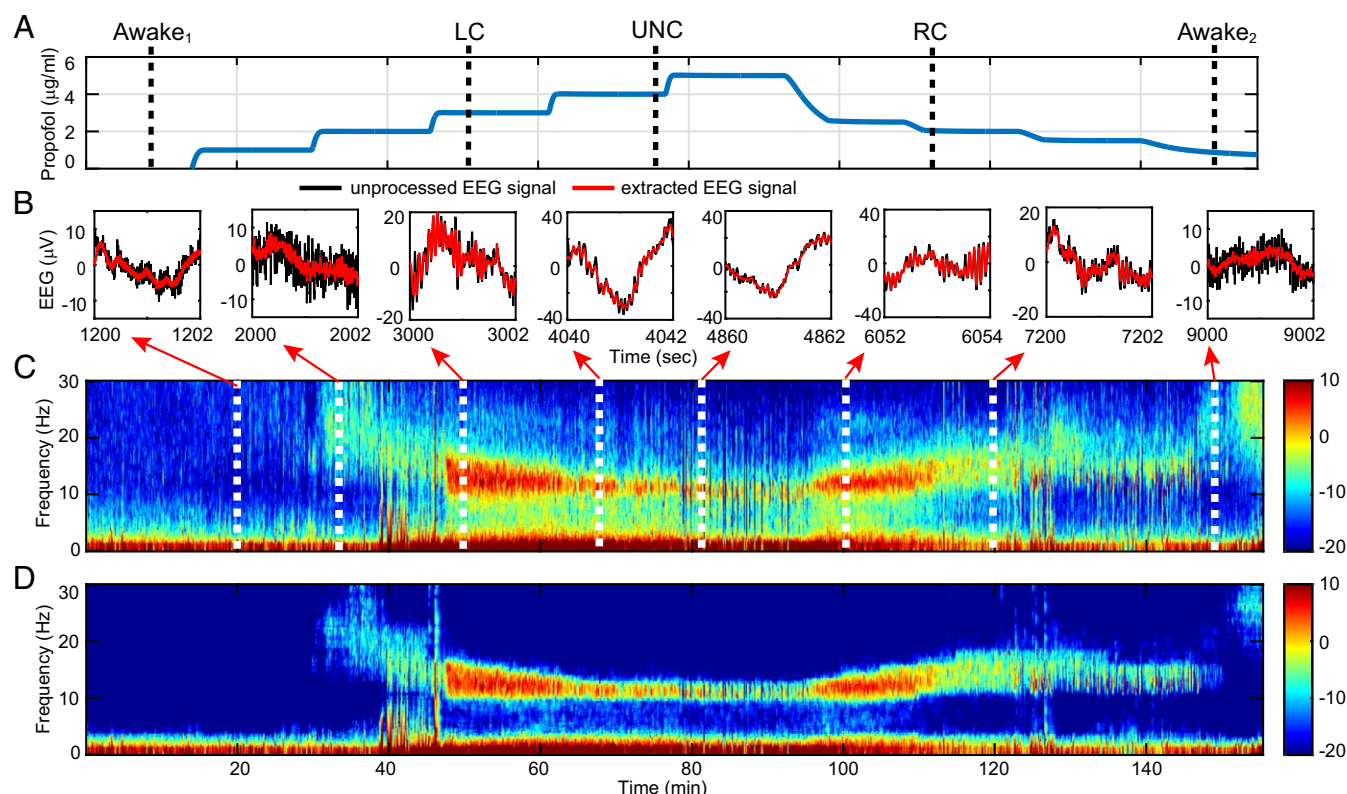


Fig. 3. Spectrogram analysis of EEG recorded in a volunteer subject receiving a computer-controlled infusion of propofol. (A) Time course of propofol target effect site concentrations based on the Schneider model (35). The black dashed vertical lines define the anesthetic states determined by the behavioral analysis: awake₁, baseline conscious state; LC; UNC; RC; and awake₂, final conscious state. (B) Two seconds of unprocessed EEG (black curves) and of EEG extracted from the SS-MT analysis (red curves) [14] at different target effect site concentrations. (C) MT method spectrogram. (D) SS-MT spectrogram. The color scale is in decibels.

Based on the analyses of the subject's responses to yes–no questions administered every 4 s, we identified five distinct behavioral or anesthetic states: conscious, loss of consciousness (LC), UNC, recovering consciousness (RC), and conscious (Fig. 3*A*, black dashed vertical lines) (35). The important scientific question to answer is whether these distinct anesthetic states are associated with distinct EEG signatures.

We take $T = 2,331,000$ and set $J = 1,000$, $K = 2,331$, and $M = 5$. This gives a spectral resolution of 1.5 Hz for the MT method. Because we expect the EEG frequency content to change substantially with changes in target effect site concentration, we estimated, using the EM algorithm, a new set of model parameters from the first 5 min of EEG data recorded at each level. The effects of changing the propofol infusion rate are apparent in the unprocessed EEG (Fig. 3*B*, black curves), the denoised time-domain signal (Eq. 14 and Fig. 3*B*, red curves), the MT spectrogram (Fig. 3*C*), and the SS-MT spectrogram (Fig. 3*D*). At baseline, moderate-amplitude slow oscillations dominate the EEG. Low-amplitude beta-gamma oscillations appear midway through level 2 and transition into narrow-band, high-amplitude alpha oscillations by level 4. At the same time, the slow oscillations transition to high-amplitude, slow-delta oscillations. By level 5, the alpha oscillations have nearly dissipated, and the EEG is dominated by slow-delta oscillations. As the propofol infusion rate is decreased, EEG dynamics are recapitulated in reverse. As in the previous examples, the SS-MT spectrogram shows substantial spectral denoising and increased resolution relative to the MT spectrogram. The denoised time-domain signals and the SS-MT spectrogram strongly suggest that different oscillatory components are present in the EEG when the subject is in different anesthetic states.

Inferring Differences in Spectral Power Between Anesthetic States.

To make a formal statistical inference about the relationship between these anesthetic states and EEG signatures, we compare the power across the spectrogram from 0.1 to 30 Hz among

representative 100-s intervals during each of the anesthetic states (Fig. 3*A*). The 100-s intervals are baseline awake period awake₁, 500–600 s; LC, 3,100–3,200 s; UNC, 4,600–4,700 s; RC, 6,600–6,700 s; and final awake state awake₂, 9,000–9,100 s. To compare two 100-s intervals for each frequency ω in a given frequency range, we compute the average difference spectrogram between two intervals:

$$\Delta \bar{f}_{r,s}(\omega) = \frac{1}{100} \left[\int_r f_t^{SS-MT}(\omega) dt - \int_s f_t^{SS-MT}(\omega) dt \right], \quad [19]$$

where r and s are two distinct 100-s intervals. To determine if there is a significant change in the spectrogram between any two of the anesthetic states, we use a Monte Carlo procedure to compute an approximate empirical Bayes' 95% confidence interval (95% CI) for $\Delta \bar{f}_{r,s}(\omega)$ (30). Together, the Kalman filter [7 and 8], the Kalman smoothing [16], and the covariance smoothing [17] algorithms define the multivariate complex Gaussian joint posterior density of $\Delta Z_{k|K}$ for $k = 1, \dots, K$, conditional on the parameter estimates. The quantity $\Delta \bar{f}_{r,s}(\omega)$ is a function of the $\Delta Z_{k|K}$, so that given a random sample of the $\Delta Z_{k|K}$, we can compute $\Delta \bar{f}_{r,s}(\omega)$. By drawing a large number of the $\Delta Z_{k|K}$, say 1,000, we can, therefore, compute 95% CIs for $\Delta \bar{f}_{r,s}(\omega)$ (Fig. 4). A significant difference in power is observed if the zero is not in the 95% CI.

We show 6 of 10 possible comparisons of differences in power among the five anesthetic states (Fig. 4). The LC and UNC states show significant increases in power in the slow-delta and alpha bands relative to awake₁, the baseline awake state (Fig. 4*A* and *B*). There is also a significant increase in power in the upper part of the slow-delta and in the alpha bands between RC and awake₁ (Fig. 4*C*) and between LC and the UNC state (Fig. 4*E*). In contrast, there are no appreciable differences in power between awake₂ and awake₁ (Fig. 4*D*) or between LC and RC (Fig. 4*F*). These findings are in complete agreement with the structure of the power in the spectrogram (Fig. 3*D*). We can conclude that there are significant quantifiable differences in

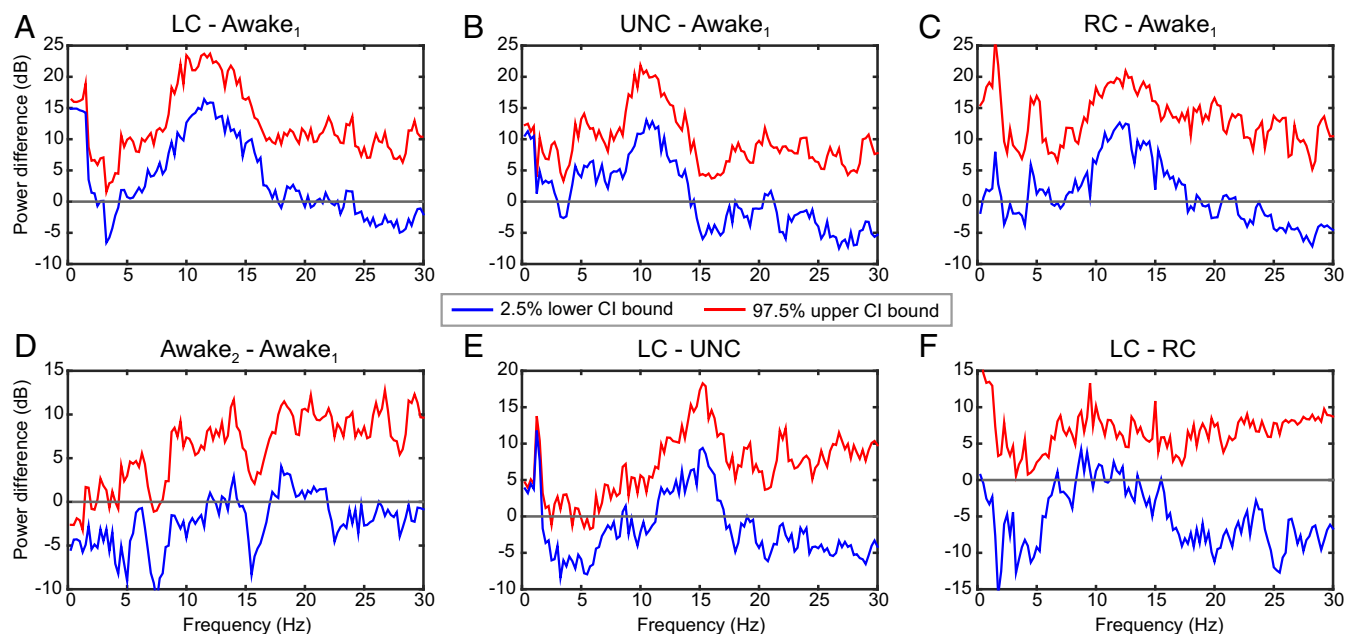


Fig. 4. The 95% CIs for comparisons of the average power differences between the anesthetic states in Fig. 3*A*. Each panel shows the 95% empirical Bayes confidence interval (CI) for the average power difference defined by the lower 2.5% CI bound (blue curves), and the upper 97.5% CI bound (red curves). (A) LC-awake₁: LC compared with baseline conscious state. (B) UNC-awake₁: UNC compared with baseline conscious state. (C) RC-awake₁: RC compared with baseline conscious state. (D) Awake₂-awake₁: final conscious state compared with baseline conscious state. (E) LC-UNC: LC compared with the UNC state. (F) LC-RC: LC compared with RC.

EEG power between different anesthetic states and that those differences can range from 10 to 20 dB (95% CI). These findings also agree with and go beyond the original analyses of these data, in which hypothesis testing methods with Bonferroni corrections were used to compare the anesthetized states with just awake₁ (35). Using the model in Eq. 18, we assessed the coverage probability of the empirical Bayes' CIs in a simulation study (SI Appendix) and found that the actual and nominal coverage probabilities are in good agreement.

Spectrogram Denoising. The SS-MT spectrogram has greater denoising than the MT spectrogram (Figs. 1–3) because of the stochastic continuity constraints (Eq. 3) and the eigenspectrogram averaging (SI Appendix, Figs. S3 and S4). The stochastic continuity constraint has a different independent effect at each frequency. In both the theoretical and the real data examples, the state variances, $\sigma_{v,l}^{2,(m)}$, are small (0.05–4 dB) for frequencies with low power and large (27–38 dB) for frequencies with high power (SI Appendix, Fig. S5, blue curves). The Kalman gains, $C_{k,l}^{(m)}$, reached steady-state values within 5–10 updates, and like the state variances, the Kalman gains were small (0.1–0.4) for frequencies with low power and large (0.7–0.95) for frequencies with high power (SI Appendix, Fig. S5, red curves). Rewriting Eq. 7c as

$$\Delta Z_{k|k}^{(m)}(\omega_l) = (1 - C_{k,l}^{(m)})\Delta Z_{k-1|k-1}^{(m)}(\omega_l) + C_{k,l}^{(m)}Y_{k,l}^{(m),F} \quad [20]$$

shows that the increment difference estimate on interval k is a weighted average between the increment difference estimate on interval $k - 1$ and the Fourier transform of the tapered data on interval k . In particular, frequencies with low power weight $Z_{k-1|k-1}^{(m)}(\omega_l)$ more than $Y_{k,l}^{(m),F}$. This weighting favors suppressing increases or fluctuations in the low-power or noise frequencies. In contrast, frequencies with high power weight more the new information in $Y_{k,l}^{(m),F}$. These differential effects denoise the spectrogram by heightening the contrast between frequencies with high power and those with low power in the analysis of the simulated data (Fig. 1F and SI Appendix, Fig. S6 C and D) and in the analysis of the actual EEG recordings (Fig. 2F and SI Appendix, Figs. S6C and S9C).

Averaging the MT eigenspectrograms reduces the variability in the MT spectrogram (SI Appendix, Fig. S3) (11). Each SS-MT eigenspectrogram (SI Appendix, Fig. S4 A–C) has variability comparable with the SS periodogram (Fig. 2E) (29). Averaging the SS-MT eigenspectrograms reduces the variability of the SS-MT spectrogram at each frequency relative to the SS periodogram (SI Appendix, Fig. S4D) by M^{-1} [9], thus giving the SS-MT spectrogram a further denoising enhancement.

Spectral Resolution and Leakage Reduction. Kalman filter updating [7c and 20] enhances the spectral resolution of the SS-MT spectrogram relative to the MT spectrogram by reducing leakage. To see why, assume that $f_k(\omega_j)$ and $f_k(\omega_l)$ are the true spectrograms on time interval k at two frequencies ω_j and ω_l , respectively, and that $f_k(\omega_j) \gg f_k(\omega_l)$. Let $\Delta\omega_r$ be the frequency resolution chosen for the MT analysis. If $|\omega_j - \omega_l| < \Delta\omega_r$ ($|\omega_j - \omega_l| > \Delta\omega_r$), then in the MT analysis, the narrow (broad)-band power at ω_j leaks into the power at ω_l (12). The extent of the leakage is governed by the power spectral density of each taper (SI Appendix, Figs. S7, S8, S10, and S11). In the SS-MT analysis, because ω_l has low power, $\Delta Z_{k|k}^{(m)}(\omega_l)$ weights $\Delta Z_{k-1|k-1}^{(m)}(\omega_l)$ much more than $Y_{k,l}^{(m),F}$, the term in Eq. 20 that carries the leakage from ω_j . Hence, broad- and narrow-band power leakage from ω_j into the power at ω_l is reduced, because the Kalman gain at ω_l is small.

For example, at 70 min (Fig. 2 D and F and SI Appendix, Fig. S9 A and C), the MT and SS-MT spectrograms agree in the high-power frequency bands (i.e., 0.1–5 and 9.5–15.5 Hz) and disagree in the low-power frequency bands (5.1–9.4 and 15.6–30 Hz); 6 Hz is just on the border of the narrow-band leakage from 5 Hz for the MT spectrogram (SI Appendix, Fig. S9A). The 12-dB difference between the MT and the SS-MT spectrograms at 6 Hz results, because the former has leakage from the power at 5 Hz, whereas the latter has enhanced denoising and reduced leakage. A 10- to 15-dB power difference persists between the MT and SS-MT spectrograms beyond 15 Hz because of the small values of the Kalman gain in this frequency band (SI Appendix, Fig. S9 C and D).

At 80 min (Fig. 2 D and F and SI Appendix, Fig. S9 B and D), the MT and SS-MT spectrograms also agree in the high-power frequency bands (0.1–5 and 10.5–15 Hz) and disagree in the low-power frequency bands (i.e., 5.1–9.4 and 15.1–30 Hz). A similar argument explains the 7-dB difference in power at 16 Hz between the MT and the SS-MT spectrograms at minute 80. The same argument also explains in the analysis of the simulated data example the 7-dB difference in power at 11 Hz in the MT and SS-MT spectrograms at 25 min (SI Appendix, Fig. S6D). The differences between MT and SS-MT tapering are discussed in SI Appendix.

Discussion

Time series are a prominent big data class with growth that is being spurred by innovations in sensing and recording technologies. These data track dynamic processes, making accurate real-time analyses an important objective. Aside from simply extracting important features from the series, the analysis should provide associated measures of uncertainty, so that formal statistical inference can be conducted the same way that it would be conducted for questions arising from smaller, simpler datasets. No current time-frequency analysis method provides an inference framework applicable to the entire series (8, 11–16). To address the inference problem for nonstationary time series, we combined the MT and the SS approaches into an empirical Bayes' paradigm for frequency- and time-domain analyses of nonstationary time series. We showed the SS-MT inference paradigm by analyzing differences in EEG power between different propofol-induced anesthetic states (Fig. 4). By reporting the results in terms of 95% empirical Bayes' CIs, we measure directly the effect size (i.e., how much EEG power differs among propofol's anesthetic states). We base our inferences on 95% CIs derived from the empirical Bayes' estimate of the joint posterior distribution of the power across all of the frequencies and all times in the time series rather than on tests of multiple hypotheses. Our small simulation study (SI Appendix) suggests that the nominal and actual coverage probabilities of the empirical Bayes' CIs are in good agreement. The empirical Bayes' paradigm has been suggested as a practical approach to solving large-scale inference problems (37).

Our analyses offer scientific insights. The SS-MT spectrograms show denoising and spectral resolution that more clearly define the frequency content of anesthetic EEG signals. As a consequence, the results in Fig. 2 suggest that most of the theta oscillation power in the sevoflurane spectrogram could be caused by power shifts from both the alpha band above and the slow-delta band below. The results in Fig. 4 allow us to make formal inferences about EEG power difference as a function of the level of unconsciousness in a single individual.

SS-MT Time-Frequency Analysis of Nonstationary Time Series. In addition to providing a statistical inference framework, the SS-MT paradigm has other desirable features. By the spectral representation theorem, the orthogonal increment differences are

the fundamental process that underlies a stationary time series (27, 38). Hence, we defined nonstationarity by starting with the common practice of choosing a minimal interval on which the time series is assumed stationary (*SI Appendix, Table S1*) and then applying a stochastic continuity constraint [3 and 6] to link the increment differences across the minimal intervals. We constructed the SS model by taking the observed data to be the Fourier transforms of the M tapered data series [5]. For a given taper, the Fourier transform of the tapered data is J independent, complex Gaussian observations in the frequency domain. Hence, to estimate the increment differences, we implemented in parallel J independent 1D complex Kalman filters [7a–7d]. Given the M tapers, the $M \cdot J$ algorithms run in parallel, and the SS-MT spectrogram (cross-spectrogram) estimates are computed by summing the M eigenspectrograms (eigen cross-spectrograms) in Eq. 9 (*SI Appendix, Eq. S14*) at each Kalman filter update. Parallel computation makes SS-MT spectrogram estimation attractive for real-time applications. Each 1D complex Kalman filter has an associated Kalman smoother [16], covariance smoothing [17], and an EM algorithm for parameter estimation at each frequency (*SI Appendix*).

Both the SS and the MT components of SS-MT analysis contribute significantly to spectrogram denoising. The state variances and Kalman gains are high (low) at frequencies with high (low) power (*SI Appendix, Fig. S5*). Therefore, the Kalman filter updating [7c and 20] denoises the spectrogram by heightening the contrast between high- and low-power spectrogram ordinates (Figs. 1 *D* and *F*, 2 *D* and *F*, and 3 *C* and *D*). The MT component of the SS-MT algorithm further contributes to the denoising by averaging the eigenspectrograms to reduce the variance at all frequencies by M^{-1} (*SI Appendix, Figs. S3 and S4*). In addition, SS estimation [7c and 20] enhances the spectral resolution in the SS-MT spectrogram relative to the MT spectrogram by reducing both narrow-band and broad-band leakage (*SI Appendix, Figs. S6 and S9*). Because the Kalman gains at low-power frequencies are small, leakage from even nearby frequencies with high power is reduced (*SI Appendix, Figs. S6 and S9*). In our simulated and real data examples, the effect of SS updating on denoising and spectral resolution was a 10- to 15-dB difference between the SS-MT and the MT spectrograms in the low-power frequency bands (Figs. 1 *D* and *F*, 2 *D* and *F*, and 3 *C* and *D* and *SI Appendix, Figs. S6 and S9*).

By applying the spectral representation theorem to the estimated increment differences [14 and 15], we extracted the time-domain signals within specified frequency bands as well as instantaneous phase and amplitude (Fig. 3*B* and *SI Appendix, Figs. S1 and S2*). The SS-MT paradigm is highly flexible, because arbitrary combinations of frequency components can be chosen to construct the time-domain signal. Time-domain signal extraction is not possible with standard nonparametric spectral methods, which only estimate power as a function of frequency. Estimating instantaneous phase and amplitude with conventional approaches requires a separate analysis (14). The SS-MT paradigm conducts spectral analysis, signal extraction, and instantaneous phase and amplitude estimation as parts of a unified framework.

Theoretical and Problem-Specific Methods Development. Our local stationarity definition differs from the Dahlhaus definition, which assumes time-invariant increment differences with deterministic continuity of the time-dependent transfer function in the spectral representation (33). The lengths of the local intervals are either assumed (33) or estimated (39). The local spectral estimates are computed as the squared modulus of the Fourier transform of the data after applying a single taper or by fitting a parametric model using local Whittle likelihoods (40). Like the MT methods, these local estimation approaches do not combine information across the local stationary intervals.

In contrast, the stochastic continuity constraint imposed by the random walk model enables recursive estimation of the time-dependent increment differences and the spectrogram. The current form of the continuity constraint has a theoretical drawback. It implies that, at each frequency, spectral power grows with time, since the theoretical spectrum on stationary interval k at frequency ω_j is

$$f_k(\omega_j) d\omega_j = f_{k-1}(\omega_j) d\omega_j + \sigma_j^2. \quad [21]$$

In practice, the spectrogram estimate does not explode, because on every stationary interval k , $C_{k,l}^{(m)}$, the Kalman gain is bounded between zero and one [8]. If the Kalman gain is zero, then the spectrogram estimate on interval k is the SS-MT spectrogram estimate on interval $k-1$, whereas if the Kalman gain is one, then the spectrogram estimate on interval k is the MT spectrogram estimate on interval k [20]. Nevertheless, a possible correction to Eq. 21 is to assume that the increment differences follow a stationary model, such as

$$\Delta Z_k(\omega_j) = \rho_j \Delta Z_{k-1}(\omega_j) + v_k(\omega_j), \quad [22]$$

where we assume that $0 < \rho_j < 1$. Hence, we have

$$\begin{aligned} f_k(\omega_j) d\omega_j &= E \|\Delta Z_k(\omega_j)\|^2 \\ &= \rho_j^2 E \|\Delta Z_{k-1}(\omega_j)\|^2 + \sigma_j^2 \\ &= \rho_j^2 f_{k-1}(\omega_j) d\omega_j + \sigma_j^2. \end{aligned} \quad [23]$$

Eq. 22 means that the nonstationary time series has an underlying stationary increments process. The parameter ρ_j can easily be estimated in the EM algorithm. Our SS approach falls into the class of regularization methods for solving big data problems (41). Thus, the current wealth of regularization research can be applied to the SS-MT paradigm.

In our data analyses, we followed the common practice of setting the stationary interval a priori (*SI Appendix, Table S1*). Our analyses (*SI Appendix, Fig. S14 and Table S2*) suggest that the spectrogram estimates can be sensitive to interval choice and that different stationary intervals could be optimal for different frequencies. Therefore, in future work, we will incorporate interval choice into the estimation by evaluating the likelihood as a function of stationary interval length.

At present, both our model-fitting algorithms and inference framework depend critically on the Gaussian observation and state models [2 and 3]. Description of signal frequency content and inferences may be inaccurate when these assumptions do not hold. As an alternative, model-fitting and inference using non-Gaussian SS models can be readily carried out using sequential Monte Carlo methods (42). This extension will be the topic of a future investigation.

Application of the SS-MT paradigm in time-frequency analyses of different types of nonstationary time series is a productive way to extend our methods by allowing the question under investigation to guide problem-specific development of the framework. We will continue to study EEG time series recorded under different anesthetic drugs (36). The EEG recorded during sevoflurane general anesthesia (Fig. 2) suggests a smooth process defining continuity among the time intervals. Therefore, higher-order stationary models could be chosen to impose a greater degree of smoothness on the increment differences and the spectrogram. In contrast, the EEG recorded during induction of and recovery from propofol-induced unconsciousness (Fig. 3) suggests that a process with jump discontinuities may be a more appropriate state model for these data. *SI Appendix, Table S1* summarizes problems from different fields of science that have used MT spectral methods to study nonstationary processes. These several applications suggest a rich testbed for further development of the SS-MT paradigm.

Adaptive model parameter estimation using local likelihood, SS, or method of moments techniques can be combined with different continuity models. A key decision in using adaptive estimation is defining the timescale of the parameter updates. We used the target anesthetic levels as covariates—subject matter constraints—also be used to reduce the number of parameters. We limited state variance estimation to frequencies below 30 Hz based on knowledge of the frequency range relevant for tracking anesthetic states (36). The naïve empirical Bayes' CIs had good coverage probabilities based on our small simulation study of the model in Eq. 18. These intervals can be further calibrated by taking account of the uncertainty in the maximum likelihood

parameter estimates obtained from the EM algorithm (43). Computation of the SS-MT cross-spectrogram (*SI Appendix, Fig. S12*) suggests that our paradigm can be extended to inference problems for multivariate nonstationary time series.

The SS-MT paradigm provides a computationally efficient framework for spectrogram estimation, time-domain signal extraction, and statistical inference for nonstationary time series.

ACKNOWLEDGMENTS. This work was partially supported by National Research Foundation of Korea Grant 2017R1C1B5017254 from the Korean Government (to S.K.), a Guggenheim Fellowship in Applied Mathematics (to E.N.B.), NIH Transformative Research Award R01-GM104948 (to E.N.B.), NIH Award P01-GM118629 (to E.N.B.), funds from MGH (E.N.B.), and funds from the Picower Institute for Learning and Memory (E.N.B.).

- Chatterji S, Blackburn L, Martin G, Katsavounidis E (2004) Multiresolution techniques for the detection of gravitational-wave bursts. *Classical Quantum Gravity* 21:S1809–S1818.
- Haigh JD, Winning AR, Toumi R, Harder JW (2010) An influence of solar spectral variations on radiative forcing of climate. *Nature* 467:696–699.
- Haykin SS, Steinhardt AO (1992) *Adaptive Radar Detection and Estimation* (Wiley-Interscience, New York), Vol 11.
- Emery WJ, Thomson RE (2001) *Data Analysis Methods in Physical Oceanography* (Elsevier Science, Amsterdam).
- Misra P, Enge P (2006) *Global Positioning Systems: Signals, Measurements, and Performance* (Granga-Jamuna, Lincoln, MA).
- Zheng X, Chen BM (2013) *Stock Market Modeling and Forecasting* (Springer, London), Vol 442.
- Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN (2005) A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J Neurophysiol* 93:1074–1089.
- Mitra P, Bokil H (2007) *Observed Brain Dynamics* (Oxford Univ Press, New York).
- Yilmaz T, Foster R, Hao Y (2010) Detecting vital signs with wearable wireless sensors. *Sensors* 10:10837–10862.
- Quatieri TF (2008) *Discrete-Time Speech Signal Processing: Principles and Practice* (Prentice Hall, Englewood Cliffs, NJ).
- Percival DB, Walden AT (1993) *Spectral Analysis for Physical Applications* (Cambridge Univ Press, Cambridge, UK).
- Babadi B, Brown EN (2014) A review of multitaper spectral analysis. *IEEE Trans Biomed Eng* 61:1555–1564.
- Koornwinder TH (1995) *Wavelets: An Elementary Treatment of Theory and Applications* (World Scientific, River Edge, NJ).
- Huang NE, et al. (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc Math Phys Eng Sci* 454:903–995.
- Daubechies I, Lu J, Wu H-T (2011) Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool. *Appl Comput Harmon Anal* 30:243–261.
- Daubechies I, Wang YG, Wu H-T (2016) ConceFT: Concentration of frequency and time via a multitapered synchrosqueezed transform. *Philos Trans A Math Phys Eng Sci* 374:20150193.
- Fahrmeir L, Tutz G (2001) *Multivariate Statistical Modeling Based on Generalized Linear Models* (Springer, New York).
- Durbin J, Koopman SJ (2012) *Time Series Analysis by State Space Methods* (Oxford Univ Press, Oxford), Vol 38.
- Brown EN, Frank LM, Tang D, Quirk MC, Wilson MA (1998) A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *J Neurosci* 18:7411–7425.
- Smith AC, Brown EN (2003) Estimating a state-space model from point process observations. *Neural Comput* 15:965–991.
- Shumway RH, Stoffer DS (1982) An approach to time series smoothing and forecasting using the EM algorithm. *J Time Ser Anal* 3:253–264.
- Bohlin T (1977) Analysis of EEG signals with changing spectra a short-word Kalman estimator. *Math Biosci* 35:221–259.
- Kitagawa G, Gersch W (1996) *Smoothness Priors Analysis of Time Series* (Springer, Berlin), Vol 116.
- Tarvainen MP, Hiltunen JK, Ranta-aho PO, Karjalainen PA (2004) Estimation of non-stationary EEG with Kalman smoother approach: An application to event-related synchronization (ERS). *IEEE Trans Biomed Eng* 51:516–524.
- Qi Y, Minka TP, Picard RW (2002) Bayesian spectrum estimation of unevenly sampled nonstationary data. *Proc IEEE Int Conf Acoust Speech Signal Process* 2:1473–1476.
- Ba D, Babadi B, Purdon PL, Brown EN (2014) Robust spectrotemporal decomposition by iteratively reweighted least squares. *Proc Natl Acad Sci USA* 111:E5336–E5345.
- Shumway RH, Stoffer DS (2011) *Time Series Analysis and Its Applications* (Springer, New York).
- Prerau MJ, Brown RE, Bianchi MT, Ellenbogen JM, Purdon PL (2017) Sleep neurophysiological dynamics through the lens of multitaper spectral analysis. *Physiology* 32:60–92.
- Brillinger DR (2001) *Time Series: Data Analysis and Theory* (SIAM, Philadelphia), Vol 38.
- Czanner G, Eden UT, Wirth S, Yanike M, Suzuki WA, Brown EN (2008) Analysis of between-trial and within-trial neural spiking dynamics. *J Neurophysiol* 99:2672–2693.
- Carlin BP, Louis TA (2009) *Bayesian Methods for Data Analysis* (CRC, Boca Raton, FL), 3rd Ed.
- Dempster NM, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B Stat Methodol* 39:1–38.
- Dahlhaus R (1997) Fitting time series models to nonstationary processes. *Anal Stat* 25:1–37.
- Cimenser A, et al. (2011) Tracking brain states under general anesthesia by using global coherence analysis. *Proc Natl Acad Sci USA* 108:8832–8837.
- Purdon PL, et al. (2013) Electroencephalogram signatures of loss and recovery of consciousness from propofol. *Proc Natl Acad Sci USA* 110:E1142–E1151.
- Purdon PL, Sampson A, Pavone KJ, Brown EN (2015) Clinical electroencephalography for anesthesiologists part I: Background and basic signatures. *Anesthesiology* 123:937–960.
- Efron B (2013) *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction* (Cambridge Univ Press, Cambridge, UK).
- Priestley MB (1981) *Spectral Analysis and Time Series* (Academic, London).
- Adak S (1998) Time-dependent spectral analysis of nonstationary time series. *J Amer Statist Assoc* 93:1488–1501.
- Dahlhaus RD (2000) A likelihood approximation for locally stationary processes. *Annal Stat* 28:1762–1794.
- Fan J, Han F, Liu H (2014) Challenges of big data analysis. *Natl Sci Rev* 1:293–314.
- Doucet A, De Freitas N, Gordon N (2001) *Sequential Monte Carlo Methods in Practice* (Springer, New York).
- McLachlan GJ, Krishnan T (2008) *The EM Algorithm and Extensions* (Wiley, Hoboken, NJ).