

Capturing Spike Variability in Noisy Izhikevich Neurons Using Point Process Generalized Linear Models

Jacob Østergaard

ostergaard@math.ku.dk

University of Copenhagen, 2100 Copenhagen, Denmark

Mark A. Kramer

mak@math.bu.edu

Uri T. Eden

tzvi@math.bu.edu

Boston University, Boston, MA 02215, U.S.A.

To understand neural activity, two broad categories of models exist: statistical and dynamical. While statistical models possess rigorous methods for parameter estimation and goodness-of-fit assessment, dynamical models provide mechanistic insight. In general, these two categories of models are separately applied; understanding the relationships between these modeling approaches remains an area of active research. In this letter, we examine this relationship using simulation. To do so, we first generate spike train data from a well-known dynamical model, the Izhikevich neuron, with a noisy input current. We then fit these spike train data with a statistical model (a generalized linear model, GLM, with multiplicative influences of past spiking). For different levels of noise, we show how the GLM captures both the deterministic features of the Izhikevich neuron and the variability driven by the noise. We conclude that the GLM captures essential features of the simulated spike trains, but for near-deterministic spike trains, goodness-of-fit analyses reveal that the model does not fit very well in a statistical sense; the essential random part of the GLM is not captured.

1 Introduction ---

As recordings of neural activity become increasingly sophisticated, the resulting data become increasingly complex. Making sense of these data often requires more sophisticated approaches than visualization and simple summary statistics. One more advanced approach is the development and application of a model. Models serve both to characterize observed data and summarize the collective scientific knowledge of the brain. In neuroscience, as in many other fields, these models are typically segregated into two categories: dynamical or mechanistic models, and statistical models.

Dynamical models arise as an application of mathematical rules motivated by biophysical laws. These models tend to be deterministic—if not in practice, at least in spirit—and provide a mechanistic explanation for many dynamic brain activities. Statistical models are typically designed to capture data structure; these models often do not rely on neuronal biophysics.

The generalized linear model (GLM) has been an essential part of modern statistics since its introduction by Nelder and Wedderburn (1972), and today these models are ubiquitous in statistical analyses in many differing fields. This includes biological processes, where in recent years, GLMs have been used to describe coding properties and history dependence in neural spiking data (Kass & Ventura, 2001; Pillow et al., 2008; Sarma et al., 2012; Truccolo, Eden, Fellows, Donoghue, & Brown, 2005). The popularity of GLMs stems from the many advantageous features of this model class. The model defines a link function between the predictors and responses that ensures that the negative log likelihood is convex, allowing for fast optimal estimation of parameters while still admitting flexible model capabilities, by utilizing general basis functions (McCullagh & Nelder, 1989). As with statistical models in general, the GLM quantifies both predictable and unpredictable structure in data and possesses efficient procedures for parameter estimation and model diagnostics and powerful tools to assess goodness of fit. An advantage of including uncertainty as an intrinsic model component in the design stage is that this random component can, to some extent, compensate for an incomplete specification of the factors influencing the observed data.

Dynamical models have been a pillar in data modeling for neuroscientists since Lapicque's 1907 integrate-and-fire model (Brunel & van Rossum, 2007). Contrary to statistical models, these models typically focus on deterministic processes and are often built to explicitly represent mechanistic features of the data of interest. As such, dynamical models are often biophysically meaningful by construction, which may not be the case for a statistical model. Dynamical models are diverse, ranging from the biophysically realistic model of Hodgkin and Huxley (1952) to the more abstract model of Izhikevich (2003). The latter implements a simple mathematical design, while still producing realistic neural behaviors, such as bursting. Because of these properties, the Izhikevich model is a common choice, for example, in large-scale simulations of neurons with different behavior (Izhikevich & Edelman, 2008) and to simulate the impact of current input to large cortical networks (Ali, Sellers, & Fröhlich, 2013).

Historically, dynamical models have been used to examine deterministic or aggregate features of observed neuronal behavior. While dynamical models are often interpretable in terms of biophysical features, estimating parameters from individual spike trains is a complicated task. Most mechanistic models possess multiple parameters, which typically remain experimentally unconstrained. For example, the Hodgkin-Huxley (1952) model possesses at least seven parameters (e.g., capacitance, reversal potentials,

and maximal conductances for the sodium, potassium, and leak currents) with wide ranges of possible values. In computational neuroscience, a common procedure for estimating these parameters is hand-tuning to produce simulated model dynamics that qualitatively match the desired neuronal activity (Prinz, Billimoria, & Marder, 2003). Hand-tuning approaches usually require a great deal of time and expertise (Traub, Wong, Miles, & Michelson, 1991; Nadim, Olsen, de Schutter, & Calabrese, 1995; Vanier & Bower, 1999). Moreover, once a set of suitable parameters is found, it is often unclear whether the solution is unique or whether other model formulations exist that are compatible with the data (Prinz, Bucher, & Marder, 2004). Statistical models are often used to estimate features and describe associations but may not be directly physically interpretable.

Both modeling paradigms—dynamical and statistical—provide distinct advantages and disadvantages. Ideally, methods would exist to leverage the advantages of both approaches while mitigating their weaknesses. For example, a unified approach would allow researchers to better interpret statistical model results in terms of mechanistic features while exploiting the mathematical theory of statistical models. This theory includes tools of parameter estimation, but also rigorous procedures to validate the model against data. The aim of this letter is to investigate how simple GLMs perform when used to describe the spiking patterns obtained from simulated Izhikevich neurons with noisy input. The noise is included to resemble realistic physical observations and, more important, introduce and control variability of the data.

Previous work has analyzed the relationship between integrate-and-fire models and statistical techniques (Brunel & Latham, 2003; Paninski, 2006; Hertäg, Durstewitz, & Brunel, 2014). In these approaches, integrate-and-fire dynamical models were augmented by a stochastic component, and a statistical approach was used to estimate the firing rate. Here we pursue a different strategy, starting with a well-known class of history-dependent statistical models and analyzing the models' capabilities in capturing general features of various spike train patterns generated from a dynamical model. The class of GLMs in this exposition is very flexible, and, contrary to a dynamical model, there are no model assumptions regarding the specific mechanisms related to the physical processes that generate the spike train data, such as refractoriness. Recently, Weber and Pillow (in press) showed that the GLM framework replicates multiple spiking patterns from the Izhikevich model, but to the authors' knowledge, no work has been presented that analyzes how well the GLM class captures features of different spike patterns under the varying influence of noise.

In this letter, we show that the performance of the particular class of multiplicatively separable, history-dependent GLMs varies with the amount of noise added to the input and with the large-scale firing properties of the Izhikevich neuron. This class of GLMs is commonly used in spike train modeling (Pillow et al., 2008; Macke et al., 2011; Ahmadian, Pillow, &

Paninski, 2011; Latimer, Chichilnisky, Rieke, & Pillow, 2014). We examine how well the GLMs capture the intrinsic features of the individual neuron types, such as tonic spiking and bursting, as well as the variability in the individual spike trains originating from the noisy input. We first present the simulation model and define how variation is injected through both the input and the model parameters. Then the GLM design is outlined along with tools used for model assessment. In a series of numerical analyses, we demonstrate how GLMs behave for various levels of noise and different deterministic model features. We discuss in what ways the noise may affect interpretation of the GLM and the use of this model class for spike train modeling by evaluating the goodness of fit of the models in terms of capturing both predictable and unpredictable structure. Finally, we highlight some of the shortcomings of simple multiplicatively separable GLMs and discuss extensions to the standard model that can mitigate some of the limitations identified in this letter.

2 Methods

2.1 Simulation of Neuron Activity. In this letter, we implement the Izhikevich neuron to simulate neural spiking activity (Izhikevich, 2003). The Izhikevich neuron is a relatively simple, dynamical model capable of reproducing many types of neural spiking behavior. We note that mechanistic models of neuron spiking can often be reduced to two-dimensional systems having a fast voltage variable and a slower recovery variable. For the neuron model considered here, we may interpret v as a voltage and u as a recovery current, a phenomenological variable that represents the sum of all slow currents that modulate the spike generation mechanism (Izhikevich, 2010). The model includes four parameters a , b , c , and d and a further input current I_t . Equation 2.1 shows the deterministic dynamical model for $t \geq 0$:

$$\begin{aligned} \frac{dv_t}{dt} &= 0.04v_t^2 + 5v_t + 140 - u_t + I_t, \\ \frac{du_t}{dt} &= a(bv_t - u_t), \\ \text{if } v_t \geq 30 \text{ then } &\begin{cases} v_{t+} = c \\ u_{t+} = u_t + d \end{cases}. \end{aligned} \tag{2.1}$$

The dynamics in equation 2.1 generates spikes whenever v_t passes a fixed threshold of 30. Immediately following a spike at time t (where $v_t \geq 30$), both variables are reset at time $t+$; v_t is reset to a fixed value given by the parameter c , and u_t is increased by an amount d . The parameters a , b , c , d

determine the spike behavior; the parameters a and b act as a decay rate and a sensitivity parameter, respectively, and the parameters c and d determine how the variables are reset following a spike.

In order to introduce variation in simulations of equation 2.1, we let $I_t = I_0 + \sigma W_t$, where I_0, σ are constants and W_t denotes a standard white noise process (Øksendahl, 2007). The σ parameter controls the level of noise in the process and, in turn, the variability of spiking behavior. We may interpret σ as the noise in the input driving the neuron. Another source of variability comes from the different parameter values used to simulate the spike trains. In this letter, we model six types of neurons that produce distinct firing patterns (Izhikevich, 2003) when the input is constant. When $\sigma = 0$, the simulation model 2.1 is deterministic, and the neurons behave completely regularly, while for $\sigma > 0$, the model evolves stochastically. Assuming an Itô interpretation, model 2.1 was simulated using a Euler-Maruyama scheme with time step $\Delta t = 0.1$ ms, initial conditions $v_0 = -70$, $u_0 = bv_0$, and varying parameters a, b, c, d , and I_0 . We note that the 0.1 ms discretization is less than that used in Izhikevich (2003) and ensures that at most one spike occurs per discrete time bin. These simulations were then used to obtain spike trains. Although the Izhikevich neuron lacks the detailed biophysical mechanisms of more complex neuron models, this simple model serves as an excellent test bed for fitting the GLMs to diverse, realistic spike train behaviors and exemplifying the issues in interpreting the results.

2.2 Model Design. Let $N(t)$ denote the cumulative number of spikes observed up to time $t \in \mathbb{R}$. Define $\lambda(t|H_t, \theta)$ as the conditional intensity of $N(t)$, where H_t indicates the spiking history up to time t , and θ the model parameters. This conditional intensity defines the instantaneous probability of a spike given past spiking:

$$\lambda(t|H_t, \theta) = \lim_{\Delta t \rightarrow 0} \frac{P(\text{spike in } (t, t + \Delta t]|H_t, \theta)}{\Delta t}.$$

Given a set of observed spike times $\{s_j\}_{j=1}^n$ in an interval $[0, T]$, the log likelihood

$$\log L(s_1, \dots, s_n; \theta) = \int_0^T \log \lambda(t|H_t, \theta) dN(t) - \int_0^T \lambda(t|H_t, \theta) dt, \quad (2.2)$$

can be approximated as a discrete sum over individual time bins by assuming that the number of spikes in each bin is Poisson distributed with a rate parameter that depends on past spiking for sufficiently small Δt (see

Truccolo et al., 2005). The discrete likelihood approximation is

$$\log L(s_1, \dots, s_n; \theta) \approx \sum_k \log \lambda(t_k | H_{t_k}, \theta) \Delta N(t_k) - \sum_k \lambda(t_k | H_{t_k}, \theta) \Delta t, \quad (2.3)$$

where $\Delta N(t_k)$ counts the spike activity in $(t_k, t_k + \Delta t]$. Since the Poisson and binomial distribution converge in the limit $\Delta t \rightarrow 0$, either could be assumed in the individual time bins. However, choosing the Poisson distribution leads to the canonical log-link function (McCullagh & Nelder, 1989).

The approximate log likelihood, equation 2.3, shows that choosing the form of the intensity $\lambda(t_k | H_{t_k}, \theta)$ is a crucial step of the model design. To include history dependence, we include a trailing history of the m past spike times, and in order to reduce the dimensionality of the model, we also introduce indicator basis functions that account for the spike windows in preceding time bins, wider than the observation bins of width Δt . In order to simplify the model formulation, we assume that the influence of the previous spikes on the intensity is multiplicatively separable, a common model assumption when working with GLMs. Thus, when modeling the log intensity process, we deal with a sum of components.

Defining the $(m \times p)$ matrix B with the (i, j) th entry,

$$b_{ij} = \mathbf{1}_{[(j-1)w+1, jw]}(i) = \begin{cases} 1 & \text{for } i \in \{(j-1)w+1, \dots, jw\} \\ 0 & \text{else} \end{cases}, \quad (2.4)$$

where the parameter w denotes the number of Δt -wide time bins grouped, the model becomes

$$\log(\lambda(t_k | H_{t_k})) = \beta_0 + \sum_{j=1}^p \beta_j \sum_{i=1}^m y_{k-i} b_{ij}. \quad (2.5)$$

Here $y_k = \sum_{j=1}^n \mathbf{1}_{(t_k, t_k + \Delta t]}(s_j) = \Delta N(t_k)$, and the parameter vector is $\theta = (\beta_0, \beta_1, \dots, \beta_p)$. With $\Delta t = 0.1$ ms and $w = 10$, each β_j in equation 2.5 relates to the total spiking activity in 1 ms time bins.

Defining the (exponential) filter $\mathcal{F}(\tau) = \exp\left(\sum_{j=1}^p \beta_j \mathbf{1}_{[(j-1)w, jw]}(\tau)\right)$, this acts as the modulation of $\lambda_0 = \exp(\beta_0)$ at lagged time bins $t_{k-\tau}$.

As the randomness of the input vanishes, the y_k series becomes increasingly deterministic and tends to produce spiking at only a small set of interspike intervals. In this case, only a few of the y_{k-i} 's in equation 2.5 are nonzero at regularly lagged time bins. Since $\lambda(t | H_t, \theta) \Delta t \approx P(\text{spike in } (t, t + \Delta t] | H_t, \theta)$, then $\lambda(t_k | H_{t_k}) = \exp(\beta_0 + \sum_{j=1}^p \beta_j \sum_{i=1}^m y_{k-i} b_{ij}) \rightarrow 0$ in intervals with no spike activity. Denoting the estimate of β_j by $\hat{\beta}_j$,

this implies some estimates $\hat{\beta}_j$ must diverge to $-\infty$. In bins of spike activity, $P(\text{spike in } (t, t + \Delta t] | H_t, \theta) \rightarrow 1$, when the spiking is deterministic. Hence, the estimated $\hat{\beta}_j$'s in these bins will amplify the model's probability of spiking in $(t, t + \Delta t]$ in order to match the observations and approximate the relation $\lambda(t | H_t) \rightarrow \frac{1}{\Delta t}$. To summarize, for a given interval $(t, t + \Delta t]$,

$$\lim_{\Delta t \rightarrow 0} \lambda(t | H_t) = \begin{cases} \infty & \text{given spike activity} \\ 0 & \text{given no spike activity} \end{cases},$$

implies that $\lim_{\Delta t \rightarrow 0} \lambda(t | H_t)$ describes a set of Dirac delta functions centered at the spike times. We note that an example of this finding will be illustrated in Figure 4, which displays approximate delta functions at the periodic interval of spiking, when the data generating process is nearly deterministic.

2.3 Penalized GLM Regression. Data with negligible variability can lead to convergence issues of parameter estimates in equation 2.5. If there is little or no variation in the data, there is a high chance of including perfect predictors in the model that will cause instability in the estimation of model parameters since the likelihood surface will be close to flat in certain directions (see Wedderburn, 1976). Using a penalized regression is one way to handle these convergence issues and keep the parameter estimates finite.

Given a log-likelihood function $L(\theta)$ with parameters $\theta = (\beta_0, \beta_1, \dots, \beta_p)$, a penalized regression will maximize $L(\theta)$ subject to a constraint

$$L_\kappa(\theta) = L(\theta) - \kappa \sum_{j=0}^p \|\beta_j\|^q, \quad (2.6)$$

where $\|\cdot\|^q$ denotes the q -norm. For $q > 0$, the penalization in equation 2.6 will shrink the parameters toward 0. The choice of $q = 1$, corresponding to an L^1 penalization (LASSO regression), will, in addition to shrinkage, promote sparsity (Hastie, Tibshirani, & Friedman, 2001). The penalization parameter¹ κ determines how strong the shrinkage effect is, and $L_\kappa(\theta)$ implies that the optimization depends on some fixed κ . Choosing κ is a crucial part of using penalized regression, and therefore the choice should agree with the aim of the analysis. For the purposes presented in this exposition, we settled on two options that both depend on a goodness-of-fit test of an estimated model (see section 2.5). The first choice was simply to choose κ such that the KS statistic was minimized; we refer to this choice as *optimal KS*, where KS denotes the Kolmogorov-Smirnov test presented in

¹We chose here to denote the penalization by κ , rather than the standard λ , in order not to confuse the penalization factor in the model-fitting procedure with the intensity process, denoted $\lambda(\cdot)$.

section 2.5. The other choice was to choose the maximum penalization, κ , such that the goodness-of-fit p -value was insignificant with respect to a pre-determined threshold denoting the significance level. In what follows, we chose a threshold significance level of 0.05. We refer to the second choice for κ as *maximum κ* . Obviously the two differ in behavior. The first choice of κ selects models with a better goodness of fit, whereas the second selects sparser models due to the choice of $L1$ penalization in equation 2.6. We use the second choice (maximum κ) as the default to impose sparsity. However, to illustrate certain points, we also refer to models chosen by optimal KS in some parts. In the case that no p -values were above the threshold or if an optimal value could not be determined, the value of κ was set as the smallest values tried in the model estimation procedure: $\kappa \approx 0$. This occurred for 66% of the simulations for nonbursting neurons and 12% of bursting neurons. We note that $\kappa \approx 0$ still includes a small penalty so that the divergence of β_j estimates is avoided, but fewer parameters are set to zero with the LASSO regression.

2.4 Approximate Covariance Matrix. The observed Fisher information matrix $\hat{\mathbf{F}}$ was used to estimate the covariance matrix for the penalized GLM. The observed Fisher information for an unrestricted GLM model is given as

$$\hat{\mathbf{F}} = X' \hat{W} X,$$

where X is the so-called design matrix of the GLM that contains the predictors of the model and \hat{W} is a model-specific diagonal matrix that is computed as part of the estimation procedure (McCullagh & Nelder, 1989). For the models considered here, $X = HB$, where B is the matrix of indicator basis vectors described in section 2.2 and H is the matrix of m lagged observations, such that the k th row of H is $H_{(k)} = (y_{k-1}, \dots, y_{k-m})$. For a Poisson GLM, the diagonal matrix \hat{W} consists of the predicted intensities $\hat{\lambda}(t_k | H_{t_k})$. The inverse of the observed Fisher information, $\hat{\mathbf{F}}^{-1}$, was then used as an estimate of the covariance matrix of $\hat{\beta}$. Due to the penalization in equation 2.6, parameters were restricted from converging toward $\pm\infty$, but because of this convergence issue, the unrestricted model would not admit a sensible covariance estimates. The estimate $\hat{\mathbf{F}}^{-1}$ does not account for a penalization of the likelihood, and therefore the matrix $\widehat{\text{Cov}}(\hat{\beta}) = \hat{\mathbf{F}}^{-1}$ should be interpreted as only a rough approximation for the penalized model. Recent work addresses the bias in the covariance estimator due to inclusion of a penalization term in the model formulation (van de Geer, Bühlmann, Ritov, & Dezeure, 2014; Taylor & Tibshirani, 2015); however, standard software packages for LASSO estimation (e.g., `glmnet`, used here; Friedman, Hastie, & Tibshirani, 2010) have not yet implemented a correction for this bias. In this (letter), we focus on the sign of the covariance estimate and not the

specific value. Thus, although biased, $\widehat{\mathbf{F}}^{-1}$ can still reveal trends in the parameters. Defining $\Lambda = \sqrt{\text{diag}(\widehat{\mathbf{F}}^{-1})}$, then

$$\widehat{\text{Cor}}(\hat{\beta}) = \Lambda^{-1} \widehat{\mathbf{F}}^{-1} \Lambda^{-1} \quad (2.7)$$

is an estimate of the correlation matrix for $\hat{\beta}$. In section 3, we present correlation matrix estimates based on equation 2.7.

2.5 Goodness of Fit. In order to evaluate a statistical model, one must take into account both its structural and the random components. Intuitively one can think of this as measuring the model's ability to capture both the structure of the observed data (structural component) and generalizability by compensating for features that are not accounted for in the structural component through the statistical distribution (random component). Thus, with deterministic input, it is possible that a statistical model can describe the observed features to near perfection. However, such a model will rarely predict the features of some other data with only slight variations to the first. This lack of generalizability, caused by an inadequate fit of the random component in deterministic settings, will lead to poor statistical model diagnostics.

To assess the models' goodness of fit, the time rescaling theorem (Brown, Barbieri, Ventura, Kass, & Frank, 2002) was applied to the observed spike times $\{s_j\}_{j=1}^n$, using the estimated intensity process $\hat{\lambda}$, in order to obtain rescaled spike times $\{z_j\}_{j=1}^n$. The empirical distribution of the z_j 's was then compared to the theoretical $\text{Exp}(1)$ distribution using the Kolmogorov-Smirnov (KS) test statistic (Kass, Eden, & Brown, 2014). This statistic is given by

$$D_n = \sup |\hat{F}_n(x) - F_0(x)|, \quad (2.8)$$

where $\hat{F}_n(x)$ and $F_0(x)$ denote the empirical and theoretical cumulative distribution functions (CDFs) respectively. Plotting $\hat{F}_n(x)$ against $F_0(x)$ with approximate 95% confidence bounds $\pm \frac{1.36}{\sqrt{n}}$, as suggested by Kass et al. (2014), provides a visual assessment of the KS test. (For a detailed discussion of these goodness-of-fit procedures, we refer readers to Brown et al., 2002, and Kass et al., 2014.)

In addition to the KS statistic, equation 2.8, the relative deviance was used to evaluate the model fit. While the statistic D_n measures how the empirical model deviates from a theoretical model, the relative deviance can reveal the amount of structure present in the data. The relative deviance measures how an estimated model captures data features. The smallest model deviance for a given data set occurs with the saturated model, where the number of parameters equals the number of data points. Due to the

Table 1: Model Parameters a , b , c , d Used to Simulate the Six Types of Neurons and the Mean Input I_0 .

Neuron Type	a	b	c	d	I_0
Tonic spiking	0.02	0.20	-65	6	14
Phasic spiking	0.02	0.25	-65	6	1
Tonic bursting	0.02	0.20	-50	2	10
Phasic bursting	0.02	0.25	-55	0.05	1
Mixed mode	0.02	0.20	-55	4	10
Spike frequency adaptation	0.01	0.20	-65	8	20

Source: Parameter values and neuron type from Izhikevich (2004).

equal number of parameters and data points, there are no data left to estimate data variability. Therefore, the saturated model will completely describe the observed data set but will not generalize well to another data set. As such, we could say that the saturated model is the “most structural” we can define for a specific data set because it is purely descriptive. The other extreme is the null model, which is the one-parameter model including an intercept, β_0 , only. For the null model, only a single parameter is estimated, in this case defining a homogeneous Poisson process with no influence of past spiking. It is thus an example of the “most stochastic” model we can define, in the sense that it maximizes the entropy of the spike counting process. The relative deviance measures where an estimated model lies in the spectrum between these two extremes, with a relative deviance of zero corresponding to a fit equal to the saturated model and a relative deviance of one corresponding to a fit equal to the null model.

It is important to note that for a traditional deviance analysis, the goodness of fit of multiple models is compared by assessing their deviance on the same data set. Here, we examine for one particular class of GLMs the relative deviance across different data sets with different levels of input current noise. As such, a lower relative deviance of a different noise level should not be interpreted as an improved fit, but as one whose description of the data structure is closer to that of a saturated model.

3 Results

Spike trains for six types of Izhikevich neurons were simulated with the parameters displayed in Table 1. We note that the six types of neurons considered here mimic spiking behavior observed in in vivo and in vitro neural recordings (Izhikevich, 2004).

For each type of neuron, 10 spike trains were simulated for 29 values of σ evenly distributed in the interval $(0, 20]$. Hence, the simulations ranged from almost deterministic ($\sigma = 0.1$) to almost completely random spiking ($\sigma = 20$). As σ is increased, the intrinsic features of the individual neuron

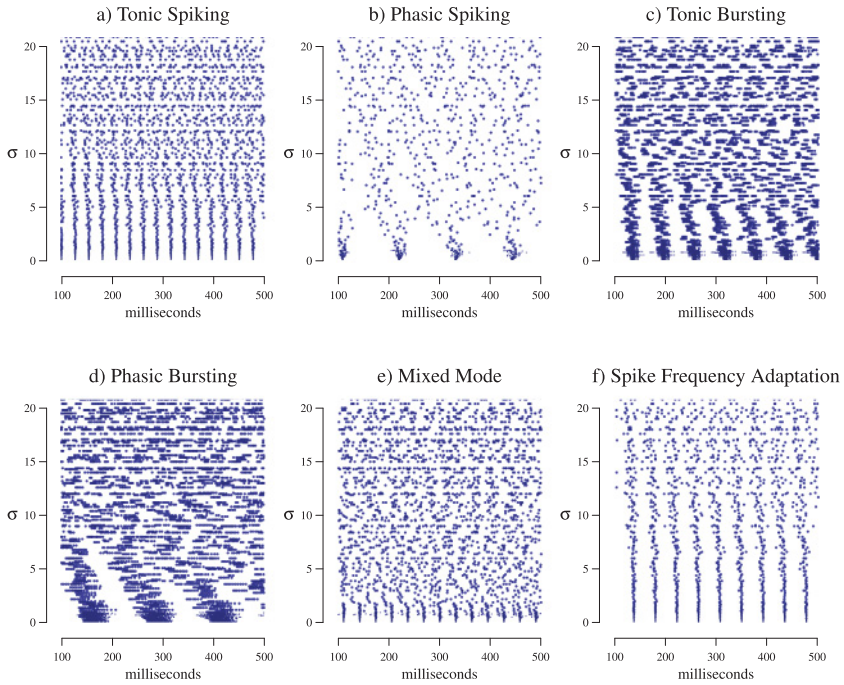


Figure 1: Simulated spike trains for various neuron types and values of σ . Each type displays nearly regular behavior as σ approaches 0 and almost completely random patterns as σ approaches 20. Distortion of the individual neuron-type behavior varies from rapid (phasic spiking and mixed mode) to gradual (tonic spiking and spike frequency adaption). Burst periods remain present at high levels of noise, and the spike frequency adaption neuron retains features of its spike pattern for σ as high as 15.

types become progressively noise driven and indistinguishable. When $\sigma \gg 0$, the spiking activity is based more on the randomness of the input than the model parameters. Each simulation consisted of $2 \cdot 10^5$ observations at times $t_k, k = 1, \dots, 2 \cdot 10^5$, with time step $\Delta t = 0.1$ ms, corresponding to 20 seconds of observations for each simulation. The spike trains were derived from the simulated voltage trajectories by determining when $v_{t_k} \geq 30$. Figure 1 presents the simulated spike trains for each type of neuron for varying σ values. An L^1 penalized regression for a Poisson GLM of the form 2.5 was fitted for each simulated spike train, where the value of κ in equation 2.6 was based on either optimal KS or maximum κ (see section 2.3), depending on the analysis.

The history dependence of the GLM was set to 100 ms, corresponding to $m = 1000$ time steps for a discretization step of $\Delta t = 0.1$ ms. With the

indicator width set to 1 ms ($w = 10$), the model, equation 2.5, had a maximum of $100 + 1$ parameters. Some of the parameters are shrunk to zero by the penalized regression procedure, and as such, the resulting filters consisted of fewer than 101 nonzero parameters.

The numerical analysis was carried out with the statistical programming environment *R* (R Core Team, 2016) using the package *glmnet* to perform penalized regression for GLMs.

3.1 Tonic Spiking and Bursting with Intermediate Noise. We first present a more detailed analysis of the tonic spiking and tonic bursting neurons for a single value of σ to illustrate the model fitting and the goodness-of-fit analyses. These neuron types were chosen specifically to investigate how the GLM handles the regular behavior of a tonic spiking neuron, which possesses a unimodal interspike interval (ISI) distribution, versus the switching behavior of a bursting neuron, which produces a bi-modal ISI distribution. The value of $\sigma = 5$ was chosen such that the neurons display both a predictable structure of interest as well as variability, as evident from Figures 1a and 1c. The analyses for the two types of neurons are presented in Figures 2 and 3 for the tonic spiking and tonic bursting neurons, respectively. These figures display an interval of simulated spiking activity for the neuron, with the estimated intensity function $\hat{\lambda}(t|H_t)$ below, the corresponding histogram of ISIs, the histogram of rescaled spike times, the estimated filter, a KS plot of $\hat{F}_n(x)$ versus $F_0(x)$ with approximate 95% confidence bounds, and a plot of the residual process $N(t) - \int_0^t \hat{\lambda}(t|H_t)dt$. Penalization was set by maximum κ with a KS statistic significance threshold above 0.05 for both analyses.

The estimated intensity for the tonic spiking neuron (see Figure 2b) displays peaks that coincide with the spikes of the potential (see Figure 2a). There is a rise in the intensity prior to spiking and an instantaneous drop immediately following an observed spike, which implies that the model captures the predictable structure well. The ISI histogram for the tonic spiking neuron (see Figure 2d) is unimodal with a mean of 26.6 ms, and the width of the histogram implies some variability between spike times, which is expected for the choice of $\sigma = 5$. The estimated filter (see Figure 2e) displays a refractory period up to 20 ms and shows peaks around multiples of 26.6 ms, indicating that the model captures the regularity of spiking; this result agrees well with the type of neuron considered. The multiple peaks in the filter suggest that at this level of σ , the current probability of spiking is well predicted by the most recent spike as well as the previous two spikes. The rescaled spike time histogram (see Figure 2f), overlaid with an $\text{Exp}(1)$ distribution (red line), shows that the model agrees well with the theoretical distribution, which is further supported by the KS plot (see Figure 2g). Furthermore, the KS statistic, $D_n^{\text{spiking}} = 0.046$, yields a p -value of 0.088, in accordance with the choice of κ such that the p -value is > 0.05 . For reference,

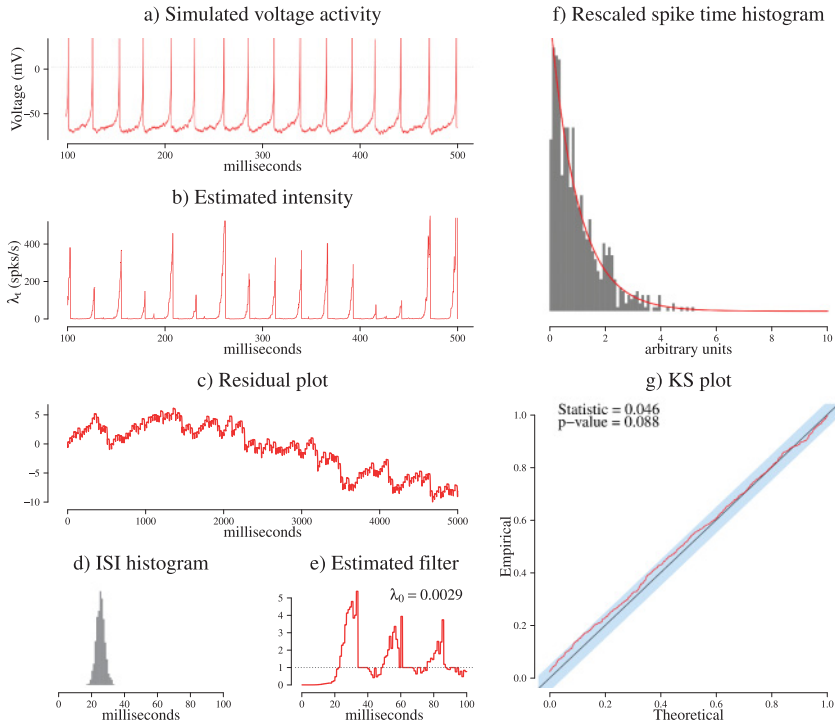


Figure 2: At a moderate level of noise, the GLM captures features of the tonic spiking neuron. (a) The simulated voltage activity and (b) the corresponding estimated intensity process. (c) The residual plot does not show any trending behavior. The ISI histogram (d) possesses a broad peak, and the peaks in the estimated filter (e) are consistent with the approximate interval between spikes. The rescaled spike time histogram (f) is well approximated by an $\text{Exp}(1)$ probability density function (red curve) and the KS plot (g) indicates a decent fit.

using optimal KS instead to set κ , the KS statistic was 0.022 with a p -value of 0.88. Inspection of the residuals (first quarter of the total residual process is shown in Figure 2c) does not give rise to concerns regarding time-dependent trends.

For the tonic bursting model in Figure 3, the estimated intensity (see Figure 3b) shows that bursting is well captured; we note that increases in the estimated intensity occur just before the times of action potential bursts (see Figure 3a). The rise of the intensity before a burst varies slightly between individual bursts, and the drop after a period of bursting is not as sharp as for the case for tonic spiking. These observations suggest that it is more complicated to capture the beginnings and ends of bursts than single spikes. We note that the filter (see Figure 3e) does not possess a second peak between

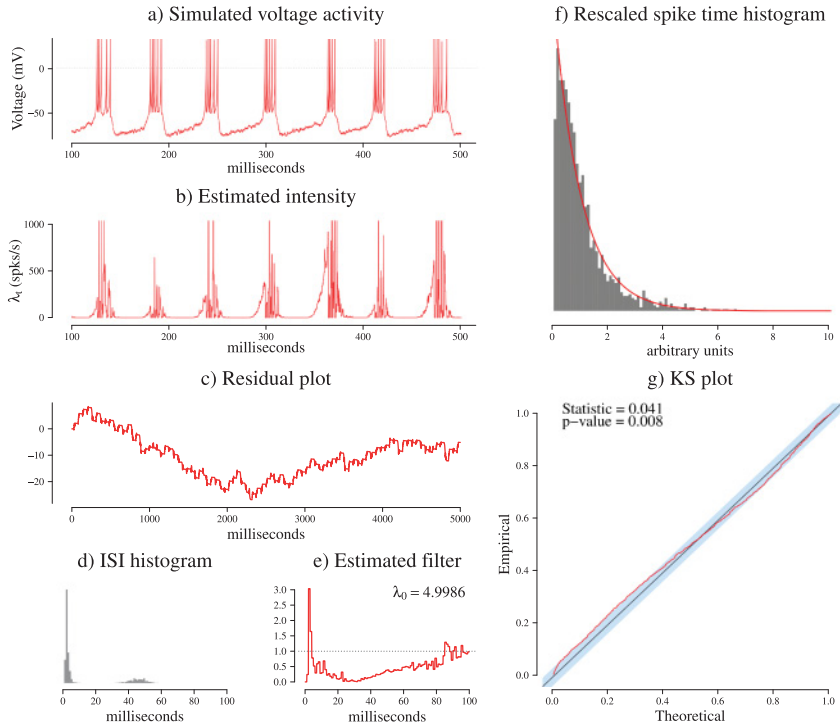


Figure 3: At a moderate level of noise, the GLM captures features of the tonic bursting neuron. Example of the (a) simulated voltage activity and (b) the corresponding estimated intensity process. The residual process (c) does not display trends, and the bimodal ISI histogram (d) is approximated by the filter (e). The rescaled spike time histogram (f) is well approximated by an Exp(1) probability density function (red curve), while the KS plot (g) shows slight overestimation of short ISIs and underestimation of longer ISIs.

40 ms and 50 ms, as seen in the ISI histogram (see Figure 3d). Instead, there is an inflection in the inhibitory trough of the filter starting around 40 ms. When convolved with the past bursting activity this inflection leads to a rapid rise in intensity approximately 40 ms after the final spike in the previous burst. The KS plot in Figure 3g and the residual process in Figure 3c provide further evidence that the GLM captures the structure in the bursting data well.

The rescaled ISI histogram in Figure 3d agrees with the Exp(1) distribution, and the KS statistic is $D_n^{\text{burst}} = 0.041$, which is less than the value for the tonic spiking neuron: $D_n^{\text{spiking}} = 0.046$. However, the corresponding p -value is significantly lower at 0.008, because for the tonic bursting

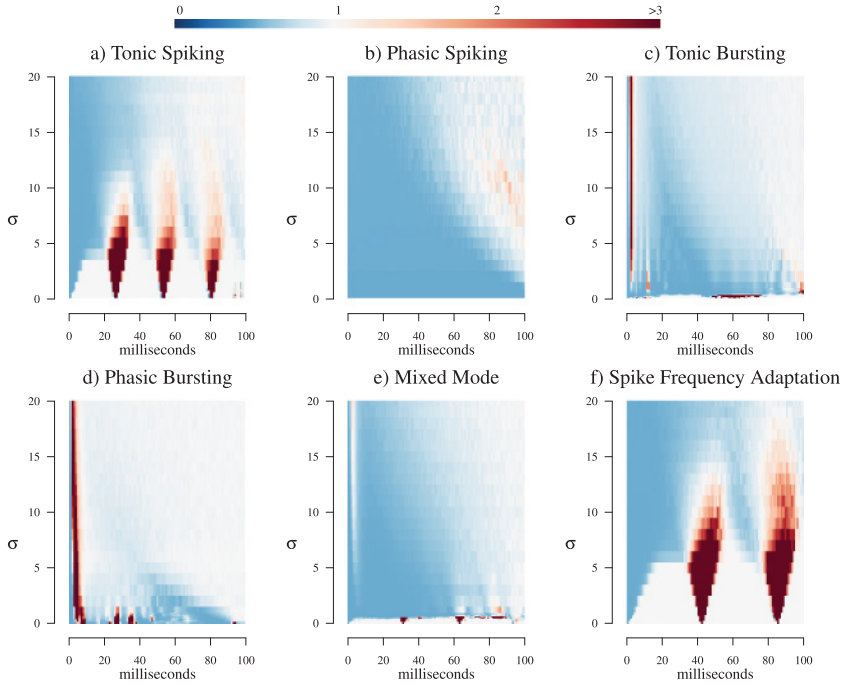


Figure 4: Estimated filters \mathcal{F} using maximum κ penalization with threshold 0.05, as functions of lagged time (in ms) and σ . Blue colors indicate refractory periods ($\mathcal{F} < 1$), and red colors indicate excitatory effects ($\mathcal{F} > 1$). White indicates $\mathcal{F} \approx 1$. Peaks are clear for tonic spiking and spike frequency adaption neurons in panels a and f, as well as for bursting neurons in panels c and d. Refractory periods are present for all types and vary with σ . Note that the \mathcal{F} 's converge toward collections of delta functions as σ approaches 0.

neuron, the number of observed spikes is $N^{\text{bursting}}(T) = 1668$ compared to $N^{\text{spiking}}(T) = 757$ for the tonic spiking neuron. The larger number of spikes results in a much lower p -value for tonic bursting, since the KS test is sensitive to the number of observed spikes. Using optimal KS to set κ for the tonic bursting did not improve the model fit; the KS statistic was 0.040 with a p -value of 0.009.

3.2 Spike History Filters. We now turn to a more general analysis of the models, across a wider range of firing properties and noise levels, σ . Figure 4 presents the estimated GLM filters for combinations of simulated neuron types and σ values. The penalization parameter κ was set by maximum κ with a threshold of 0.05 (see section 2.3). Filters for regularly spiking neurons, such as the tonic spiking neuron, Figure 4a, and the spike frequency

adaption neuron, Figure 4f, display peaks in the filters for low to intermediate values of σ . For $\sigma \rightarrow 0$ the filters approach collections of delta functions, with mass concentrated around multiples of the deterministic (ISI) period. For example, for the deterministic tonic spiking neuron, the ISI is ≈ 27 ms; hence, the peaks in the filter \mathcal{F} appear near 27, 54, and 81 ms in agreement with the regular ISIs of the tonic spiking neuron. For the bursting neurons, Figures 4c and 4d, there is a narrow peak at approximately 2 ms to 5 ms, except for very small σ . Decreases in the filter follow the peaks at short times, indicating that refractory periods are captured by the model for all combinations of type and σ . However, the refractory period varies as σ increases: either increasing (for tonic spiking) or decreasing (phasic spiking), or remaining approximately unchanged (for tonic bursting). This indicates a minimal absolute refractory period for bursting neurons that is independent of the input noise. Common to all filters is a flattening of \mathcal{F} as σ increases, most visibly for regularly spiking neurons (tonic spiking and spike frequency adaption).

This implies that by increasing the injected noise σ , the spiking approaches a homogeneous Poisson process, and it becomes progressively difficult to extract mechanistic structure, such as the interval of regular spiking (see Figure 4a) or interbursting activity.

3.3 Goodness of Fit. KS tests were performed for combinations of neuron type and σ to assess each model's goodness of fit. Figure 5 presents an overview of these results. The plot displays the $-\log(D_n)$, where D_n is the KS statistic (see equation 2.8) and the relative deviance, respectively. The statistics, D_n , were log-transformed to emphasize trends as σ varies. Furthermore, in order to present implications of changing σ , the penalization parameter κ was set by optimal KS, such that different σ 's could lead to different penalizations. Had we instead chosen penalization by maximum κ , the KS statistics would not reveal an optimal range of σ .

The $-\log(D_n)$ for all simulated neuron types in Figure 5 increases quickly as σ increases from 0 to 2, suggesting that for low values of σ , the GLM has difficulty in capturing the near-deterministic spiking distribution but does well at capturing this distribution once some variability is present. The corresponding p -values are above 0.05 for $\sigma > 1$ for all neuron types, except phasic bursting and tonic bursting. For phasic bursting, the p -values exceed 0.05 for $\sigma \in [5, 6]$, whereas for tonic bursting, the p -values never exceed 0.05. Except for mixed mode and spike frequency adaption, all types display optimal values in a range of σ 's, where the mentioned types seem to settle at a plateau for $\sigma \geq 4$. The optimal range, both the width and location, of σ depends on the type of neuron.

For the relative deviances, we chose to use maximum κ . The increasing trend in the relative deviance suggests that for low values of σ , there is more predictable structure captured by the model, whereas for higher noise levels, the estimated model approaches a homogeneous Poisson model, which

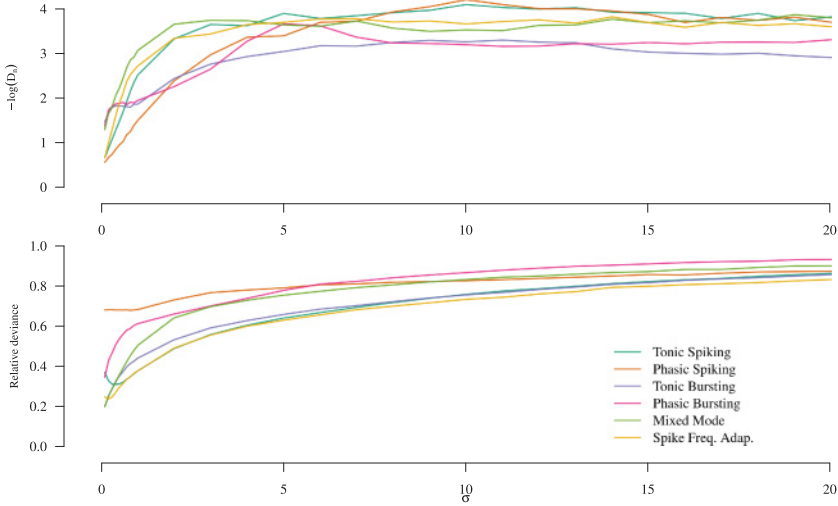


Figure 5: Two methods to assess goodness of fit for varying σ . (Top) For the log-transformed KS statistics D_n , larger values correspond to smaller KS statistics. (Bottom) Relative deviance of the estimated model. A value of zero indicates a fit equal to the saturated model, and a value of one indicates a fit equal to the null model.

includes only a baseline firing parameter. This general trend is consistent with our intuition for the model behavior; as the randomness of the input overshadows features of the Izhikevich model dynamics, the spike trains become better described by a homogeneous Poisson process.

We conclude from these two assessments that an optimal range for σ exists in which the GLM captures both the variability and predictable structure of the spike trains simulated using the noisy Izhikevich neurons. For very low values of σ , there is little variability to be captured by the GLM, while for very high values of σ , there is little predictable structure beyond the baseline firing rate. However, a general optimal range of σ cannot be explicitly defined as it depends on the type of neuron. Thus, multiple goodness-of-fit measures should be used when analyzing data to assess the degree to which the model fits the variability (e.g., KS test) and the predictable structure in the model (e.g., relative deviance).

3.4 Model Structure for Extreme σ Values. The results in Figure 4 display the estimated filter parameters as $\sigma \rightarrow 0$. We note that the maximum estimated filter values are capped at a value of 3 in Figure 4, so the full extent of the peak values is not directly visualized. The peaks, increasing in height and decreasing in width, resemble delta functions as the noise level

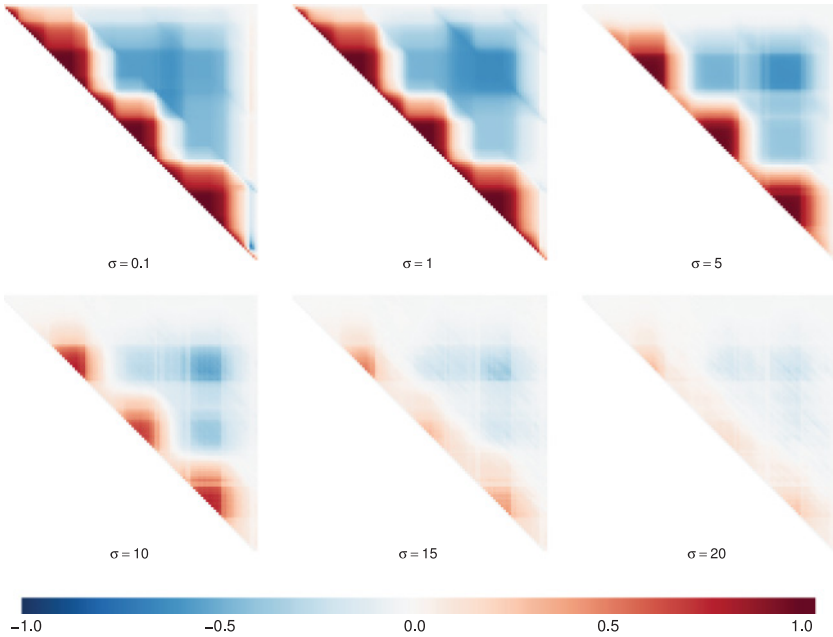


Figure 6: Estimated correlation matrices of the tonic spiking neuron with $\sigma = 0.1, 1, 5, 10, 15, 20$. Notice how the correlation structure vanishes as σ increases. The case $\sigma = 5$ (top, right) displays three noticeable bumps, corresponding to the three peaks of the filter in Figure 2. There are clear positive correlation among lags close to each other and negative correlation among lags farther apart.

approaches zero. In the other extreme of large σ , most of the visible structure in the filter vanishes, except for a refractory period. Except in the case of bursting neurons, the filters flatten (i.e., approach 1) for large σ , implying that less neuron-specific structure is captured by the models. This finding is consistent with the results in Figure 5, where the relative deviances approach 1 (the null model) with low values of the corresponding KS statistics, indicating that the simulated spike trains are well approximated by a GLM with an intercept (baseline) term only, which corresponds to a homogeneous Poisson process.

The estimated correlation matrices (see section 2.4) for the tonic spiking neuron in Figure 6 reveal that for low σ values, there is a positive correlation (red) between parameters that are close to each other in the temporal dimension along the diagonal, but at the three peaks, the positive correlation extends further in time. There is negative correlation (blue) among variables that are far apart in the temporal dimension, but most

interesting, the clusters of variables associated with each peak exhibit negative correlation.

The negative correlation between peak variables suggests that past spiking at these lags is highly correlated and that as predictors, they are largely redundant. We note that the neuron spike probability can be affected by the first, second, or third last-recorded spike, but it is the total accumulated effect of the past 100 ms that determines the current spike probability. As such, the filter should be evaluated as a whole rather than at individual points in time.

As $\sigma > 5$, the correlation structure starts to disappear, and for higher values of σ , the correlation structure in the model is greatly reduced. Once more, this is in line with the findings in Figure 5 as models for very noisy data capture mostly baseline activity.

For the tonic bursting neuron, the correlation matrices (not shown) showed little or no structure beyond a few milliseconds off-diagonal. Based on Figure 4, this lack of correlation structure is not surprising. For higher values of σ , the filter captured only the increased probability of a spike with short delay, but as the remaining part of the filter approaches 1, the (irregular) interburst periods were not captured. Thus, any relatively strong dependence beyond the intraburst interval is not expected as σ increases. As in the case of the tonic spiking filters, the entire filter should be interpreted for tonic bursting. Thus, it is not only the time of the last spike that determines the current probability but the 100 ms history of spikes, which, in the case of bursting, can include multiple spikes in a burst or multiple burst periods, that affects the probability of spiking.

4 Discussion

In this letter, we have examined how well a commonly used class of GLMs performs when used to capture both predictable structure and variability of spike trains derived from simulated, noisy Izhikevich neurons. A useful model is one that allows for clear, simple interpretations, which depends critically on the form of the model. When referring to GLMs in this letter, we implicitly mean the specific class of multiplicatively separable history-dependent GLMs. These GLMs were designed to capture the influence of past spiking on the current firing intensity using a simple set of indicator basis functions and assuming that the influences of previous spikes are multiplicatively separable. This indicator basis is well suited to display the approximation of delta functions, when the noise level converges to 0 and the spike trains become increasingly deterministic. However, using indicator basis functions increases the number of parameters compared to other choices such as splines' bases and can lead to the problem of perfect separation. Therefore, it was necessary to use a penalized regression to fit the models with indicator basis functions for low noise levels where only a few parameters suffice to describe the structure. For this reason, we opted for

an L^1 penalization, corresponding to LASSO regression, which shrinks redundant parameters to 0 by promoting sparsity. The implementation of the LASSO regression we used here utilized cyclical coordinate descent (Friedman et al., 2010), when optimizing the likelihood, whereas the standard GLM routine in R uses Fisher scoring to find maxima. Even when the penalization parameter was set to zero (i.e., unconstrained regression), these two methods can yield rather different results when the likelihood surface is nearly flat. However, even in such cases, the overall model predictions and diagnostics were indistinguishable.

As expected, the estimated filters resembled delta functions for low values of σ (see Figure 4). In general, the GLMs captured both predictable structure and variability from the input data. For near-deterministic spike trains, the variation is negligible, and as such, the interpretation and quality of the fit become highly sensitive to the choice of basis functions. Other commonly used basis functions, such as splines, which impose the smoothness of the influence of past spikes, are likely to lead to poorly fit models and incorrect interpretations when the spiking process is exactly or nearly deterministic. Although we focus here on simulations, we expect to encounter similar issues when applying the GLM (see equation 2.5) to analyze regular spike train activity recorded from a real neuron.

In contrast to low-noise regimes, spike trains simulated with a strong, noisy input signal produced spike patterns consistent with a homogeneous Poisson process (see Figure 1). In the corresponding filters for these highly stochastic spike trains, any structure besides a refractory period following an observed spike vanished. For tonic spiking, there was no positive modulation in very noisy regimes, whereas for tonic bursting, the intraburst interval at short delay was captured, while the interburst interval at longer delays was missing. This loss of structure was also evident from the relative deviance in Figure 5, which approached 1 as σ increased.

Our analyses suggest that multiple goodness-of-fit diagnostics are necessary to determine the extent to which statistical models capture the predictable structure and variability of neural spiking. For very low-noise regimes, measures that relate primarily to variability, such as the KS statistic, may indicate lack of fit, even for models that capture the predictable structure of the data well. When the random component is ill fitting, the model does not generalize well to data of similar, but slightly mismatching, structure, thus resulting in poor values of the KS measure. In general, goodness-of-fit measures should account for how well a model captures both the variability (e.g., KS test) and predictable structure (e.g., relative deviance) in the data.

Analyzing tonic spiking and tonic bursting neurons, we found that the GLM captures both types quite well, although the KS statistic remained significant for tonic bursting. This was despite the fact that the KS statistics for both types were of equal magnitude and driven by the higher number of observed spikes for the tonic bursting neuron compared to the tonic spiking neuron. However, by interpreting the estimated filters, we found that

both exhibited a clear structure inherent to the neuron type with features that are intuitively understandable.

There are a number of ways we might extend the models discussed here to better capture structure and variability in the data or to allow for more interpretable parameter estimates. A straightforward model restriction would be to replace the multiplicatively separable effects of multiple previous spikes in the GLM with a renewal model structure,

$$\log(\lambda(t|H_t)) = \beta_0 + g(s^*, t), \quad (4.1)$$

where $g(s^*, t)$ is a function of the previous observed spike time,

$$s^* = \max_j \{s_j | s_j < t\},$$

and time t . With this formulation, estimating a filter \mathcal{F} for equation 4.1 of a tonic spiking neuron would not produce three peaks as in Figures 2 and 4, only a single peak around the mean ISI at 26.6 ms, since it would take into account only the time since the previous spike. Besides only a single peak for tonic spiking, the estimated filter for equation 4.1 is expected to display similar behavior with regard to convergence and dependence, like the filters for tonic spiking examined in this letter. We found that a history-dependent filter for tonic bursting does not repeat itself for filters accounting for only 100 ms of past spiking history. However, an estimated filter for equation 4.1 would not integrate the past history, as was the case in Figure 3e. This might lead to a more intuitive filter that would display peaks at both bursting and interbursting ISIs.

It was clear from the analysis of tonic bursting that the estimated filter should be interpreted as a whole and not at individual time points. For the tonic bursting neuron, the ISI histogram was bimodal due to the two timescales involved. This was well captured by the model, as evident from the estimated intensity (see Figure 3b); however, the separation of the timescales was not clear from the filter (see Figure 3e). A possible way to extend the model to explicitly account for this separation is to include a latent state that determines when the neuron is bursting. This could be formulated as

$$\log(\lambda(t|X_t, H_t)) = \beta_0 + \mathbf{1}_{\{X_t=1\}}[\log(\mu(t|H_t))] + \mathbf{1}_{\{X_t=0\}}[\log(\gamma(t|H_t))], \quad (4.2)$$

where

$$X_t = \begin{cases} 1 & \text{neuron is bursting at time } t \\ 0 & \text{else} \end{cases}$$

and $\mu(\cdot|\cdot)$, $\gamma(\cdot|\cdot)$ are individual intensities describing intraburst and interburst times, respectively. A model such as equation 4.2 could potentially capture both intrabursting and interbursting better than the models considered here due to the explicit modeling of the dual timescales present. Estimation of the state X_t could also possibly link parameters of such a model to the parameters of the Izhikevich model that controls bursting behavior. The goal in this presentation was to examine a commonly used class of GLMs and analyze their performance in relation to simulated Izhikevich neurons. As such, we leave an analysis of the described model extensions to future work.

In conclusion, the results of this letter suggest that for the commonly used GLM for spike train data with multiplicatively separable history dependence, there is a range of input noise values for the Izhikevich neurons in which the GLM optimally captures both predictable structure and variability. This range depends on which properties are intrinsic to the spike train: intraburst intervals are captured even at high noise levels, while regular spiking and interburst spiking intervals are not captured at high noise levels. At low noise levels, the basis functions implemented here are delta functions, which can predict data structure but do not generalize well to data from slightly altered generative models. Thus, the choice of basis functions becomes crucial to the credibility of the model for spike trains with little or no variation. However, although an optimal range is evident from the results in this letter, extensions to the model formulation 2.5, such as equations 4.1 and 4.2, could possibly lead to improved model descriptions of specific neuron properties such as bursting and thus extend the optimal range. Expanding the optimal working range of the GLMs would be of interest, for instance, in order to classify neurons according to both type and noise level.

Acknowledgments

J.Ø. is a participant of the Dynamical Systems Interdisciplinary Network, University of Copenhagen. M.A.K. was partially supported by the NSF CAREER grant NSF-DMS 1451384. U.T.E. was partially supported by a grant from the Simons Foundation (320136).

References

- Ahmadian, Y., Pillow, J. W., & Paninski, L. (2011). Efficient Markov chain Monte Carlo methods for decoding neural spike trains. *Neural Comput.*, 23(1), 46–96.
- Ali, M. M., Sellers, K. K., & Fröhlich, F. (2013). Transcranial alternating current stimulation modulates large-scale cortical network activity by network resonance. *Journal of Neuroscience*, 33(27), 11262–11275.

- 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 Comput.*, 14(2), 325–346.
- Brunel, N., & Latham, P. E. (2003). Firing rate of the noisy quadratic integrate-and-fire neuron. *Neural Computation* 15(10), 2281–2306.
- Brunel, N., & van Rossum, M. C. W. (2007). Lapicque's 1907 paper: From frogs to integrate-and-fire. *Biological Cybernetics*, 97(5), 337–339.
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1–22.
- Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning*. New York: Springer.
- Hertäg, L., Durstewitz, D., & Brunel, N. (2014). Analytical approximations of the firing rate of an adaptive exponential integrate-and-fire neuron in the presence of synaptic noise. *Frontiers in Computational Neuroscience*, 8, 116.
- Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, 117(4), 500–544.
- Izhikevich, E. M. (2003). Simple model of spiking neurons. *Trans. Neur. Netw.*, 14(6), 1569–1572.
- Izhikevich, E. M. (2004). Which model to use for cortical spiking neurons? *Trans. Neur. Netw.*, 15(5), 1063–1070.
- Izhikevich, E. M. (2010). *Dynamical systems in neuroscience: The geometry of excitability and bursting*. Cambridge, MA: MIT Press.
- Izhikevich, E. M., & Edelman, G. M. (2008). Large-scale model of mammalian thalamocortical systems. *Proceedings of the National Academy of Sciences*, 105(9), 3593–3598.
- Kass, R. E., Eden, U. T., & Brown, E. N. (2014). *Analysis of neural data*. New York: Springer.
- Kass, R. E., & Ventura, V. (2001). A spike-train probability model. *Neural Computation*, 13(8), 1713–1720.
- Latimer, K., Chichilnisky, E., Rieke, F., & Pillow, J. (2014). Inferring synaptic conductances from spike trains under a biophysically inspired point process model. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems*, 14 (pp. 954–962). Cambridge, MA: MIT Press.
- Macke, J. H., Buesing, L., Cunningham, J. P., Yu, B. M., Shenoy, K. V., & Sahani, M. (2011). Empirical models of spiking in neural populations. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems*, 24 (pp. 1350–1358). Red Hook, NY: Curran.
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman and Hall/CRC.
- Nadim, F., Olsen, Ø. H., de Schutter, E., & Calabrese, R. L. (1995). Modeling the leech heartbeat elemental oscillator. I: Interactions of intrinsic and synaptic currents. *Journal of Computational Neuroscience*, 2(3), 215–235.
- Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A (General)*, 135(3), 370–384.
- Øksendahl, B. (2007). *Stochastic differential equations*. New York: Springer.

- Paninski, L. (2006). The spike-triggered average of the integrate-and-fire cell driven by gaussian white noise. *Neural Computation*, 18(11), 2592–2616.
- Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., & Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207), 995–999.
- Prinz, A. A., Billimoria, C. P., & Marder, E. (2003). Alternative to hand-tuning conductance-based models: Construction and analysis of databases of model neurons. *Journal of Neurophysiology*, 90(6), 3998–4015.
- Prinz, A. A., Bucher, D., & Marder, E. (2004). Similar network activity from disparate circuit parameters. *Nat. Neurosci.*, 7(12), 1345–1352.
- R Core Team. (2016). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.
- Sarma, S. V., Cheng, M. L., Eden, U., Williams, Z., Brown, E. N., & Eskandar, E. (2012). The effects of cues on neurons in the basal ganglia in Parkinson's disease. *Frontiers in Integrative Neuroscience*, 6, 40.
- Taylor, J., & Tibshirani, R. J. (2015). Statistical learning and selective inference. *Proceedings of the National Academy of Sciences*, 112(25), 7629–7634.
- Traub, R. D., Wong, R. K., Miles, R., & Michelson, H. (1991). A model of a CA3 hippocampal pyramidal neuron incorporating voltage-clamp data on intrinsic conductances. *Journal of Neurophysiology*, 66(2), 635–650.
- Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., & Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of Neurophysiology*, 93(2), 1074–1089.
- van de Geer, S., Bühlmann, P., Ritov, Y., & Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.*, 42(3), 1166–1202.
- Vanier, M. C., & Bower, J. M. (1999). A comparative survey of automated parameter-search methods for compartmental neural models. *Journal of Computational Neuroscience*, 7(2), 149–171.
- Weber, A. I., & Pillow, J. W. (in press). Capturing the dynamical repertoire of single neurons with generalized linear models. *Neural Computation*.
- Wedderburn, R. W. M. (1976). On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. *Biometrika*, 63(1), 27.