HHS Public Access

Author manuscript

IEEE Trans Biomed Eng. Author manuscript; available in PMC 2019 February 01.

Published in final edited form as:

IEEE Trans Biomed Eng. 2018 February ; 65(2): 241–253. doi:10.1109/TBME.2017.2762687.

Developing a Nonstationary Computational Framework with Application to Modeling Dynamic Modulations in Neural Spiking Responses

Amir Akbarian,
Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT 84112, USA

Kaiser Niknam,
Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT 84112, USA

Moahammadbagher Parsa,
Neuralynx Inc., Bozeman, MT 59715, USA

Kelsey Clark,
Department of Cell Biology and Neuroscience, Montana State University, Bozeman, MT 59717, USA

Behrad Noudoost, and
Department of Ophthalmology and Visual Sciences, University of Utah, Salt Lake City, UT 84132, USA

Neda Nategh [Member, IEEE]
Department of Electrical and Computer Engineering and the Department of Ophthalmology and Visual Sciences, University of Utah, Salt Lake City, UT 84132, USA

Abstract

Objective—This paper aims to develop a computational model that incorporates the functional effects of modulatory covariates (such as context, task, or behavior), which dynamically alter the relationship between the stimulus and the neural response.

Methods—We develop a general computational approach along with an efficient estimation procedure in the widely used generalized linear model (GLM) framework to characterize such nonstationary dynamics in spiking response and spatiotemporal characteristics of a neuron at the level of individual trials. The model employs a set of modulatory components, which nonlinearly interact with other stimulus-related signals to reproduce such nonstationary effects.

Results—The model is tested for its ability to predict the responses of neurons in the middle temporal cortex of macaque monkeys during an eye movement task. The fitted model proves successful in capturing the fast temporal modulations in the response, reproducing the spike

Corresponding authors: N. Nategh and B. Noudoost.

response temporal statistics, and accurately accounting for the neurons' dynamic spatiotemporal sensitivities, during eye movements.

Conclusion—The nonstationary GLM framework developed in this study can be used in cases where a time-varying behavioral or cognitive component makes GLM-based models insufficient to describe the dependencies of neural responses on the stimulus-related covariates.

Index Terms

point process models; nonstationary models; generalized linear model; response modulatory covariates; neural signal processing

I. Introduction

Task or context-dependent change in the processing of sensory inputs or modulation of stimulus-evoked responses is an important function of the brain and is essential in forming our sensory perception. For example, acoustic filter properties of primary auditory cortex neurons can dynamically adapt to stimulus statistics, classical conditioning, instrumental learning and the changing auditory attentional focus [1]. In the retina, when the visual scene changes from a low to high contrast, temporal filtering quickly accelerates, sensitivity decreases, and the average response increases in retinal neurons [2]. Rapid eye movements (saccades) influence many aspects of visual processing, including suppression of overall sensitivity [3], as well as spatial [4]–[6], temporal [7], [8] and chromatic [9] perception. Studies in the lateral geniculate nucleus (LGN) and the parietal cortex, including the middle temporal (MT), medial superior temporal (MST) and lateral intraparietal (LIP) areas have revealed suppression of neural activity and thus visual sensitivity before saccades and enhancement afterward [10]–[14]. The fact that the transformation of sensory information to neuronal responses is influenced by these cognitive variables poses a difficulty for understanding the neural code in sensory areas.

By characterizing the influence of different covariates on the stimulus-response relationship, statistical model-based approaches provide a powerful means to identify the effect of cognitive variables on the neural code of sensory processing. The point process generalized linear model (GLM) has been widely used for describing the encoding of stimuli in neuronal spike trains as a function of several extrinsic (e.g., sensory stimuli [15], motor variables [16] or behavior [17]) or intrinsic covariates (e.g., spike refractoriness or burstiness [18], or network states [19], [20]). Although quite powerful in encoding and decoding neuronal responses, the GLM framework faces challenges pertaining to modulatory stimulus processing. The classical GLM cannot accommodate changes in the spatiotemporal properties of the underlying system over time, i.e., time-variant or nonstationary systems. Thus, the classical GLM structure fails to capture time-varying characteristics of neural systems, for example due to eye movements or adapting stimuli, which may dynamically control the spatiotemporal integration of sensory inputs to the system [21]–[23]. Indeed, this scenario can no longer benefit from the well-behaved likelihood-based estimation, which is guaranteed by the structure of the classical GLM. Therefore, extending the GLM framework to account for dynamic and nonlinear modulatory effects on multiple stimulus components is

essential to study task- or context-induced changes in the neural representation of the sensory environment.

To allow for point process models to capture the temporal nonstationarity of the system up to some temporal and spatial resolution, current approaches use adaptive filter algorithms [24]–[27]. Existing solutions mostly extend methods based on the Least Mean Squares and Recursive Least Squares algorithms for dynamic parameter estimation to track changes in neural receptive fields (RFs) over time. Although successful in analyzing receptive field dynamics, methods based on adaptive filters should trade off accuracy and robustness with spatial and temporal resolutions of changes in receptive field structure and spike response modulations for parameter estimation purposes. Moreover, existing adaptive filtering solutions do not provide an explicit model of how modulatory covariates may interact with other covariates to produce the neurons response in nonstationary settings.

To address this issue, we develop a general probabilistic framework to extend the GLM approach in order to account for time-varying information about a stimulus from single-trial spike trains and incorporate multiple nonlinear subunit inputs. Each input can in principle implement different additive or multiplicative modulations over time providing a powerful framework for analyzing nonstationary neural systems under a broad range of stimuli. We also provide a direct and efficient method to estimate the model parameters from spiking data and validate it by fitting the model to neuronal responses. We both employ the classical point process goodness-of-fit measures and also extend new measures based on dynamic stimulus-related correlation to assess the performance of our nonstationary GLM (NSGLM) in accurately predicting novel data from a nonstationary system and to compare it with the existing solutions for the nonstationary settings. The fitted NSGLM, with biophysically interpretable components, is shown to be capable of accurately describing the encoding mechanism of neural responses in visual cortex during a visually guided saccade task. Specifically, using the NSGLM combined with a high spatiotemporal resolution experimental design we were able to account for the detailed modulation of neuronal responses within the MT cortex when influenced by saccadic eye movements. The pseudorandom noise patterns used for visual stimulation enabled an unbiased estimation of the model parameters and nonlinearities. Taken together, our new data-driven model framework provides a general platform to track the modulation of sensory representations by both extrinsic and intrinsic variables during brain behavioral and cognitive functions, both on a finer time scale and also using much richer nonlinear computations than was possible with previous studies.

II. Statistical Framework

The neural spike trains are commonly modeled as point processes. A point process can be defined as a sequence of discrete events taking place in continuous time. In a neural spike train over a time interval $(0, T]$, a sequence of spiking events occurring at times $0 < e_1 < e_2 < \ldots < e_M < T$ forms a point process [28]. The counting function associated with this point process, $N(t)$, can be defined as the number of times the neuron fired a spike between time 0 and time t, where $t \in (0, T]$. A point process can be fully characterized by its conditional intensity function (CIF), $\lambda(t|H(t))$, defined as:

$$\lambda\left(t|H\left(t\right)\right)=\lim_{\Delta\to 0}\frac{p\left\{N\left(t+\Delta\right)-N\left(t\right)=1|H\left(t\right)\right\}}{\Delta} \quad (1)$$

where $H(t)$ includes spiking history of events up to time t and also other related covariates. For the sufficiently small time bin Δ such that at most one spike falls in each time bin, it follows that the probability of a spike in the time interval $(t, t+\Delta)$ can be approximated as:

$$p\left(\text{spike in }\left(t,t+\Delta\right)|H\left(t\right)\right)\approx\lambda\left(t|H\left(t\right)\right)\Delta \quad (2)$$

The CIF specifies the rate function of a conditionally Poisson process given the covariates and spike history, which represents the instantaneous firing rate of the neuron given the history of the process and the covariates. Using the fact that the point process is completely defined by its CIF enables modeling of the neural spike train in terms of a point process by defining its conditional intensity as a function of different covariates [29]. First, by binning the counting process, $N(t)$, over the entire time interval $(0, T]$, we construct a discrete-time representation of the point process, r_t, defined as $r_t \triangleq N(t\Delta) - N((t-1)\Delta)$, where Δ is the time bin size and $t \in \{1, 2, \ldots, \left\lfloor \frac{T}{\Delta} \right\rfloor\}$. Second, by choosing a value of Δ small enough such that at most one spike falls in each time bin, we define the spike train as a binary sequence of zeros and ones, i.e., a Bernoulli process with parameter defined by (2). Finally, the joint probability of the discretized spike train is expressed as a product of probability mass functions of the Bernoulli events as follows:

$$p\left(\left\{N\left(t\Delta\right)\right\}_{t=1}^{\left\lfloor\frac{T}{\Delta}\right\rfloor}\right)=\prod_{t}\left[\lambda_t\Delta\right]^{r_t}\left[1-\lambda_t\Delta\right]^{1-r_t} \quad (3)$$

where $\lambda_t \triangleq \lambda(t\Delta|H_t)$ and $H_t \triangleq H(t\Delta)$. For small Δ, $[1-\lambda_t\Delta] \approx \exp(-\lambda_t\Delta)$ and (3) can be expressed as [29]:

$$p\left(\left\{N\left(t\Delta\right)\right\}_{t=1}^{\left\lfloor\frac{T}{\Delta}\right\rfloor}\right)\approx\exp\left\{\sum_{t}r_t\log\left(\lambda_t\Delta\right)-\sum_{t}\lambda_t\Delta\right\} \quad (4)$$

Note that assuming the CIF is constant over any interval $((t-1)\Delta, t\Delta]$, then by (4), the distribution of each r_t represents the probability mass function of a Poisson random variable. To model the effects of different covariates on generating the spikes, we define the conditional intensity of the spiking point process in discrete time as a parametric function of different covariates.

A widely used computational framework to model the CIF is the so-called GLM framework. The GLM framework has proven successful in relating the neural spiking responses to extrinsic covariates like sensory stimuli as well as intrinsic ones like the spiking history of

the neuron [26], [29]–[31]. The structure of this framework also guarantees a computationally tractable method for estimating the model parameters. In this framework, the instantaneous spiking probability of the neuron as a function of a given input signal $\boldsymbol{s}$ is described as:

$$p\left(\text{spike}|\boldsymbol{s}\right)=f_{\eta}\left(K\boldsymbol{s}\right) \quad (5)$$

where K is a linear operator projecting the high dimensional input vector $\boldsymbol{s}$ onto a lower-dimensional subspace, and f is a nonlinear function representing nonlinear properties of the neuron, parameterized by a set of parameters η, which maps the output of the linear stage to the neuron's instantaneous firing rate [30], [32]. It has been shown that the parameters of this model (η, K) can be estimated efficiently in a maximum likelihood (ML) framework under some benign conditions on the form of the model nonlinearity and a global maximum can be attained using efficient gradient ascent methods [30]. Note that (4) has the same form of the likelihood function as a GLM under a Poisson probability model and a log link function [29].

Although very powerful in relating the spiking responses of neurons to sensory stimuli, the GLM framework is challenged when the relationship between the response and the stimulus changes over time due to non-sensory covariates such as behavior or cognition, i.e., a nonstationary system. In this paper, we extend the classical GLM framework to capture the time-varying stimulus-response relationship and provide an efficient procedure to estimate the parameters of this nonstationary GLM, fit to the real data in a computationally tractable way. The following sections describe how the NSGLM can generalize the current GLM-based approaches enabling us to trace the dynamic spatiotemporal properties of the neural response in the presence of modulatory non-sensory factors robustly and at the sampling resolution of spike trains.

A. NSGLM framework

The NSGLM is an extension of the GLM framework described above (more details can be found in [30]). Similar to the classical GLM, the NSGLM assumes that the neuron's spikes are generated according to a nonhomogeneous Poisson point process with instantaneous firing rate λ_t. The NSGLM is comprised of five stages (Fig. 1): (1) the stimulus kernels, $\boldsymbol{k}_i$, which characterize the temporal sensitivity of the neuron at the i^{th} spatial dimension of the stimulus; (2) the gain kernels, $\boldsymbol{\omega}_i$, corresponding to each $\boldsymbol{k}_i$'s output, which determine how sensitive the response is to each spatiotemporal feature across time; (3) the offset kernel, $\boldsymbol{b}$, which determines the time-varying baseline activity; (4) the post-spike kernel, $\boldsymbol{h}$, which captures response dependencies on the neuron's recent spiking history $\boldsymbol{r}$; (5) the static nonlinearity, $f(\cdot)$, which generates the neuron's instantaneous firing rate. The predicted firing rate λ_t is then given as:

$$\lambda_t=f\left(\sum_i \boldsymbol{\omega}_i\left(t\right)\cdot\left(\boldsymbol{k}_i*\boldsymbol{s}_i\right)\left(t\right)+\left(\boldsymbol{h}*\boldsymbol{r}\right)\left(t\right)+\boldsymbol{b}\left(t\right)\right) \quad (6)$$

where $\boldsymbol{s}_i$ is the time course of the stimulus at the i^{th} spatial dimension, $\boldsymbol{\omega}_i(t)$ is the instantaneous gain factor at time t associated with the i^{th} spatial dimension and $\boldsymbol{b}(t)$ is the instantaneous offset factor at time t. The * and · denote respectively the linear convolution operation and the element-wise multiplication over time. The nonlinearity function, $f(\cdot)$, has been chosen to be a fixed exponential function satisfying the conditions for efficient optimization [30]. In our case, a comparison of the exponential nonlinear function with the empirical mapping of the neuron's actual firing rate to the filtered stimulus also shows the adequacy of this choice of nonlinearity for our spiking data. The exponential nonlinearity not only provides an interpretation of how the outputs of model kernels influence each other like gain factors in generating the neuron's firing rate, but is also useful for a computationally efficient estimation of the model's parameters. However, our model is not limited to this choice of nonlinearity and any nonlinear function that satisfies the GLM optimization requirement can be used for the NSGLM as well. Also note that (6) reduces to a classic GLM when the $\boldsymbol{\omega}_i$s are equal to 1 and $\boldsymbol{b}$ is a constant over time.

B. Model Estimation

A maximum likelihood estimation based method has been used to estimate the model's parameters. The probability of a spike train under the model is given by a Poisson process (similar to (4)) as follows:

$$p\left(\boldsymbol{r}|\boldsymbol{s}\right)=\prod_{t=1}^{T}p\left(r_t|\boldsymbol{s}\right)\propto\prod_{t=1}^{T}\left(\Delta\lambda_t\right)^{r_t}e^{-\Delta\lambda_t} \tag{7}$$

where $\boldsymbol{s}$ is the sequence of stimuli driving the spike train, $\boldsymbol{r}=\{r_t\}_{t=1}^{T}$, Δ is the size of time bin used to compute the spike count at time bin t, r_t, and T is the number of time bins in the trial. Therefore, the log-likelihood of the observed spike data given the model parameters is given by the point process log-likelihood [33]:

$$LL\left(\boldsymbol{\theta}\right)=\sum_{t=1}^{T}r_t\log\left(\Delta\lambda_t\right)-\Delta\lambda_t \tag{8}$$

where $\boldsymbol{\theta}=\{\boldsymbol{\theta}_k,\boldsymbol{\theta}_\omega,\boldsymbol{\theta}_b,\boldsymbol{\theta}_h\}$ is the set of parameters used to parameterize the model kernels for stimulus, gain, offset, and post-spike components (ordered as they appear in the set). The time bin size, Δ, is chosen such that at most one spike can occur in each time bin (Δ was chosen for discretizing the spike trains). Each kernel was represented as a weighted sum of basis functions and parameterized by the weight parameters, $\boldsymbol{\theta}$.

To fit the model in (6), we optimized the parameters by maximizing the log-likelihood of observed spike trains given the stimulus according to (8). Note that with fixed gain kernels, $\boldsymbol{\omega}_i$s, over time, the likelihood function, $LL(\boldsymbol{\theta})$, can be optimized in the context of a classical GLM. Moreover, for a given set of stimulus kernels, $\{\boldsymbol{k}_i\}$, the likelihood function will be a linear function of $\boldsymbol{\omega}_i$s and thus can be again optimized in the context of the GLM. However, with the addition of modulatory kernels $\boldsymbol{\omega}_i$s (for capturing the nonstationary effects), the

likelihood function becomes more complex and unlike GLMs cannot be optimized using the regular gradient ascent method. Indeed, the resulting likelihood function no longer has a globally optimal solution.

As a result, we developed an optimization strategy to efficiently estimate the model parameters as follows. As explained above for a given set of $\{\boldsymbol{k}_i\}$, the model reduces to a GLM framework, which has a global maximum. Thus, the important task for optimizing this nonconcave function is to determine the best choice for the initial guess of the stimulus kernels, $\{\boldsymbol{k}_i\}$. We tried different initialization strategies and found the following method to give a stable solution for the data and a well-behaved solution path on the cross-validated data through regularization. First, the values for the gain kernel parameters, $\boldsymbol{\theta}_\omega$, are set to 1, resulting in a uniform gain kernels $\boldsymbol{\omega}_i$s, which thereby reduces the model to a GLM [30]. The reduced model can be optimized for the GLM terms including the stimulus kernel, post-spike kernel, and offset kernel. Optimization of $\boldsymbol{\omega}_i$s then proceeds by maximizing the LL with the choice of GLM fits for other kernels using a numerical gradient ascent, i.e., finding a local maximum by iteratively maximizing the log-likelihood of the model by searching along the ascent direction in the parameter space. We found that $\boldsymbol{\omega}_i$ and $\boldsymbol{k}_i$ specifying each probe location do not need to be optimized jointly. Instead, alternating between the optimization of $\boldsymbol{\omega}_i$ and $\boldsymbol{k}_i$ (in which the output of each stage is used to initialize the next stage to update the other set of parameters) turns out to be more computationally efficient, producing a stable solution with regard to different initializations of the GLM terms $\{\boldsymbol{\theta}_k, \boldsymbol{\theta}_b, \boldsymbol{\theta}_h\}$ (similar to [34]), and benefits the convergence time of the fitting procedure. Specifically, we fit the model using coordinate descent [35], alternating between fitting the gain kernels and other kernels as a whole, and optimizing over each set of parameters with gradient ascent.

Using smooth basis functions to parametrize the kernels results in smooth fitted kernels, and so the overfitting problem due to possible sharp fluctuations in the kernels is not an issue here. To verify this, we ensured that the solution path was well behaved and the model was not overfitted to the training data through regularization, using a cross-validated ridge prior for the gain kernel and offset kernel and a cross-validated L1 regularization for the stimulus kernel. Specifically, we add general smoothness and sparseness penalty terms of the following form to the likelihood function $LL(\boldsymbol{\theta})$ in (8):

$$\mu_b \|\nabla \boldsymbol{b}\|_2 + \sum_i (\mu_\omega \|\nabla \boldsymbol{\omega}_i\|_2 + \mu_k \|\boldsymbol{\theta}_{\boldsymbol{k}i}\|_1) \quad (9)$$

where $\boldsymbol{\theta}_{\boldsymbol{k}_i}$ is the set of parameters used for estimating the i^{th} stimulus kernel, $\boldsymbol{\omega}_i$ and $\boldsymbol{b}$ represent respectively the i^{th} gain kernel and the offset kernel, and μ_ω, μ_b, μ_k are hyperparameters which determine the strength of gain kernel and offset kernel smoothness over time, and sparseness regularization of the stimulus kernel, respectively. We estimated the hyperparameters by maximizing the likelihood using a separate cross-validation dataset; however, our results were not very sensitive to the selection of hyperparameters. Evaluating the performance of the model on the cross-validated data (which was withheld from the training set) showed similar accuracy and prediction power as for the unregularized model;

accuracy and prediction power on cross-validated data were also comparable with those for the training data, demonstrating the generalization power of the model and that the parameters have not been overfitted to the training data. Therefore, the reported results here and the corresponding Fig. 4, 5, 6 and 7 are based on the model's performance on the test data (not used for fitting) for a model with no regularizing term.

C. Model Evaluation

Since our model works on spiking data represented by a binary point process at the resolution of single trials for individual neurons, regular goodness-of-fit methods, such as mean squared error, used for continuous-valued processes are insufficient to provide a quantitative measure for the performance of our point process model to its full capacity. Several solutions have been proposed to measure the goodness of fit of a point process framework, which quantify how much congruency exists between the observed response and the prediction of the model [29], [36]. In order to examine how well the model reproduces and predicts the neuron's spiking activity we employed and extended different goodness-of-fit measures to quantify different aspects of our model's precision: (1) using a Kolmogorov-Smirnov test, we confirmed the congruency between the neuron's response and the model prediction at a fine timescale in terms of the interspike intervals statistics; (2) we verified that the model sufficiently describes most of structure in the data by showing that the point process residual between the model prediction and the neuron's response does not contain information about the external variables; and lastly (3) we verified the trial-by-trial congruency between the model and the neuronal response in terms of their correlation pattern over time using a joint peristimulus time histogram (JPSTH) method.

1) Kolmogorov-Smirnov (K-S) Goodness-of-fit Analysis—To assess how accurately the model can predict the spike train data, we used a similarity measure designed for point process data based on the time-rescaling theorem [36]. This method has been used to assess goodness of fit in several studies [29], [37] to test model goodness of fit for spike train data.

Here we briefly describe this method. Considering a sequence of spiking events represented as a point process occurring at times $e_1, e_2, \ldots e_M$, let $\lambda(t|H(t), \boldsymbol{\theta})$ be the conditional intensity function of the spiking process estimated using a model fit to the spiking data and parametrized by $\boldsymbol{\theta}$. Using the estimated CIF, rescaled times z_k can be computed as follows,

$$z_k = 1 - e^{-\tau_k}, k = \{1, 2, \ldots, M-1\} \quad (10)$$

where

$$\tau_k = \int_{e_k}^{e_{k+1}} \lambda\left(t|\boldsymbol{\theta}, H\left(t\right)\right) dt \quad (11)$$

The z_k values obtained by this transformation will form an independent uniform distribution over the unit interval if and only if the estimated CIF corresponds to the true CIF underlying the spiking process [36]. The so-called K-S plots are used to measure this agreement by plotting the ordered quantiles of z_k values versus the values of the cumulative distribution

function of the uniform density function [29]. To evaluate how well the model CIF approximates the true CIF, the points' deviation from an expected 45° line is measured and the K-S statistic is used to construct the corresponding confidence intervals [38].

To assess how well the model performs in terms of the original interspike intervals (ISIs), we computed a mean ratio (R) quantity for each bin of ISI values, which is defined as the mean of all the ratios of the empirical density of the z_k values corresponding to that ISI to the expected uniform density in the related bin [29]. A mean ratio $R = 1$ implies that the model estimated CIF perfectly estimates the distribution of the original ISIs. To also quantify the divergence of the model predicted distribution of the rescaled ISIs from the expected 'independent' distribution, we assessed the temporal correlations measured by the autocorrelation function of z_ks. The correlation values should be zero for independent z_ks. For visualization purposes z_ks are plotted versus z_{k+1}s, which demonstrates the second-order correlation values.

2) Point Process Residual Analysis—To evaluate the full account of the model in explaining the relationship between task and behavioral variables and spiking activity, it is also important to examine structure in the data that is not described by the model. For this purpose, we measured the correlations among the point process residuals (the difference between actual and predicted responses) and the visual stimulus.

The standard approach of residual analysis is used to analyze the structure in the data not described by the model. The point process residual [29], [39] over non-overlapping moving time windows was used to evaluate the difference between the actual and model predicted data for spike trains as follows:

$$res\,(B_k) = N_k - \int_{B_k} \lambda\,(t|\boldsymbol{\theta}, H\,(t))dt \quad (12)$$

where $res(B_k)$ denotes the value of the residual signal in time bin $B_k = ((k-1)\Delta, k\Delta]$, where Δ is the bin size, N_k denotes the spike count in the time bin B_k and $\lambda(t|\boldsymbol{\theta}, H(t))$ is the model estimated CIF given the model parameters $\boldsymbol{\theta}$ and the spike history and the covariates $H(t)$. We then computed the cross-correlation function between the residual signal and the corresponding stimulus covariates to quantify their relationship. If the correlations were nonzero, there would be some structure in the data not explained by the model and so left in the residual. These correlations were also compared to the cross-correlation between the neuron's response and the stimulus covariates to measure their significance.

3) JPSTH Analysis—The JPSTH analysis is a method for visualizing the relationship between the spike trains of two simultaneously recorded neurons, and can reveal the dynamic correlation between the neurons' responses [40], [41]. As the name implies, JPSTH provides the two-dimensional peristimulus time histogram (PSTH) of two neurons with respect to an event where the histogram is computed over different time shifts of action potentials of the two neurons. We extended this concept to assess the similarity of the model's prediction and the neuron's response by computing a JPSTH between actual spike trains and the simulated spike trains generated according to the model's predicted firing rate

as locked to the presentation of each stimulus probe. A JPSTH is computed by constructing a matrix of spike incidences where each entry of the matrix is incremented by one for the cases when the neuron fired a spike and the model predicted a spike at the corresponding time incidences of that matrix entry, relative to probe onset. If the model is able to predict the timing of actual spikes, this will be reflected in high values along the diagonal of the JPSTH. In the ideal JPSTH, all other entries will be zero, showing that the model's predicted spike times successfully coincide with the neuron's actual spikes.

The JPSTH enabled us to infer the dynamics of the actual and predicted spike response's covariation over time, which is not possible using the ordinary cross correlation based on the PSTH or relative time-shifted versions of the two responses. While the conventional cross correlation measures average covariation over the entire length of data, the JPSTH measures dynamic covariation associated with repeated presentation of the stimulus. Thus, the JPSTH gives us a more detailed picture of the correlation structure than the conventional PSTH- or shift-based correlation methods [41].

III. Results

We sought to develop a computational framework that could accurately predict the responses of neurons and also capture the dynamic changes in their spatiotemporal characteristics in a single model in important scenarios which a classical GLM could not describe. We applied our NSGLM framework to the responses of 40 MT neurons recorded in two macaque monkeys during a visually guided saccade task. Neurons in MT area are visually selective to spatiotemporal stimulus features such as motion direction. This area also contains both motor and visual information and thus is a good candidate for investigating the impact of eye movements on processing the visual signals [42]. Previous studies have also shown evidence of saccadic suppression in MT neurons [13].

In this section, we demonstrate the prediction power of the NSGLM by applying the model to the responses of neurons in the MT cortex despite the fact that a change in the eye position alters the spatiotemporal sensitivity of the neuron during the time course of a trial. This allows us to quantitatively characterize perisaccadic changes in MT visual processing, which is also applicable to other visually selective brain areas that are involved in processing eye movements [43]–[45]. Moreover, this general computational framework is applicable to other brain areas undergoing nonstationary changes under arbitrary task or behavioral conditions. Using different measures of goodness of fit on cross-validated data we quantitatively assess how precisely the NSGLM can describe the temporal modulation of the spike response, spike timing statistics, and the statistical structure in the spike data. We also quantitatively compare the performance of the NSGLM in describing the real spiking response of a sample neuron during an eye movement task to other, widely used or state-of-the-art methods for point process filter estimation.

A. Application to real data: The NSGLM can precisely describe the neuron's spiking response during a behavioral task

1) Visually guided saccade task—Two adult male rhesus monkeys (Macaca mulatta) were used in this study. All experimental procedures were in accordance with the National

Institutes of Health Guide for the Care and Use of Laboratory Animals, the Society for Neuroscience Guidelines and Policies. The protocols for all experimental, surgical, and behavioral procedures were approved by the Montana State University Institutional Animal Care and Use Committee. In our experiment the monkey performed a visually guided saccade task while the activity of MT neurons was recorded (Fig. 2A, B). In this task, a fixation point appeared at the center of the screen. After the monkey fixated on the initial fixation point, a target point appeared 10 degrees away horizontally. While the monkey kept its eyes on the fixation point, probes flashed on the screen in a 9 by 9 grid of possible locations (Fixation 1). The grid was positioned such that it covered the estimated pre-saccadic and post-saccadic receptive fields of the neuron as well as the initial fixation point and the saccade target point. At each time point, there was just one probe on the screen, and its position changed every 7 ms (at the frame rate of the monitor). The precise timing of the probes was verified using a photodiode. Each probe was a white square (100% contrast), 0.5 by 0.5 degrees of visual angle (dva), against a black background. The grid of possible probe locations was scaled in each recording session based on the estimated RF positions; spacing between adjacent grid locations ranged from 1.5 dva – 2.5 dva. The locations of consecutive probes followed a pseudorandom order. A condition was defined by the complete sequence of probes presented throughout the length of the trial (81 conditions were presented). Conditions were designed so that each probe appeared at every time point. After a randomized interval of approximately 600 ms – 750 ms, the initial fixation point disappeared as a cue for the monkey to make a saccade to the target point. After the saccade, the monkey had to hold fixation on the target point for 600 ms (Fixation 2). Probes continued flashing during the saccade and during Fixation 2. Each trial lasted a total of 2100 ms – 2300 ms; at the end of the trial the monkey was rewarded with a drop of juice. Fig. 2C shows, from top to bottom, the presented probe sequences, the eye position, and the recorded spikes in a sample trial.

2) Fitting the NSGLM to experimentally recorded neural data—To fit the model, kernels were represented as a weighted sum of basis functions and parameterized by the weight parameters. The basis functions were chosen as smooth temporal functions in the form of shifted raised cosines separated by $\pi/2$ radians spanning the trial time. The specified length and total duration of basis functions were as follows: 7-ms raised cosine covering a 100-ms window after probe onset for stimulus kernels, and 100-ms raised cosine covering a 1100-ms window around the time of saccade for gain and offset kernels. The length of basis functions for the stimulus kernel was in accordance with the monitor's 144 Hz frame rate. To parameterize the post-spike kernel, a set of both high and low temporal resolution basis functions were used including ten 1-ms uniform functions to capture rapid changes immediately following spike generation (e.g. refractory effects), and ten 7-ms raised cosine functions to capture longer-timescale dependencies on the response history (similar to [20]).

The input to the model includes the spatiotemporal patterns of the flashing probes on the screen. At each instant of time the flashing probes can take one of the 81 probe locations (on the 9 by 9 grid of possible locations). $\boldsymbol{s}_1$, $\boldsymbol{s}_2$, …, $\boldsymbol{s}_{81}$ inputs in Fig. 3 denote the pattern of each probe appearance on the screen in a sample trial. Appearance of the probes in different locations evoked different patterns of neural responses. During fixation 1, probes in the

neuron's initial RF (RF1) evoke a higher firing rate, while during fixation 2, after the eye has shifted to the target point, probes in the new location of the neuron's RF (RF2) evoke a stronger response. A classical GLM is insufficient to describe the changes in the neural response due to the dynamic spatiotemporal properties of the neuron caused by the shift of gaze. However, using the NSGLM, the fitted stimulus kernels and gain kernels capture the neuron's dynamic spatiotemporal sensitivity across eye movements. Sample fits are shown in Fig. 3.

In our model, the visual stimulus is passed through a bank of temporal stimulus linear kernels corresponding to each probe location, which together construct the neuron's spatiotemporal RF. Specifically, the visual stimulus is convolved with the stimulus kernels and the filtered stimulus is then combined multiplicatively with a bank of gain kernels associated with each probe location, which modulates the neuron's sensitivity to probes at each location and at each time point relative to the time of the saccade. The set of gain kernel outputs captures the neuron's spatial sensitivity changes over the course of a trial due to the change in eye position. The set of saccade-modulated outputs are then summed across spatial positions to obtain a 'generator potential' (e.g., time-varying membrane potential as a result of the sum of synaptic currents) in response to the visual stimulation during an eye movement. Another signal that contributes to the generator potential fluctuations is the neuron's baseline firing rate, which is modulated by saccade events and combined with the outputs of the linear kernels via an offset kernel. Finally, the spike aftercurrent is incorporated in the form of the spike train history filtered by a post-spike kernel. For each neuron, the summed outputs of all the kernels are passed through an exponential nonlinearity to obtain an instantaneous firing rate underlying the predicted spike train (Fig. 3).

Here, we verify the model's performance in predicting responses to test stimuli, held out from the data used for training, in terms of different aspects of response characteristics by evaluating the model's ability to capture (1) the interspike interval statistics using the K-S test (Fig. 4), (2) the statistical structure in spiking activity using the residual analysis (Fig. 5), and (3) the temporal precision of the response using the JPSTH analysis (Fig. 6).

It should be noted that all these model performance measures were evaluated on the test data, which was held out from the data used to train the model. This is essential because spiking responses contain both a signal component driven by the task variables and a noise component reflecting the inherent variability of the neuronal responses or the effects of unmodeled covariates (e.g., the state of the brain). Therefore, a good performance on the test data demonstrates that the model was fit to the signal and not the noise. These measures are described below.

B. Model's ability to predict temporal patterns of spiking activity using the K-S goodness-of-fit analysis

We used a K-S test to quantitatively assess how the model describes the spike train data at the level of single trials [29]. Fig. 4A shows the K-S plots for two representative sample neurons showing different degrees of agreement between the model's predicted rescaled ISIs distribution (denoted by quantiles along the x-axis) and the expected 'uniform' distribution under the true spiking process (denoted by CDF (cumulative distribution function) along the

y-axis). For the example neuron on the left, the points lie within the 95% confidence interval of the 45° line, indicating that the model-predicted CIF corresponds to the true CIF underlying the spike process; however, for the example neuron on the right, the model tends to overestimate the true CIF, reflected in the points lying below the 95% confidence error bounds.

To assess how well the model performs in terms of the original ISIs, we computed the mean ratio (R) values as the mean of the ratios of the empirical probability density of the time-rescaled ISIs to the expected uniform density over bins of the ISI values. Fig. 4B shows the R values across the values of ISI for the same two neurons in panel (A). Under the model, for the neuron on the left, the spike rate is underestimated ($R > 1$) for lower ISI values and the estimation improves (R reaches to 1) for larger ISI values, indicating the agreement between the model estimated rate underlying the spike train and the true spiking process. However, for the neuron on the right, the spike rate is overestimated ($R < 1$) under the model for all ISI values. For the population of 40 neurons, the average of mean ratios over ISI values was 0.94 ± 0.008 indicating that overall the model was successful in reproducing the ISI statistics of the spike train data (Fig. 4C). To check for the independence of rescaled times under the model, we measured the temporal correlation between every consecutive time-rescaled ISIs. Fig. 4D shows scatter plots for consecutive ISI values for the two sample neurons above. The correlation between consecutive ISIs is 0.008 ± 0.001 for the population, indicating that overall the model was successful in capturing the temporal structure of the spike train data (Fig. 4E). The result for correlations at different lags (i.e., higher-order correlations) was consistent with the result for lag 1 (i.e., second-order correlation).

C. Model's ability to capture statistical information in spiking activity about stimulus using the point process residual analysis

Fig. 5A shows the temporal correlation functions between the probe stimulus inside the neuron's RF and the residuals, compared to the correlation of the same stimulus and the spike data for a typical MT neuron. The response window is defined as a window centered at the time of the maximum stimulus-response correlation value with the width defined as the time lags whose associated correlations are higher than that of the shuffled data. As shown in the Fig. 5A, in the response window of the neuron, correlation between the stimulus and the response rises, showing that for this neuron the stimulus-response correlation conveys information about the probe stimulus inside the RF. Most of the correlation is captured by the model predicted response, and the remaining stimulus-residual correlation is only a small fraction of the original correlation.

Fig. 5B shows the correlation between the response and the stimulus, and also the correlation between the residual and the stimulus, for the population of 40 MT neurons over the time lags used for computing the correlation. The correlation between stimuli and the residuals is very small and significantly less than the correlation between stimuli and the responses (mean stimulus-response correlation = 0.078 ± 0.004, $p < 0.001$; mean stimulus-residual correlation = 0.003±0.003, $p = 0.18$; mean difference between response and residual = 0.074 ± 0.003, $p < 0.001$). These results indicate that the model's predicted response captures most of the correlation between the stimulus and the neural response.

D. Model's ability to capture temporal precision of spiking response using the JPSTH analysis

We computed the relative timing of actual and predicted spikes by constructing JPSTH matrices based on the response to probes in the RF1. Using the JPSTH measure, we see that the NSGLM captured the fine temporal features of the neural response as shown in the scatter diagram for stimuli presented inside the RF for a sample MT neuron in Fig. 6A (left panel). The high density cloud parallel to the principal diagonal represents the coincidence of actual and predicted spikes at the visual latency and over the response window of the neuron, indicating an agreement between the actual and predicted response latencies. In Fig. 6A (right panel) the histograms along the abscissa and ordinate axes approximate the ordinary PSTH of the actual and predicted spikes, respectively. The peristimulus time (PST) coincidence histogram along the principal diagonal represents the probability of coincidences in the actual and predicted trains of spikes. As expected, the probability of the coincidences rises rapidly around the time of neuron's response latency and decays slowly subsequently. The histogram on the upper right in Fig. 6A (right panel) represents the ordinary cross-correlogram of the actual and predicted spike trains, which is obtained by summing along the para-diagonal bins of the JPSTH matrix and shows a central peak, consistent with the coincident spikes in the actual and predicted trains. The time precision of the model exists across the population of recorded neurons, as visualized using the average normalized JPSTH matrices over all the neurons in Fig. 6B (n=40). Fig. 6C shows the values of the cross-correlation coefficient between the actual and predicted responses (corresponding to the central peak of the cross-correlogram) for 40 neurons.

E. Comparison with other studies: NSGLM outperforms existing approaches

There exist several approaches for incorporating time-varying properties and the effects of modulatory covariates when modeling nonstationary neural systems. A major difference between the existing ideas, however, is the temporal and spatial resolution at which each model can operate robustly. There have been extensions to the GLM framework by incorporating time-varying modulatory signals to describe nonstationarity in the data. However, the capabilities of these extensions have been limited in several respects. Some are limited to low temporal resolution, where the firing rate is modulated by a multiplicative interaction with a positive, smooth, and slowly varying signal as state variables [46]. Moreover, this form of modulation structure will limit the possible underlying mechanisms driving response nonstationarities. Some other extensions to the GLM framework allow higher temporal resolution in capturing nonstationarity by introducing time-dependent multiplicative and additive factors modulating the stimulus drive, however, their effects act so globally that they may distort the sensitivity of the response to stimulus [47]. Non-GLM-based approaches such as nonstationary dynamics models have been proposed to capture slow modulations in firing rates across trials [48]. Lastly, adaptive filtering approaches, although successful in capturing receptive field plasticity [24]–[27], are challenged when trying to track fast changes in the system. The higher the learning rate of these algorithms are, the fewer observations will be available to estimate the filter parameters. This will affect the precision of these algorithms in tracking smaller changes and also their robustness when very few spikes are available during short periods of observation. No existing models

capture nonstationary dynamics with the combination of high spatial and high temporal resolution of the NSGLM described here.

In this section, we analyze the performance of the NSGLM versus a few of the existing methods representing the widely-used or state-of-the-art approaches that have proven successful for modeling point processes or estimating the system's spatiotemporal characteristics. These methods include the classical GLM for static filter estimation [30], the steepest descent point process filter (SDPPF) for adaptive filter estimation [27], and a recent GLM-based method by Zanos et al. with modulatory factors to account for saccadic suppression effects [47]. The classical GLM provides an optimal estimation in ML sense of the fixed spatiotemporal filters describing the neuron's spiking response over the entire trial. The SDPPF-based model provides an optimal estimation of the adaptive filters representing the spatiotemporal sensitivity of the neuron obtained by a steepest descent procedure. The Zanos model employs a two-step procedure for first estimating the neuron's static spatiotemporal RFs and second estimating additional time-dependent gain and offset terms given the RF estimates from the first step to account for saccadic suppression effects.

Fig. 7 shows the comparison between the performance of these methods for a sample MT neuron over test data (not used for training the models). First these models were evaluated in terms of how well a model can reproduce the measured firing rate of the neuron as obtained by averaging over spike trains from several repeated trials. Fig. 7A shows the average stimulus-evoked response on the trials where a probe has been presented inside the RF1 of the neuron before and after the eye movement, respectively, along with the model-predicted firing rate response for GLM, NSGLM, Zanos, and SDPPF models. While all the models can predict the firing rate during fixation 1, the classical GLM and the SDPPF-based model fail to predict the different response to the same stimulus during fixation 2 due to the eye movement. The classical GLM with static components is not capable of capturing nonstationarity in the system, e.g., the neuron's changing spatiotemporal RF with respect to the saccade time. The adaptive filtering approach, although providing a dynamic estimation of the neuron's spatiotemporal receptive fields, shows insufficient to capture the very fast dynamics of the system induced by the saccade due to its limited robustness in filter estimation when the observation window reflecting those fast dynamics is too short. The Zanos model successfully follows the overall firing rate response, both before and after the saccade due to the presence of a modulatory gain component, enabling to capture the global effects of eye movements on the neuron's RF.

To further analyze the capabilities of the models with an intrinsic ability to account for nonstationarity of the system (i.e., NSGLM, Zanos, and SDPPF models) we compare their performance on the timescale of spiking events at the level of single spike trains. The NSGLM outperforms the two other nonstationary models in describing both the statistical structure in the spike data and spike timing statistics as shown by the temporal correlation function between stimulus and residual and K-S plots, respectively (Fig. 7B, C). The adaptive filter approach falls short in accurately tracking the response temporal statistics on the timescale of the saccade execution and saccade-induced modulations for filter estimation. This is due to the adaptive filter's limited precision when its observation window is too short to capture the fast response modulation or robustly estimate the neuron's

changing spatiotemporal sensitivity across the saccade. Although the results here are shown for the SDPPF algorithm, other adaptive filtering algorithms suffer from similar shortcoming in terms of limited robustness and accuracy in capturing and estimating fast changes. The Zanos model with single temporal modulatory component for all locations makes it insufficient to track the change in the neuron's sensitivity across space with respect to the instantaneous position of the eye, which has been the case for our experiment. Moreover, the Zanos approach does not enable an optimization framework for simultaneous estimation of the model parameters and solves for the time-dependent components assuming a fixed RF envelope for the course of the trial.

Thus, by enabling dynamic estimation of the neuron's spatiotemporal sensitivity with a high temporal and spatial resolution, the NSGLM can capture the neural dynamics on the scale of saccade-induced spatiotemporal modulations and single spike trains where the existing approaches prove insufficient.

IV. Discussion

Our goal was to develop a model that includes time-varying covariates that modulate the relationship between the stimulus and response. We extended the widely used GLM framework to incorporate these nonstationary components. GLMs have been successfully used to characterize the stimulus sensitivity of neurons in early sensory areas (e.g. the retina [20], thalamus [49], primary auditory cortex [50], primary visual cortex [51], primary somatosensory cortex [50], or primary motor cortex [29]). However, classical GLMs are unable to accommodate modulatory factors that change over time, altering the nature of the relationship between the stimulus and neural activity. Previous attempts to incorporate modulatory factors into the GLM framework have not taken into account how these factors interact with other covariates to alter the stimulus-response relationship. The NSGLM introduced in this paper incorporates these interactions, extending the GLM framework to include time-varying modulatory components. One aspect of the GLM framework that makes it popular is its computational tractability in terms of efficiently finding the unique optimal solution. Introducing the modulatory components, however, makes a simultaneous maximization of the parameters to no longer be computationally tractable. Instead, by alternating between separate sets of parameters, we maintained the concavity of the likelihood function over each parameter set, enabling the algorithm to converge to an optimal set of parameters. Combining this alternating optimization method with an efficient initialization procedure proved successful, and the fitted model precisely captured the time course of the stimulus-response relationship in the presence of time-varying non-sensory covariates.

This model can be used to describe neural responses in different sensory modalities and cognitive tasks in which the non-sensory variables could change the stimulus-response relationship. In our case, we validated the model by fitting it to the spiking responses of MT neurons recorded during a visually guided saccade task. The fitted model successfully describes the neural response while the spatiotemporal sensitivity of neurons changes over time due to the movement of the eye. The model (1) predicts the instantaneous firing rates for individual trials, evaluated by the distribution of interspike intervals and a KS test; (2)

captures the majority of the correlation between the stimulus and the neural response, based on the residual analysis; and (3) matches the temporal pattern of the neural response, measured using the joint distribution of the predicted and actual responses.

Whereas the model itself provides a phenomenological description of the neuronal responses, its components can also provide plausible interpretations in terms of the underlying biophysical mechanisms. The stimulus kernel represents the spatiotemporal features that the afferent neurons are most sensitive to. The gain kernel reflects a time-dependent multiplicative control signal along each stimulus dimension. Incorporating this gain kernel provides a means to capture the changes in the neuron's sensitivity induced by a time-varying factor (in this case, an eye movement). The offset kernel captures any additive effects induced by stimulus-independent global changes of response sensitivity across time. The post-spike kernel reflects temporal changes in neuronal excitability due to spike history dependent effects such as refractoriness, burstiness, or adaptation.

The NSGLM offers several advantages over existing approaches, some of which we discuss here. By introducing time-dependent modulatory components, the NSGLM is able to capture the nonlinear dependencies of neural responses on multiple covariates, as well as the dynamic relationship between the response and the stimulus. In the NSGLM, we do not impose any assumptions about the stimulus statistics or the filter arrangement. The parameterization of the model using smooth basis functions allows kernels to take arbitrary forms, improving the ability of the model to generalize across a broad range of response characteristics. By defining explicit interactions between the modulatory components and stimulus-driven signals, the NSGLM provides a means to describe the functional role of each modulatory component in the stimulus processing; whereas models employing adaptive filtering to describe the nonstationary nature of the data do not provide such a functional understanding of these interactions. Moreover, the NSGLM can be easily modified or expanded to suit particular experimental questions. Future extensions of the model could incorporate the population level information, which can be reflected in the correlated activity of neurons or synchrony with local field potential oscillations. In the NSGLM framework, these types of extensions are straightforward and can be implemented by adding corresponding kernels to account for population level information. Moreover, although we used the exponential nonlinearity for Poisson spike generation, other choices of nonlinear functions or non-Poisson distributions could also be substituted into this framework without sacrificing the computational tractability. Lastly, the general probabilistic framework of the NSGLM provides a means to design a statistically optimal decoder, enabling us to readout the visual scene based on the responses.

V. Conclusion

In this article, we introduced a nonstationary GLM framework, which can be used in cases where a time-varying behavioral or cognitive component makes GLM-based models insufficient to describe the dependencies of neural responses on the interactions between internal and external covariates. Moreover, we developed a maximum likelihood based method for the model fitting, which proved computationally tractable for robustly estimating the model parameters. We validated the NSGLM approach by fitting the model to the spike

trains of 40 MT neurons recorded during a visually guided saccade task, where a regular GLM model could not account for the neurons' changing spatiotemporal sensitivity due to the shift in the eye position. Using multiple goodness-of-fit measures, we have shown that the fitted NSGLM model successfully reproduced different aspects of the neural response, including the average firing rate to repeated stimulus presentations, the interspike interval statistics, the correlation between the stimulus and response, and the temporal precision of the response. The fitted model, with biologically-plausible components, provides a descriptive means to understand the functional effects of modulatory signals in sensory processing. Furthermore, this approach will provide a model-based decoder, which will enable us to have a readout of the sensory stimulus during behavioral tasks. This combined encoding and decoding approach, enabled by the general framework of the NSGLM, can provide a powerful tool to study a variety of context- or task-dependent effects on sensory processing.

Acknowledgments

The work was supported by Montana State University and University of Utah startup fund to BN and NN. NN lab is supported by NSF1566621 and BN lab is supported by Whitehall 2014-5-18, NIH R01EY026924, and NSF1439221 and 1632738 grants. This work was also supported by National Institutes of Health (EY014800), and an Unrestricted Grant from Research to Prevent Blindness, Inc., New York, NY, to the Department of Ophthalmology and Visual Sciences, University of Utah.

References

1. Fritz J, et al. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nature neuroscience. 2003; 6(11):1216–1223. [PubMed: 14583754]
2. Baccus SA, Meister M. Fast and slow contrast adaptation in retinal circuitry. Neuron. 2002; 36(5): 909–919. [PubMed: 12467594]
3. Volkmann FC. Human visual suppression. Vision research. 1986; 26(9):1401–1416. [PubMed: 3303665]
4. Cai RH, et al. Perceived geometrical relationships affected by eye-movement signals. Nature. 1997; 386(6625):601. [PubMed: 9121582]
5. Lappe M, et al. Postsaccadic visual references generate presaccadic compression of space. Nature. 2000; 403(6772):892–895. [PubMed: 10706286]
6. Ross J, et al. Compression of visual space before saccades. Nature. 1997; 386(6625):598. [PubMed: 9121581]
7. Burr DC, Morrone C. Temporal impulse response functions for luminance and colour during saccades. Vision research. 1996; 36(14):2069–2078. [PubMed: 8776473]
8. Reppas JB, et al. Saccadic eye movements modulate visual responses in the lateral geniculate nucleus. Neuron. 2002; 35(5):961–974. [PubMed: 12372289]
9. Burr DC, et al. Selective suppression of the magnocellular visual pathway during saccadic eye movements. Nature. 1994; 371(6497):511. [PubMed: 7935763]
10. Zanos TP, et al. A sensorimotor role for traveling waves in primate visual cortex. Neuron. 2015; 85(3):615–627. [PubMed: 25600124]
11. Bremmer F, et al. Neural dynamics of saccadic suppression. Journal of Neuroscience. 2009; 29(40):12 374–12 383.
12. Royal DW, et al. Correlates of motor planning and postsaccadic fixation in the macaque monkey lateral geniculate nucleus. Experimental Brain Research. 2006; 168(1–2):62–75. [PubMed: 16151777]
13. Thiele A, et al. Neural mechanisms of saccadic suppression. Science. 2002; 295(5564):2460–2462. [PubMed: 11923539]

14. Ibbotson M, et al. Enhanced motion sensitivity follows saccadic suppression in the superior temporal sulcus of the macaque cortex. Cerebral Cortex. 2007; 17(5):1129–1138. [PubMed: 16785254]
15. Pillow JW, et al. Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. Journal of Neuroscience. 2005; 25(47):11 003–11 013.
16. Fu Q, et al. Temporal encoding of movement kinematics in the discharge of primate primary motor and premotor neurons. Journal of Neurophysiology. 1995; 73(2):836–854. [PubMed: 7760138]
17. Rorie AE, et al. Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (lip) of the macaque monkey. PloS one. 2010; 5(2):e9308. [PubMed: 20174574]
18. Keat J, et al. Predicting every spike: a model for the responses of visual neurons. Neuron. 2001; 30(3):803–817. [PubMed: 11430813]
19. Maynard E, et al. Neuronal interactions improve cortical population coding of movement direction. Journal of Neuroscience. 1999; 19(18):8083–8093. [PubMed: 10479708]
20. Pillow JW, et al. Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature. 2008; 454(7207):995–999. [PubMed: 18650810]
21. Noudoost B, et al. Top-down control of visual attention. Current opinion in neurobiology. 2010; 20(2):183–190. [PubMed: 20303256]
22. Albright TD, Stoner GR. Contextual influences on visual processing. Annual review of neuroscience. 2002; 25(1):339–379.
23. Noudoost B, et al. Stimulus context alters neural representations of faces in inferotemporal cortex. Journal of neurophysiology. 2016 jn–00 667.
24. Brown EN, et al. An analysis of neural receptive field plasticity by point process adaptive filtering. Proceedings of the National Academy of Sciences. 2001; 98(21):12 261–12 266.
25. Eden UT, et al. Dynamic analysis of neural encoding by point process adaptive filtering. Neural computation. 2004; 16(5):971–998. [PubMed: 15070506]
26. Sheikhattar A, et al. Recursive sparse point process regression with application to spectrotemporal receptive field plasticity analysis. IEEE Transactions on Signal Processing. 2016; 64(8):2026–2039.
27. Frank LM, et al. Contrasting patterns of receptive field plasticity in the hippocampus and the entorhinal cortex: an adaptive filtering approach. Journal of Neuroscience. 2002; 22(9):3817–3830. [PubMed: 11978857]
28. Brown EN, et al. Likelihood methods for neural spike train data analysis. Computational neuroscience: A comprehensive approach. 2003:253–286.
29. Truccolo W, et al. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. Journal of neurophysiology. 2005; 93(2):1074–1089. [PubMed: 15356183]
30. Paninski L. Maximum likelihood estimation of cascade point-process neural encoding models. Network: Computation in Neural Systems. 2004; 15(4):243–262.
31. Simoncelli EP, et al. Characterization of neural responses with stochastic stimuli. The cognitive neurosciences. 2004; 3:327–338.
32. McCullagh P, Nelder JA. Generalized linear models, no. 37 in monograph on statistics and applied probability. 1989
33. Paninski L, et al. Maximum likelihood estimation of a stochastic integrate-and-fire neural encoding model. Neural computation. 2004; 16(12):2533–2561. [PubMed: 15516273]
34. Ahrens MB, et al. Inferring input nonlinearities in neural encoding models. Network: Computation in Neural Systems. 2008; 19(1):35–67.
35. Byrne CL. Alternating minimization and alternating projection algorithms: A tutorial. Sciences New York. 2011:1–41.
36. Brown EN, et al. The time-rescaling theorem and its application to neural spike train data analysis. Neural computation. 2002; 14(2):325–346. [PubMed: 11802915]

37. Haslinger R, et al. Discrete time rescaling theorem: determining goodness of fit for discrete time statistical models of neural spiking. Neural computation. 2010; 22(10):2477–2506. [PubMed: 20608868]
38. Johnson, NL., et al. Distributions in statistics: continuous univariate distributions, vol. 2. NY: Wiley; 1970.
39. Andersen, PK., et al. Statistical models based on counting processes. Springer Science & Business Media; 2012.
40. Gerstein GL, Perkel DH. Simultaneously recorded trains of action potentials: analysis and functional interpretation. Science. 1969; 164(3881):828–830. [PubMed: 5767782]
41. Aertsen A, et al. Dynamics of neuronal firing correlation: modulation of" effective connectivity. Journal of neurophysiology. 1989; 61(5):900–917. [PubMed: 2723733]
42. Bremmer F, et al. Eye position effects in monkey cortex. i. visual and pursuit-related activity in extrastriate areas mt and mst. Journal of neurophysiology. 1997; 77(2):944–961. [PubMed: 9065860]
43. Noudoost B, et al. A distinct contribution of the frontal eye field to the visual representation of saccadic targets. Journal of Neuroscience. 2014; 34(10):3687–3698. [PubMed: 24599467]
44. Han X, et al. Dynamic sensitivity of area v4 neurons during saccade preparation. Proceedings of the National Academy of Sciences. 2009; 106(31):13 046–13 051.
45. Neupane S, et al. Two distinct types of remapping in primate cortical area v4. Nature communications. 2016; 7
46. Rabinowitz NC, et al. A model of sensory neural responses in the presence of unknown modulatory inputs. 2015 arXiv preprint arXiv:1507.01497.
47. Zanos TP, et al. Mechanisms of saccadic suppression in primate cortical area v4. Journal of Neuroscience. 2016; 36(35):9227–9239. [PubMed: 27581462]
48. Park M, et al. Unlocking neural population non-stationarities using hierarchical dynamics models. Advances in Neural Information Processing Systems. 2015:145–153.
49. Butts DA, et al. Temporal precision in the visual pathway through the interplay of excitation and stimulus-driven suppression. Journal of Neuroscience. 2011; 31(31):11 313–11 327.
50. Calabrese A, et al. A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. PloS one. 2011; 6(1):e16104. [PubMed: 21264310]
51. McFarland JM, et al. Inferring nonlinear neuronal computation based on physiologically plausible inputs. PLoS Comput Biol. 2013; 9(7):e1003143. [PubMed: 23874185]

Significance

In addition to being quite powerful in encoding time-varying response modulations, this general framework also enables a readout of the neural code while dissociating the influence of other non-stimulus covariates. This framework will advance our ability to understand sensory processing in higher brain areas when modulated by several behavioral or cognitive variables.

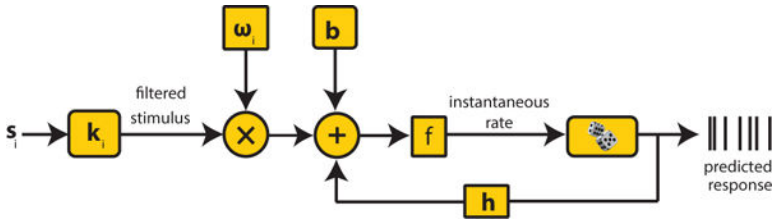

Fig. 1.

NSGLM structure. A schematic of the model structure illustrating the order in which the stimulus is filtered and modulated by the fitted kernels to generate the model's prediction. The model describes the probability of a spike train response given a set of input stimulus variables along time and space dimension. The temporal sequence of the stimulus along each spatial dimension i, $\boldsymbol{s}_i(t)$ passes through the stimulus kernels $\boldsymbol{k}_i(t)$ and is then modulated multiplicatively by temporal gain kernels for each spatial dimension i, $\boldsymbol{\omega}_i(t)$. The output of all kernels are summed with a temporal offset kernel $\boldsymbol{b}(t)$ and also a feedback signal, i.e. the neural response filtered by a post-spike kernel. The summed output is then passes through the nonlinearity function f to generate the instantaneous spike rate prediction.

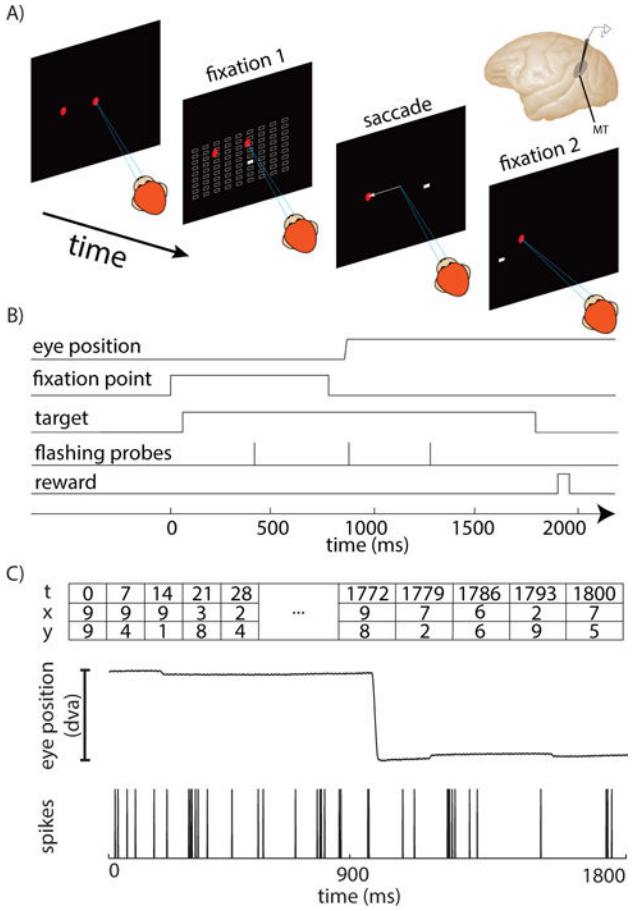

Fig. 2.

Illustration of the experimental paradigm. (**A**) Illustrates visually guided saccade task. Inset shows recording area in the brain, MT area. The pseudorandom white noise stimulus in space and time and the intrinsic timing variability of the saccade are important in order to model their effects on generating neural responses independently. (**B**) The sequence of different events in the visually guided saccade task. (**C**) The sequence of presented probes, actual eye position and the neural response in a sample trial (from top to bottom). '*t*' in the table shows the time of each probe presentation while '*x*' and '*y*' show the index of each probe along x and y coordinates on the 9 by 9 grid of possible probe locations.

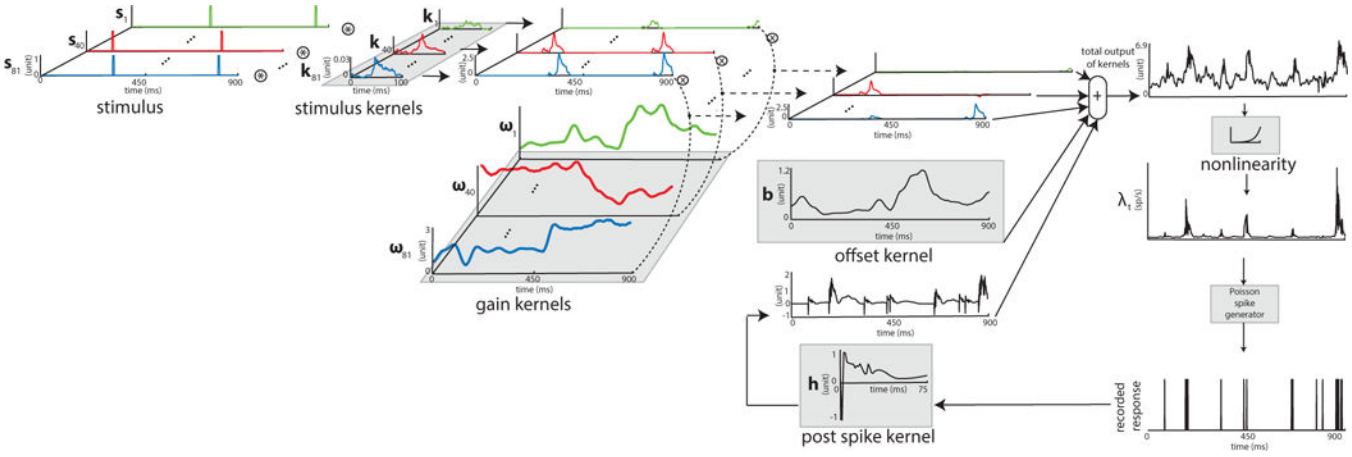

Fig. 3.

Visualization of the fitted kernels of the NSGLM for a sample MT neuron. Illustrates the fitted model components: stimulus kernels, gain kernels, offset kernel, post-spike kernel, and exponential nonlinearity. The input stimulus is filtered through the stimulus kernels and then scaled multiplicatively by the temporal gain kernels. The outputs of all kernels are summed with an offset kernel as well as a feedback signal via the post-spike kernel, and passed through an exponential nonlinear function to produce the instantaneous spike rate. The spiking response is generated according to a conditionally Poisson process under the model predicted time-varying spike rate. Example of the model-predicted and actual response for a single trial from test data has been illustrated. Black vertical bars mark actual spike times.

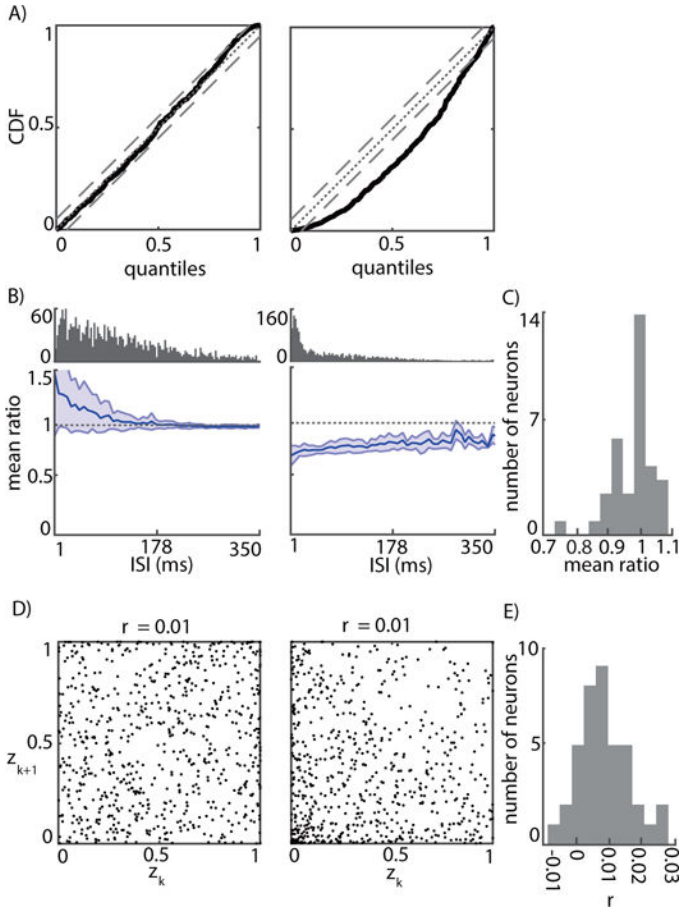

Fig. 4.

The K-S analysis for model goodness of fit. (**A**) K-S goodness-of-fit plots for two example MT neurons. Quantiles refer to the time-rescaled ISIs and CDF refers to the expected uniform distribution when the model-estimated CIF corresponds to the true one. For the neuron on the left, the K-S plot shows that the estimated model passed the goodness-of-fit test: the points lie within the 95% confidence interval of the 45° line (dotted); two-sided 95% confidence error bounds of the K-S statistics are denoted by the parallel dashed lines. The model for the neuron on the right tends to overestimate the true CIF. (**B**) Mean ratio R shows that the estimated model for the neuron on the left underestimated the intensity function ($R > 1$) for lower ISI values and reaches to 1 for larger ISI values, indicating the agreement between the model and data; and the estimated model for the neuron on the right overestimated the spike rate ($R < 1$) for all ISI values. The top plot shows the histogram of ISI values for each neuron. (**C**) Histogram of the mean ratios for the population (n=40) (average $R = 0.94 \pm 0.008$), indicating that overall the model was successful in reproducing the ISI statistics of the spike train data (cross-validated data). (**D**) Scatter plot of the consecutive time-rescaled ISIs; correlation value is shown at the top. Lower correlation values correspond to a more independent time-rescaled distribution, which is expected under an accurate model. (**E**) Histogram of correlation values between consecutive time-rescaled ISIs ($r = 0.008 \pm 0.001$) for the population (n=40), indicating that overall the model was successful in capturing the temporal structure of the spike train data (cross-validated data).

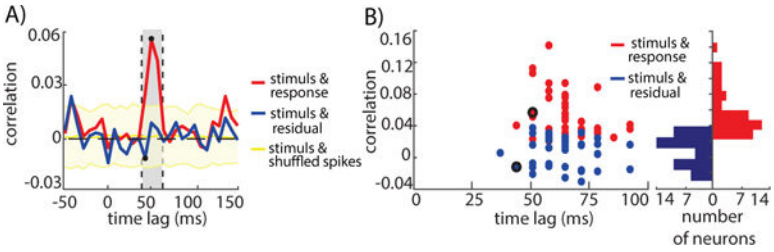

Fig. 5.

Residual analysis of the model's prediction. (**A**) Illustrates temporal correlation function between stimulus and response (blue) and between stimulus and residual (red), for an example MT neuron when the stimulus was presented inside the RF1. The gray bar indicates analysis window used in (**B**), and the yellow area shows the correlation between the stimulus and the shuffled response (mean ± standard error). The absence of correlation between the stimulus and residual indicates that there is no structure left in the data that is statistically related to the stimulus. (**B**) Correlation values for the population, between stimulus and response (blue), and between stimulus and residual (red). The correlation between stimuli and residuals is very small and significantly less than the correlation between stimuli and response (cross-validated data). The histograms show the distribution of correlation values across the population of 40 neurons.

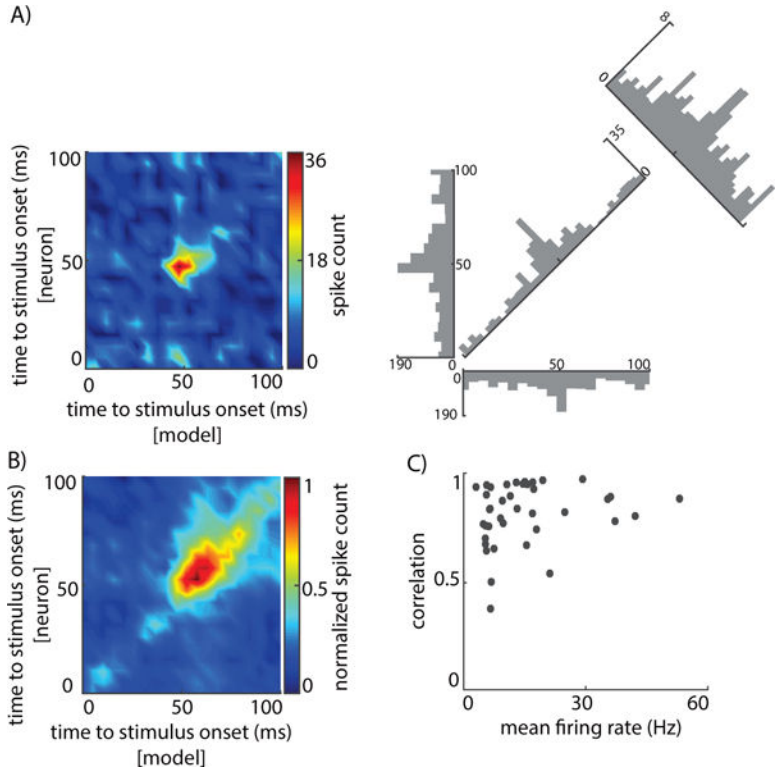

Fig. 6.

Time precision analysis of the model's prediction. (**A**) (left) Heatmap diagram of the JPSTH matrix for stimuli inside the RF of a sample neuron. Each point represents the number of times both the model and the neuron fired a spike at the corresponding time incidences, relative to probe onset (summed across all presentations of that probe). (right) Histograms of values of the JPSTH diagram on the left along different dimensions; the coordinates of histogram grouping are the same as those for the JPSTH diagram. Histograms along the abscissa and ordinate approximate the ordinary PSTH of the actual and predicted spikes, respectively. Histogram along the principal diagonal represents the probability of coincidences in the actual and predicted trains of spikes with respect to the stimulus onset. Histogram on the upper right represents the cross-correlogram of the actual and predicted spike trains. (**B**) Average of normalized JPSTHs for the population (n=40). Each JPSTH was normalized to its range of values before averaging for the population analysis. The high density cloud spans over the response window of the neurons on average. (**C**) Each point corresponds to the cross correlation coefficient between the actual and predicted PSTH responses for each of 40 neurons (cross-validated data).

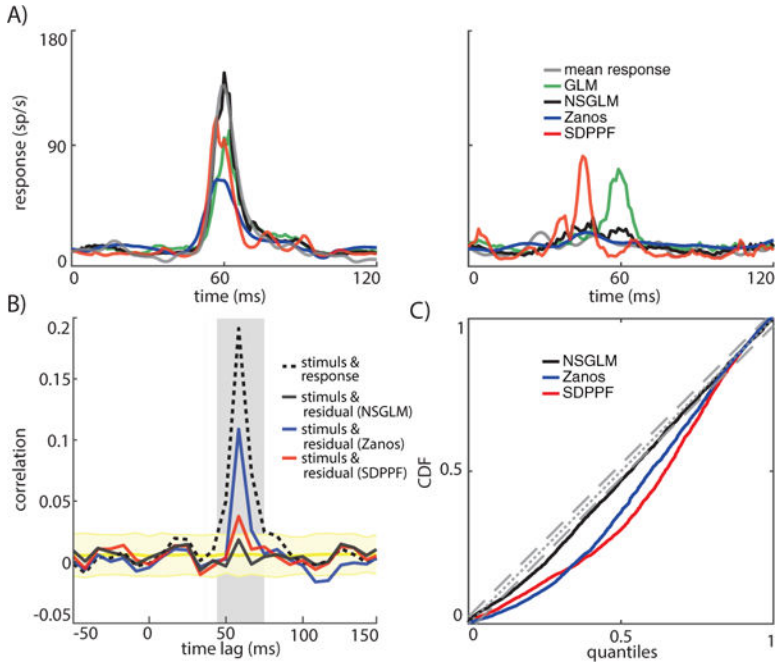

Fig. 7.

Performance comparison of the NSGLM versus widely-used existing approaches for a sample MT neuron. (**A**) Firing rate analysis, for model predicted firing rate response over repeated probe presentation in the RF1 during fixation 1 (left) and fixation 2 (right). While all the models can predict the firing rate response to probes at RF1 during fixation 1, the GLM and the SDPPF-based model fail to correctly predict the firing rate response during fixation 2. (**B**) Residual analysis of predictions from the nonstationary models: The absence of correlation between stimulus in RF1 and residual of the NSGLM prediction (solid black) indicates that the NSGLM captured all the structure in the data that was statistically related to the stimulus; while the remaining stimulus-residual correlations for Zanos (blue) and SDPPF-based (red) models are significantly nonzero in the response window (gray bar). Dashed black trace illustrates temporal correlation function between the same stimulus and the neuron's response, and the yellow area shows the correlation between the stimulus and the shuffled response (mean ± standard error). (**C**) K-S goodness-of-fit plots for the nonstationary models: NSGLM outperforms both models in reproducing the ISI statistics of the spike train data. The parallel dashed lines denote two-sided 95% confidence error bounds of the K-S statistics.