Assessing Neuron Response to External Stimuli with a Data-Driven Procedure for Spike Train Extraction and GAMLSS Regressions

Assessing neuron response to external stimuli with a data-driven procedure for spike train extraction and GAMLSS regressions

Francesco Barile[1], Silvia Dallari[2], Carlotta Pacifici[3⋆], and Bettina Grün[4]

[1] University of Milano-Bicocca, 20126 Milan, Italy
f.barile2@campus.unimib.it,
WWW home page: https://en.unimib.it/francesco-barile
[2] University of Bologna, 40126 Bologna, Italy
silvia.dallari2@unibo.it,
WWW home page: https://www.unibo.it/sitoweb/silvia.dallari2/en
[3] University of Bologna, 40126 Bologna, Italy
carlotta.pacifici@unibo.it,
WWW home page: https://www.unibo.it/sitoweb/carlotta.pacifici/en
[4] Wirtschaftsuniversität Wien, 1020 Wien, Austria
Bettina.Gruen@wu.ac.at,
WWW home page: https://statmath.wu.ac.at/~gruen

Abstract. Gaining a better understanding of the brain is an important aspect in biological research. To this end, new measurement tools are employed to collect data on calcium traces for mice to analyze the neuron response to external stimuli. Suitable statistical analysis methods for the challenging data sets emerging from these measurements are essential to eventually learn more about the functioning of the brain. We use the calcium traces collected for one mouse across three sessions to showcase a statistical analysis focusing on extracting potential effects of external stimuli on the neuron response based in particular on advanced regression modeling approaches. This statistical analysis consists of: (1) a heuristic algorithm for spike train extraction, (2) an exploratory analysis in combination with suitable pre-processing to highlight features of the data and (3) a regression analysis based on zero-adjusted generalized additive linear models for location, scale and shape to relate the spike trains to stimuli and sessions as well as inspect their evolvement over time. Results indicate how the flexibility of advanced regression modeling approaches allows to account for specific data structures as well as easily compare different modeling approaches.

Keywords: exploratory analysis, GAMLSS, semi-parametric regression model, spike train extraction, zero-adjusted distribution

⋆ The first three authors contributed equally to the work. They are alphabetically ordered by surnames.

1 Introduction

The study of the neuron response to external stimuli in awake behaving mice is nowadays possible thanks to the existence of powerful microscopes which capture calcium levels. While, at rest, neurons have a stable intracellular calcium concentration, during activity, calcium floods the cells and leads to a quick and transient rise in the concentration. Therefore, the fluorescent calcium traces of the observable neurons are recorded over time and the transient spikes in the level of calcium can be considered a response to a stimulus.

An important applied research question in the field of neuroimaging is the investigation of external stimuli effects on neuron's activations. In general, external stimuli effects can be estimated using regression-like models where predictors that can affect the outcome serve as covariates. Models that investigate the effect of a specific stimulation on the produced activity are usually referred to as encoding models, as the interest is to predict the activity of a population of neurons in response to a given stimulus and other experimental conditions. Previous work for example considered an encoding model, where the number of spikes is regressed over a set of experimental conditions [4]. Generalized linear models (GLM; [11]), which are flexible and yet interpretable, are one of the fundamental tools for investigating the response of neurons to external factors.

We use data from the data repository of the Allen Brain Observatory [2] to showcase a suitable statistical analysis for inferring effects of external stimuli on neuron response. We use data consisting of calcium concentration traces of neurons collected for a single mouse across three sessions where the mouse was exposed to different sequences of stimuli. The statistical analysis requires to extract spike trains from the calcium concentration traces which represent the neuron response and then specify suitable regression models to infer external stimuli effects as well as account for potential time effects. We indicate that for the present data the assumptions underlying GLMs are too restrictive, causing the models to be unsuitable for application and indicating the need to consider a more flexible modeling framework.

The first aim is to estimate the spike trains starting from the observed calcium traces. For this purpose, a three-step heuristic algorithm is proposed. Once the spike trains have been estimated, the focus moves to the analysis of how the mouse reacts as stimuli, sessions, neurons and time vary. To accomplish this task, we start by performing an exploratory analysis before implementing a regression analysis. The exploratory analysis shows that the estimated spikes have a large amount of zeros and a positively skewed distribution. Thus, a zero-adjusted generalized additive model for location, scale and shape (GAMLSS) which is able to account for the presence of zeros is exploited to perform the regression analysis. The GAMLSS [14] provides a lot of flexibility on how the dependent variable distribution is modeled. The GAMLSS is suitable to model a response variable where the parameters of its underlying random variable are expressed as functions of explanatory variables and the dependent variable does not necessarily follow an exponential family distribution.

This work is organized as follows. After a brief description of the data in Section 2, the algorithm implemented to estimate the spike trains is presented in Section 3. In Section 4 the exploratory analysis is shown, while the regression analysis based on zero-adjusted GAMLSS is performed in Section 5. We conclude with a discussion in Section 6.

2 Data Set

We analyze a data set obtained from the data repository of the Allen Brain Observatory [2]. The data contain the cortical activity of neurons at different depths from different visual areas in response to a range of visual stimuli in mice. The experiments performed result in recordings of the fluorescence traces of neurons while the mice are placed in front of a screen that shows different types of visual stimuli (more information can be found, for example, in [2] and [5]).

The calcium traces over time of a single mouse are observed for several neurons across different imaging sessions. Each imaging session consists of a sequence of different stimuli. As outlined in a technical report (originally available from the Allen Brain Observatory website; [1]), there are 5 types of stimuli:

- *drifting gratings*: This stimulus consists of a full-field sinusoidal grating that drifts in a direction perpendicular to the orientation of the grating; it is labeled as *drift_grat* in the following.
- *static gratings*: This stimulus consists of a full-field sinusoidal grating, that varies in orientation (the angle of the grating), spatial frequency (the width of the grating), and phase (the position of the grating); its label is *stat_gratings*.
- *natural scenes*: This stimulus involves a long sequence of images belonging to nature and to the animal world; it is labeled as *nat_scenes*.
- *natural movie*: This stimulus involves three different scenes taken from the film "Touch of Evil" by Zugsmith & Welles in 1958. The labels of the three scenes are *nat_mov_1*, *nat_mov_2* and *nat_mov_3*, respectively.
- *locally sparse noise*: This stimulus presents a sequence of multiple black and white spots moving in different locations. Depending on the spot size used, this stimulus is called *locally sparse noise 4 degree*, labeled as *l_spar_n_4d*, or *locally sparse noise 8 degree*, labeled as *l_spar_n_8d*.

The stimuli are separated by an inter-stimuli grey period, which is labeled as *none* in the following analysis. In addition, for each imaging session, there are at least five minutes with no structured visual stimulus given to the mouse (what is called spontaneous activity and is also labeled *none* in the analysis that follows).

The imaging sessions considered in this work are sessions A, B and C2. Session A is characterized by *drifting gratings, natural movie 1, natural movie 3, spontaneous activity* and *inter-stimuli grey period*. Session B contains *static gratings, natural images, natural movie 1, spontaneous activity* and *inter-stimuli grey period*, while session C2 contains *locally sparse noise 4 degree, locally sparse noise 8 degree, natural movie 1, natural movie 2, spontaneous activity* and *inter-stimuli*

grey period. For simplicity, session C2 is referred to as session C. In the regression analysis that follows a session-specific effect is included that captures potential mean shifts across sessions. In this way the model considered is able to capture any unexplained differences between the sessions and this would also include potential ordering effects.

The data set can be acquired exploiting the Jupyter notebooks [10] accessible at https://allensdk.readthedocs.io/en/latest/examples.html. We filter the data by selecting only one mouse, only the primary visual cortex area and only three different depths (200 μm, 275 μm, 375 μm). Furthermore, the first observations of some traces are cut in order to have the frame of the first stimulus align.

Traces are not available for all neurons in all the sessions. There seems to exist a spatial and temporal pattern linking the neuron's activity to the stimuli [1]. This could explain the fact that not all neurons are present in each session and suggests that the absence is not at random. In this work, we proceed by ignoring this issue and by retaining and analyzing the neurons where traces appear in each session. In the regression analysis, we include neuron-specific main effects to alleviate this issue.

To fix a bit of notation, note that in the following the neurons are indexed with i, the sessions with j, the stimuli with s and the time point with t or $\tilde{t}$ depending on the time unit used. The index j corresponds to session A, B or C, while i takes values from 1 to n_j or to $\tilde{n}$, with n_j indicating the total number of neurons analyzed in session j and $\tilde{n}$ the total number of neurons present in all three sessions. In particular, in the following analysis 47 neurons are studied in session A, 44 in session B and 39 in session C, while just 27 neurons are considered in all the three sessions. Finally, t goes from 1 to T_j and $\tilde{t}$ from 1 to $\tilde{T}_j$, where T_j and $\tilde{T}_j$ are the total number of observations in session j when the time unit is the frequency of 30 Hz (also denoted as frames) and seconds, respectively. Specifically, for these data T_j is 115,481, 113,854 and 124,060 for sessions A, B and C respectively, while $\tilde{T}_j$ is 3,817, 3,763 and 4,104.

3 Data-driven spike train extraction

The neuron response is captured by spike trains which contain the information on the locations of a neuron response in form of a spike as well as the size of the response in form of the amplitude of the spike. These spike trains are not directly observable but only noise measurements in form of the calcium concentration traces are available. We propose a heuristic algorithm to extract the spike trains from the calcium traces in a data-driven way.

The proposed procedure is based on the following generative model which is popular in neuroscience. This model considers the observed traces as noisy realizations of the true calcium concentration and the dynamics of the calcium concentration as an autoregressive process with jumps in correspondence to the firing events caused by the neural response (see [16]). Thus, for each imaging

session j and neuron i, one has:

$$y_{i,j,t} = b_{i,j} + c_{i,j,t} + \epsilon_{i,j,t}, \qquad \epsilon_{i,j,t} \sim N(0, \sigma^2_{i,j}), \quad t = 1, \ldots, T_j, \tag{1}$$

$$c_{i,j,t} = \gamma_{i,j}\, c_{i,j,t-1} + a_{i,j,t} + v_{i,j,t}, \quad v_{i,j,t} \sim N(0, \tau^2_{i,j}), \quad t = 2, \ldots, T_j, \tag{2}$$

where $y_{i,j,t}$ is the observed calcium trace, $c_{i,j,t}$ is the calcium concentration, $b_{i,j}$ is a baseline parameter, $\gamma_{i,j}$ is a decay parameter, $\sigma^2_{i,j}$ and $\tau^2_{i,j}$ are the variances of the Gaussian errors and $a_{i,j,t}$ is the neuron's response to be estimated, which is 0 if there is no spike, i.e., no firing event, and which takes a positive value when a spike is detected. In the following, a simplified version of this model is adopted, in which the baseline parameter $b_{i,j}$ is set equal to zero and the variance $\tau^2_{i,j}$ is assumed to be negligible and, thus, $v_{i,j,t}$ is considered equal to zero.

The three different steps of the proposed procedure which results in estimates for $a_{i,j,t}$ representing the neuronal response consist of:

Step 1: Identification of the longest time interval without stimuli where presumably no spikes are contained.
Step 2: Estimation of the decay parameter, of the calcium concentration and of the time points where spikes occur.
Step 3: Estimation of the amplitude of the spikes.

In more detail, each of the steps can be described as follows.

Step 1. It is assumed that in absence of stimuli spikes are less likely to occur compared to when a stimulus is present. This is the reason why the *spontaneous activity* time intervals are considered in this step to find an interval without potential spikes. Thus, the first step of the algorithm involves:

1. Select in a data-driven way the largest time interval in which no stimuli occur. Figure 1 highlights the time interval selected by the algorithm at this point for a specific imaging session.
2. Fit a LOWESS smoother (i.e., a locally-weighted polynomial regression function; [3]) to the calcium time series for each neuron within each session to help detecting the potential spikes in this time interval. In particular, a smoother span of 0.01 is chosen.
3. Address the presence of potential measurement errors by setting within each session a neuron-specific threshold defined as:

$$\text{threshold}_{i,j} = \text{median}_{i,j} + 2RSD_{i,j},$$

where $\text{median}_{i,j}$ and $RSD_{i,j}$ are the corresponding median and robust estimator of the standard deviation of the calcium traces for neuron i and session j. At this point it is assumed that in absence of stimuli the calcium distribution is approximately Gaussian. Thanks to this assumption $RSD_{i,j}$ can be taken as $IQR_{i,j}/1.349$, where IQR is the interquartile range, that is a robust statistic, and 1.349 corresponds to the length of the IQR of the standard Gaussian distribution.

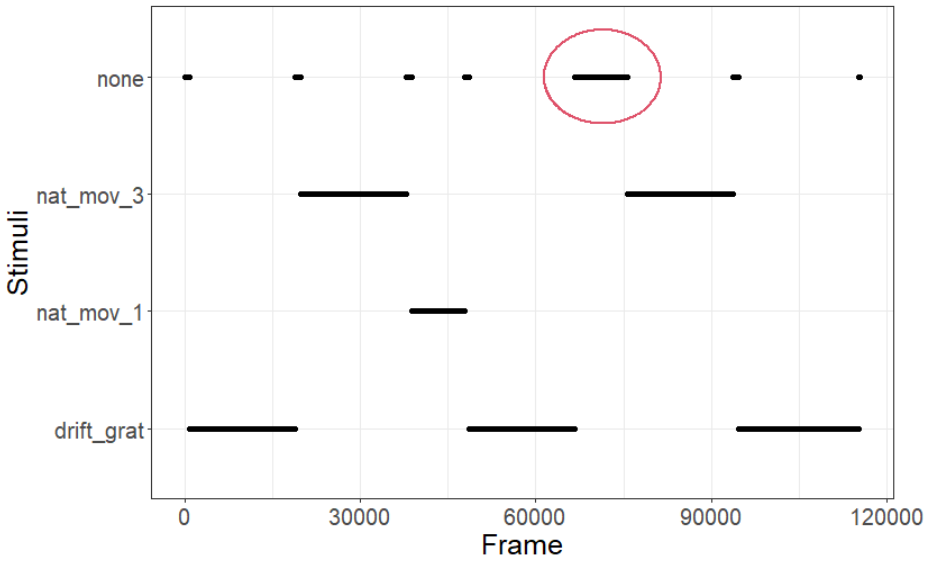

Fig. 1: Stimuli duration for session A. The circle highlights the selected time interval.

4. Select the largest time sub-interval in which the smoothed values are below the threshold by combining points 2 and 3. Figure 2 shows the observed and the smoothed time series for a specific neuron and session, together with the threshold obtained with the formula of point 3. The highlighted sub-interval is the one chosen by the algorithm in this step.

Step 2. The observed calcium time series in the sub-interval identified in Step 1 is used to estimate within each session a neuron-specific parameter $\gamma_{i,j}$ by fitting an AR(1) model. Then, the estimated $\gamma_{i,j}$ parameter enters the constrained l_0 penalization procedure presented below as a fixed parameter.

To estimate the calcium concentration from the noisy measurements, the constrained l_0 penalization procedure developed by [9] is applied using the R [12] package `FastLZeroSpikeInference` [8]. This procedure estimates the concentration, for each neuron i and session j, by solving:

$$\underset{c_{i,j,1},\ldots,c_{i,j,T_j},a_{i,j,2},\ldots,a_{i,j,T_j}}{\text{minimize}} \left\{ \frac{1}{2} \sum_{t=1}^{T_j} (y_{i,j,t} - c_{i,j,t})^2 + \lambda_{i,j} \sum_{t=2}^{T_j} \mathbb{I}_{\mathbb{R}^+ \setminus \{0\}}(a_{i,j,t}) \right\} \quad (3)$$

$$\text{subject to } a_{i,j,t} = c_{i,j,t} - \gamma_{i,j}\, c_{i,j,t-1} \geq 0.$$

The hyperparameter $\lambda_{i,j}$ controls the degree of penalization to identify spikes, i.e., positive values of $a_{i,j,t}$. To choose $\lambda_{i,j}$, the following procedure has been considered in the literature (see [5]): One performs the procedure for different values of the tuning parameter and selects the smallest value that leads to

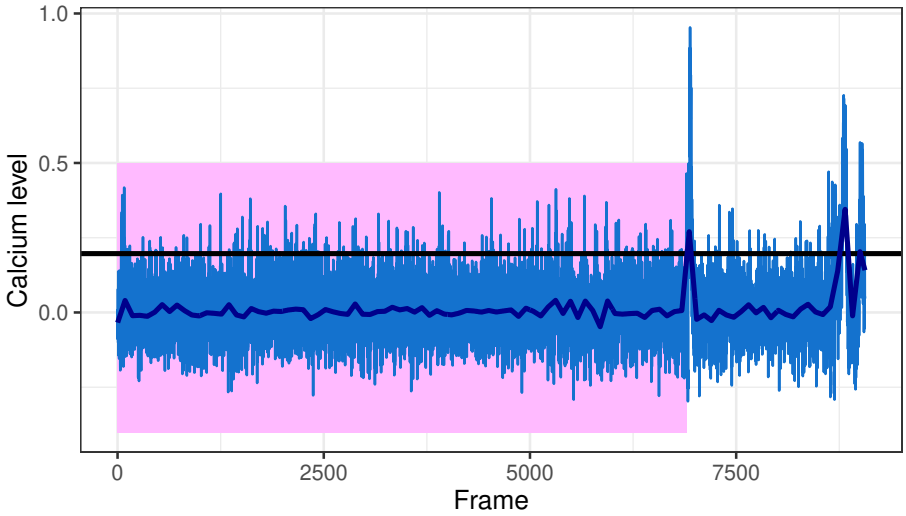

Fig. 2: The lighter colored time series represents the observed calcium time series, the darker one the smoothed time series and the horizontal line the threshold for a specific neuron and session.

a minimal number of identified events which are smaller than two times the standard deviation of the noise. Following this idea, we determine the session and neuron-specific hyperparameter $\lambda_{i,j}$ as the smallest value minimizing the number of identified events smaller than $2RSD_{i,j}$, where an event is defined as $c_{i,j,t} - c_{i,j,t-1} > 0$, with t an estimated time at which the neuron spikes and $c_{i,j,t}$, $c_{i,j,t-1}$ the estimated calcium concentration at time t and $t-1$ respectively. When an event is smaller than $2RSD_{i,j}$, the corresponding change in $c_{i,j,t}$ is considered a random increase, which does not constitute a spike, i.e., a neuronal response. Note that in this step, calcium concentrations and spike times are estimated, for each of the different values of $\lambda_{i,j}$, using the constrained l_0 penalization procedure described in Equation (3) and considering the observed calcium time series in the sub-interval found in Step 1 for a specific neuron and session.

In more detail, the following sub-steps are executed:

1. Consider the interval $[0, 1]$ for $\lambda_{i,j}$.
2. Shift the interval, if necessary, to ensure it contains a value of $\lambda_{i,j}$ such that the number of events smaller than $2RSD_{i,j}$ is zero, which means the estimated calcium concentrations do no contain any random increases. Indeed, the purpose is to choose the smallest $\lambda_{i,j}$ according to which no random increases are present.
3. Divide the interval in 10 parts (sub-intervals) of equal length and choose the sub-interval where at the lower bound a positive number of events smaller

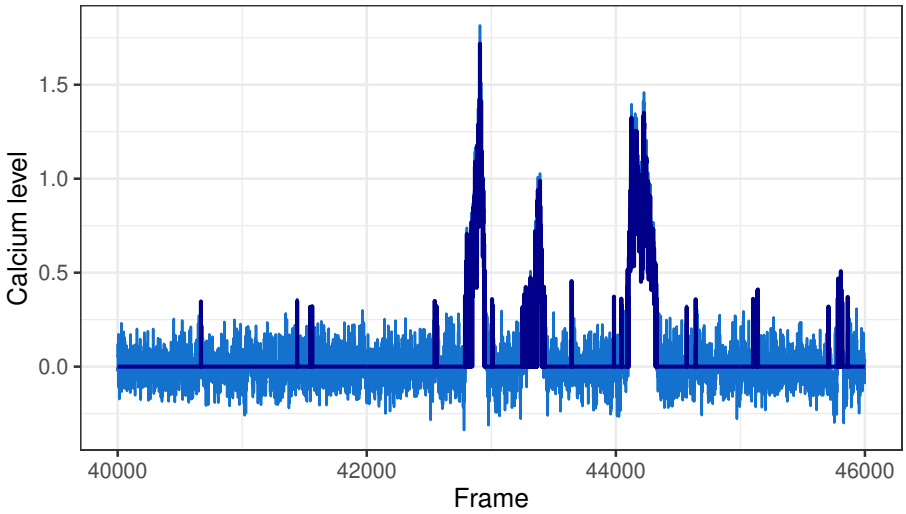

Fig. 3: The lighter colored time series is the observed calcium time series, while the darker one represents the estimated spike trains for a portion of the frame of a specific neuron and session.

than $2RSD_{i,j}$ is detected and where at the upper bound no random increases are detected.

4. Stop when at the lower bound of this sub-interval the number of random increases detected is equal to 1, or the length of the sub-interval considered is smaller than a pre-specified ϵ. This sub-step ensures that the chosen $\lambda_{i,j}$ is the smallest among all the ones that lead to zero random increases.
5. Choose as hyperparameter $\lambda_{i,j}$ the upper bound of the final interval obtained at the previous sub-step.

Once $\gamma_{i,j}$ and $\lambda_{i,j}$ are estimated using the data only from the sub-interval which presumably contains no spikes, the calcium concentration and the spike times are estimated considering the complete data for a specific neuron of a given session using the l_0 optimization problem described in Equation (3).

Step 3. The spikes' amplitudes are estimated for each neuron i within session j using the following formula:

$$a_{i,j,t} = c_{i,j,t} - \gamma_{i,j}\, c_{i,j,t-1}, \tag{4}$$

where t is a time point at which the neuron spikes. At the time points in which the neuron does not spike, $a_{i,j,t} = 0$. In Figure 3 the spike trains estimated by the algorithm are represented together with the fluorescence traces for a specific neuron and session. Only a sub-interval of time is shown to be able to easily compare the observed time series to the estimated spike train time series.

4 Exploratory Analysis

The aim of the exploratory analysis is to characterize and summarize the estimated spike trains, available for each neuron and potentially varying by visual stimulus and imaging session. To reduce the influence of extreme values, the initial and final spikes with no associated stimuli have been removed from the data set for each session and neuron before performing the following exploratory analysis.

The heatmap presented in Figure 4 shows the neuron-specific estimated average spike values disaggregated by session and stimulus. Given a specific session j and stimulus s, the plotted value $\bar{a}_{i,j,s}$ for neuron i is obtained as the average of the estimated spikes $a_{i,j,t}$ over the number of observations specific of that session and stimulus:

$$\bar{a}_{i,j,s} = \frac{1}{|T_{j,s}|} \sum_{t \in T_{j,s}} a_{i,j,t} \quad \forall\, i = 1, \ldots, \tilde{n},\ j \in \{A, B, C\} \text{ and } s = 1, \ldots, S_j, \quad (5)$$

where $T_{j,s}$ is the set of observations belonging to session j and stimulus s and $|T_{j,s}|$ represents its size, equal for each neuron i. as we already mentioned in Section 2, $\tilde{n}$ is the number of neurons appearing in all three sessions and S_j is the number of stimuli in session j.

Figure 4 provides useful information on the composition of the sessions and the underlying neuronal activity. The only stimulus that appears in all three sessions is natural movie 1 (labeled "nat_mov_1"). The sessions share also a period of absence of stimuli, indicated with the label "none", where the neurons are simply left to spontaneous activity. The stimuli for which there are the highest spike values all refer to natural movie, because the heatmap cells for these columns appear lighter. In particular, for natural movie 1 the highest values of estimated spikes are registered throughout all neurons. As opposed to the natural movie stimuli, those based on gratings produce low spike levels. In addition, neuron "696325683" seems to remain quite active regardless of the stimuli and the sessions, since even during the absence of stimuli its spikes are higher than those of other neurons.

Figure 4 provides a static information on the behavior of the estimated spikes. The following analysis still focuses on the estimated spikes but now observed over time. The estimated spikes have a frequency of 30 Hz, meaning that they are observed every $0.0\bar{3}$ seconds. This results in a very high frequency and we thus re-scale the spike trains to a bigger time unit that is the second. Using this time unit implies that 30 to 31 observations are aggregated to obtain the second-specific measurement. Given a specific neuron i and session j, the newly aggregated spikes associated to the same second are denoted as $\tilde{a}_{i,j,\tilde{t}}$ and are aggregated as follows:

$$\tilde{a}_{i,j,\tilde{t}} = \sum_{t=1}^{\Delta} a_{i,j,t} \qquad \forall\, i = 1, \ldots, \tilde{n},\ j \in \{A, B, C\} \text{ and } \tilde{t} = 1, \ldots, \tilde{T}_j, \quad (6)$$

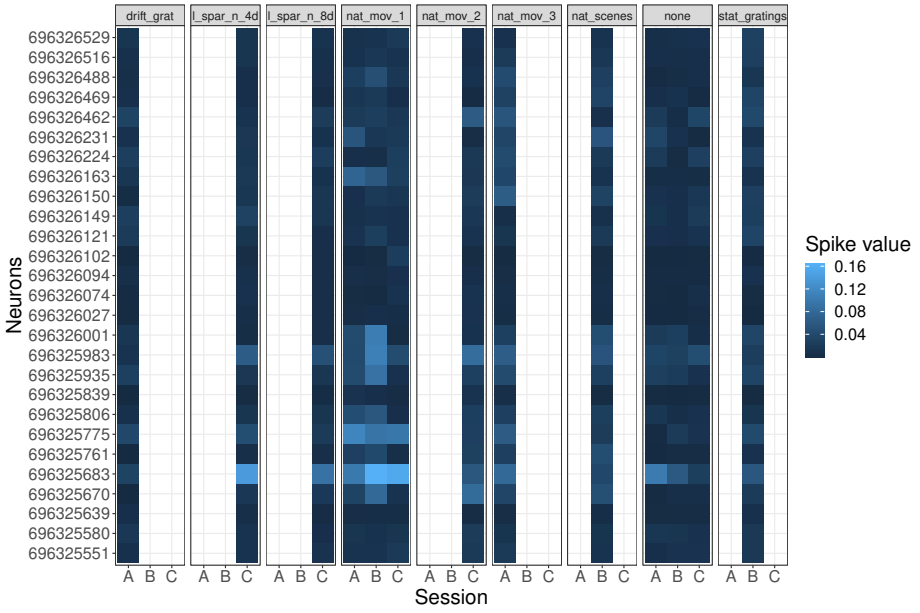

Fig. 4: Estimated mean spikes by session (x-axis), stimulus (columns) and neuron (y-axis).

where $\Delta \in \{30, 31\}$ is the number of observations within the aggregation window; as already mentioned in Section 2, $\tilde{n}$ is the number of neurons appearing in all the three sessions and $\tilde{T}_j$ is the total number of observations expressed in seconds in session j. The stimulus associated to $\tilde{a}_{i,j,\tilde{t}}$ is defined as the modal stimulus occurring in the aggregation window.

Figure 5 shows the time series of the estimated spike trains for a specific neuron in session A. Blocks represent the visual stimuli occurring in the session. In general, the spikes are higher at the beginning of the session and as time progresses the level of the highest spikes decreases until the end of the interval where no stimuli occurs (block labeled 4). After this spontaneous activity interval, responses to the stimulus "natural movie 3" (block with label 2) are slightly higher than the first time the stimulus was transmitted; similar considerations can be made for the stimulus "drifting gratings" (block with label 1) at the end of the session, since the spikes are higher than those corresponding to the same stimulus transmitted right before the period of spontaneous activity.

Figure 6 shows the time series of the estimated spike trains for a specific neuron in session B. Similarly to the previous case, the first stimulus (i.e., "static gratings", block with label 1) presents the highest levels of spikes. The spikes are slightly more persistent given that they remain high also at the beginning of the second stimulus "natural images" (block with label 2). This behavior may be motivated by the fact that the type of stimulus is changed. Lastly, also for

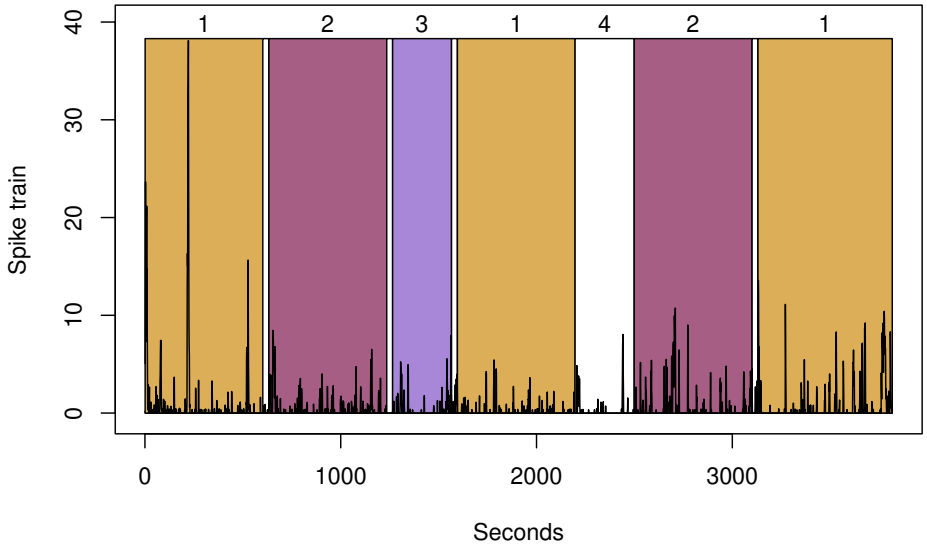

Fig. 5: Spike train (y-axis) aggregated by seconds (x-axis) in session A. Blocks represent the stimuli characterizing the session, labeled as: 1 = drifting gratings, 2 = natural movie 3, 3 = natural movie 1, 4 = spontaneous activity.

session B it is confirmed that the amplitude of the spikes becomes higher after the period of spontaneous activity.

Figure 7 shows the time series of the estimated spike trains for a specific neuron in session C, which consists of a greater number of stimuli belonging to two main dimensions: sparse local noise and movie. The series appears more noisy, especially for what concerns the first type of stimuli (blocks with label 1 and 2) which are transmitted exclusively in this session. The neuronal activity seems persistent and it remains high even in the first and second interval of spontaneous activity (block with label 3). Only towards the end of the session, where the neuron has been exposed to the locally sparse noise stimuli more than once, the activity seems to decrease in terms of the amplitude of the spikes. Concerning the stimuli "natural movie 1" (block with label 4) and "natural movie 2" (block with label 5), the neuron reacts similarly in terms of the amplitude of the spikes.

Lastly, it is interesting to analyze the distribution of the estimated spikes according to the three imaging sessions. Figure 8 provides information on the shape of the distribution of the spikes grouped by sessions. Each histogram reflects the values of the estimated spikes for all the neurons, stimuli and seconds in each session. In all three sessions, the distribution of the estimated spikes is clearly heavily positively skewed and of a non-Gaussian form. The vast majority of the observations are also close to zero, and indeed a substantial proportion of values in the estimated spike trains is equal to zero.

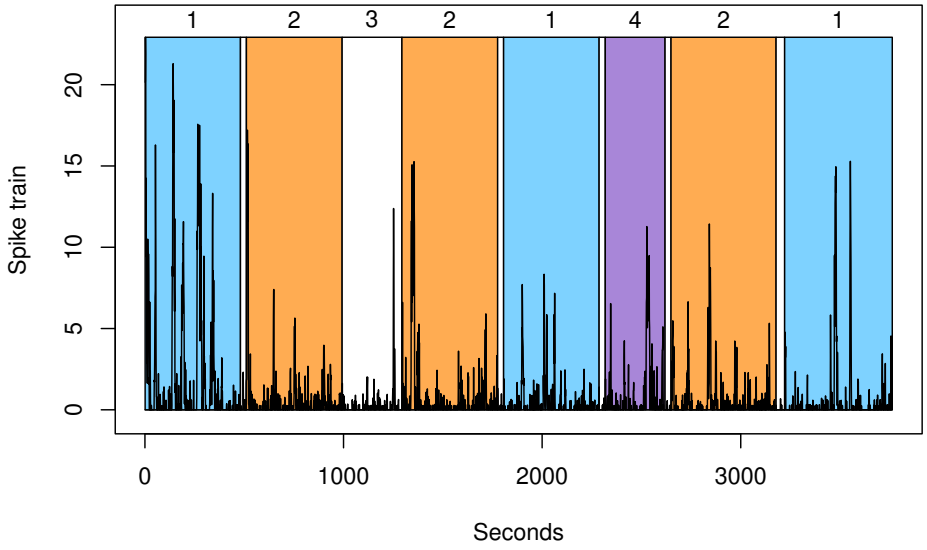

Fig. 6: Spike train (y-axis) aggregated by seconds (x-axis) in session B. Blocks represent the stimuli characterizing the session, labeled as: 1 = static gratings, 2 = natural images, 3 = spontaneous activity, 4 = natural movie 1.

The exploratory analysis suggests that a regression analysis to infer stimuli effects needs to take the data characteristics of the spike trains into account. A suitable distribution needs to combine a point-mass distribution allowing for the zero observations with a continuous positive distribution which is required to capture the positive skewness.

5 Regression Analysis Using GAMLSS

We employ a regression approach to infer stimuli effects on the neuron response using the estimated spike trains as dependent variable. We also investigate potential time effects on neuron response and account for potential ordering effects and neuron-specific effects by also including sessions and neurons as covariates.

The characteristics of the dependent variable require a more flexible modeling framework for a univariate response variable compared to what the popular GLM and generalized additive models (GAM; [7]) provide. Generalized additive models for location, scale and shape (GAMLSS; [14]) are able to provide the flexibility needed because they do not assume that the response variable y is distributed according to an exponential family distribution. GAMLSS only require a parametric distribution assumption for the response variable. Also, within the GAMLSS framework, the systematic part of the model is expanded to allow not only the mean, or in general location, but all the parameters of the conditional distribution of y to be modeled as functions of explanatory variables x.

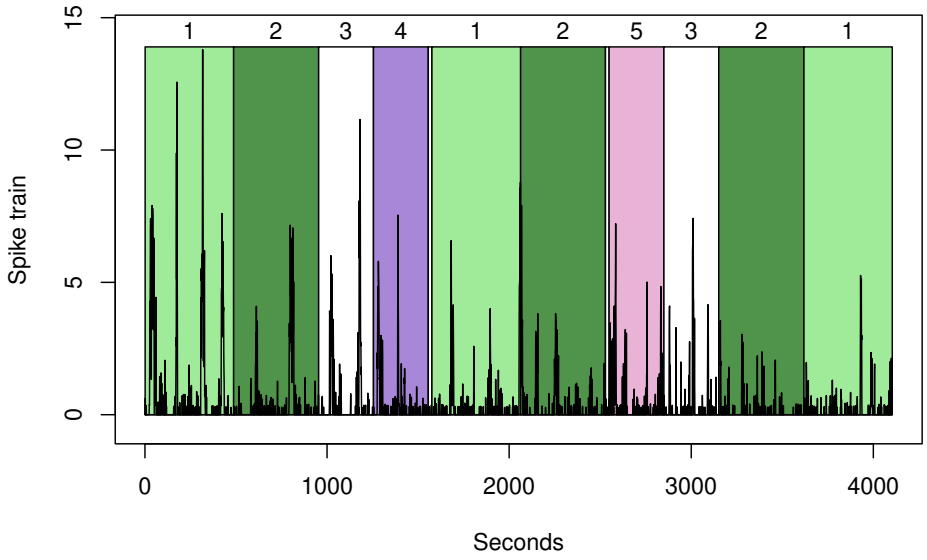

Fig. 7: Spike train (y-axis) aggregated by seconds (x-axis) in session C. Blocks represent the stimuli characterizing the session, labeled as: 1 = locally sparse noise (4 deg), 2 = locally sparse noise (8 deg), 3 = spontaneous activity, 4 = natural movie 1, 5 = natural movie 2.

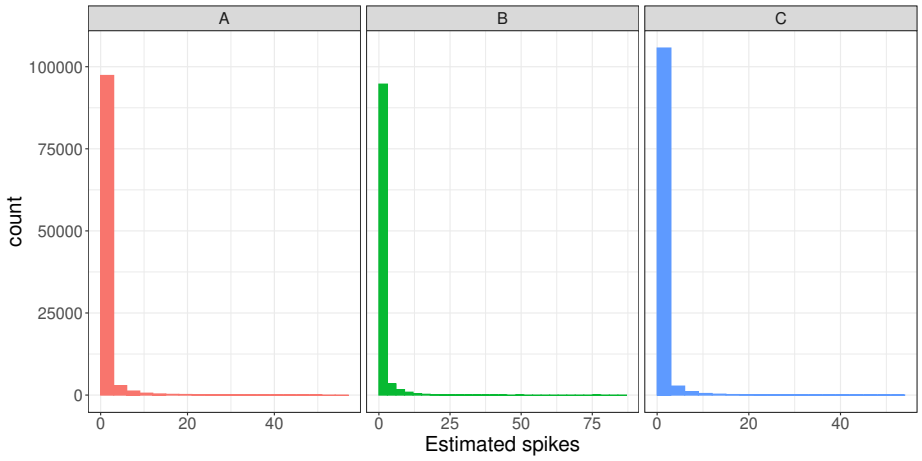

Fig. 8: Histogram of estimated spike train values grouped by sessions.

A semi-parametric modeling approach is also investigated for the time effects in order to relax the assumption of a linear effect of time on the neuron response. We thus exploit two extensions of GAMLSS compared to GLM which consist of

the increased flexibility in selecting the distribution for the dependent variable as well as including additive effects for covariates.

5.1 Considered Zero-Adjusted GAMLSS

For our analysis, we consider as response variable the estimated spikes $\tilde{Y}_i = \text{sum}$ of spikes per second, defined in Equation (6) as $\tilde{a}_{i,j,\tilde{t}}$, where, after re-indexing, observation index i ranges from $1, \ldots, n$ with $n = \sum_{j \in \{A,B,C\}} \tilde{T}_j$.

To account for the presence of zeros in the estimated spike train series, a zero-adjusted GAMLSS is considered. This allows to model non-negative response variables (whose data points include 0's) by combining a point-mass distribution at zero with a continuous distribution defined on the positive real line. This results in a model with one extra parameter for the mass point at zero and is equivalent to fitting two separate models, a GAMLSS for the $(0, \infty)$ part, and a logit model for the zero part versus the non-zero part.

Specifically, let $\mathcal{D}$ be a distribution defined on the positive real line parameterized by the vector of parameters $\boldsymbol{\theta}_i = (\mu_i, \sigma_i, \nu_i, \tau_i)^\top$. Moreover, let $\tilde{Y}_i$ be a non-negative response variable. Then,

$$\tilde{Y}_i = \begin{cases} 0, & \text{with probability } \pi_i, \\ Y_i, & \text{with probability } 1 - \pi_i, \end{cases}$$

leading to a mixture of an absolute continuous and a point-mass distribution. In particular:

$$\tilde{Y}_i \overset{ind}{\sim} \pi_i \delta_0 + (1 - \pi_i)\, \mathcal{D}\left(\boldsymbol{\theta}_i\right)$$

where δ_x denotes the Dirac measure centered on x and the mixing weight $\pi_i = \text{logit}^{-1}(\eta_0^i)$. Note that $\mathbb{I}_{\{0\}}(\tilde{Y}_i) \sim \textit{Bernoulli}(\pi_i)$, where $\mathbb{I}_{\{0\}}(\tilde{Y}_i)$ is an indicator function for the event $\{\tilde{Y}_i = 0\}$.

We fit two different zero-adjusted GAMLSS to investigate how the stimuli in the three different sessions affect the spike trains of each neuron through time.

Linear parametric zero-adjusted GAMLSS. For the continuous part of the mixed distribution, we consider $Y_i \sim \textit{Gamma}(\boldsymbol{\theta}_i)$, where the population distribution vector of parameters boils down to the bi-dimensional vector $\boldsymbol{\theta}_i = \left(g_1^{-1}\left(\eta_1^i\right), g_2^{-1}\left(\eta_2^i\right)\right)^\top$ encoding *location* and *scale* parameters. The *shape* and *scale* parameters of a Gamma distribution are simply reparametrizations of the former pair of population parameters. In particular:

- we consider the *log-link* function for the *location* parameter, i.e., $g_1\left(\mu_i\right) = \log\left(\mu_i\right)$, and four explanatory variables, namely:

  $$\boldsymbol{x}_1^i = \left(x_{11}^i = \textit{session}_i,\ x_{12}^i = \textit{stimuli}_i,\ x_{13}^i = \textit{time}_i,\ x_{14}^i = \textit{neuron}_i\right)^\top$$

  with x_{12}^i being the mode of the stimuli in that second and the time x_{13}^i being in seconds;

– we consider a constant *scale* parameter, with *identity-link* function, i.e., $g_2(\sigma) = \sigma$, that is:
$$\boldsymbol{x}_2^i = 1,\ \forall i = 1, \ldots, n$$
implying that the model matrix across all observations corresponds to $\mathbf{X}_2 = \mathbf{1}_n$, with $\mathbf{1}_n$ denoting the unit vector of dimension n;
– we do not consider additive terms.

For the point-mass part of the mixed distribution:

– we consider $\boldsymbol{v}_i = \boldsymbol{x}_1^i,\ \forall i = 1, \ldots, n$, so that the model matrix across all observations is given by $\boldsymbol{V} = \mathbf{X}_1$;
– we do not consider additive terms.

This leads to a simple logistic regression model. Overall, we have:

$$\text{logit}(\pi_i) = \boldsymbol{\alpha}^\top \boldsymbol{v}_i \ \text{ and } \ Y_i \sim Gamma(\exp(\boldsymbol{\beta}_1^\top \boldsymbol{x}_1^i), \sigma). \tag{7}$$

Semi-parametric zero-adjusted GAMLSS. The linear parametric zero-adjusted GAMLSS imposes that an estimated time effect induces approximately a constant rate of change. For the model in Equation (7), it holds that

$$\frac{\partial}{\partial x}\mathbb{E}(\tilde{Y} \mid x) = \frac{\partial}{\partial x}\left[1 - \pi(x)\right]\mu(x) = \frac{((\beta - \alpha)e^{\alpha x} + \beta)e^{\beta x}}{(1 + e^{\alpha x})^2} \approx \frac{2\beta - \alpha}{4}$$

as, for sufficiently small α and β, $e^{\alpha x} \approx e^{\beta x} \approx 1\ \forall x$ lying in the bounded time domain. This implies that the expected spike intensity evolves approximately linearly over time x, which could be a restrictive framework.

To allow for a general evolvement over time, we consider an alternative model where we let the regressor $x_{13} = \textit{time}$ (in seconds) enter the model in Equation (7) in a semi-parametric fashion considering penalized smoothing splines:

$$\text{logit}(\pi_i) = \boldsymbol{\alpha}^\top \boldsymbol{v}_i + h_0\left(\mathbf{x}_{13}^i\right) \ \text{ and } \ Y_i \sim Gamma(\exp(\boldsymbol{\beta}_1^\top \boldsymbol{x}_1^i + h_1\left(\mathbf{x}_{13}^i\right)), \sigma). \tag{8}$$

with $\boldsymbol{v}_i = \boldsymbol{x}_1^i = \left(x_{11}^i = \textit{session}_i,\ x_{12}^i = \textit{stimuli}_i,\ x_{14}^i = \textit{neuron}_i\right)^\top$.

5.2 Estimation and Inference

We use the `gamlssZadj` function of the **gamlss.inf** package [6] to obtain the maximum likelihood estimates for the parameters in the two models specified. Package **gamlss.inf** is an add-on package to package **gamlss** [15] and it estimates the two components of the mixture distribution separately. However, the resulting fitted object behaves like a typical object returned from fit functions in package **gamlss**. For the semi-parametric specification, we use function `pb()` from package **gamlss**, so that, to estimate $h_k\left(\mathbf{x}_{13}^i\right)\ \forall\ k = 0, 1$, it simply suffices to insert $\texttt{pb}\left(\mathbf{x}_{13}^i\right)$ in the regression model formulas.

Once the model has been estimated, expected spike trains can be predicted as a function of the explanatory variables to investigate their partial effect on

it. Specifically, expected spike trains are computed as point estimates of $\mathbb{E}(\tilde{\boldsymbol{Y}} \mid \boldsymbol{X}_1, \boldsymbol{V}) = (1 - \boldsymbol{\pi})\,\mathbb{E}(\boldsymbol{Y} \mid \boldsymbol{X}_1)$ with $\boldsymbol{\pi} = \mathbb{E}(\mathbb{I}_{\{0\}}(\tilde{\boldsymbol{Y}}) \mid \boldsymbol{V})$, where $\boldsymbol{X}_1$ and $\boldsymbol{V}$ are the model matrices collecting all the explanatory variables of the continuous and point-mass part of the model respectively, regardless of the fashion in which they enter the model (e.g., *linear parametric*, *semi-parametric*).

In order to have some measure of reliability of the point estimates, confidence bounds are necessary. However, function `gamlssZadj` of the **gamlss.inf** package does not provide any methods to combine estimates of the two separate models in order to get standard errors. Thus, a simulation-based method is exploited to this end. In particular, the asymptotic normality of the maximum likelihood estimates is exploited to simulate a distribution for $\mu_i = g_1^{-1}(\eta_1^i)$ and $\pi_i = \text{logit}^{-1}(\eta_0^i) = \left(1 + \exp\left(-\eta_0^i\right)\right)^{-1}$. Specifically, $\forall i = 1, \ldots, n$, we sample B realizations from $Z_1^i \sim \mathcal{N}\left(\hat{\eta}_1^i,\ \hat{\sigma}_{1,i}^2\right)$ and $Z_0^i \sim \mathcal{N}\left(\hat{\eta}_0^i,\ \hat{\sigma}_{0,i}^2\right)$ with $\hat{\eta}_1^i$, $\hat{\sigma}_{1,i}$ and $\hat{\eta}_0^i$, $\hat{\sigma}_{0,i}$ being, respectively, the estimated linear predictor and approximate standard error of the continuous and point-mass part of the mixed distribution which can be easily extracted from the output of the `predict` method for '`gamlss`' objects. Then we define $\tilde{Z}_1^i = g_1^{-1}(Z_1^i)$ and $\tilde{Z}_0^i = \left(1 + \exp\left(-Z_0^i\right)\right)^{-1}$ and let $\tilde{Y}_i^* = (1 - \tilde{Z}_0^i)\ \tilde{Z}_1^i$; finally, a Monte-Carlo estimate of the prediction standard error for $\mathbb{E}(\tilde{Y}_i \mid \boldsymbol{x}_1^i, \boldsymbol{v}_i)$ is computed as the standard deviation of $\hat{F}_{\tilde{Y}_i^*}(\cdot)$, the latter being the empirical distribution function of $\tilde{Y}_i^*$ based on the B Monte-Carlo replicates, with B sufficiently large.

5.3 Results

Table 1 reports the estimated effects for the *linear parametric* zero-adjusted GAMLSS denoted by $\mathcal{M}_1$ and the *semi-parametric* model denoted by $\mathcal{M}_2$. The latter model is to be preferred according to the *Bayesian information criterion* (BIC), where we obtain $\text{BIC}(\mathcal{M}_1) = 598,673.1$ and $\text{BIC}(\mathcal{M}_2) = 593,580.7$. According to [13], this provides strong evidence for the semi-parametric model based on a uniform model prior. In Table 1, for the sake of exposition, the coefficients of $x_{14} = \textit{neuron}$ have not been inserted for both models. Table 1 provides the effect estimates for the external stimuli separately for the zero-component as well as the continuous distribution, i.e., how the external stimuli impact on the linear predictor scale on the mean of the zero-component as well as the continuous distribution.

To complement the separate effects provided in Table 1, Figure 9 visualizes the total effects on the neural response estimated by indicating the predictions of the expected spike intensities for a specific neuron in dependence of stimuli and time for the two models. Note that for the sake of representation, some initial and final seconds have been discarded and not included in the plot as they led to point estimates with wide confidence intervals. The plot suggests that the two models mostly imply similar findings. Indeed, stimuli *4. Natural movie 1* and *6. Natural movie 3* exhibit the highest expected spike intensities for both models. Moreover, it is worth noting that, the *spontaneous activity* (*9. None*) has the lowest expected spike intensity for the *linear parametric* model and is among the

Explanatory	**Parametric model**		**Semi-parametric model**	
variable	$\tilde{y} = 0$ ($\boldsymbol{\alpha}$)	$\tilde{y} > 0$ ($\boldsymbol{\beta}_1$)	$\tilde{y} = 0$ ($\boldsymbol{\alpha}$)	$\tilde{y} > 0$ ($\boldsymbol{\beta}_1$)
drift_grat	–	–	–	–
l_spar_n_4d	$-2.0 \times 10^{-1}{}^{***}$ (0.0291)	$1.5 \times 10^{-1}{}^{***}$ (0.0276)	$-7.6 \times 10^{-2}{}^{**}$ (0.0292)	$9.7 \times 10^{-2}{}^{***}$ (0.0271)
l_spar_n_8d	$-2.5 \times 10^{-1}{}^{***}$ (0.0291)	$-7.8 \times 10^{-2}{}^{**}$ (0.0276)	$-4.2 \times 10^{-1}{}^{***}$ (0.0292)	$1.1 \times 10^{-1}{}^{***}$ (0.0272)
nat_mov_1	$-4.9 \times 10^{-1}{}^{***}$ (0.0233)	$2.3 \times 10^{-1}{}^{***}$ (0.0220)	$-8.9 \times 10^{-1}{}^{***}$ (0.0235)	$6.8 \times 10^{-1}{}^{***}$ (0.0216)
nat_mov_2	$-4.6 \times 10^{-1}{}^{***}$ (0.0367)	$3.1 \times 10^{-1}{}^{***}$ (0.0341)	$-8.5 \times 10^{-1}{}^{***}$ (0.0367)	$6.4 \times 10^{-1}{}^{***}$ (0.0336)
nat_mov_3	$-4.8 \times 10^{-1}{}^{***}$ (0.0171)	$3.1 \times 10^{-1}{}^{***}$ (0.0161)	$-7.9 \times 10^{-1}{}^{***}$ (0.0171)	$6.2 \times 10^{-1}{}^{***}$ (0.0158)
nat_scenes	$-4.9 \times 10^{-1}{}^{***}$ (0.0294)	-9.1×10^{-3} (0.0276)	$-9.5 \times 10^{-1}{}^{***}$ (0.0294)	$4.3 \times 10^{-1}{}^{***}$ (0.0272)
stat_gratings	$-3.0 \times 10^{-1}{}^{***}$ (0.0295)	$1.2 \times 10^{-1}{}^{***}$ (0.0278)	$-4.0 \times 10^{-1}{}^{***}$ (0.0296)	$1.6 \times 10^{-1}{}^{***}$ (0.0273)
none	$1.3 \times 10^{-1}{}^{***}$ (0.0233)	$-1.5 \times 10^{-1}{}^{***}$ (0.0223)	$-7.3 \times 10^{-2}{}^{**}$ (0.0232)	$1.1 \times 10^{-1}{}^{***}$ (0.0220)
session A	–	–	–	–
session B	$-9.5 \times 10^{-2}{}^{***}$ (0.0246)	$7.2 \times 10^{-2}{}^{**}$ (0.0231)	2.9×10^{-2} (0.0247)	-4.1×10^{-2} (0.0228)
session C	$-4.8 \times 10^{-2}{}^{*}$ (0.0238)	$-1.0 \times 10^{-1}{}^{***}$ (0.0226)	$-6.0 \times 10^{-2}{}^{*}$ (0.0238)	$-8.8 \times 10^{-2}{}^{***}$ (0.0222)
time	$5.8 \times 10^{-7}{}^{***}$ (1.3×10^{-7})	$-3.3 \times 10^{-6}{}^{***}$ (1.2×10^{-7})	–	–

Table 1: Estimated effects of stimuli, sessions, and time. Figures in parentheses refer to standard errors. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

lowest expected spike intensities for the *semi-parametric* model. These overall results may be compared to the separate effect estimates as shown in Table 1. These indicate for the spontaneous activity, i.e., the none stimulus, that the α coefficient is positive and the β_1 coefficient is negative for the linear parametric model capturing the effects compared to *1. Drifting Gratings*. In contrast, they have opposite signs for the semi-parametric model, which is in line with the none stimulus showing slightly higher expected spike intensities than *1. Drifting Gratings* for the semi-parametric model in Figure 9.

As for $x_{13} = time$ (in seconds), one obtains a positive effect on the probability (π) to have no spike for the *linear parametric* model. In contrast, one obtains a (small) negative effect (given its low magnitude) on the *location* parameter of the Gamma distribution, hence on the spike intensity. Furthermore, the top of Figure 9 reveals the linear behavior of the expected spike intensity through time. In particular, it is decreasing, as $\frac{2\beta-\alpha}{4} = -1.8 \times 10^{-6}$. The bottom of Figure 9

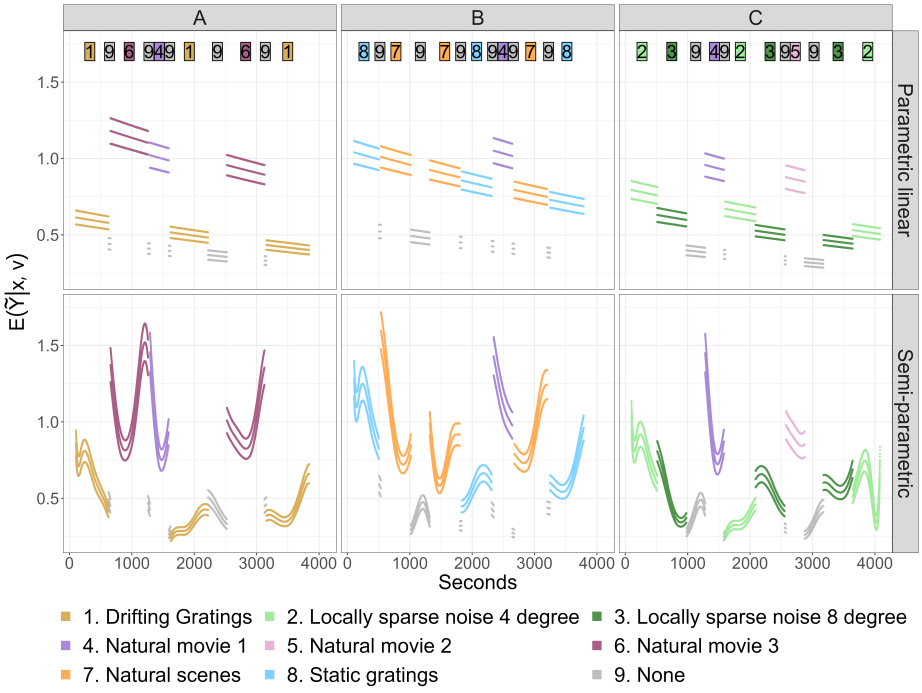

Fig. 9: Expected spike intensity for *neuron* "696326231" per *session* and *stimulus* by *time (in seconds)*. The first row refers to the linear parametric zero-adjusted GAMLSS, while the second to the semi-parametric zero-adjusted GAMLSS. Columns refer to sessions.

shows the higher flexibility of the *semi-parametric* model. Despite the *semi-parametric* model being preferred according to the BIC, the wiggly behavior of the predicted response observed may be a symptom of overfitting, resulting in poor predictability when trying to generalize, a cost that should be taken into account along with interpretability when comparing competing models.

6 Discussion

The present work provides a statistical analysis of effects of external stimuli on neuron response using calcium concentration traces measured for a single mouse across three sessions. The work is structured in three main self-contained sections. After describing the data set, the first empirical part of the work concerns determining the neuron response based on extracting spike trains and is detailed in Section 3. This section is based on a generative model popular in neuroscience, and the outcome of this section is used in the subsequent stages of this work which focus on analyzing neuron response. We introduce a heuristic algorithm which identifies in a data-driven way the actual spike trains based on the time se-

ries of observed calcium traces that is only a proxy of the true neuronal activity. For each neuron and session, the algorithm selects in a data-driven way taking robustness into account the longest time interval in which no stimuli occur, estimates the calcium concentration using a constrained l_0 penalization procedure and derives the amplitude of the spikes corresponding to the spike train values.

The estimated spike trains are then analyzed in an exploratory way in Section 4. Spikes are analyzed statically and dynamically and their intensity is shown to vary with neurons, stimuli, sessions and also with time. Another important insight that arises from this part of the analysis is regarding the shape of the distribution of the spikes. The distribution clearly is positively skewed and of a non-Gaussian form, and exhibits a point mass at zero. This motivates the choice of the zero-adjusted model described in Section 5, where we perform a regression analysis to infer effects of the external stimuli on neuron response.

We use for the regression a zero-adjusted generalized additive model for location, scale and shape (GAMLSS) that allows for greater flexibility in the specification of the conditional distribution chosen for the response variable. This distribution is a mixture of a point-mass and a continuous distribution: a logit model is specified for the zero values and a Gamma distribution is fitted for positive values. Through this model it is possible to investigate the effects of sessions, stimuli and time on the expected spike intensity.

Two separate models have been fitted, a parametric one without additive terms on the time variable and another semi-parametric one where time is expressed through the use of penalized smoothing splines. Interesting insights arise from the estimation results: the stimuli belonging to the *Natural movie* category produce among the highest expected spike intensity, a finding that both models share; moreover, it can be noticed that in absence of stimuli, the expected spike intensity takes the lowest values in the *linear parametric* model and among the lowest values in the *semi-parametric* model. In addition, concerning the *linear parametric* model, time has a positive effect on the probability to have no spike, and in contrast it has a (small) negative effect on the spike intensity.

The specific models estimated can also help investigate the presence of non-linear effects of some covariates on the response variable. Based on the *semi-parametric* model, we were able to infer that time indeed has a non-linear effect. Thus, even though the *linear parametric* model is suitable for interpretation, it may be too restrictive when it comes to describing the effect time has on the expected spike intensity. On the other hand, the wiggly behavior of the response variable highlighted in the *semi-parametric* model is perhaps too complex and may eventually be seen as a sign of overfitting or having the time covariate capture effects of other omitted variables, which may also have led to the smaller value of $\mathrm{BIC}(\mathcal{M}_2)$. A possible alternative model to consider would be to express the response variable as a specific non-linear function of time or add new covariates that would better capture the structure in the data.

The analysis considered only data from a single mouse. The generalizability of results across several mice is thus unclear and assessing this aspect would require additional data and analysis. In particular considering several mice si-

multaneously does not only require to allow for a mouse-specific effect but also requires to align neurons across mice which poses additional challenges in the statistical data analysis. The current results, however, clearly indicate the need for moving beyond GLM regression to account for the characteristics in the distribution of the dependent variable as well as investigate potential non-linear effects of time on neuron response. Using a mixture distribution for the dependent variable required additional inference steps not available in standard software to infer the partial effect of the external stimuli on the mean estimate of the neuron response. We used a Monte-Carlo simulation approach to estimate confidence intervals and visualize the partial effects estimated together with uncertainty bands. This allows to easily compare the partial effects estimated for the different external stimuli across the three sessions allowing also for different effects of time.

As a further future development, it could be worth to investigate in detail the interaction effects between sessions, time and stimuli on the expected spike intensity and thus gain additional insights into the effects of these aspects on activation.

References

[1] Allen Brain Observatory. http://help.brain-map.org/display/observatory/Documentation (2017)

[2] Allen Institute for Brain Science: Allen Brain Observatory. http://observatory.brain-map.org/visualcoding (2016)

[3] Cleveland, W.S.: Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association **74**(368), 829–836 (1979). DOI 10.1080/01621459.1979.10481038

[4] D'Angelo, L., Canale, A.: Efficient posterior sampling for Bayesian Poisson regression. Journal of Computational and Graphical Statistics **32**(3), 917–926 (2023). DOI 10.1080/10618600.2022.2123337

[5] de Vries, S.E.J., Lecoq, J.A., Buice, M.A., Groblewski, P.A., Ocker, G.K., Oliver, M., Feng, D., Cain, N., Ledochowitsch, P., Millman, D., Roll, K., Garrett, M., Keenan, T., Kuan, L., Mihalas, S., Olsen, S., Thompson, C., Wakeman, W., Waters, J., Williams, D., Barber, C., Berbesque, N., Blanchard, B., Bowles, N., Caldejon, S.D., Casal, L., Cho, A., Cross, S., Dang, C., Dolbeare, T., Edwards, M., Galbraith, J., Gaudreault, N., Gilbert, T.L., Griffin, F., Hargrave, P., Howard, R., Huang, L., Jewell, S., Keller, N., Knoblich, U., Larkin, J.D., Larsen, R., Lau, C., Lee, E., Lee, F., Leon, A., Li, L., Long, F., Luviano, J., Mace, K., Nguyen, T., Perkins, J., Robertson, M., Seid, S., Shea-Brown, E., Shi, J., Sjoquist, N., Slaughterbeck, C., Sullivan, D., Valenza, R., White, C., Williford, A., Witten, D.M., Zhuang, J., Zeng, H., Farrell, C., Ng, L., Bernard, A., Phillips, J.W., Reid, R.C., Koch, C.: A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. Nature Neuroscience **23**(1), 138–151 (2020). DOI 10.1038/s41593-019-0550-9

[6] Enea, M., Stasinopoulos, M., Rigby, B., Hossain, A.: gamlss.inf: Fitting Mixed (Inflated and Adjusted) Distributions (2019). URL https://CRAN.R-project.org/package=gamlss.inf. R package version 1.0-1

[7] Hastie, T.J., Tibshirani, R.J.: Generalized Additive Models. Chapman and Hall, London, United Kingdom (1990). DOI 10.1201/9780203753781

[8] Jewell, S., Hocking, T.D.: FastLZeroSpikeInference: Fast Nonconvex Deconvolution of Calcium Imaging Data (2018). URL https://CRAN.R-project.org/package=FastLZeroSpikeInference. R package version 2018.12.10

[9] Jewell, S.W., Hocking, T.D., Fearnhead, P., Witten, D.M.: Fast nonconvex deconvolution of calcium imaging data. Biostatistics **21**(4), 709–726 (2019). DOI 10.1093/biostatistics/kxy083

[10] Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., Jupyter development team: Jupyter notebooks – a publishing format for reproducible computational workflows. In: F. Loizides, B. Schmidt (eds.) Positioning and Power in Academic Pub-

lishing: Players, Agents and Agendas, pp. 87–90. IOS Press (2016). DOI 10.3233/978-1-61499-649-1-87

[11] Nelder, J.A., Wedderburn, R.W.M.: Generalized linear models. Journal of the Royal Statistical Society. Series A (General) **135**(3), 370–384 (1972). DOI 10.2307/2344614

[12] R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2022). URL https://www.R-project.org/

[13] Raftery, A.: Bayesian model selection in social research. Sociological Methodology **25**, 111–163 (1995). DOI 10.2307/271063

[14] Rigby, R.A., Stasinopoulos, D.M.: Generalized additive models for location, scale and shape. Journal of the Royal Statistical Society. Series C (Applied Statistics) **54**(3), 507–554 (2005). DOI 10.1111/j.1467-9876.2005.00510.x

[15] Stasinopoulos, D.M., Rigby, R.A.: Generalized additive models for location scale and shape (GAMLSS) in R. Journal of Statistical Software **23**(7), 1–46 (2007). DOI 10.18637/jss.v023.i07

[16] Vogelstein, J.T., Packer, A.M., Machado, T.A., Sippy, T., Babadi, B., Yuste, R., Paninski, L.: Fast nonnegative deconvolution for spike train inference from population calcium imaging. Journal of Neurophysiology **104**(6), 3691–3704 (2010). DOI 10.1152/jn.01073.2009