Received: 18 February 2021 | Accepted: 13 January 2022

DOI: 10.1111/biom.13626

BIOMETRIC PRACTICE

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

Bayesian nonparametric analysis for the detection of spikes in noisy calcium imaging data

Laura D'Angelo[1] | **Antonio Canale**[2] | **Zhaoxia Yu**[3] | **Michele Guindani**[3]

[1]Department of Economics, Management and Statistics, University of Milano-Bicocca, Milan, Italy

[2]Department of Statistical Sciences, University of Padova, Padova, Italy

[3]Department of Statistics, University of California, Irvine, Irvine, California, USA

Correspondence
Laura D'Angelo, Department of Economics, Management and Statistics, University of Milano-Bicocca, 20126 Milan, Italy.
Email: laura.dangelo@unimib.it

[Correction added on May 16, after first online publication: CRUI-CARE funding statement has been added.]

Abstract

Recent advancements in miniaturized fluorescence microscopy have made it possible to investigate neuronal responses to external stimuli in awake behaving animals through the analysis of intracellular calcium signals. An ongoing challenge is deconvolving the temporal signals to extract the spike trains from the noisy calcium signals' time series. In this article, we propose a nested Bayesian finite mixture specification that allows the estimation of spiking activity and, simultaneously, reconstructing the distributions of the calcium transient spikes' amplitudes under different experimental conditions. The proposed model leverages two nested layers of random discrete mixture priors to borrow information between experiments and discover similarities in the distributional patterns of neuronal responses to different stimuli. Furthermore, the spikes' intensity values are also clustered within and between experimental conditions to determine the existence of common (recurring) response amplitudes. Simulation studies and the analysis of a dataset from the Allen Brain Observatory show the effectiveness of the method in clustering and detecting neuronal activities.

KEYWORDS

Dirichlet process, mixture of finite mixtures, model-based clustering, nested Dirichlet process, spike and slab

1 | INTRODUCTION

In recent years, calcium imaging has become a popular technique to measure the neuronal activity in awake, freely moving, and behaving animals over time. Due to the development of miniaturized and flexible microendoscopes for fluorescence microscopy, this technique has enabled the study of how individual neurons and neuronal networks encode external stimuli and cognitive processes (Li *et al.*, 2015; Nakajima and Schmitt, 2020). Calcium ions generate intracellular signals that determine a large variety of functions in all neurons (Grienberger and Konnerth, 2012). The mechanism at the basis of calcium imaging is a physiological process of the cells: when a neuron fires, calcium floods the cell and produces a transient spike in its concentration. By using genetically encoded calcium indicators, which are fluorescent molecules that react when binding to the calcium ions, it is possible to optically measure the level of calcium ions by analyzing the observed fluorescence trace. The outcome of this technique is a movie of time-varying fluorescence intensities, from which the spike trains of the observable neurons are often extracted through a complex preprocessing phase. In general, this phase is meant to deal with two issues: identifying the spatial location of each neuron in the optical field and deconvolving the temporal signals to extract their spike trains. Several

methods are employed to extract the fluorescence traces, for example, by using independent component analysis (Mukamel *et al.*, 2009; Dombeck *et al.*, 2010), nonnegative deconvolution (Vogelstein *et al.*, 2010), and nonnegative matrix factorization (Maruyama *et al.*, 2014).

The resulting processed data consist of a fluorescent calcium trace for each observable neuron in the targeted area (see Figure 3 for an example). Although the observed fluorescence trace can be analyzed directly (Shen *et al.*, 2021), it is a proxy of the underlying cellular activity and the information relevant to many studies often requires the precise spike times and the intracellular calcium concentration of the observable neurons when the animal is subjected to external stimuli (Vogelstein *et al.*, 2009). Extracting the neuronal activity from these series is not trivial: the calcium imaging technology has several limitations, including the presence of measurement noise, the nonlinearity between fluorescence transient and calcium concentration, and the slow decay of the fluorescence trace compared to the underlying neuronal activity (Rose *et al.*, 2014; Friedrich *et al.*, 2017; Dana *et al.*, 2019). Moreover, the large scale of the time series introduces additional complexity to the analysis. Therefore, a precise estimation of the spike times and amplitudes is a fundamental step toward the understanding of the neurons' behavior.

As a motivating application, here we consider a publicly available data set from the Allen Brain Observatory (Allen Institute MindScope Program, 2016; de Vries *et al.*, 2020) of calcium imaging data obtained through two-photon microscopy in behaving mice. This study is an extended in vivo survey of physiological activity in the mouse visual cortex in response to a range of visual stimuli (Allen Brain Observatory, 2017). Each mouse is placed in front of a screen where different types of visual stimuli are shown, while the mouse's neuronal activity is recorded. The stimuli vary from simple synthetic images such as locally sparse noise or static gratings, to complex natural scenes and movies. The goal of the study is to investigate how neurons at different depths in the visual areas respond to stimuli of different complexity. Specifically, each neuron in the visual cortex can be characterized by its *receptive field*, that is, the features of the visual stimulus that trigger the signaling of that neuron. Hence, it is of critical interest to devise methods that allow inferring how the neuronal response varies under the different types of visual stimuli. We expect that the neuronal activity will vary across all the experimental settings and that some variations in its intensity will be observed based on the stimulus.

Several approaches have been proposed to accurately and efficiently estimate the neuronal activity in calcium imaging data from single neurons. For example, Friedrich and Paninski (2016) and Friedrich *et al.* (2017) have proposed an online algorithm based on a lasso penalty to enforce sparsity of signal detection. Jewell and Witten (2018) and Jewell *et al.* (2019) have proposed using an L_0 penalty in lieu of the L_1 penalization and an efficient algorithm to identify the presence or absence of spikes. In a Bayesian framework, Pnevmatikakis *et al.* (2013) have proposed to conduct inference on spike trains by estimating posterior probabilities of a latent binary indicator of spike presence at each time point. However, the model in Pnevmatikakis *et al.* (2013) does not explicitly assume sparsity of the spikes. Moreover, it is expected the rate and the distribution of spikes to be stimulus dependent (Brenner *et al.*, 2002), but none of the previous approaches accounts for the heterogeneity of spikes' behaviors as a function of the stimulus. As Figure 3 clearly shows for the Allen Brain Observatory data, the spikes' intensities vary greatly according to the type of stimulus; see also the discussion in Shibue and Komaki (2020) where they employ a marked point processes for the deconvolution of calcium imaging data.

In this article, we introduce a coherent nested Bayesian finite mixture model that allows the estimation of the spiking activity of each neuron—which could be seen as a first step for the analysis of larger brain activity combining multiple neurons in a region. In addition, our model *simultaneously* allows us to reconstruct the distributions of spikes under various experimental conditions; for example, in response to different types of visual stimuli in the Allen Brain Observatory dataset. More specifically, our modeling framework estimates and clusters the distributions of the calcium transient spikes' amplitudes via a nested formulation of mixture of finite mixtures (Miller and Harrison, 2018; Argiento and De Iorio, 2019) and, in particular, exploiting the generalized mixture of finite mixtures (gMFM) prior recently proposed by Frühwirth-Schnatter *et al.* (2021).

The proposed model further adopts the use of a common atom specification as in Denti *et al.* (2021) for estimating the distribution of the spikes' amplitudes under each experimental condition. The proposed common atom gMFM has several advantages with respect to typical Bayesian nonparametric models for nested data. With respect to models based on Dirichlet process priors, the gMFM provides increased flexibility to estimate partitions characterized either by many, well-balanced, clusters or by a small set of large clusters. The common atom model allows us to obtain nested inference on densities without incurring the degeneracy issues pointed out by Camerlenghi *et al.* (2019) for the widely used nested Dirichlet process of Rodríguez *et al.* (2008). At the same time, the common atom formulation still leverages two nested layers of random discrete mixture priors to borrow information between experiments and to identify similarities in the distributional patterns of the neuronal responses to different stimuli. In addition, differently than in the nested Dirichlet process, the common atom model also allows us

to cluster the inferred spikes' intensity values both within and between experimental conditions, so to infer common (recurring) response amplitudes. Finally, we allow our model to enforce sparsity of neuron firing over time by assuming a spike-and-slab prior specification on the marginal distribution of the amplitudes.

2 | BAYESIAN MIXTURE MODEL FOR CALCIUM IMAGING DATA

2.1 | Model and prior specification

The observed fluorescence is often considered as a noisy realization of the underlying true calcium concentration. To model a neuron's activity, we adopt a popular model in the neuroscience literature, where the decay in fluorescence is modeled through an autoregressive process and the spikes are modeled as jumps in correspondence to the neuron's firing events (Vogelstein *et al.*, 2010). Denoting with y_t the observed fluorescence trace of a neuron and with c_t the underlying calcium concentration, for $t = 1, \dots, T$, one can assume

$$\begin{aligned} y_t &= b + c_t + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0, \sigma^2), \\ c_t &= \gamma c_{t-1} + A_t + w_t, \quad w_t \sim \mathrm{N}(0, \tau^2), \end{aligned} \tag{1}$$

where b models the baseline level of the observed trace and ϵ_t is a measurement error. In the absence of neuronal activity, the true calcium concentration c_t is considered to be centered around zero. The parameter A_t captures the neuronal activity: in the absence of a spike ($A_t = 0$), the calcium level follows an AR(1) process controlled by the parameter γ ; when a spike occurs, the concentration increases instantaneously with the spike amplitude $A_t > 0$.

We are interested in characterizing the neuronal activity under different experimental conditions. For each time point $t = 1, \dots, T$, let g_t be a discrete categorical variable, taking values in $\{1, \dots, J\}$, where J is the number of distinct experimental settings, so that $g_t = j$ indicates that the neuronal activity at time t is observed under condition j. The experimental conditions are often designed to capture variations in neuronal activity with respect to a baseline process, which may represent a "typical" brain process. For example, in the Allen Observatory data, the interest is to investigate visually evoked functional responses of neurons in the mouse's visual cortex. Therefore, some neurons associated with visual decoding should be expected to activate in all conditions. It is then of interest to study not only *if* but also *how* the neurons differentially respond to the presentation of a variety of visual stimuli.

In this paper, we propose a hierarchical Bayesian approach to investigate similarities and differences in the distribution of spikes over time and conditions. In order to borrow information across different experimental conditions, one option is to fit a parametric hierarchical random effect model and obtain a post-MCMC clustering of the estimated spikes A_t by grouping together those spikes with similar magnitudes. This approach has several limitations: on the one hand, the distribution of the random effects is constrained into a specific parametric form; on the other hand, the clustering of, say, the posterior mean estimates of the parameters A_ts do not allow the model to fully describe stimulus-specific distributional differences and to take into account the posterior uncertainty in the spikes.

In order to allow flexible modeling of distributions and to describe the heterogeneity of distributional features, we assume a nested Bayesian finite mixture specification. More specifically, we rewrite (1) as

$$y_t \mid b, \gamma, c_{t-1}, A_t, \sigma^2, \tau^2 \sim \mathrm{N}(\mathrm{b} + \gamma\, \mathrm{c_{t-1}} + \mathrm{A_t}, \sigma^2 + \tau^2),$$

and we assume that the spikes A_t are from stimulus-specific distributions, that is, $(A_t \mid g_t = j, G_j) \sim G_j$, $j = 1, \dots, J$, to account for the observed variety of neuronal activity under different experiment settings. We further allow clustering the distributions across conditions, in order to capture similar patterns of neuronal activity. Indeed, one may typically expect $K < J$ distributional clusters. For example, a neuron may respond to general visual stimulation and not specifically to the type of stimulus considered. More specifically, we assume the following generalized mixture of finite mixtures structure:

$$G_1, \dots, G_J \mid Q \sim Q, \qquad Q = \sum_{k=1}^{K} \pi_k \delta_{G_k^*} \tag{2}$$

where $\pi_1, \dots, \pi_K \mid K \sim \mathrm{Dirichlet_K}(\alpha/\mathrm{K}, \dots \alpha/\mathrm{K}), \alpha > 0$, and $G_1^*, \dots, G_K^*$ are a set of cluster-defining distributions, obtained as realizations of an underlying random probability measure, specified further below. Equation (2) implies that the G_js, $j = 1, \dots, J$, have a positive probability of clustering together, thereby giving rise to *distributional clusters*. In practice, the number of mixture components, K, is typically larger than the number of clusters, K_+, and some of the atoms G_k^* are not assigned to any of the G_js (empty components). The prior on the number of mixture components K is a translated beta-negative-binomial distribution as in Frühwirth-Schnatter *et al.* (2021). Including a prior $p(K)$ leads to both K_+ and K being random a priori. Finally, the distributional atoms G_k^*, $k = 1, \dots, K$ are also obtained as a realization from an underlying generalized mixture of

finite mixtures,

$$G_k^* = \sum_{l=1}^{L} \omega_{l,k} \delta_{A_l^*} \tag{3}$$

with $\omega_{1,k}, \ldots, \omega_{L,k} \mid L \sim \text{Dirichlet}_L(\beta/L)$, for some positive real number $\beta > 0$. The set of atoms A_l^* is common across all distributions $G_1^*, \ldots, G_K^*$, and they are obtained as independently and identically distributed draws from a centering measure, $A_l^* \sim G_0(A_l^*)$. Therefore, Equation (3) defines a clustering of the inferred spike intensities both within a given condition (i.e., for fixed G_k^*) and across conditions (i.e., across the G_k^*s; hence, across the G_js). In the following, we adopt common terminology in the literature on nested Bayesian nonparametric priors and indicate the clustering induced on the A_t through the proposed two-layers prior as *observational clustering*. The nested gMFM formulation requires the specification of a prior on the number of components that specify the lower level distributional atoms $G_k^*, L \sim p(L)$. Once again, some of the components may be empty.

We enforce sparsity in the detection of the spikes by modeling the base measure G_0 for the parameters A_l^* with a spike-and-slab specification (Mitchell and Beauchamp, 1988), which is a convex mixture between a Dirac mass at zero—representing the absence of neuronal response—and a diffuse density on the positive real numbers—representing the intensity of the neuronal response. More specifically, we assume

$$G_0 = (1 - p)\, \delta_0 + p\, \text{Ga}(\text{h}_{\text{A1}}, \text{h}_{\text{A2}}), \tag{4}$$

where the slab is a gamma distribution, Ga(a, b) with mean a/b and variance a/b^2. The choice of a gamma distribution in (4) is particularly relevant for sparsity-inducing purposes, as the gamma density belongs to the set of moment nonlocal prior densities, as defined by Johnson and Rossell (2010). Therefore, a negligible probability density is assigned to values in a neighborhood of zero, thus inducing a clear separation between the baseline neuronal activity and the neuronal responses. In particular, the higher the shape parameter h_{A1}, the larger is the separation. We assume a $\text{Beta}(\text{h}_{1\text{p}}, \text{h}_{2\text{p}})$ prior for the proportion of spikes p with h_{1p} much smaller than h_{2p} in order to favor sparsity of detections. A spike-and-Gamma model has also been used for the analysis of calcium imaging data by Wei *et al.* (2019), although in a two-stage setup for modeling the distribution of the (already deconvolved) estimated spikes.

The proposed formulation can be seen as a special case of *inner* spike-and-slab nonparametric priors, following a terminology introduced by Canale *et al.* (2017, 2021). In the following, we will refer to the proposed specification as a finite common atom model (fCAM).

The Bayesian model elicitation is completed by assuming conjugate priors for the underlying calcium level concentration parameters, that is, the baseline calcium level b, and the variances σ^2 and τ^2. Specifically, the following conjugate prior distributions are assumed:

$$c_0 \sim \text{N}(0, \text{C}_0), \quad \text{b} \sim \text{N}(\text{b}_0, \text{B}_0)$$

$$1/\sigma^2 \sim \text{Ga}(\text{h}_{1\sigma}, \text{h}_{2\sigma}), \quad 1/\tau^2 \sim \text{Ga}(\text{h}_{1\tau}, \text{h}_{2\tau}).$$

Finally, under the assumption that the process is stationary with a positive correlation between the calcium level at consecutive times, we constrain $\gamma \in (0, 1)$ and let $\gamma \sim \text{Beta}(\text{h}_{1\gamma}, \text{h}_{2\gamma})$, a priori.

Consistently with our aim of modeling the activity of a single neuron, each unknown parameter is neuron specific. However, the hyperparameters of the priors above can be specified at the population level, especially in the case of independent analyses for multiple neurons.

2.2 | Posterior inference

For computational purposes, it is often convenient to rewrite the likelihood for an observation y_t under condition $g_t = j$ by introducing two latent cluster allocation variables, $S_j = S_{g_t}$ and M_t, indicating the distributional cluster for the group j and the observational cluster for y_t, respectively.

Given K and $\{\pi_k\}_{k=1}^K$, the distributional allocation variable $S_j \sim \text{Multin}_\text{K}(\boldsymbol{\pi})$, where Multin_K denotes a multinomial distribution with K categories and event probabilities $\boldsymbol{\pi}$. Similarly, given L and $\{\omega_l\}_{l=1}^L$, the observational allocation variable $M_t \sim \text{Multin}_\text{L}(\boldsymbol{\omega})$. For notation simplicity, we have indicated the parameter vectors using bold letters in the equations above. Therefore, conditionally on the other model parameters, the joint distribution of the observed data and the latent cluster allocations can be written as

$$f(\boldsymbol{y}, \boldsymbol{M}, \boldsymbol{S} \mid \boldsymbol{\pi}, \boldsymbol{\omega}, \boldsymbol{A}^*) = \prod_{j=1}^{J} \pi_{S_j} \prod_{t:\, g_t = j} \omega_{M_t, S_j}\, p(y_t \mid A_{M_t}^*),$$

which facilitates posterior inference.

More specifically, posterior inference for the proposed fCAM can be carried out quite straightforwardly by means of Markov chain Monte Carlo (MCMC) techniques. The sampling of the latent calcium level c_t uses an iterative approach based on the Kalman filter and a forward filtering backward sampling algorithm (Prado and West, 2010). Full conditional posteriors for b, p, σ^2, and τ^2 are

available in closed form thus leading to straightforward Gibbs sampling steps. For the autoregressive parameter γ, we use a Metropolis–Hastings within the Gibbs step. The sampling of A_t exploits a combination of the nested slice sampler of Denti *et al.* (2021) and the telescoping sampler of Frühwirth-Schnatter *et al.* (2021). A detailed description of the latter step is reported in the Appendix. Here, we just present a schematic description of the MCMC steps:

(1) Sample the calcium level c_t, for $t = 0, \dots, T$, using a forward filtering backward sampling:
   (a) Run Kalman filter: set $a_0 = m_0 = 0$, $R_0 = C_0 = \text{var}(c_0)$. For $t = 1, \dots, T$ let

$$a_t = \gamma\, m_{t-1} + A_t$$

$$R_t = \gamma^2\, C_{t-1} + \tau^2.$$

   Compute the filtering distribution's parameters, m_t and C_t, for $t = 1, \dots, T$, where

$$m_t = a_t + R_t\,(R_t + \sigma^2)^{-1}\,(y_t - b - a_t)$$

$$C_t = R_t - R_t^2\,(R_t + \sigma^2)^{-1}.$$

   (b) Draw $c_T \sim \text{N}(m_T, C_T)$;
   (c) For $t = T-1, \dots, 0$, draw $c_t \sim \text{N}(h_t, H_t)$, with

$$h_t = m_t + \gamma\, C_t\, R_{t+1}^{-1}(c_{t+1} - a_{t+1})$$

$$H_t = C_t - \gamma^2\, C_t^2\, R_{t+1}^{-1}.$$

(2) Sample a new value for the baseline level b:

$$b \sim \text{N}\left(\frac{b_0}{B_0} + \frac{1}{\sigma^2}\sum_{t=1}^{T}(y_t - c_t), \sqrt{\frac{1}{B_0} + \frac{T}{\sigma^2}}\right).$$

(3) Sample the variance on the output equation σ^2 and the variance on the state equation τ^2:

$$1/\sigma^2 \sim \text{Ga}\left(h_{1\sigma} + \frac{T}{2},\, h_{2\sigma} + \frac{1}{2}\sum_{t=1}^{T}(y_t - c_t - b)^2\right)$$

$$1/\tau^2 \sim \text{Ga}\left(h_{1\tau} + \frac{T}{2},\, h_{2\tau} + \frac{1}{2}\sum_{t=1}^{T}(c_t - \gamma\, c_{t-1} - A_t)^2\right).$$

(4) Update the autoregressive parameter γ using a Metropolis–Hastings step.

(5) Update the parameter p of the spike-and-slab base measure from

$$p \sim \text{Beta}\,(h_{1p} + T - n_0,\, h_{2p} + n_0),$$

   where n_0 is the number of y_t assigned to the spike component.

(6) Update the cluster allocations variables $\boldsymbol{S}$ and $\boldsymbol{M}$, the number of mixture components K and L, and the cluster parameters $\boldsymbol{A}^*$ using the nested telescoping sampling for the finite common atom model reported in the Appendix.

3 | SIMULATION STUDY

The performances of the proposed method are assessed through a simulation study. The purpose of this section is twofold, namely to assess both the ability to correctly identify the spike times and the accuracy of the inferred clustering structure.

We simulated synthetic data exhibiting a baseline level and a number of spikes representing the effect of the response of a neuron to a stimulus, thus mimicking the characteristics of real series of calcium imaging following the structure of model (1).

Specifically, we first divided the time frame into J hypothetical experimental conditions of equal length, with J varying in the different scenarios described below. Consistent with our motivating assumption that the neuronal response depends on the type of stimulus, each experimental condition is assumed to belong to one of the K distributional clusters. Then, for each experimental condition, we generated the neuronal activity: first, we generated the presence or absence of a neuron response uniformly in time, where the spike probability can vary across groups. Then, conditionally on the obtained activations, we generated some additional spikes in a short subsequent interval, so that it is very likely to observe close or even successive spikes.

In this way, the data mimic a real calcium imaging time series. Moreover, we are able to conduct a careful assessment of the ability of the model to distinguish the presence of a single high spike versus the convolution of several spikes at consecutive times. Finally, the values A_t, conditionally on their distributional cluster, are generated from one of the finite sets of spike amplitudes described below.

We simulated 50 independent datasets for each of the three scenarios described henceforth. In Scenario 1, we assumed $J = 6$ experimental conditions, generated from $K = 4$ distributional clusters.

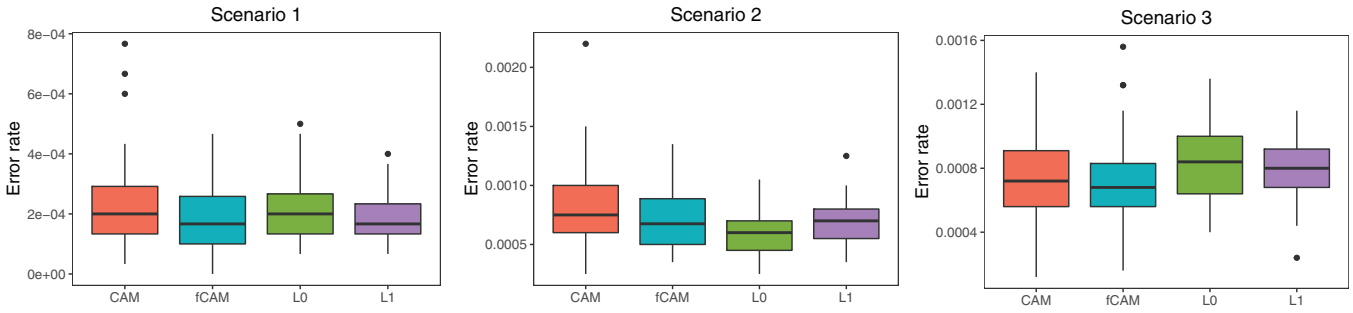

FIGURE 1 Distribution of the misclassification error rate in the simulation study for the four considered methods: CAM, fCAM, and the methods of Jewell *et al.* (2019) "L0" and Friedrich *et al.* (2017) "L1." This figure appears in color in the electronic version of this article, and any mention of color refers to that version

The spike amplitudes in the distributional clusters are (0.35, 0.89, 1.15, 1.80, 2.20), (0.65, 0.89, 1.40, 1.80), (0.35, 0.65, 1.15), and (0.35, 0.89, 1.60).

Scenario 2 assumes $J = 4$ experimental conditions and $K = 3$ distributional clusters with spike amplitudes equal to (0.3, 0.5, 0.7, 0.9, 1.1, 1.5), (0.3, 0.9, 1.5, 1.8), and (0.5, 0.9, 1.5). Finally, Scenario 3 sets $J = 5$ and $K = 3$ with the spike amplitudes in the distributional clusters being (0.3, 0.5, 0.7, 0.9, 1.1), (0.3, 0.9, 1.1, 1.3), and (0.7, 0.9, 1.3). While in Scenario 1, the amplitudes of the spikes are quite large, spaced apart, and with the corresponding distributional clusters well distinct, in Scenario 3 the spike amplitudes are more homogeneous and more clustered in time. Scenario 2 represents an in-between situation. Hence, from the first to the last scenario, we are assuming an increasing degree of complexity. The R script generating these synthetic datasets is described in the Supporting Information.

The results attained by the proposed fCAM are compared to those obtained exploiting the common atom model (CAM) of Denti *et al.* (2021)—which provides a benchmark for the clustering of the spikes and the stimulus-specific distributions—and to those obtained with the L_0 penalization method of Jewell *et al.* (2019) and the L_1 penalization method of Friedrich *et al.* (2017), which provide a benchmark for the task of spikes' detection. For the latter two methods, we have assumed complete knowledge of the autoregressive constant controlling the rate of the calcium decay, since we found that the results were quite sensitive to this estimate.

To assess the sensitivity of the proposed fCAM to the prior specification, we repeated the numerical experiment for different values of the hyperparameters h_{A1} and h_{A2} in (4). In particular, the shape parameter h_{A1} was supposed to play a key role in the detection of spikes. Keeping fixed the ratio h_{A1}/h_{A2}, the parameters were set equal to 3, 4, 6, and 8: a small value implies, a priori, less separation between zero and the distribution of the positive spikes while a large value corresponds to the opposite effect.

Focusing on the classification of each time point as a spike or not, Figure 1 summarizes the misclassification rate for all competing methods under the three scenarios. The results of the 50 replications are summarized using boxplots. For our fCAM, we report only the results obtained with $h_{A1} = h_{A2} = 8$ as those obtained for the other choices are essentially equivalent. The rates are small in absolute value and broadly comparable across the different methods, thus confirming that all the competing models are effective in detecting the spikes.

However, the proposed fCAM not only enables the detection of spikes but also allows us to conduct inference on the clustering structure. Therefore, we report on its ability to identify the clustering structure. Figure 2 reports the adjusted Rand index (Rand, 1971; Hubert and Arabie, 1985) computed on both the observational and the distributional clusters for $h_{A1} = h_{A2} = 8$ (results for other settings are similar). Values of the adjusted Rand index close to 1 denote that the identified structure resembles the true clustering. While for the observational clusters the results are broadly comparable, for the distributional clusters the performance of the proposed fCAM is uniformly superior. In addition, the variability of the results generally appears to be drastically smaller for the fCAM, thus providing evidence of greater efficiency. This is consistent with the results of Frühwirth-Schnatter *et al.* (2021) where the generalized mixture of finite mixtures is compared to a standard Dirichlet process mixture model.

From a computational point of view, the proposed algorithm is clearly more demanding than the optimization methods of Jewell *et al.* (2019) and Friedrich *et al.* (2017). However, the computing time is comparable to the slice sampler adopted for the CAM, and in general a full run requires just few minutes on a Linux machine with an i7-7700HQ 3.8 GHz Intel processor, 8 GB RAM, running R 4.1.0. For example, for a calcium trace of length 50,000, the computing time of the proposed method is around 2 min. Indeed, our experience suggests that the main factor affecting the computing time is the length of the series.

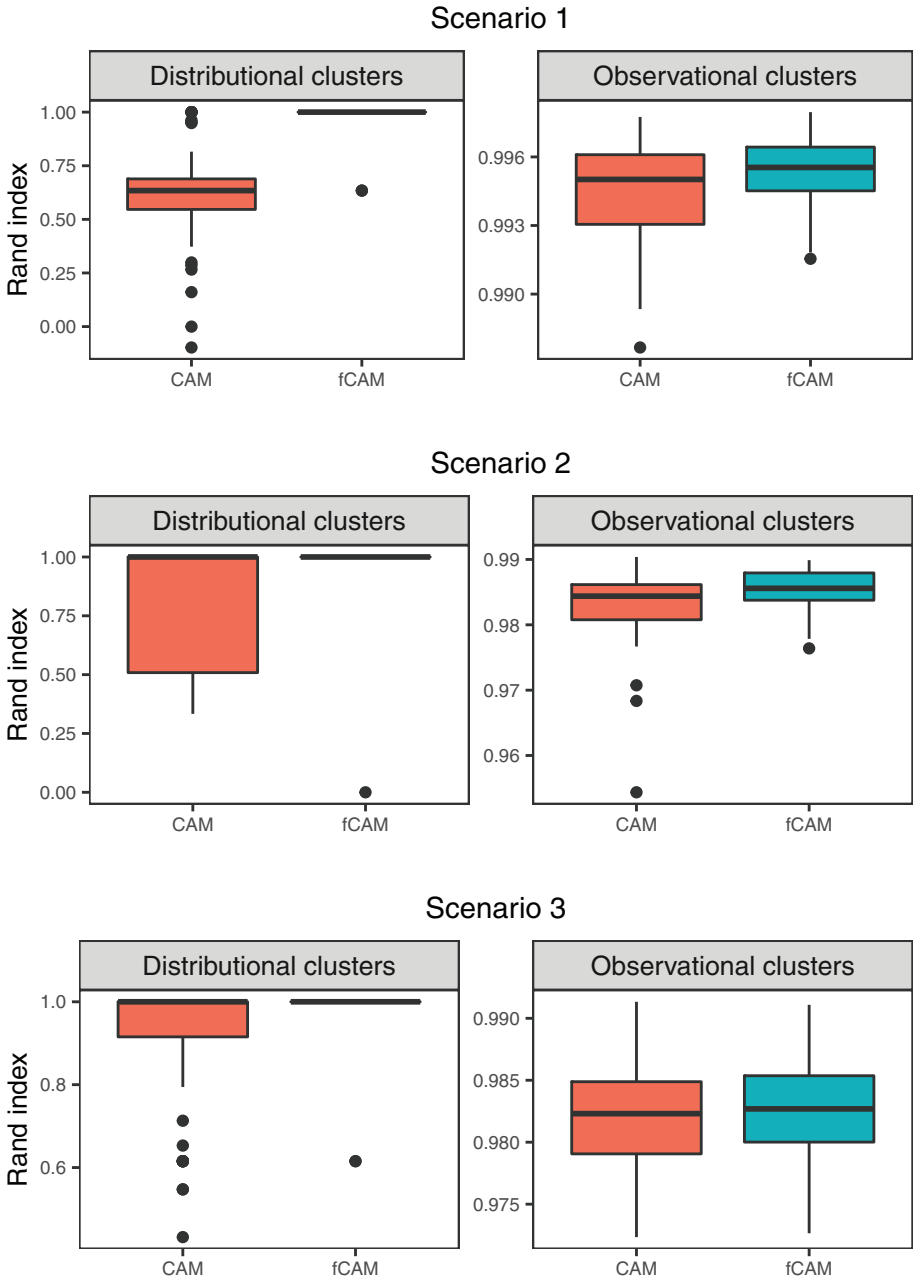

FIGURE 2 Distribution of the adjusted Rand index on the distributional and observational clusters, computed on the 50 simulations for the three scenarios of the simulated data. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

In general, in the analysis of spike activity, we expect the number of clusters to be small and—in particular—much smaller than the number of observations.

4 | ALLEN BRAIN OBSERVATORY DATA ANALYSIS

We now revert to the analysis of the data from the Allen Brain Observatory (Allen Institute MindScope Program, 2016). The data comprise the dF/F-transformed fluorescence trace for a cell during session-B of the experiment. This session comprises three types of visual stimuli (static gratings, natural scene, and natural movie) in addition to some period of spontaneous activity (absence of visual stimuli). Since the data are recorded at a frequency of 30 Hz, the resulting series consists of 113,865 time points for a total of 63.2 min. We focus our analysis on a neuron located in the primary visual area, at an imaging depth equal to 350 microns (cell id 508596945). Additional analyses for other neurons are reported in the Supporting Information.

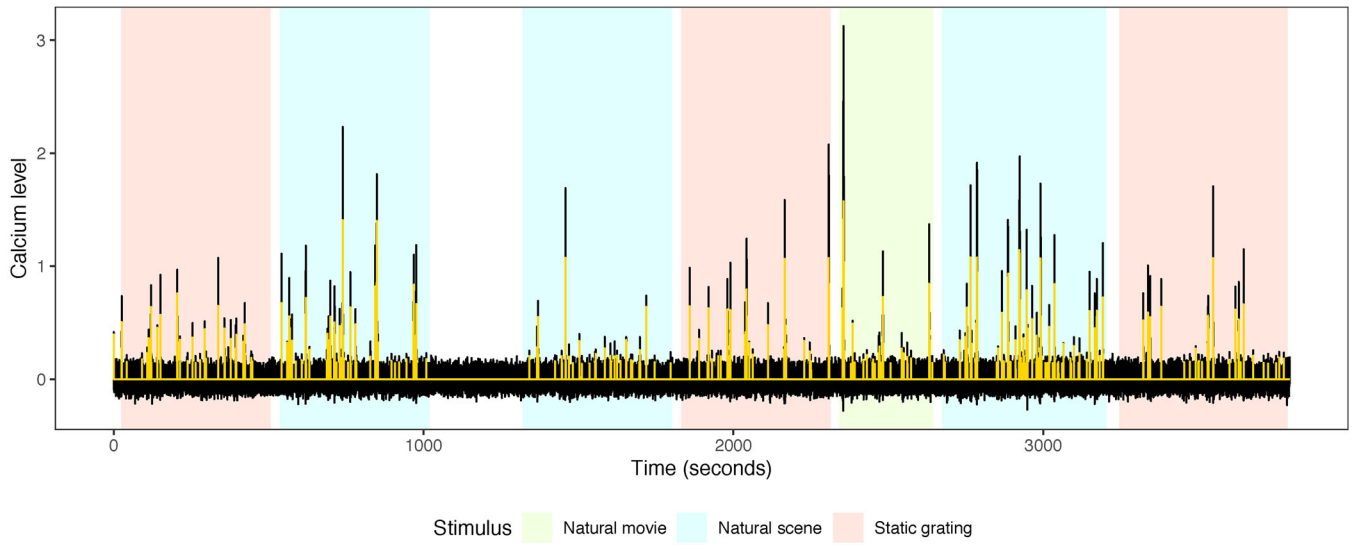

FIGURE 3 Observed fluorescence trace y_t from the Allen Brain Observatory data (dark line), and visual stimulus to which the mouse is exposed (shaded areas). The yellow line represents the estimated neuronal activity

The observed fluorescence trace is shown with a continuous black line in Figure 3. Different shaded backgrounds indicate the types of visual stimuli. Using the notation introduced in the previous sections, $J = 4$ with $j = 1, 2, 3$ corresponding to static grating, natural scene, and natural movie, respectively, and $j = 4$ indicating no stimulus presence.

We ran the MCMC algorithm of Section 2.2 using the same prior specification of Section 2.1 for 15,000 iterations discarding the first 7000 iterations as burn-in and keeping one iteration every four to improve mixing. Visual inspection of the traceplots and Geweke diagnostics showed no issues with convergence. The superimposed light line in Figure 3 represents the estimated neuronal activity in terms of the inferred amplitude A_t, that is removing the measurement errors and the result of the accumulation of calcium from the previous spikes. Here and henceforth, we identified the presence of a spike if the posterior probability of a spike at time t, say PPS_t, estimated by the proportion of nonzero A_ts over all MCMC iterations, was greater than $\kappa = 75.5\%$. This threshold allows us to control the (estimated) Bayesian false discovery rate at the preset value 0.05, that is κ solves the equation $\mathrm{FDR}(\kappa) = \frac{\sum_{t=1}^{T}(1-\mathrm{PPS}_t)I_{(\mathrm{PPS}_t>\kappa)}}{\sum_{t=1}^{T} I_{(\mathrm{PPS}_t>\kappa)}} = 0.05$. For more details, we refer to Newton *et al.* (2004) and Müller *et al.* (2007); see also Sun *et al.* (2015) for a discussion with dependent hypotheses.

As already mentioned in the Introduction, in calcium imaging it is of interest studying the distribution of the spikes in response to each experimental stimulus, and identifying similarities and differences in these distributions across stimuli.

We start by investigating the presence of similarities in the neuronal response to different types of visual stimuli. This corresponds to analyzing the clustering of the spike distributions induced by the proposed fCAM. The model clusters together the groups corresponding to the natural scene and natural movie stimuli with high posterior probability, while the static grating stimulus and the absence of stimuli are assigned to two separate distributional clusters. In other terms, the neuron appears to show similar neuronal responses in the natural scene and natural movie stimuli whereas the responses appear distinctly different under the other two conditions.

To understand whether and how the neuronal response depends on the type of stimulus, we estimated the spike amplitude distribution for each of the four types of stimuli. Figure 4 shows the histograms of posterior means of the nonzero spike amplitudes for the three types of stimuli. The distribution for the time interval between 1018 and 1319 s in Figure 3 (absence of stimuli) is not presented because no activity was detected. Despite the apparent similarities of the distributions in Figure 4, the second cluster of spike amplitude distributions (natural scene and natural movie) shows a heavier tail. Specifically, the highest observed cluster during the static grating stimulus (top plot) is centered at 1.06, while for the other two stimuli we obtained several higher values, with the largest cluster centered around 1.43.

A qualitative representation of how these spike clusters are distributed within the three groups is given in Figure 5. The three plots show a short interval of the observed calcium series, chosen in correspondence with one of the highest observed spikes. Each plot also shows a series of colored vertical lines: the lines are placed at the estimated

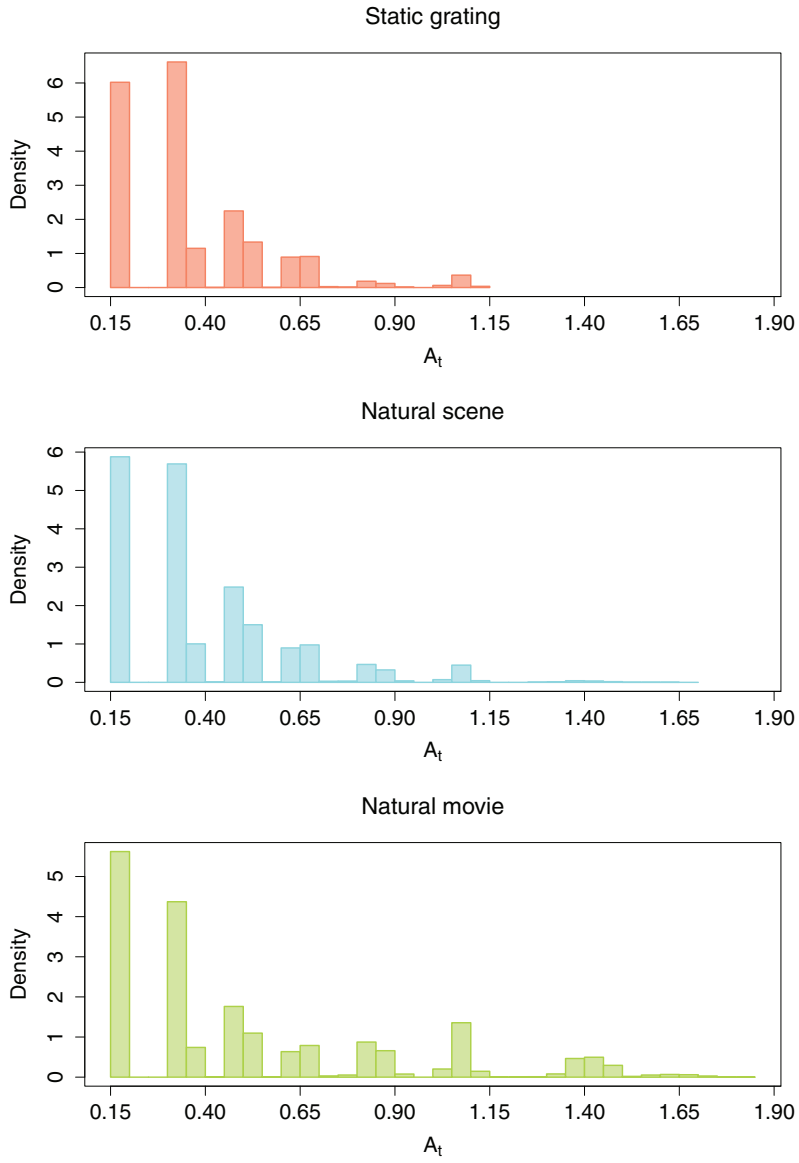

FIGURE 4 Empirical distribution of the posterior means of the observational cluster parameters A_t for the three experimental conditions of the Allen Brain Observatory data. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

spike times, and the colors correspond to the estimated spike amplitudes. The represented partition is the posterior point estimate obtained by minimizing the variation of information loss, as proposed in Wade and Ghahramani (2018). Conditionally on the obtained partition, for each cluster a representative value for the cluster parameter is obtained as follows: first, for each MCMC iteration, the group-specific average of A_t is computed keeping the partition fixed; then, these values are averaged over all the MCMC iterations. We notice that for all experiments, high values of the observed calcium level are often produced as the result of several consecutive spikes, since, individually, the spikes are characterized by a relatively low amplitude, and the observed calcium level is cumulated due to its autoregressive behavior. The autoregressive parameter γ has a posterior mean equal to 0.493 with a 95% credible interval of $(0.481, 0.505)$. This result corresponds to the understanding that the observed calcium response may be generated by high-frequency firing neurons: due to the low-sampling rate, the nonlinear calcium signal essentially captures a superimposition of multiple spikes (Hoang *et al.*, 2020).

As a matter of fact, another useful quantity we can compute to compare the neuronal activity between stimuli is the firing rate, which provides a measure of how often the neuron has activated during a specific visual stimulus. The rate computes the number of detected spikes per second, to take into account the different duration of the experiments. For the static grating stimulus, the posterior mean rate (and related 95% credible interval) is 0.223 $(0.216, 0.229)$, while for the natural scene and natural movie stimuli they are 0.419 $(0.410, 0.428)$ and 0.511 $(0.495, 0.531)$, respectively. These results highlight the role of spike-frequency adaptation, whereby some neurons show an increased

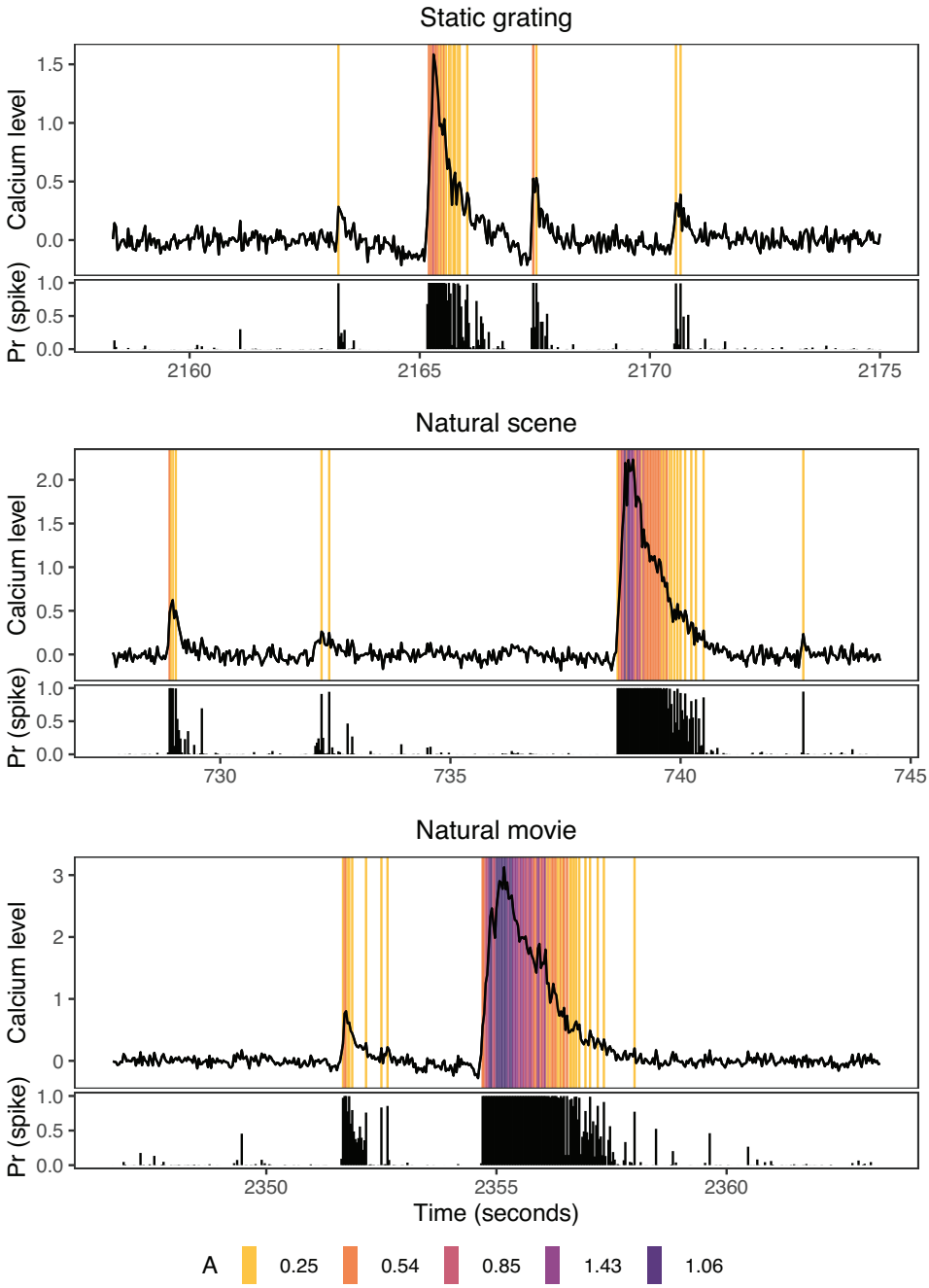

FIGURE 5 Short interval of length 500 of the Allen Brain Observatory data in correspondence of a spike, for the three stimuli. The vertical lines indicate the time of a spike and the colors correspond to the observational cluster of its amplitude. The bottom panels show the estimated posterior probability of spike presence, for each time point

activity when exposed to more complex stimuli, thus exhibiting higher firing rates and larger calcium concentration measurements (Peron and Gabbiani, 2009).

5 | DISCUSSION

Calcium imaging has become widely applied to record the neuronal activity in awake, freely moving, and behaving animals. However, reliable spike detection and spike time estimation remain challenging, due to the nonlinearity and low signal-to-noise ratio of the calcium response. We have proposed a single-stage nested Bayesian finite mixture model that allows estimating the spike activity and also characterizing how its distribution varies across stimuli.

The method shows good performances in a simulation study and captures characteristic features of neuronal

activity in an application to publicly available data from the Allen Brain Observatory.

Our approach exploited the knowledge of the stimulus types to model the partial exchangeability of the data within a Bayesian nonparametric framework. As a referee pointed out, continuous covariates are often available in the neurosciences, for example, the position of an alive animal in a two-dimensional environment. Our model cannot be directly applied with this information. The inclusion of available continuous and time-varying covariates is the subject of the ongoing investigation.

As neuroscience progresses, it is becoming more apparent that history effects can impact also seemingly low-dimensional experiments as the one considered in the Allen Brain Observatory data. Our approach does not explicitly model spike history dependence. However, it may be possible to explore the effect of history by considering appropriately defined time windows (e.g., by trials) and fix J equal to the total number of segments, regardless of our knowledge that some conditions are repeated. If history effects are present, one should expect that even segments corresponding to the same conditions may be assigned with a high probability to different clusters as time progresses. However, a proper modeling approach should incorporate further prior constraints (e.g., constraints that take into account the temporal sequence of the segments) to ensure the interpretability of the inference.

In line with the current literature, our approach is limited to the analysis of the calcium responses observed from a single neuron. Methods to infer neuronal connectivity from calcium imaging data over multiple regions of the brain remain sparse, often limited to point estimates (Mishchenko *et al.*, 2011) or the analysis of in vitro data (Rigat *et al.*, 2006). Inferences from our work could possibly be used to identify patterns across multiple neurons. For example, it is reasonable to assume that neurons exhibiting similar activity patterns may be grouped into homogeneous (spatial) clusters. Therefore, a second stage of the analysis may explicitly cluster across neurons the inferred spikes and the posterior means of the amplitudes within successive time intervals of calcium traces. A biclustering approach (see, e.g., Turner *et al.*, 2005; Chekouo *et al.*, 2015) could also be employed to describe the evolution of the neuronal patterns over time and conditions, as well as over space. Alternatively, one could apply the zero-inflated gamma model recently proposed by Wei *et al.* (2019) to study the densities of the deconvolved activity estimates and similarly heuristically compare such densities across neurons. Possible extensions of the framework presented here may focus on encoding the dynamic clustering of temporally correlated groups of neurons within the fCAM prior, as a function of external stimuli or the movement of the animal through an environment. In addition to the low signal-to-noise ratio, issues related to the dimension of the data and the accurate identification of the locations of neurons further compound the statistical and computational challenges (Petersen *et al.*, 2018).

ACKNOWLEDGMENTS

The authors thank the associate editor and an anonymous referee for very useful comments that improved the presentation of the paper.

OpenAccess Funding provided by Universita degli Studi di Milano-Bicocca within the CRUI-CARE Agreement.

OPEN RESEARCH BADGES

This article has earned an Open Materials badge for making publicly available the components of the research methodology needed to reproduce the reported procedure and analysis. All materials are available at http://github.com/laura-dangelo/fCAM_calcium_imaging.

DATA AVAILABILITY STATEMENT

The data that support the findings in this paper are available in Allen Brain Observatory - Visual Coding Amazon Web Services Public dataset at the registry.opendata.aws/allen-brain-observatory/.

ORCID

Laura D'Angelo https://orcid.org/0000-0001-5034-7414
Antonio Canale https://orcid.org/0000-0002-5403-0040
Michele Guindani https://orcid.org/0000-0002-6363-9907

REFERENCES

Allen Brain Observatory. (2017) Technical whitepaper: stimulus set and response analyses. Available at: help.brain-map.org/display/observatory/Data+-+Visual+Coding. Accessed December 01, 2021.

Allen Institute MindScope Program. (2016) Allen Brain Observatory – 2-photon visual coding [dataset]. Available at: brain-map.org/explore/circuits. Accessed December 01, 2021.

Argiento, R. and De Iorio, M. (2019) Is infinity that far? A Bayesian nonparametric perspective of finite mixture models [Preprint]. arXiv:1904.09733.

Brenner, N., Agam, O., Bialek, W. and de Ruyter van Steveninck, R. (2002) Statistical properties of spike trains: universal and stimulus-dependent aspects. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, 66, 031907.

Camerlenghi, F., Dunson, D.B., Lijoi, A., Prünster, I. and Rodríguez, A. (2019) Latent nested nonparametric priors (with discussion) *Bayesian Analysis*, 14, 1303–1356.

Canale, A., Lijoi, A., Nipoti, B. and Prünster, I. (2017) On the Pitman–Yor process with spike and slab base measure. *Biometrika*, 104, 681–697.

Canale, A., Lijoi, A., Nipoti, B. and Prünster, I. (2021) Inner spike and slab Bayesian nonparametric models. *Econometrics and Statistics*, in press. https://doi.org/10.1016/j.ecosta.2021.10.017

Chekouo, T., Murua, A. and Raffelsberger, W. (2015) The Gibbs-plaid biclustering model. *The Annals of Applied Statistics*, 9, 1643–1670.

Dana, H., Sun, Y., Mohar, B., Hulse, B.K., Kerlin, A.M., Hasseman, J.P., Tsegaye, G., Tsang, A., Wong, A., Patel, R., Macklin, J.J., Chen, Y., Konnerth, A., Jayaraman, V., Looger, L.L., Schreiter, E.R., Svoboda, K. and Kim, D.S. (2019) High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nature Methods*, 16, 649–657.

de Vries, S., Lecoq, J., Buice, M., Groblewski, P., Ocker, G., Oliver, M., Feng, D., Cain, N., Ledochowitsch, P., Millman, D., Roll, K., Garrett, M., Keenan, T., Kuan, C., Mihalas, S., Olsen, S., Thompson, C., Wakeman, W., Waters, J. and Koch, C. (2020) A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nature Neuroscience*, 23, 138–151.

Denti, F., Camerlenghi, F., Guindani, M. and Mira, A. (2021) A common atoms model for the Bayesian nonparametric analysis of nested data. *Journal of the American Statistical Association*, in press. https://doi.org/10.1080/01621459.2021.1933499

Dombeck, D.A., Harvey, C.D., Tian, L., Looger, L.L. and Tank, D.W. (2010) Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. *Nature Neuroscience*, 13, 1433–1440.

Friedrich, J. and Paninski, L. (2016) Fast active set methods for online spike inference from calcium imaging. In Lee, D., Sugiyama, M., Luxburg, U., Guyon, I. and Garnett, R., (Eds.) *Advances in Neural Information Processing Systems*. Red Hook, NY: Curran Associates, pp. 1984–1992.

Friedrich, J., Zhou, P. and Paninski, L. (2017) Fast online deconvolution of calcium imaging data. *PLOS Computational Biology*, 13, 1–26.

Frühwirth-Schnatter, S., Malsiner-Walli, G. and Grün, B. (2021) Generalized mixtures of finite mixtures and telescoping sampling. *Bayesian Analysis*, 16, 1279–1307.

Grienberger, C. and Konnerth, A. (2012) Imaging calcium in neurons. *Neuron*, 73, 862–885.

Hoang, H., Sato, M., Shinomoto, S., Tsutsumi, S., Hashizume, M., Ishikawa, T., Kano, M., Ikegaya, Y., Kitamura, K., Kawato, M. and Toyama, K. (2020) Improved hyperacuity estimation of spike timing from calcium imaging. *Scientific Reports*, 10, 17844.

Hubert, L. and Arabie, P. (1985) Comparing partitions. *Journal of Classification*, 2, 193–218.

Jewell, S. and Witten, D. (2018) Exact spike train inference via L0 optimization. *The Annals of Applied Statistics*, 12, 2457–2482.

Jewell, S.W., Hocking, T.D., Fearnhead, P. and Witten, D.M. (2019) Fast nonconvex deconvolution of calcium imaging data. *Biostatistics*, 21, 709–726.

Johnson, V.E. and Rossell, D. (2010) On the use of non-local prior densities in Bayesian hypothesis tests. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72, 143–170.

Li, N., Chen, T.-W., Guo, Z.V., Gerfen, C.R. and Svoboda, K. (2015) A motor cortex circuit for motor planning and movement. *Nature*, 519, 51–56.

Maruyama, R., Maeda, K., Moroda, H., Kato, I., Inoue, M., Miyakawa, H. and Aonishi, T. (2014) Detecting cells using non-negative matrix factorization on calcium imaging data. *Neural Networks*, 55, 11–19.

Miller, J.W. and Harrison, M.T. (2018) Mixture models with a prior on the number of components. *Journal of the American Statistical Association*, 113, 340–356.

Mishchenko, Y., Vogelstein, J.T. and Paninski, L. (2011) A Bayesian approach for inferring neuronal connectivity from calcium fluorescent imaging data. *Annals of Applied Statistics*, 5, 1229–1261.

Mitchell, T.J. and Beauchamp, J.J. (1988) Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83, 1023–1036.

Mukamel, E.A., Nimmerjahn, A. and Schnitzer, M.J. (2009) Automated analysis of cellular signals from large-scale calcium imaging data. *Neuron*, 63, 747–760.

Müller, P., Parmigiani, G. and Rice, K. (2007) FDR and Bayesian multiple comparisons rules. In Bernardo, J., Bayarri, M., Berger, J., Dawid, A., Heckerman, D., Smith, A. and West, M., (Eds.) *Bayesian Statistics 8*. Oxford, UK: Oxford University Press. 1–12.

Nakajima, M. and Schmitt, L.I. (2020) Understanding the circuit basis of cognitive functions using mouse models. *Neuroscience Research*, 152, 44–58.

Newton, M.A., Noueiry, A., Sarkar, D. and Ahlquist, P. (2004) Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics*, 5,.

Peron, S.P. and Gabbiani, F. (2009) Role of spike-frequency adaptation in shaping neuronal response to dynamic stimuli. *Biological Cybernetics*, 100, 505–520.

Petersen, A., Simon, N. and Witten, D. (2018) Scalpel: Extracting neurons from calcium imaging data. *Annals of Applied Statistics*, 12, 2430–2456.

Pnevmatikakis, E., Merel, J., Pakman, A. and Paninski, L. (2013) Bayesian spike inference from calcium imaging data. In *Signals, Systems and Computers*. Pacific Grove, CA: Asilomar Conference on Signals, Computers and Systems, pp. 349–353.

Prado, R. and West, M. (2010) *Time Series: Modeling, Computation, and Inference*, 1st edition. London: Chapman and Hall.

Rand, W.M. (1971) Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66, 846–850.

Rigat, F., de Gunst, M. and van Pelt, J. (2006) Bayesian modelling and analysis of spatio-temporal neuronal networks. *Bayesian Analysis*, 1, 733–764.

Rodríguez, A., Dunson, D.B. and Gelfand, A.E. (2008) The nested Dirichlet process. *Journal of the American Statistical Association*, 103, 1131–1154.

Rose, T., Goltstein, P.M., Portugues, R. and Griesbeck, O. (2014) Putting a finishing touch on GECIs. *Frontiers in Molecular Neuroscience*, 7, 88.

Shen, T., Lur, G., Xu, X. and Yu, Z. (2021) To deconvolve, or not to deconvolve: Inferences of neuronal activities using calcium imaging data [Preprint]. arXiv:2103.02163.

Shibue, R. and Komaki, F. (2020) Deconvolution of calcium imaging data using marked point processes. *PLOS Computational Biology*, 16, 1–25.

Sun, W., Reich, B.J., Tony Cai, T., Guindani, M. and Schwartzman, A. (2015) False discovery control in large-scale spatial multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77, 59–83.

Turner, H., Bailey, T. and Krzanowski, W. (2005) Improved biclustering of microarray data demonstrated through systematic

performance tests. *Computational Statistics & Data Analysis*, 48, 235–254.

Vogelstein, J.T., Packer, A.M., Machado, T.A., Sippy, T., Babadi, B., Yuste, R. and Paninski, L. (2010) Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of Neurophysiology*, 104, 3691–3704.

Vogelstein, J.T., Watson, B.O., Packer, A.M., Yuste, R., Jedynak, B. and Paninski, L. (2009) Spike inference from calcium imaging using sequential Monte Carlo methods. *Biophysical Journal*, 97, 636–655.

Wade, S. and Ghahramani, Z. (2018) Bayesian cluster analysis: point estimation and credible balls (with discussion) *Bayesian Analysis*, 13, 559–626.

Wei, X.-X., Zhou, D., Grosmark, A., Ajabi, Z., Sparks, F., Zhou, P., Brandon, M., Losonczy, A. and Paninski, L. (2019) A zero-inflated gamma model for post-deconvolved calcium imaging traces [Preprint]. bioRxiv: 637652.

SUPPORTING INFORMATION

Supporting Information

How to cite this article: D'Angelo, L., Canale, A., Yu, Z., Guindani, M. (2023) Bayesian nonparametric analysis for the detection of spikes in noisy calcium imaging data. *Biometrics*, 79, 1370–1382. https://doi.org/10.1111/biom.13626

APPENDIX: NESTED TELESCOPING SAMPLING

Denote with C^D the current partition on the distributions and with C^O the partition on the observations.

(1) Sample the weights on the distributions: $(\pi_1, \dots, \pi_K) \mid K, \alpha, C^D \sim \text{Dir}(e_1, \dots, e_K)$; where $e_k = \alpha/K + J_k$, and J_k is the number of groups assigned to the distribution k.

(2) Sample the weights on the observations: for all $k \in \{1, \dots, K\}$ sample a vector $\boldsymbol{\omega}_k$ from $(\omega_{1,k}, \dots, \omega_{L,k}) \mid L, \beta, C^O, C^D \sim \text{Dir}(f_{1,k}, \dots, f_{L,k})$; where $f_{l,k} = \beta/L + N_{l,k}$, and $N_{l,k}$ is the number of observations in the observational cluster l and distributional cluster k.

(3) Update the partition on the distributions C^D by sampling from the posterior distribution of the latent cluster allocation variables $\boldsymbol{S}$. For $j = 1, \dots, J$

$$\Pr(S_j = k \mid \pi, K, A^*, y, g) \propto \pi_k \prod_{t:g_t=j} \omega_{M_t,S_j}\, p(y_t \mid A^*_{M_t}),$$

with $k \in \{1, \dots, K\}$. Determine $J_k = \#\{j : S_j = k\}$, for $k = 1 \dots, K$, and the number of nonempty components $K_+ = \sum_{k=1}^{K} I\{J_k > 0\}$. Relabel the components so that the first K_+ are nonempty.

(4) Update the partition on the observations C^O by sampling from the posterior distribution of the latent cluster allocation variables $\boldsymbol{M}$. For $t = 1, \dots, T$

$$\Pr(M_t = l \mid S_{g_t} = k, \boldsymbol{S}, \boldsymbol{\omega}, L, K, \boldsymbol{A}^*, \boldsymbol{y}, \boldsymbol{g}) \propto \omega_{l,k}\, p(y_t \mid A^*_{M_t}),$$

with $l \in \{1, \dots, L\}, k \in \{1, \dots, K\}$. Determine $N_l = \#\{t : M_t = l\}$, for $l = 1 \dots, L$, and the number of nonempty components $L_+ = \sum_{l=1}^{L} I\{N_l > 0\}$. Relabel the components so that the first L_+ is nonempty. Because all the mixtures share the same atoms, the cluster parameters are sorted regardless of the distributional cluster allocation.

(5) Sample the cluster parameters for the nonempty components: $p(A^*_l \mid -) \propto p(A^*_l) \prod_{t:M_t=l} p(y_t \mid A^*_l)$.

(6) Conditional on C^D, sample the number of components K of the mixture on distributions.

(7) Conditional on C^O, sample the number of components L of the mixtures on observations. If $L > L_+$, sample a new parameter A^* for the empty components from the prior distribution.

(8) Update the hyperparameter α on the Dirichlet distribution on the mixture weights on distributions.

(9) Update the hyperparameter β on the Dirichlet distribution on the mixture weights on observations.

The posterior distributions for steps 6–9 are given in Frühwirth-Schnatter *et al.* (2021).