# Clustering activation patterns of spatially-referenced neurons

Laura D'Angelo, Antonio Canale and Michele Guindani

Statistical methods and models for complex data — Poster Session: 21 September 2022


## *Abstract*

*We describe a Bayesian nonparametric mixture model that allows for simultaneous deconvolution of series of neuronal activity and clustering of the signal based on common patterns of activation. The model makes use of a latent continuous process for the spike probabilities to identify groups of co-activating cells. Neurons' spatial dependence is also introduced through the mixture weights.*


## Calcium imaging data

### Motivating application

- **Neighboring neurons** are often organized in groups with similar functional characteristics; however, the anatomical organization of the hippocampus is still uncertain.
- Need to **identify clusters** of neurons with **similar patterns of activity over time**.
- Activity is described by *spike trains*: **binary series representing active/resting state** of each neuron at each time point.

### Technique and data

The technique of calcium imaging allows visualization of the intra-cellular **concentration of calcium over time**. This is a **proxy of the activity** of populations of neurons over time.

Transient high levels in the observed calcium indicate the neurons' activations.

A clustering of co-activating neurons is obtained by looking for similarities in the temporal patterns of these activations.
The spatial location of each neuron in the region of interest is also available.

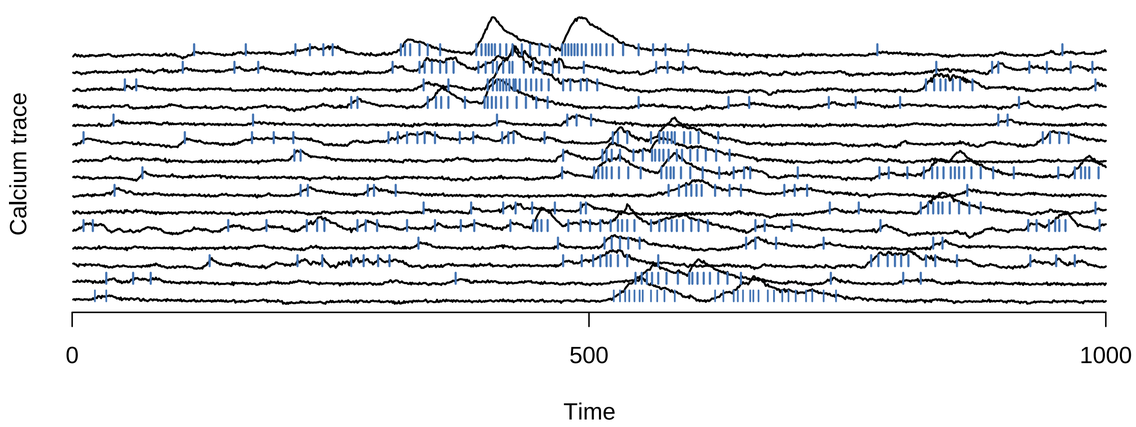


**Figure:** Calcium traces of 15 neurons (black lines). The colored segments indicate detected activations.

### Challenges

- the series of activations have to be extracted and clustered → limits of two-step approaches;
- scarcity of **clustering** methods for **binary time series**;
- spikes are not uniformly distributed over time: an increase in the observed calcium is usually the result of several consecutive spikes → interest in modeling the **temporal dependence of activations**;
- the neurons' locations provide useful information on possible existing **spatial dependence** structures between neurons → neighboring neurons should have a higher prior probability of being assigned to the same group.

## Model formulation

### Calcium dynamics

We adopt a popular model (Vogelstein et al. 2010) to describe the calcium dynamics and how it relates with the underlying activity.
For neuron $i$ at time $t$, ($i = 1, \dots, n$ and $t = 1, \dots, T$): $y_{i,t}$ observed fluorescence trace; $c_{i,t}$ true calcium concentration. We write

$$y_{i,t} = b_i + c_{i,t} + \epsilon_{i,t}, \qquad \epsilon_{i,t} \sim \mathrm{N}(0, \sigma^2),$$
$$c_{i,t} = \gamma\, c_{i,t-1} + s_{i,t} \cdot a_{i,t} + w_{i,t}, \qquad w_{i,t} \sim \mathrm{N}(0, \tau^2),$$

with $b_i$ baseline parameters; $\gamma$ a decay parameter, modeling the autoregressive calcium behavior; and $\epsilon_{i,t}$ and $w_{i,t}$ independent Gaussian errors.

- The **neuronal activations** are represented by the binary time series

$$\boldsymbol{s}_i = (s_{i,1}, \dots, s_{i,T}), \quad \text{with } s_{i,t} \in \{0, 1\}.$$

- The **spikes' amplitudes** are described by the parameters $a_{i,t} \in \mathbb{R}^+$, and they are defined conditionally on $s_{i,t} = 1$.

Neurons' locations: spatial coordinates $l_i = (l_{i,1}, l_{i,2})$. Specifically, we consider the **proximity matrix** $\Sigma = \Sigma(l_i, l_j)$, for $i, j = 1, \dots, n$, measuring the closeness between neurons.

## Prior distributions

On $a_{i,t}$ we place a DP mixture of Gamma, to account for the heterogeneity of spikes' amplitudes:

$$a_{i,t} \mid P \sim P \qquad P = \sum_{k \geq 1} \omega_k \delta_{a_k^*}$$
$$\{\omega_k\}_{k \geq 1} \sim \text{stick-breaking} \qquad a_k^* \overset{i.i.d.}{\sim} \text{Gamma}$$

On parameters ($b_i$, $\gamma$, $\sigma^2$, $\tau^2$) we place prior distributions as in D'Angelo et al. 2022.

### Prior on $s_i$: location-dependent mixture of latent Gaussian processes

Instead of working directly with the binary time series, we introduce a (transformed) **latent Gaussian process** (GP) $\tilde{\mathbf{s}}_i = \left(\tilde{s}_i(t), t \geq 0\right)$ that controls the **spike probability** over time:

$$s_{i,t} \mid \tilde{\mathbf{s}}_i \sim \text{Bernoulli}\Big(\Phi\big(\tilde{s}_i(t)\big)\Big) \quad \text{indep. for } t = 1, \dots, T$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard Gaussian distribution. We apply the probit stick-breaking process of Rodríguez and Dunson (2011) to define a $\Sigma$-informed mixture prior on $\tilde{\boldsymbol{s}}_i$:

$$\tilde{\mathbf{s}}_i \mid G_\Sigma \sim G_\Sigma \qquad G_\Sigma = \sum_{k \geq 1} \pi_k(\Sigma) \cdot \delta_{\tilde{\mathbf{s}}_k^*}$$
$$\pi_k(\Sigma) = \Phi(\alpha_k) \prod_{r<k} \big(1 - \Phi(\alpha_r)\big) \qquad \text{with } \alpha \sim \mathrm{N}_n(0, \Sigma).$$

The atoms of the mixture are independent draws from a GP over time,

$$\tilde{\mathbf{s}}_k^* \mid \theta \sim \mathrm{GP}(\mu, \Omega)$$

where $\Omega = \Omega(t, t')$ is the covariance function, modeling the temporal dependence.

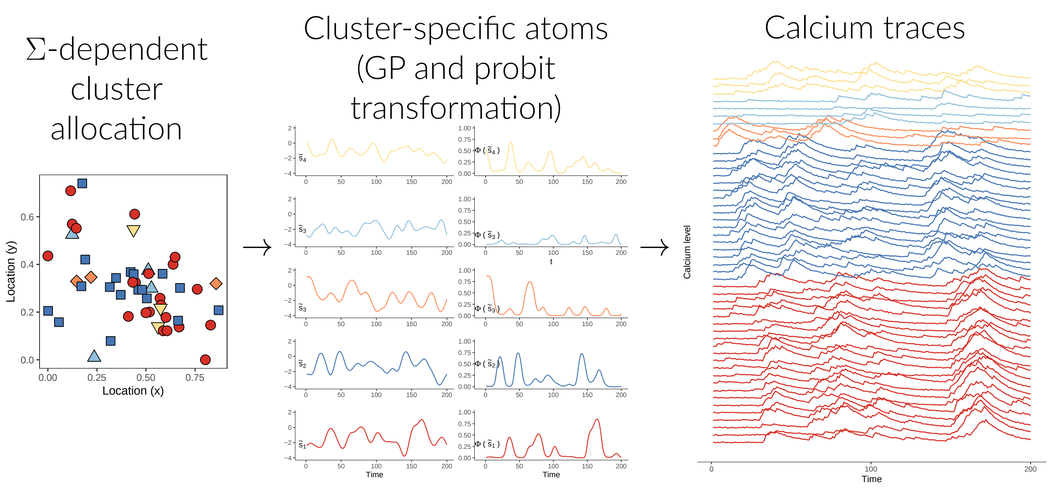


**Figure:** Graphical representation of the data generating process.

## Real data application

We considered a population of 124 hippocampal neurons, and we performed a time-varying clustering by considering 3 non-overlapping windows of 300 time points each.

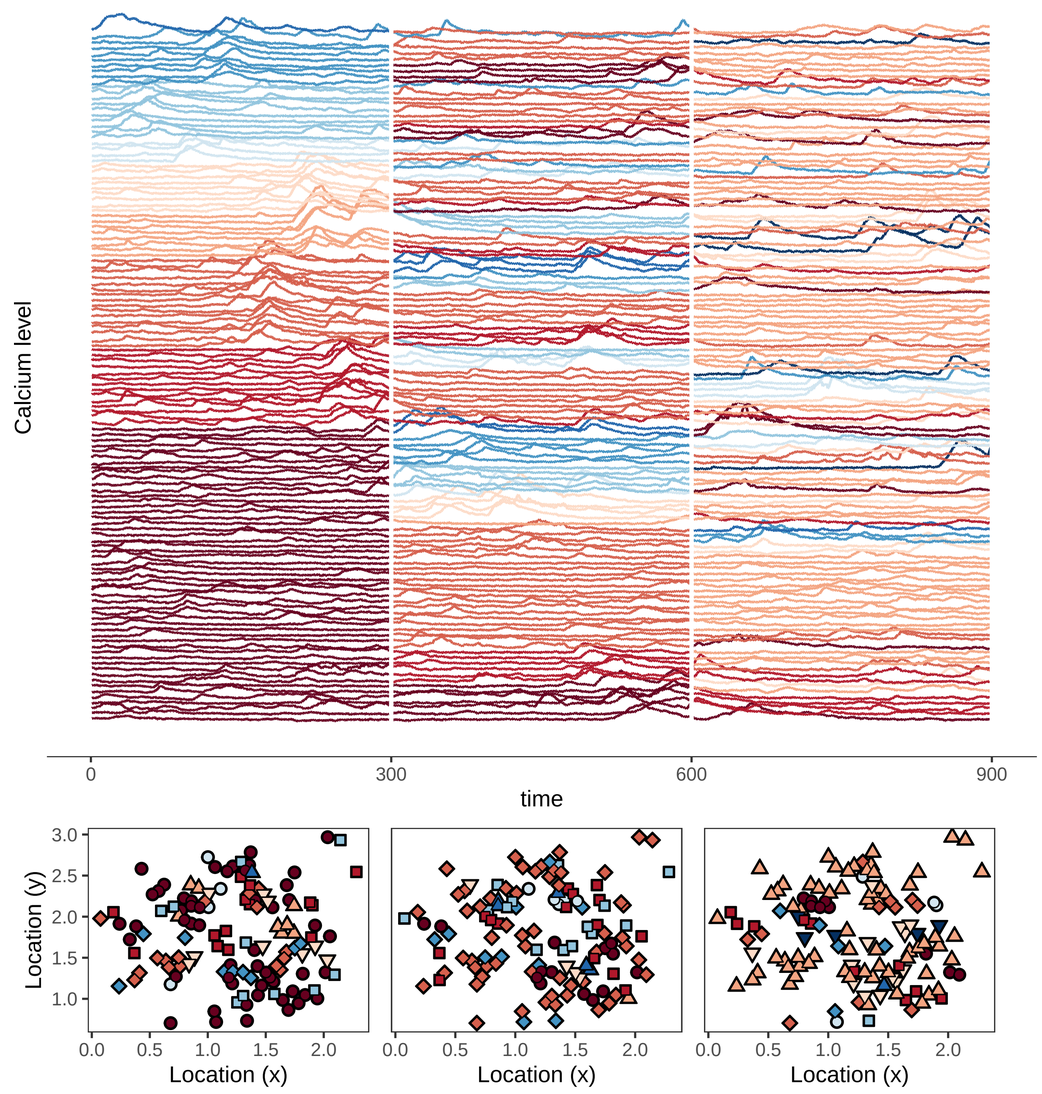


**Figure:** Top: calcium traces, sorted and colored according to the estimated cluster allocation. Bottom: neurons' locations, colored according to the cluster assignment (colors reflect the corresponding top plot).

## References


D'Angelo, L., Canale, A., Yu, Z., and Guindani, M. (2022). "Bayesian nonparametric analysis for the detection of spikes in noisy calcium imaging data". In: *Biometrics*, pp. 1–13.
Rodríguez, A. and Dunson, D. (2011). "Nonparametric Bayesian models through probit stick-breaking processes". In: *Bayesian Analysis* 6, pp. 145–178.
Vogelstein, J. T., Packer, A. M., et al. (2010). "Fast nonnegative deconvolution for spike train inference from population calcium imaging". In: *Journal of Neurophysiology* 104.6, pp. 3691–3704.