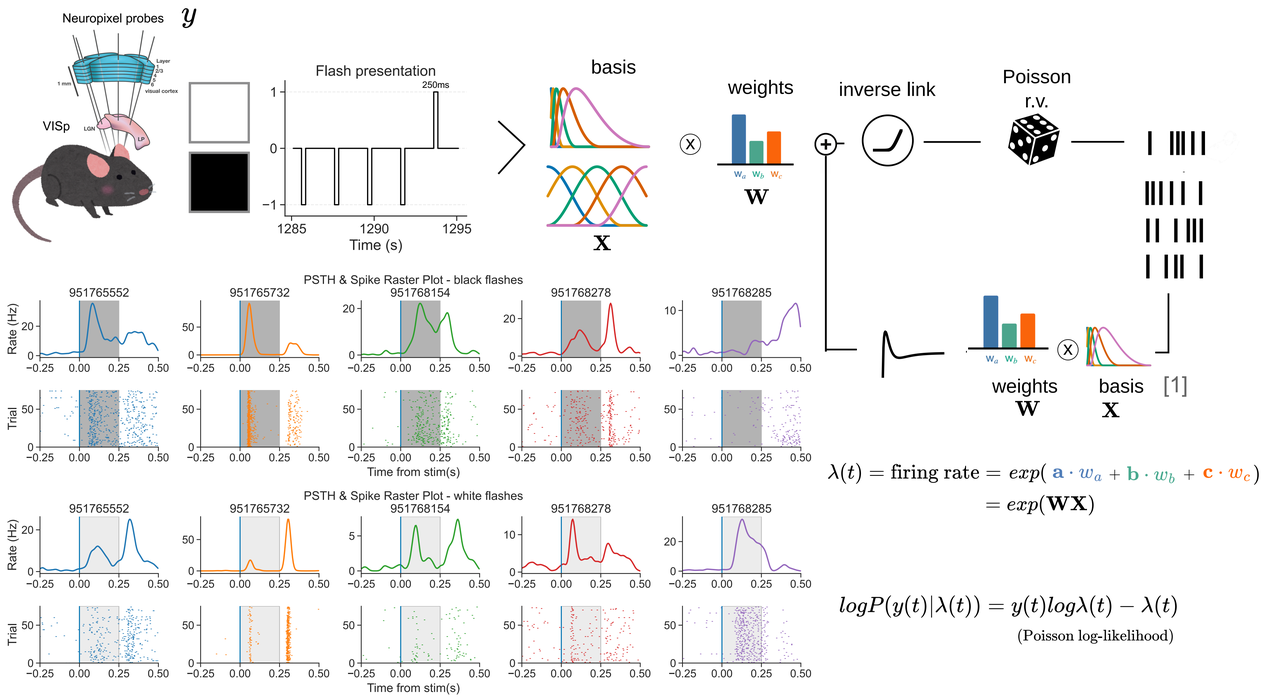
# Contributing to Open Source in Computational Neuroscience: Allen Institute's OpenScope Databook and GLM-HMMs

Camila Maura, Edoardo Balzani, Guillaume Viejo, Sarah Jo Venditto
Flatiron Institute Center for Computational Neuroscience

## Introduction

- Flatiron CCN software projects aim to provide robust tools for the community: Pynapple & NeMoS.
- We collaborated with the Open Scope Databook to build a tutorial on how to use Generalized Linear Models (GLMs) to analyze neural data from the Visual Coding - Neuropixels dataset, showcasing these tools.

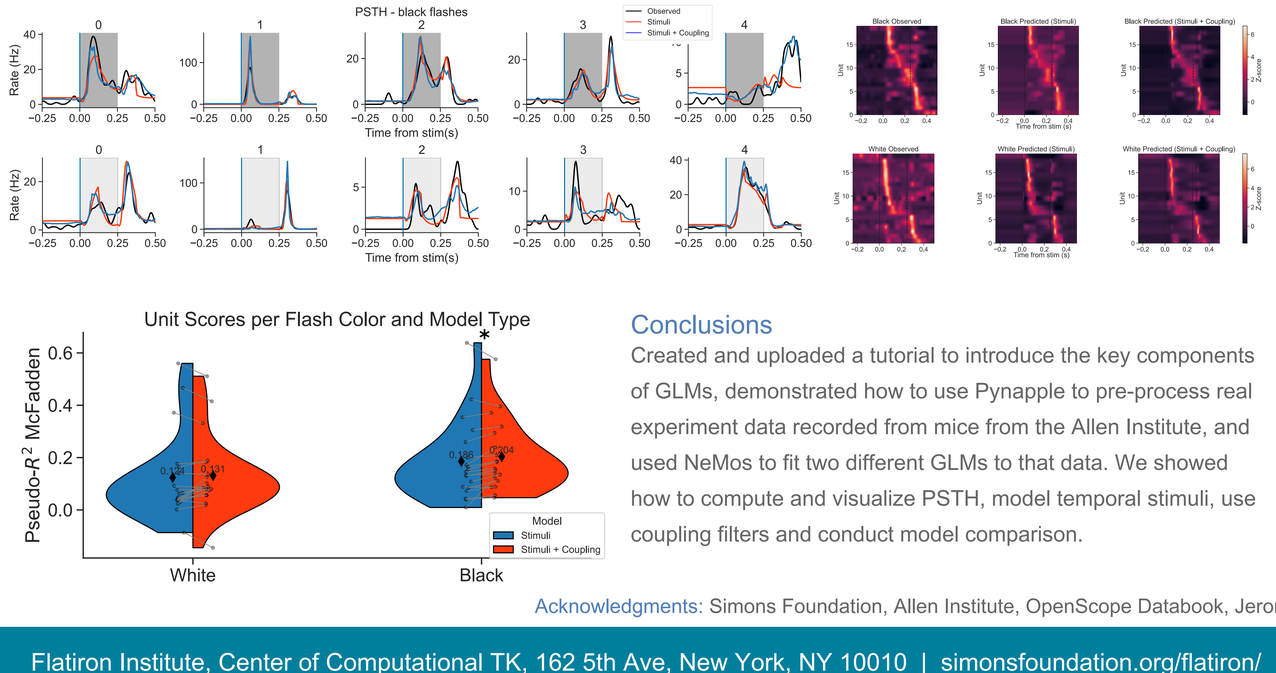
## Methods




$\lambda(t) = \text{firing rate} = exp(\ \mathbf{a} \cdot w_a + \mathbf{b} \cdot w_b + \mathbf{c} \cdot w_c\ )$
$= exp(\mathbf{WX})$

$logP(y(t)|\lambda(t)) = y(t)log\lambda(t) - \lambda(t)$
(Poisson log-likelihood)

## Results







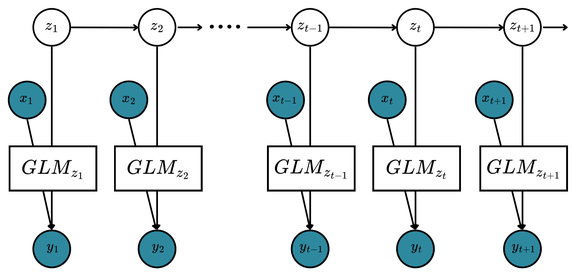
## Conclusions

Created and uploaded a tutorial to introduce the key components of GLMs, demonstrated how to use Pynapple to pre-process real experiment data recorded from mice from the Allen Institute, and used NeMos to fit two different GLMs to that data. We showed how to compute and visualize PSTH, model temporal stimuli, use coupling filters and conduct model comparison.

## Introduction

- GLM-HMMs models are useful to analyze how hidden latent states affect observable behavioral [2] [3] and neural [4] dynamics.
- No mantained open source framework for analyzing data with this model.

## Methods [5]




GLM-HMM

HMM

$p(z_t = k \mid z_{t-1} = j) = A_{jk}$ $A \in \mathbb{R}^{K \times K}$ transition matrix

$p(z_1 = k) = \boldsymbol{\pi}_k$ $\boldsymbol{\pi} \in \mathbb{R}$ initial state distribution

GLM

$y_t \mid \boldsymbol{x}, \boldsymbol{k} \sim Ber(f(-\boldsymbol{x}_t \cdot \boldsymbol{w}_k))$ $\{\boldsymbol{w}_k\}_{k=1}^K$ with $\boldsymbol{w}_k \in \mathbb{R}$ weights

$p(y_t = 1 \mid \boldsymbol{x}_t, z_t = k) = \dfrac{1}{1 + exp(-\boldsymbol{x}_t \cdot \boldsymbol{w}_k)}$

$\text{PMF} = p^k(1-p)^{1-k}$

Objective function

initial transition emission

$$p(\boldsymbol{y}, \boldsymbol{z}|\theta) = p(z_1|\boldsymbol{\pi}) \left[\prod_{t=2}^{T} p(z_t \mid z_{t-1}, A)\right] \prod_{m=1}^{T} p(y_m|z_m, \boldsymbol{w}_m)$$

$$\text{ECLL}(\theta) = \sum_z p(\mathbf{z} \mid \boldsymbol{y}, \{\boldsymbol{x}\}_{t=1}^T, \theta^{\text{old}}) \log p(\mathbf{y}, \mathbf{z} \mid \theta, \{\boldsymbol{x}\}_{t=1}^T) \quad \xi_{t,j,k} = p(z_t = k, z_{t-1} = j \mid \boldsymbol{y}, \{\boldsymbol{x}\}_{t=1}^T, \theta^{\text{old}})\ \gamma_{t,k} = p(z_t = k \mid \boldsymbol{y}, \{\boldsymbol{x}\}_{t=1}^T, \theta^{\text{old}})$$

$$\text{ECLL}(\theta) = \sum_{k=1}^{K} \gamma_{1,k} \log \boldsymbol{\pi}_k + \sum_{t=1}^{T} \sum_{j=1}^{K} \sum_{k=1}^{K} \xi_{t,j,k} \log A_{jk} + \sum_{t=1}^{T} \sum_{k=1}^{K} \gamma_{t,k} \log p(y_t \mid z_t = k, \boldsymbol{x}_t, \boldsymbol{w}_k)$$

E - Step

$\alpha_{t,k} \equiv p(\boldsymbol{y}_{1:t}, z_t = \boldsymbol{k} \mid \{\boldsymbol{x}_{t'}\}_{t'=1}^t)$ $\beta_{t,k} \equiv p(\boldsymbol{y}_{[t+1:T]} \mid z_t = k, \{\boldsymbol{x}_{t'}\}_{t'=t+1}^T)$

$$\gamma_{t,k} = \frac{\alpha_{t,k}\beta_{t,k}}{\sum_{k=1}^K \alpha_{T,k}} \qquad \xi_{t,j,k} = \frac{\alpha_{t,j} A_{jk} \beta_{t+1,k} p(\boldsymbol{y}_{t+1} \mid z_{t+1} = k, \boldsymbol{x}_{t+1}, \boldsymbol{w}_k)}{\sum_{k=1}^K \alpha_{T,k}}$$

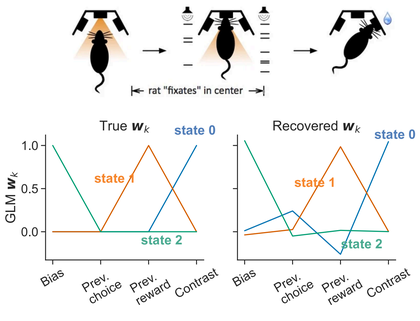
M - Step

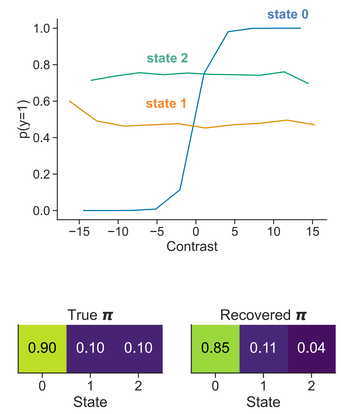
$$\pi_k^{\text{new}} = \frac{\gamma_{1,k}}{\sum_{j=1}^K \gamma_{1,j}} \qquad A_{jk}^{\text{new}} = \frac{\sum_{t=2}^T \xi_{t,j,k}}{\sum_{t=2}^T \sum_{k=1}^K \xi_{t,j,k}}$$

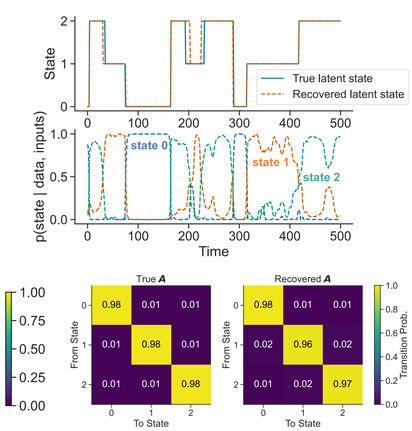
$$\boldsymbol{w}_k^{new} = \arg\min_{\boldsymbol{w}_k} \sum_{t=1}^{T} \sum_{k=1}^{K} \gamma_{t,k} \log p(y_t \mid z_t = k, \boldsymbol{x}_t, \boldsymbol{w}_k)$$
$$= \arg\min_{\boldsymbol{w}_k} \sum_{t=1}^{T} \sum_{k=1}^{K} \gamma_{t,k} \log PMF$$

## (preliminary) Results

Evidence accumulation task [6]




## Conclusions & Next Steps

We implemented a vectorized version of EM using the forward-backward algorithm to fit GLM-HMM models and conducted preliminary model recovery analysis. We are running tests, finishing the implementation of another algorithm for estimating the most probable sequence of latent states, and setting up a user interface.

Want to take a look at the Intro to GLMs notebook?
Do you want me to email you when we launch the new NeMos features?
Want to have a pdf version of this poster?

Acknowledgments: Simons Foundation, Allen Institute, OpenScope Databook, Jerome Lecoq, Carter Peene, Wolf de Wulf

Flatiron Institute, Center of Computational TK, 162 5th Ave, New York, NY 10010 | simonsfoundation.org/flatiron/

Ref: [1] Pillow, J., et al. (2008) Nature, 454 [2] Ashwood, Z. et al. (2022) Nat. Neurosci., 25 [3] Bolkan S.et al. (2022) Nat. Neurosci., 25 [4] Escola et al. (2011) Neural Computation, 23 [5] Bishop, C. (2006) Pattern Recognition and Machine Learning. Springer [6] Brunton B., et al. (2013) Science, 340