

Hidden Markov Model (HMM) and its Application to Calcium Imaging Data

June 2025

Jungsik Noh

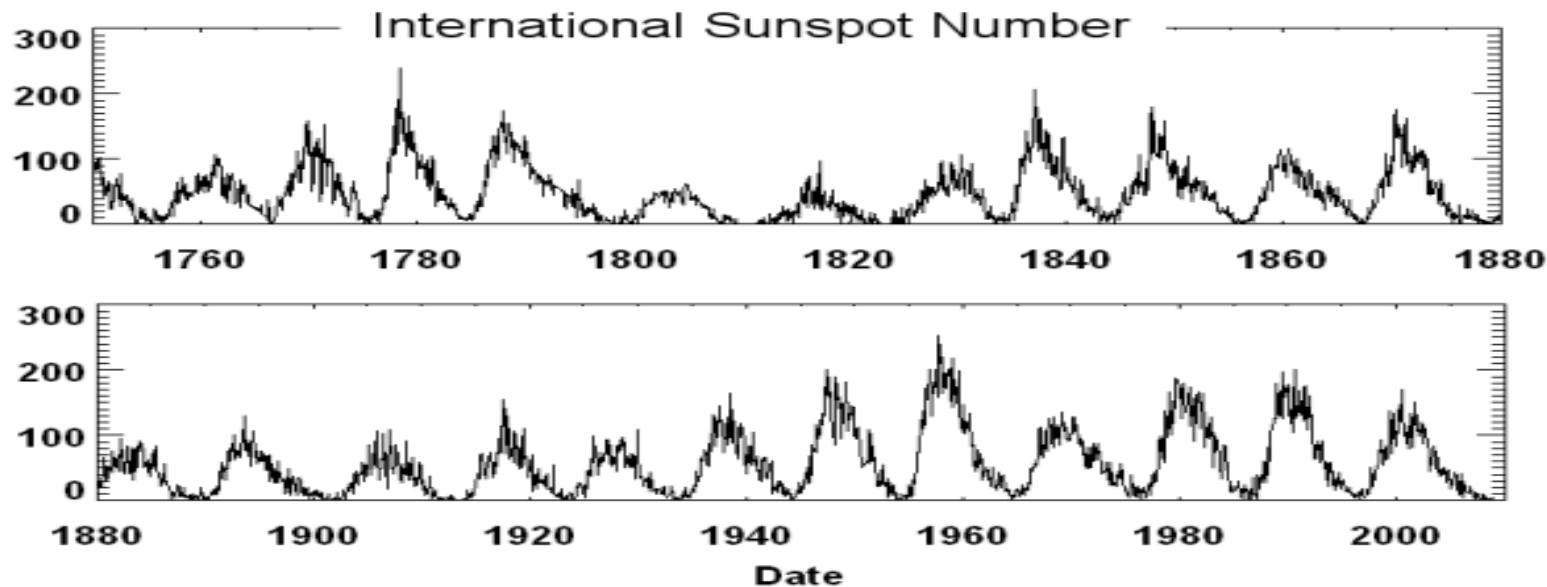
Outline

1. Time Series or Stochastic Processes
2. Markov Chain
3. Hidden Markov Model (HMM)
4. Calcium imaging data and calcium spike inference problem

Time Series (TS)

- A **time series** is a series of data points indexed in time order.
- Mostly, the data points, $\{x_t: t = 1, 2, \dots\}$, are observed at equally spaced time points.
 - ‘Data points’ can be univariate or multivariate.
 - (eg) daily numbers of sunspots, daily closing values of S&P500

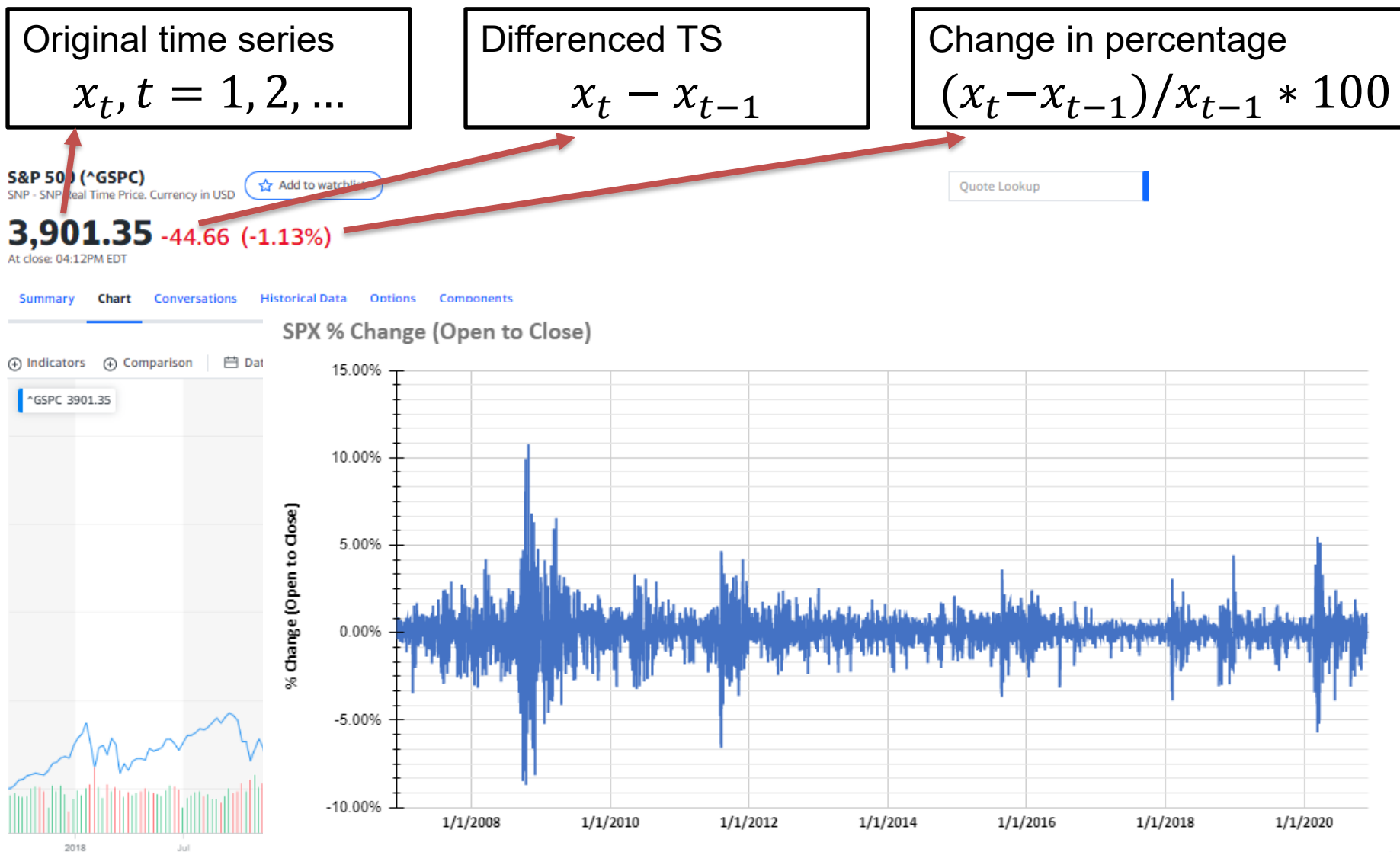
Daily sunspot numbers (annual average)



- The index has a special meaning, that is the time order.
- The random variables representing data points ($\{X_t\}$) are NOT independent.
 - X_t and X_s , for any $t, s \in \mathbb{Z}$ are dependent random variables.

Time series plots show how a TS varies over time

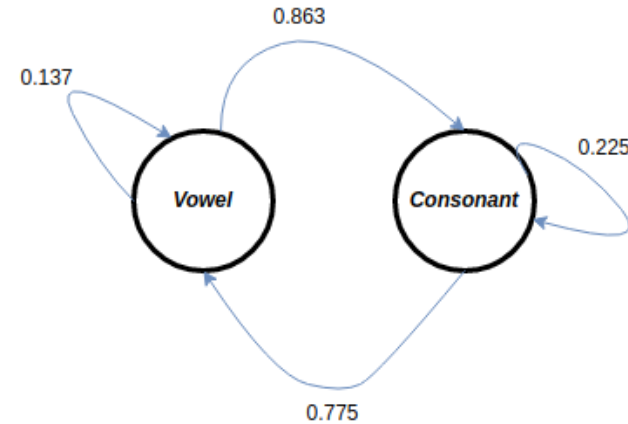
- Daily closing values of S&P500 for past 5 years



Markov chain (Discrete Markov process)

- “Andrey Markov used Markov chains to study the distribution of vowels in Eugene Onegin, written by Alexander Pushkin, and proved a central limit theorem for such chains.”

$$\{s_t; t \geq 1\} = \{vcccvvcvccvc \dots\}$$



- Markov property**

Let $S_t = \text{vowel or consonant}$.

\Leftrightarrow The future variables are independent of the past variables, given the present variable (memoryless property).

$$\Leftrightarrow \Pr(S_{t+1} = s \mid S_t = s_t, S_{t-1} = s_{t-1}, S_{t-2} = s_{t-2}, \dots) = \Pr(S_{t+1} = s \mid S_t = s_t)$$

$$\Leftrightarrow \text{concisely, } \Pr(S_{t+1} \mid S_t, S_{t-1}, S_{t-2}, \dots) = \Pr(S_{t+1} \mid S_t) \quad (\text{abuse of notation})$$

$$\Leftrightarrow \Pr(S_t \mid S_{t-1}, S_{t-2}, S_{t-3}, \dots) = \Pr(S_t \mid S_{t-1}) \text{ for any } t \in \mathbb{Z}$$

Markov processes

- A Markov process is a stochastic process satisfying the Markov property.
- For a sequence of dependent random variables that have finite state values, Markov chain has been a powerful model to describe its dynamics.
 - (eg) Corporate Credit Rating System

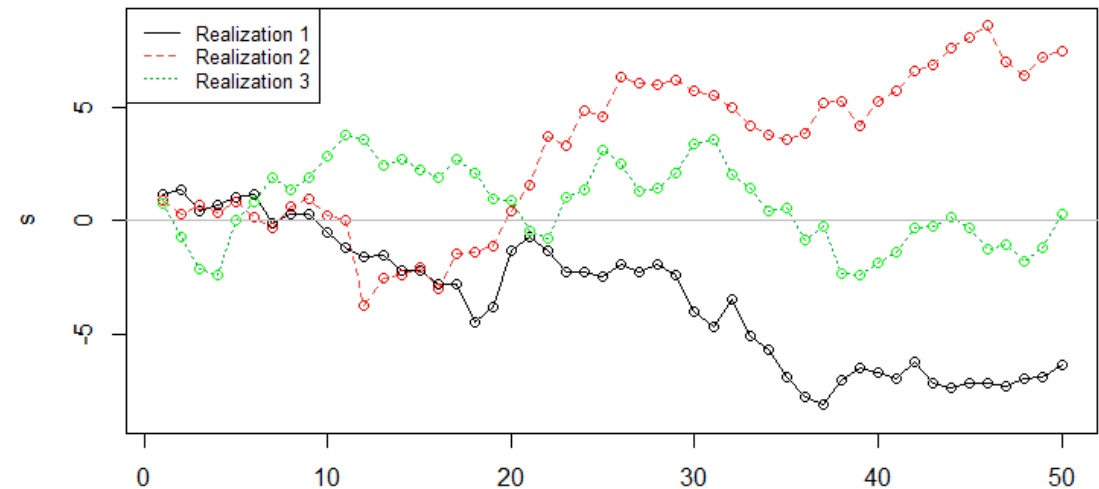
		T+1							
T		AAA	AA	A	BBB	BB	B	CCC	D
	AAA	92.29	6.96	0.54	0.14	0.06	0.00	0.00	0.000
	AA	0.64	90.75	7.81	0.61	0.07	0.09	0.02	0.010
	A	0.05	2.09	91.38	5.77	0.45	0.17	0.03	0.051
	BBB	0.03	0.20	4.23	89.33	4.74	0.86	0.23	0.376
	BB	0.03	0.08	0.39	5.68	83.10	8.12	1.14	1.464
	B	0.00	0.08	0.26	0.36	5.44	82.33	4.87	6.663
	CCC	0.10	0.00	0.29	0.57	1.52	10.84	52.66	34.030
	D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100

- Initial distribution (finite states = $\{1, 2, \dots, M\}$)
 - The distribution of S_1 , $\pi^{(1)} = (\Pr(S_1 = 1), \dots, \Pr(S_1 = M))$
- Transition probabilities, transition matrix
 - $a_{ij} = \Pr(S_{t+1} = j \mid S_t = i)$

- A random walk is a continuous Markov process.

$$S_t = \sum_{i=1}^t X_i, \quad X_i \sim \text{IID } N(0, 1), \quad S_0 = 0$$

$$E(S_t) = 0, \quad \text{Var}(S_t) = t, \quad \text{std}(S_t) = \sqrt{t}$$



Limiting distribution of a Markov chain

- $\{S_t\}$ is a Markov chain with M states.
 - Initial dist'n: $\boldsymbol{\pi}^{(1)} = (P(S_1 = 1), \dots, P(S_1 = M))$
 - Transition matrix:

$$A = [a_{ij}]_{M \times M}, \quad a_{ij} = \Pr(S_{t+1} = j \mid S_t = i)$$

- Compute $\boldsymbol{\pi}^{(t)} = (P(S_t = 1), \dots, P(S_t = M))$
 - $P(S_2 = j) = \sum_{i=1}^M P(S_1 = i) P(S_2 = j \mid S_1 = i)$ (law of total prob.)

$$= \boldsymbol{\pi}^{(1)} \cdot \begin{pmatrix} P(S_2 = j \mid S_1 = 1) \\ P(S_2 = j \mid S_1 = 2) \\ \vdots \\ P(S_2 = j \mid S_1 = M) \end{pmatrix} = \boldsymbol{\pi}^{(1)} \cdot \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{Mj} \end{pmatrix}$$

- $\boldsymbol{\pi}^{(2)} = \boldsymbol{\pi}^{(1)} \cdot \mathbf{A}$
 - $\boldsymbol{\pi}^{(t)} = \boldsymbol{\pi}^{(1)} \cdot \mathbf{A}^{t-1}$
- Suppose \mathbf{A}^t converges to \mathbf{A}^∞ .
 - (long-term behavior of S_t):

$$\boldsymbol{\pi}^{(\infty)} := \lim_{t \rightarrow \infty} \boldsymbol{\pi}^{(t)} = \lim_{t \rightarrow \infty} \boldsymbol{\pi}^{(1)} \cdot \mathbf{A}^{t-1} = \boldsymbol{\pi}^{(1)} \cdot \mathbf{A}^\infty \quad \text{(limiting distribution of a Markov chain)}$$

Limiting distribution of a Markov chain

- Properties of the limiting distribution

Since $\pi^{(t)} = \pi^{(1)} \cdot A^{t-1} = \pi^{(2)} \cdot A^{t-2} = \dots$,

$$\pi^{(\infty)} := \pi^{(1)} \cdot A^{\infty} = \pi^{(2)} \cdot A^{\infty} = \dots$$

$$\pi^{(\infty)} := \lim_{t \rightarrow \infty} \pi^{(1)} \cdot A^{t-1}$$

$$= \left(\lim_{t \rightarrow \infty} \pi^{(1)} \cdot A^{t-2} \right) \cdot A$$

$$= \pi^{(\infty)} A$$

- $\pi^{(\infty)}$ does not depend on the initial distribution.
- The transition probabilities determine the long-term behavior.
- A^t converges to A^{∞} if $\{S_t\}$ is irreducible and aperiodic.

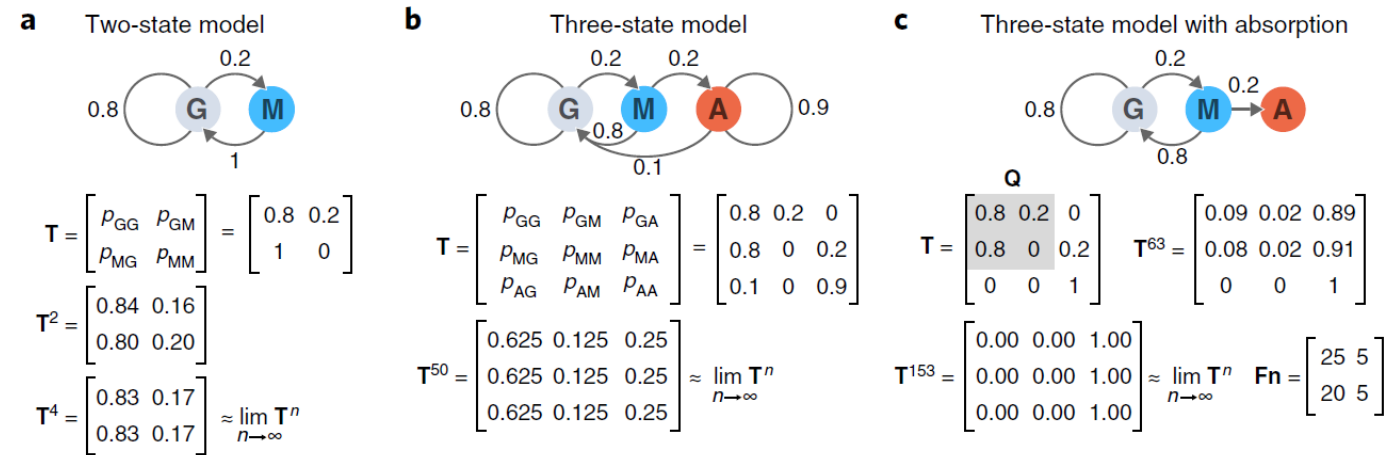


Fig. 1 | State transition models, transition matrices T , and the number of transitions required to approximate the steady-state limiting distributions, T^n ($n \rightarrow \infty$), to the displayed number of decimal places. **a, A two-state model in which a cell in the growth phase (G) can undergo mitosis (M) with a probability $p_{GM} = 0.2$. **b**, A three-state model in which the cell may enter temporary arrest (A) from M with a probability of $p_{MA} = 0.2$ but will return to G with a probability $p_{AG} = 0.1$. **c**, A three-state absorption model in which the cell remains in arrest forever. The number of time steps spent in a state before absorption is given by the fundamental matrix, $Fn = (I - Q)^{-1}$, where Q is the highlighted submatrix.**

Hidden Markov Model (HMM)

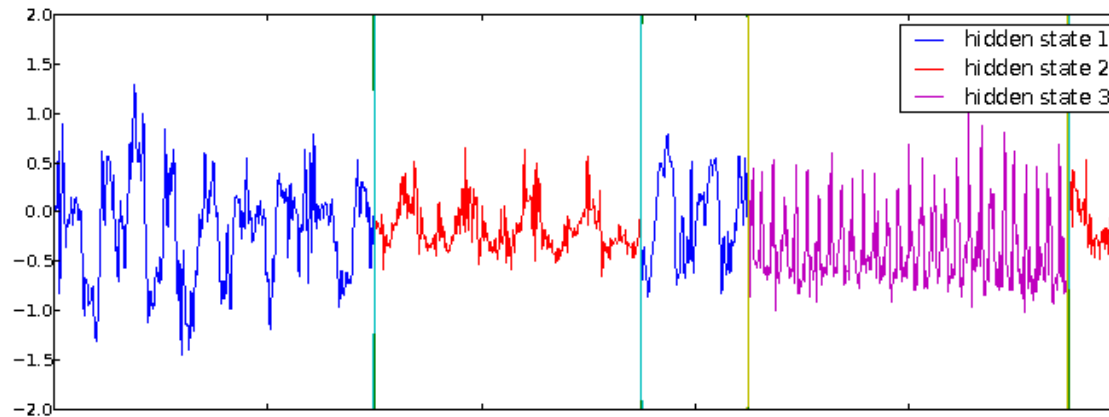
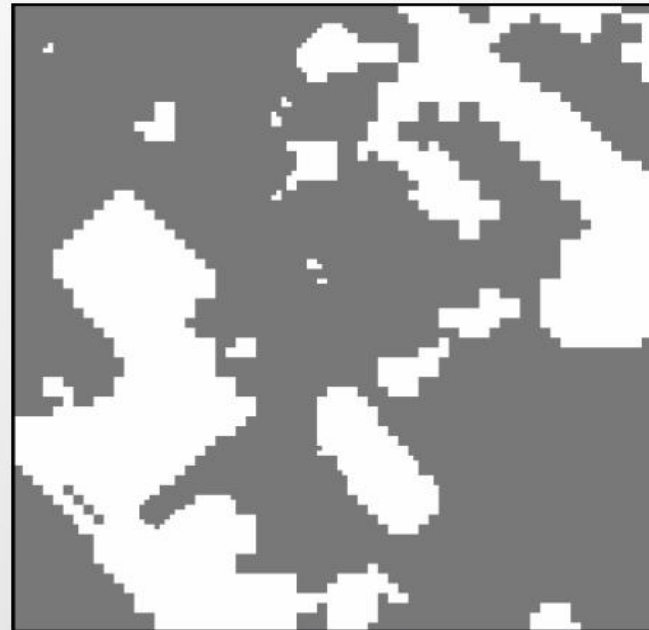


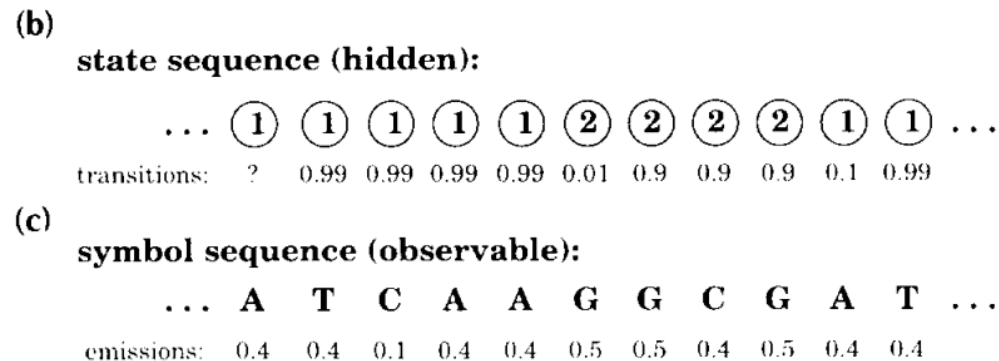
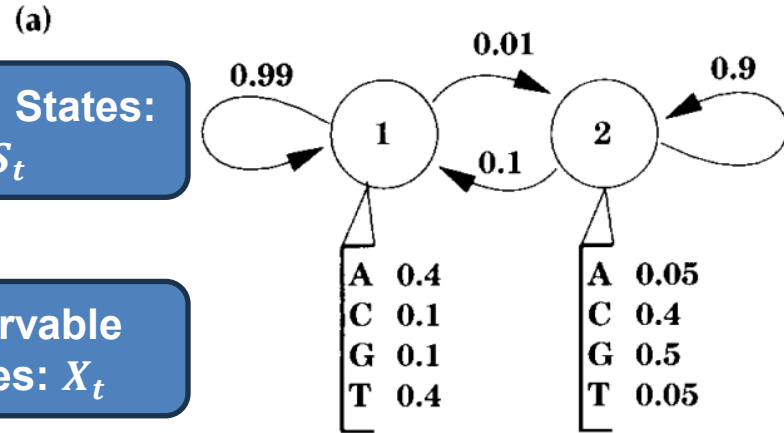
Figure 1: Human activities accelerometer data. short sequence.

For temporal data



For spatial data
- Binarize into artificial
and natural areas

Hidden Markov Model (HMM)



- A HMM assumes that the observations can be generated by multiple mechanisms (or regimes), which are latent (or unobservable) and follows a Markov chain.
 - Economic states = {recession, expansion}
 - A mixture model. EM-algorithm.
- Initial probability (say, π)
 - $\pi_i = P(S_1 = i), i = 1, 2, \dots, M$ (# of hidden states)
 - Transition probability (say, A)
 - $a_{ij} = P(S_{t+1} = j | S_t = i)$
 - PDF of X_t for a given $S_t = i$ (say, B)
 - $b_i(x) = P(X_t = x | S_t = i)$ (emission probability, discrete)
 - or $pdf_{X_t}(x | S_t = i)$ (continuous)
 - $b_i(x_t | x_{t-1}) = pdf_{X_t}(x_t | S_t = i, X_{t-1} = x_{t-1})$
(X_t may also depend on X_{t-1} . A dependent mixture model.)

Outline

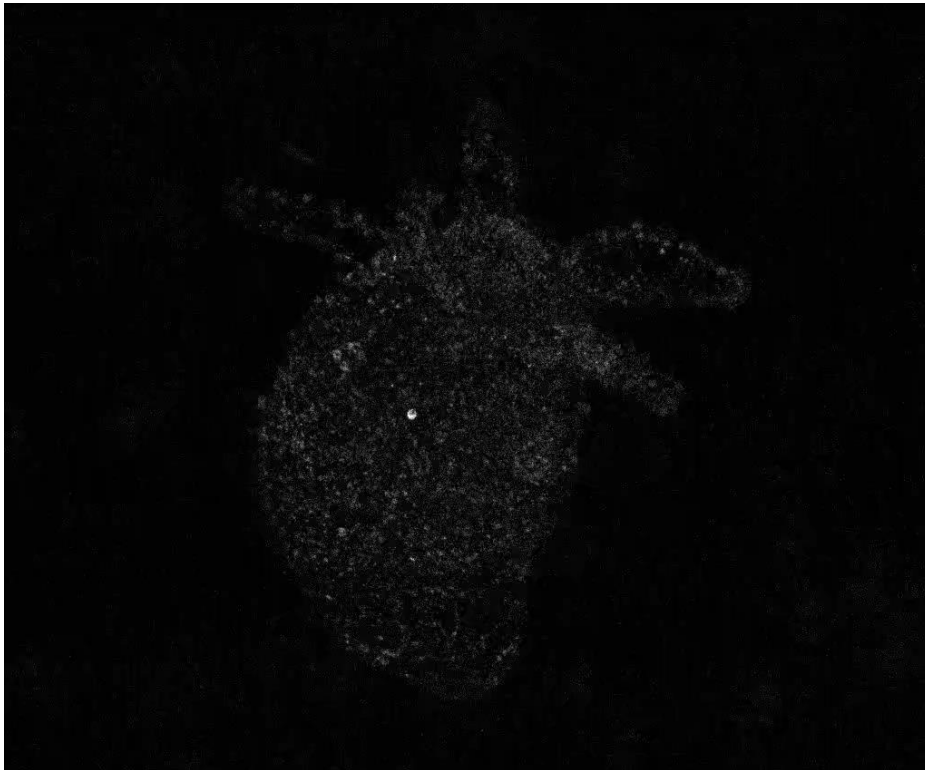
1. Time Series or Stochastic Processes
2. Markov Chain
3. Hidden Markov Model (HMM)
4. Calcium imaging data and calcium spike inference problem

Recent advances in calcium imaging

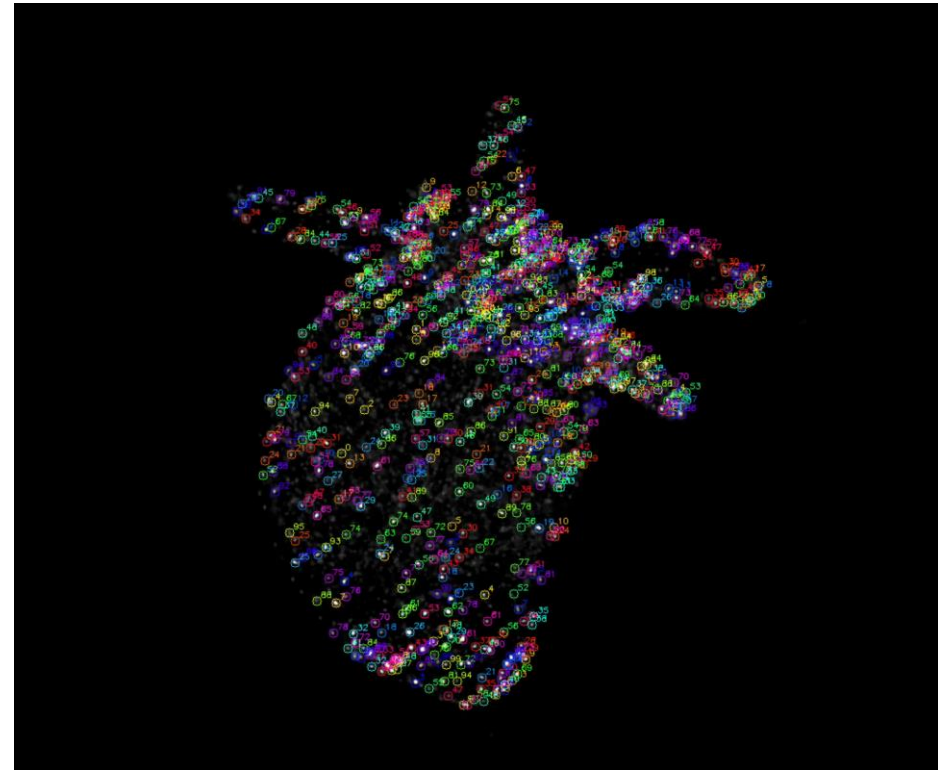
- “Every spike from every neurons in a behaving Hydra”

- Dual channel transgenic Hydra, 2D confocal

Calcium activity (GCaMP7s)



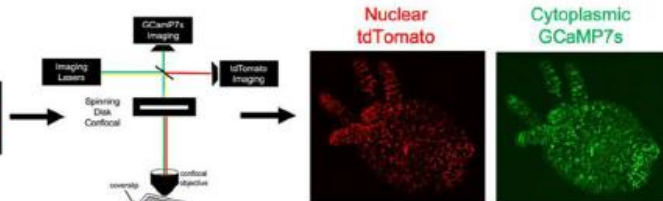
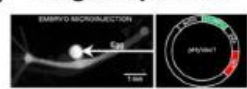
Nucleus (tracked)



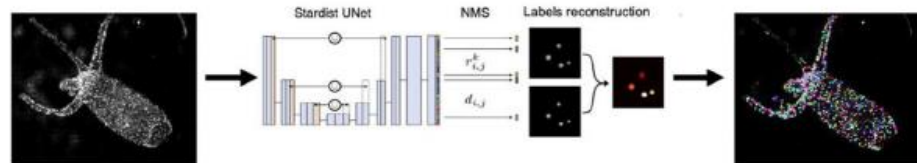
Tracking and Spike Estimation of Individual Neurons

PIPELINE OVERVIEW

I) Image acquisition

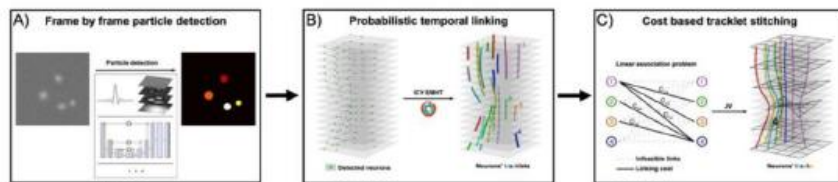


II) tdTomato nuclei segmentation



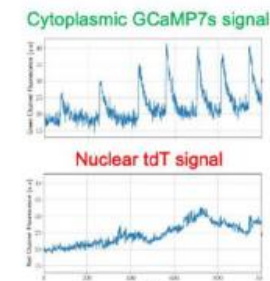
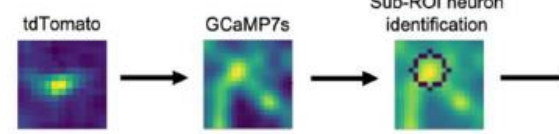
Stardist UNet

III) tdTomato nuclei tracking

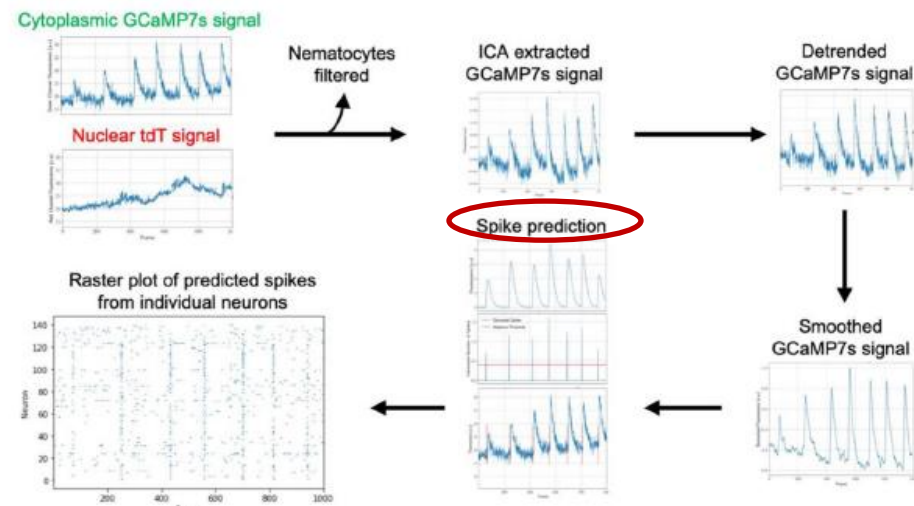


eMHT in Icy software

IV) GCaMP7s signal extraction



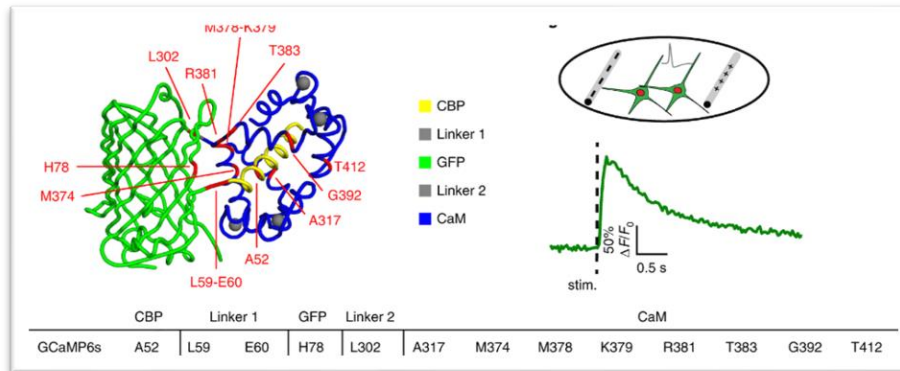
V) GCaMP7s signal processing and spike prediction



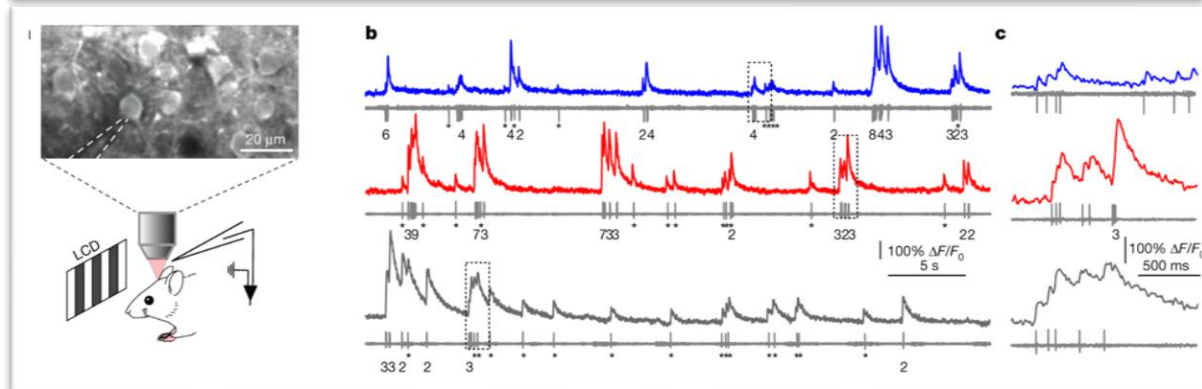
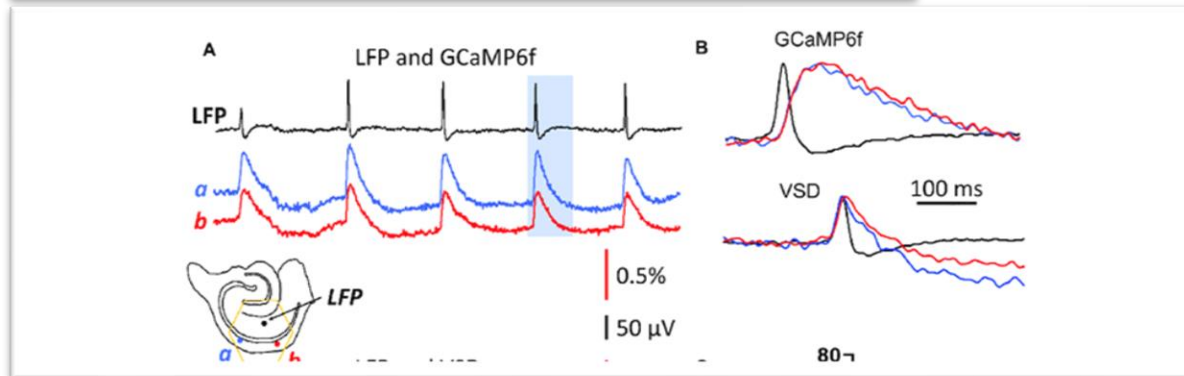
OASIS in CalmAn

Action potential and calcium sensor activity

GCaMP



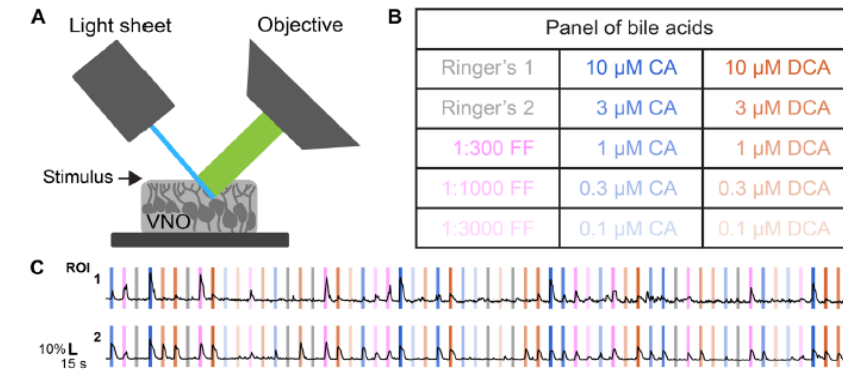
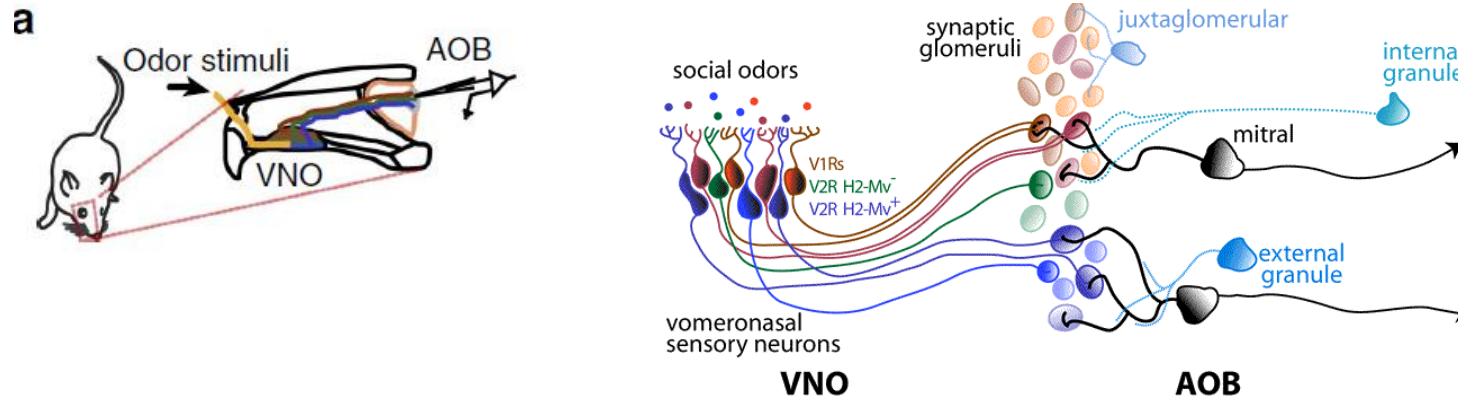
- An electric spike results in fast increase of calcium fluorescence activity followed by slower decaying.



- Spike estimation outcomes by CalmAn

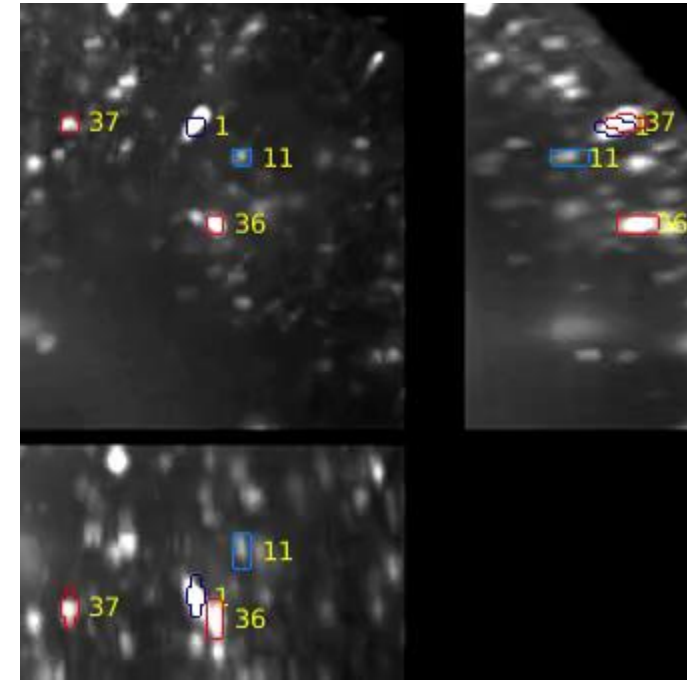
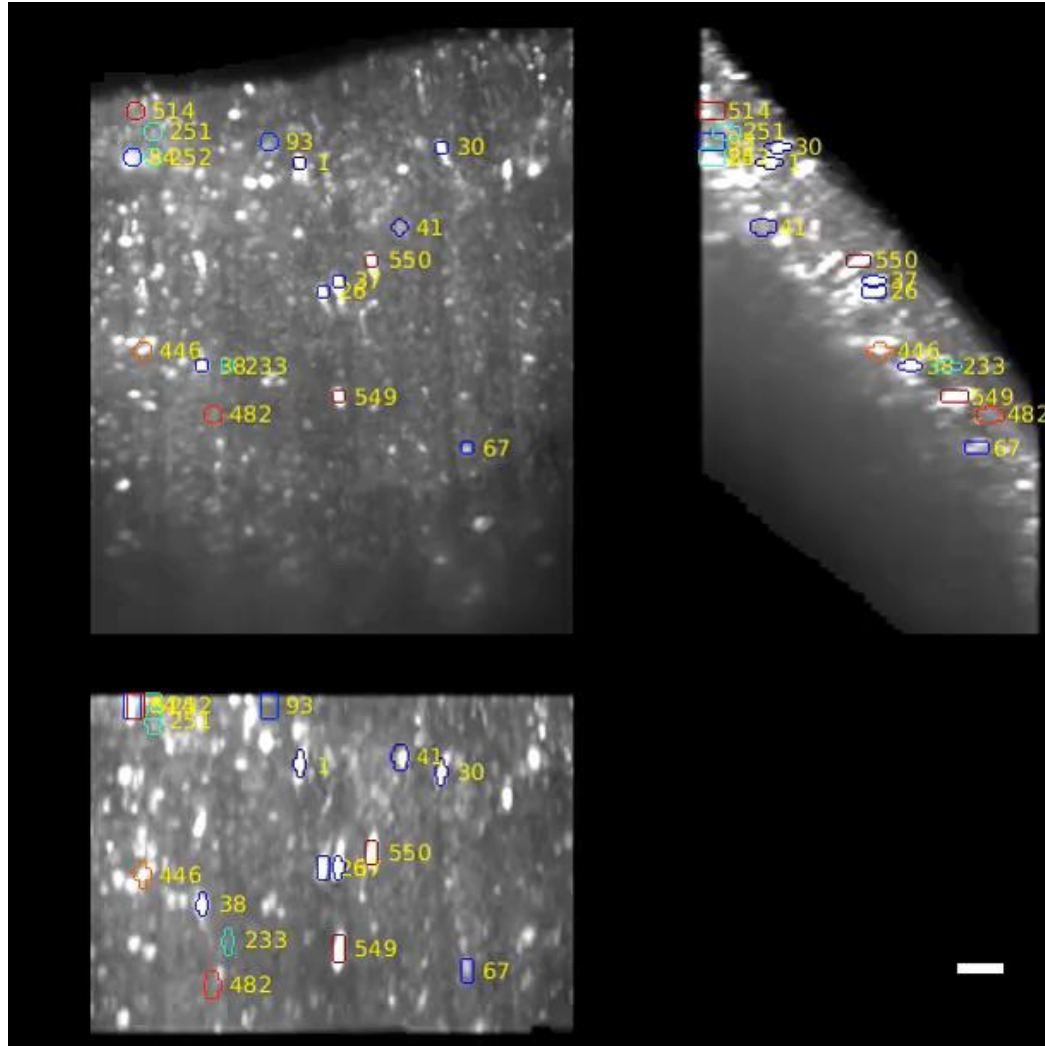
Chemosenstation of social odors (How mice detect pheromones)

Julian Meeks
@U.Rochester

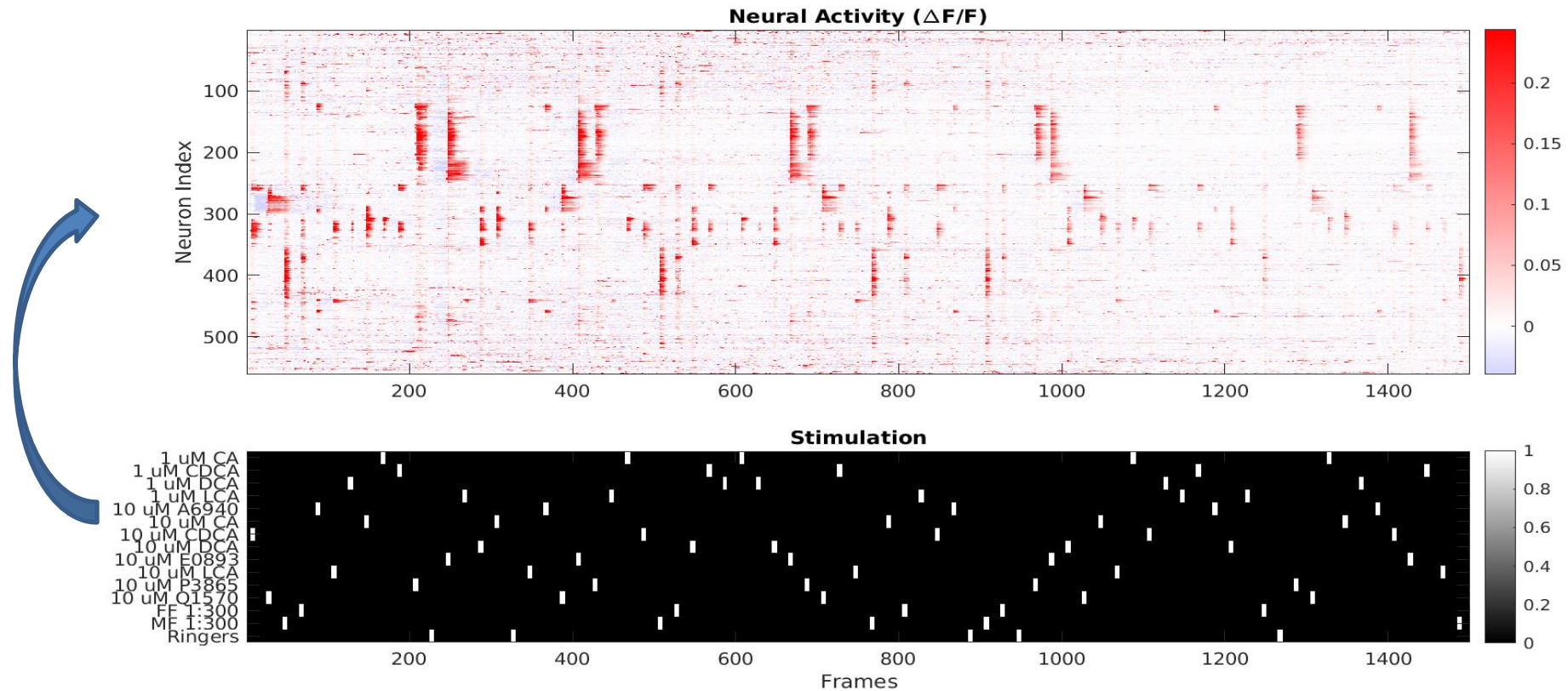


- Vomeronasal organ (VNO)
 - detects social odors (or pheromones) like urine, feces, tears, saliva, etc in many mammals.
- Vomeronasal sensory neurons (VSNs)
 - generally express 1 receptor among ~300 unique vomeronasal receptors.
- Multiple randomized stimulation of known ligands
 - 4 bile acids (CA, DCA, LCA, CDCA) in mouse feces, and 4 sulfated steroids in mouse urine (estrogen (E), androgen (A), pregnanolone (P), glucocorticoid (Q)).
- Q: How the pheromone sensing neurons encode the 8 known ligands? (ligand-receptor mapping)

Dynamic Neuron Tracker (DyNT) segments and tracks jittering single neurons in 3D imaging of deforming tissues



Ca²⁺ activity time courses in response to repeated stimulations



- Stimulations: 15 kinds (4 bile acids, 4 steroids, FF, MF (+ control), Ringer's (- control))
- 561 neurons
- A combination of stimuli can trigger a neuron.
- How to determine the relations bet'n single neuron activity TS and stimulation sequences?

Online Active Set method to Infer Spikes (OASIS) in CalmAn

- Plos comp bio, 2017.

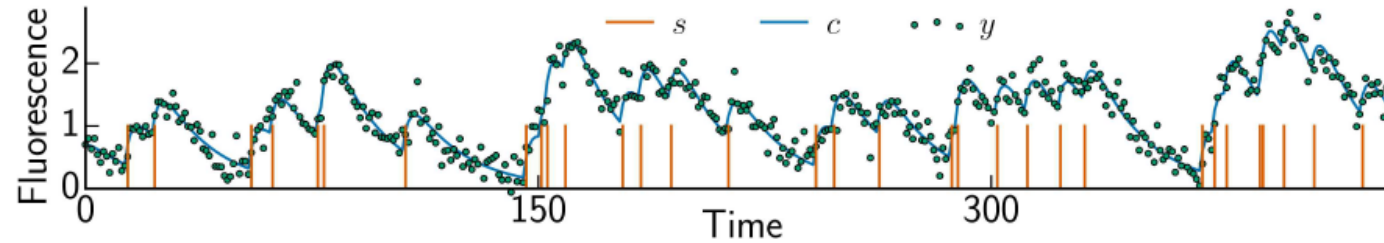


Fig 1. Generative autoregressive model for calcium dynamics. Spike train s gets filtered to produce calcium trace c ; here we used $p=2$ as order of the AR process. Added noise yields the observed fluorescence y .

- Auto-Regressive model with spike errors

$$c_t = \sum_{i=1}^p \gamma_i c_{t-i} + s_t.$$

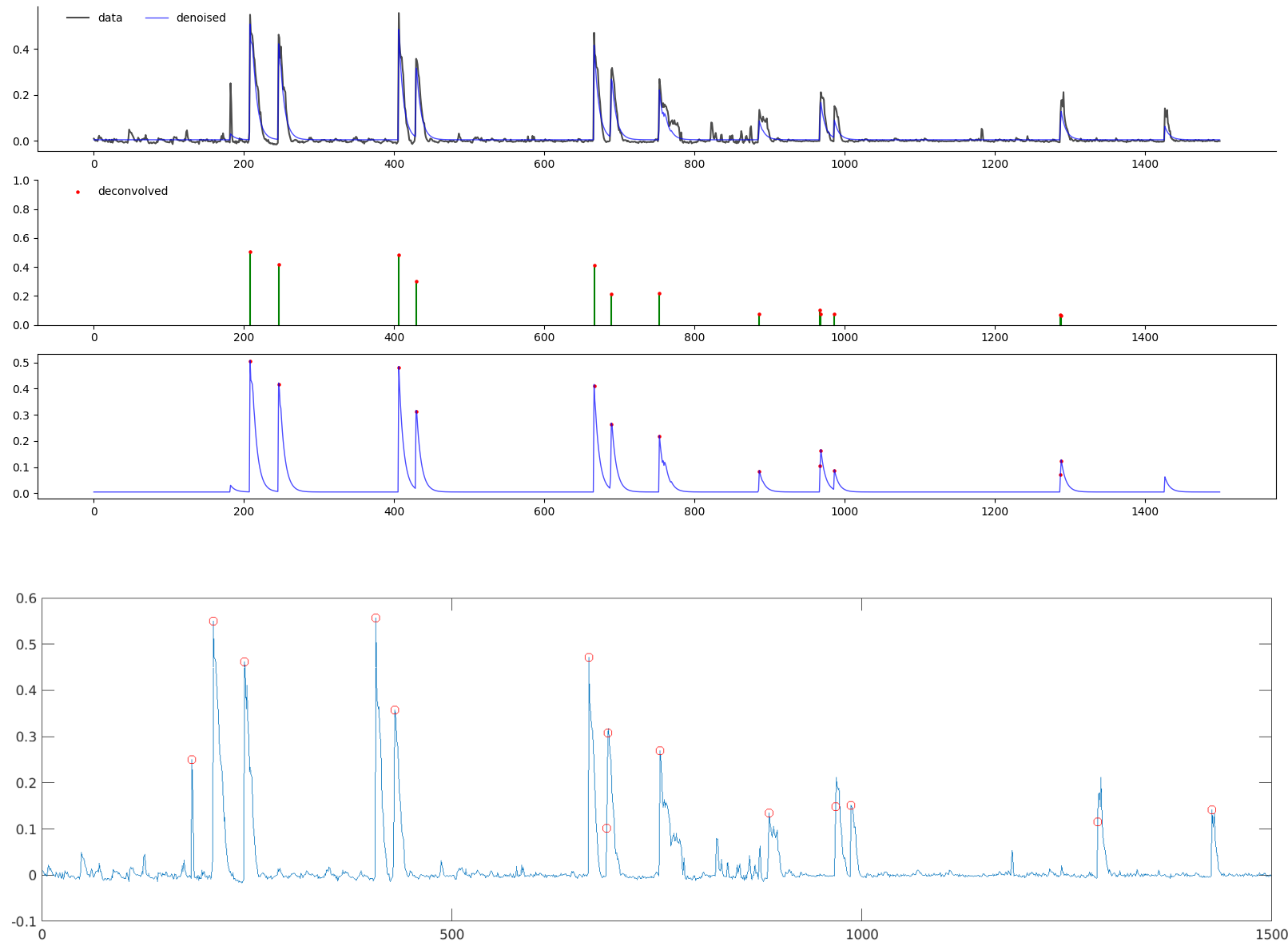
The observed fluorescence $y \in \mathbb{R}^T$ is related to the calcium concentration as [5–7]:

$$y_t = a c_t + b + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2)$$

- Constrained optimization

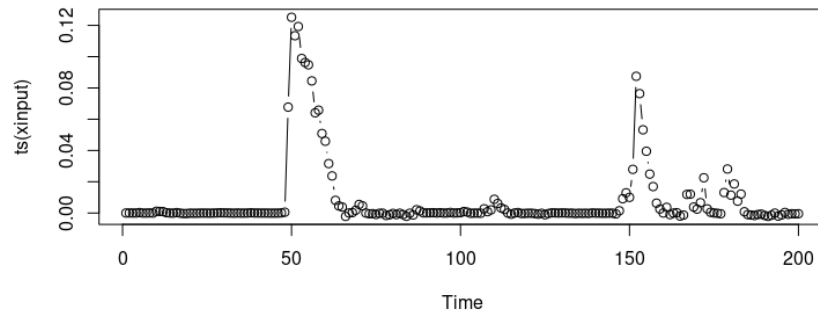
$$\underset{\hat{c}, \hat{s}}{\text{minimize}} \quad \frac{1}{2} \|\hat{c} - y\|^2 + \lambda \|\hat{s}\|_1 \quad \text{subject to} \quad \hat{s} = G\hat{c} \geq 0$$

OASIS (top) vs CASPER (Calcium Spike Estimator) (bottom)



- CASPER does not assume a constant decaying rate.
- OASIS requires a threshold to remove small spikes.

Autoregressive-HMM of calcium activity time series

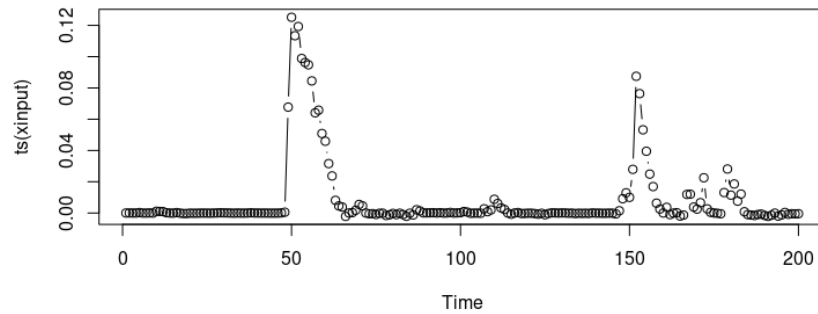


X_t : a (normalized) Ca^{2+} activity

A_t : a hidden state

$$X_t = \alpha_{A_t} + \beta_{A_t} X_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_{A_t}^2)$$

Autoregressive-HMM of calcium activity time series

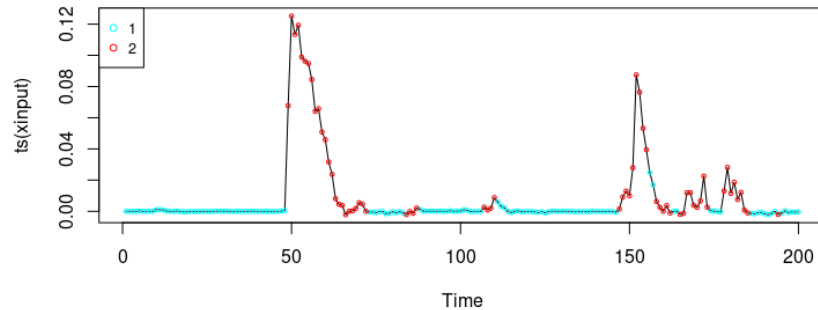


X_t : a (normalized) Ca^{2+} activity

A_t : a hidden state

$$X_t = \alpha_{A_t} + \beta_{A_t} X_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_{A_t}^2)$$

Neuron 33

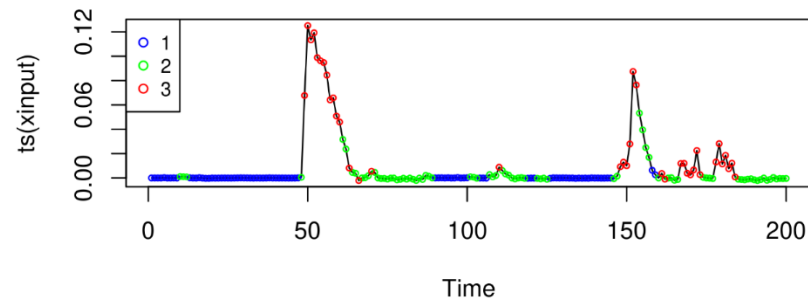


- Fitting 2-state model

	Re1.(Intercept)	Re1.n30.1	Re1.sd
St1	0.000	0.637	0.000
St2	0.003	0.889	0.016

- Degree of freedom (d.f.) = $1 + 2 \cdot 1 + 2 \cdot (3) = 9$

Neuron 33

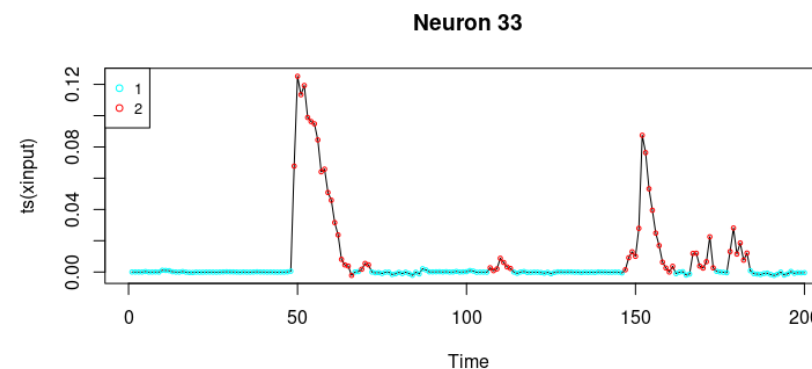
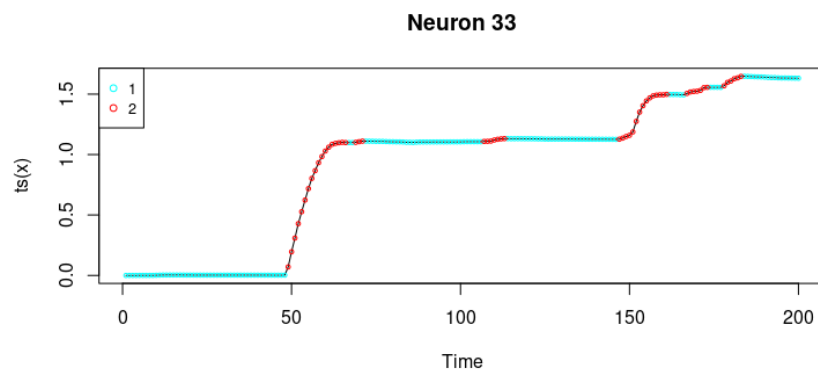


- Fitting k-state model, $k=\{1, 2, \dots, 6\}$
- BIC selected $k=3$ state model.
 - d.f. ($k=3$) =

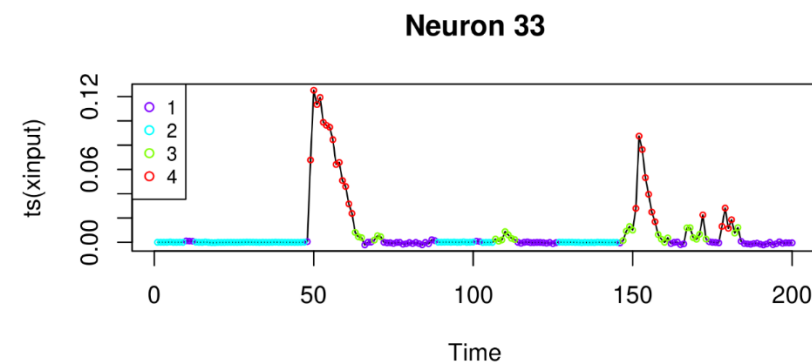
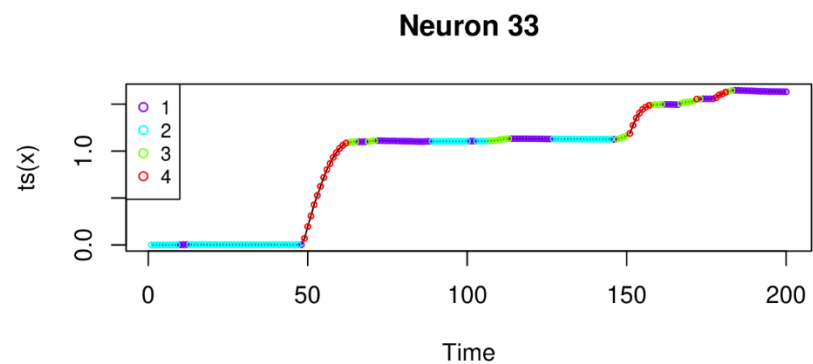
* When AR-HMM is fitted for derivative time series, it is too sensitive to noises.

Autoregressive-HMM of calcium activity time series (cumulative sum)

- Fitting 2-state model



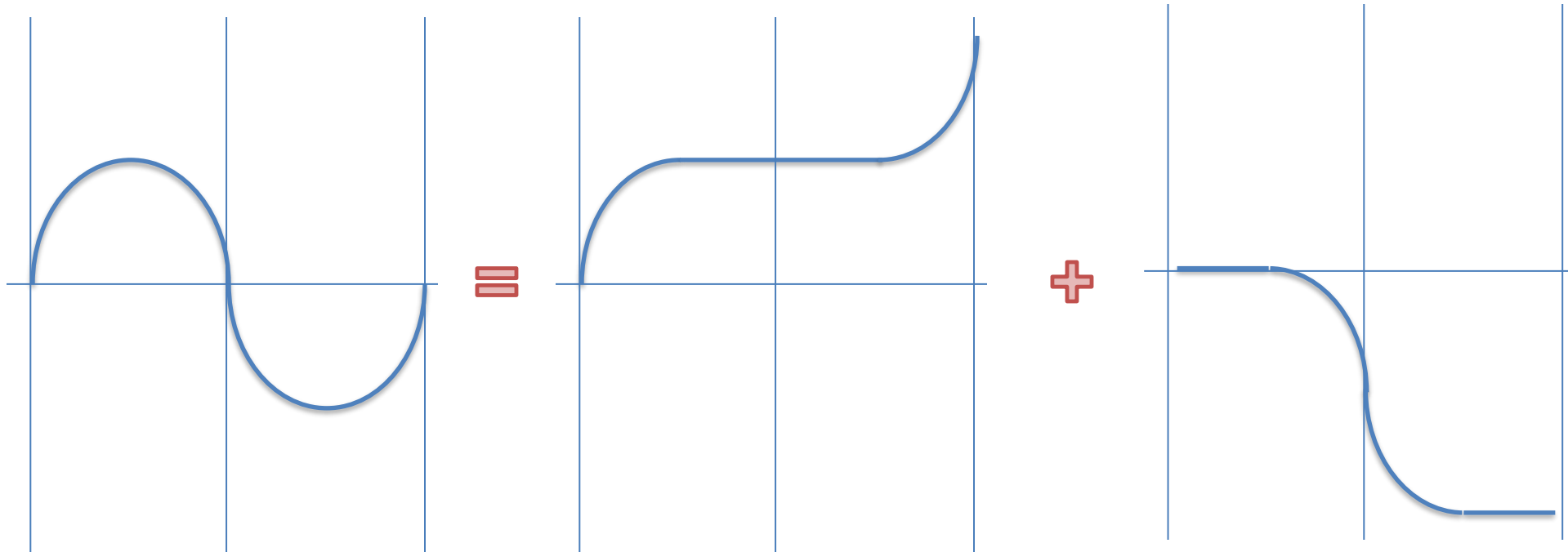
- BIC selected k=4 state model.



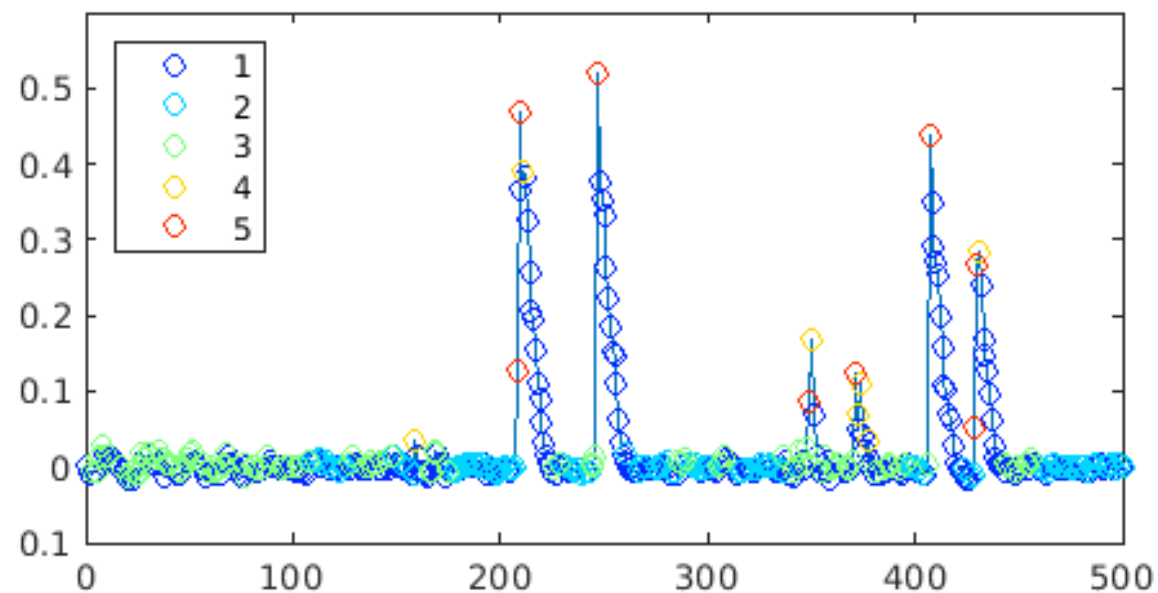
- Detects not spikes but pulses.

Jordan decomposition of a function

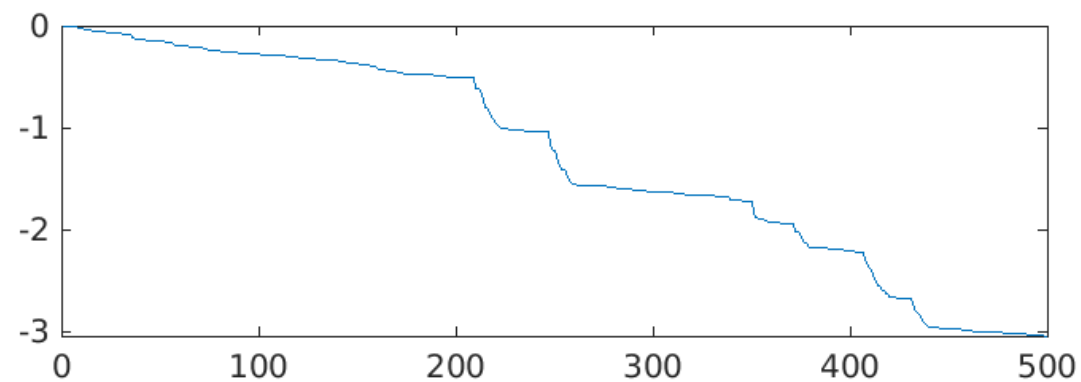
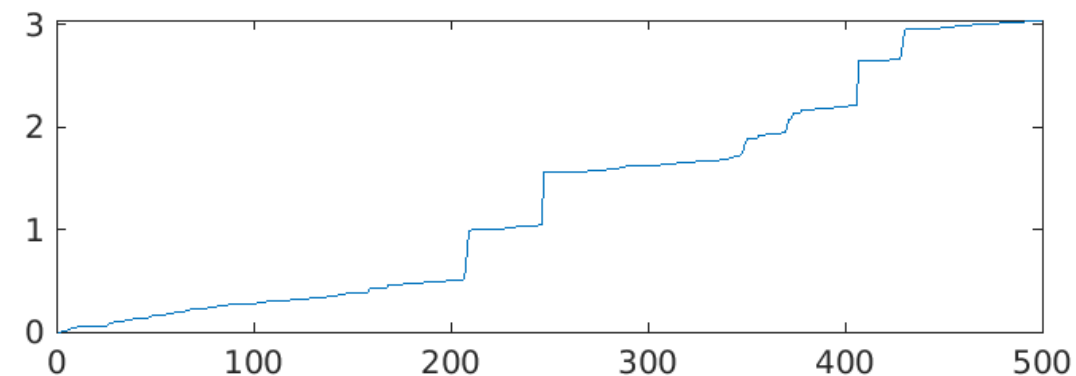
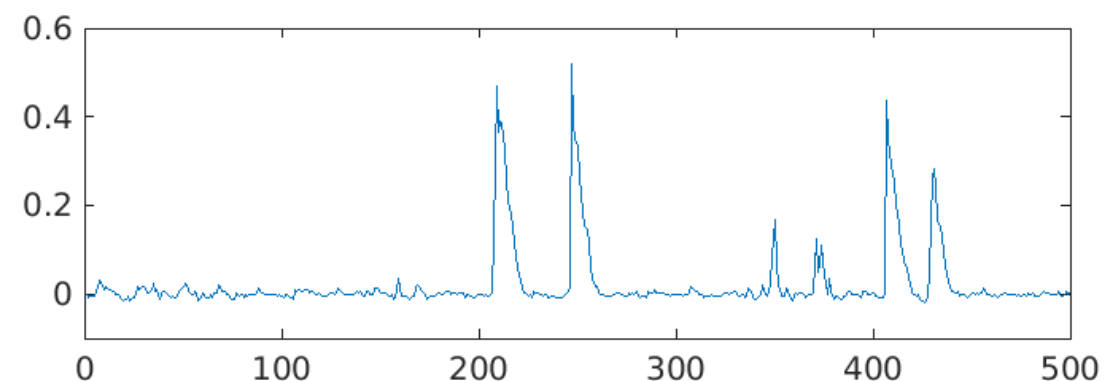
- A concept from measure theory
- If $f: [a, b] \rightarrow \mathbb{R}$ is a function of bounded variation, then there exists a pair of nondecreasing functions f^+ and f^- such that $f(x) = f^+(x) - f^-(x)$. The pair is unique up to addition of a constant.
- Roughly speaking, f^+ represents the cumulative sum of increments of f , and $-f^-$ decrements. (exactly, for a sequence).



AR-HMM after Jordan decomposition

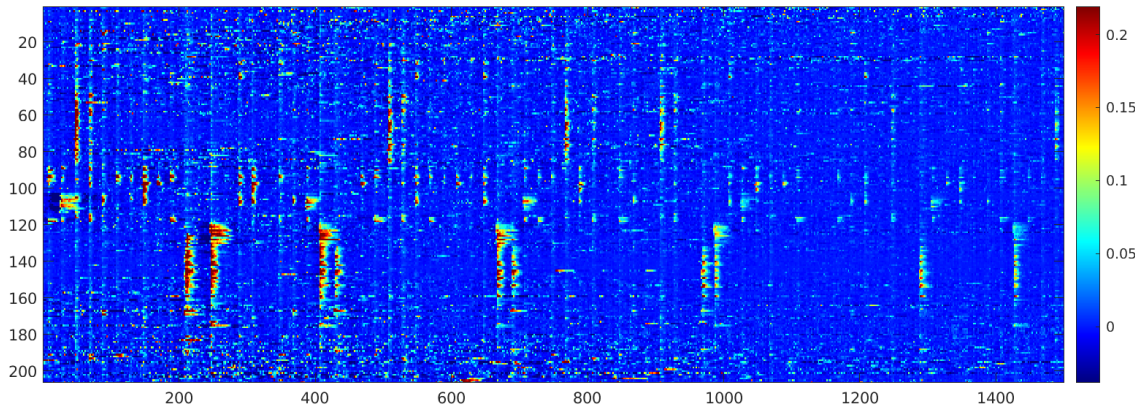


- $X_t = \alpha_{A_t} + \beta_{A_t} X_{t-1} + \epsilon_t$
- State 1: decrements
- State 5: spikes

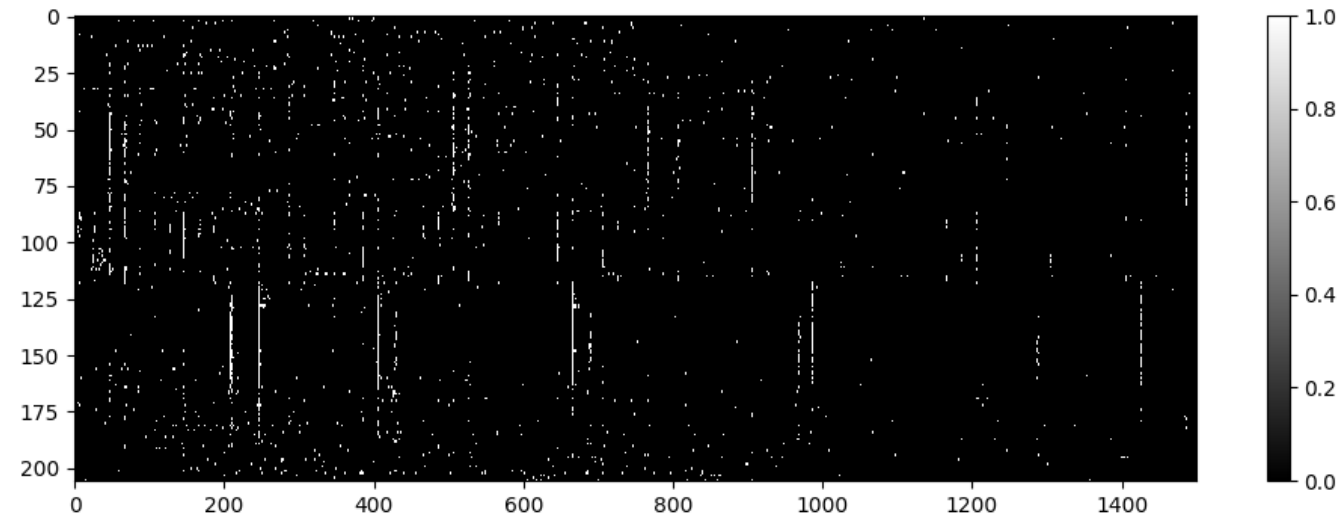


Choice of the number of states

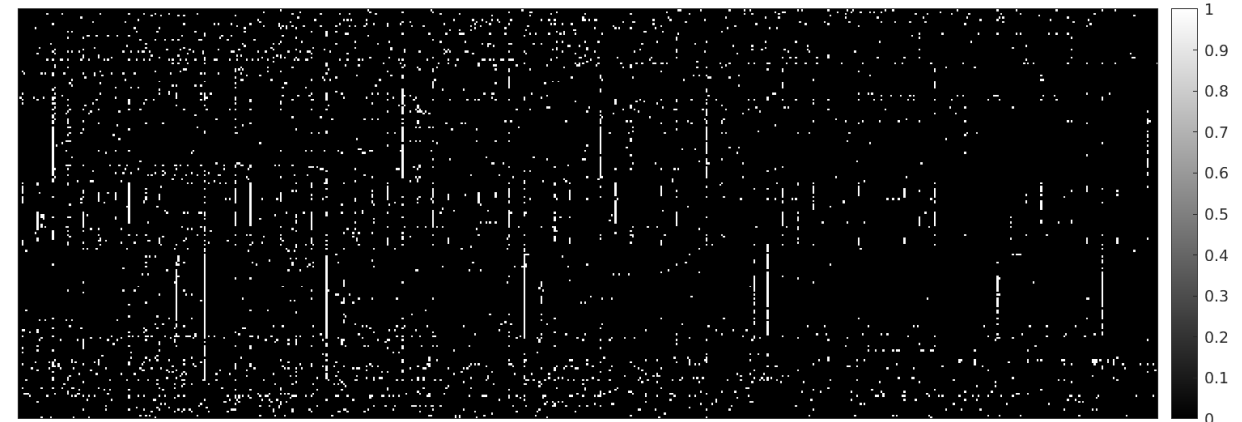
- Users specify a set of candidate numbers of states, and then BIC selects the optimal model for each neuron, while also comparing with no spike model.
- Eg) Input numbers of states (K) = [2, 4, 5].
 $K=2$ is the case of no spike.



• OASIS



• CASPER



No efficient library to implement EM algorithm for AR-HMM

- In Python and Matlab, we can easily find libraries or personal codes to estimate 'simple' HMMs, but not 'dependent' HMMs.
- R package: 'depmix' (dependent mixture)
 - slow
- **Matlab codes** to estimate AR(1)-HMMs
- Coding exercise
 - Complete `Estep_AR1HMM_template.m` under `./days2-3` folder (deadline Tuesday, 1PM).
 - Run the AR-HMM analysis for the sample dataset of 206 neurons' activity time courses.
 - Present your code and interesting analysis outcomes on Wed (~10 min).
- Math exercise
 - Resolve numerical instability issues in the E-step by introducing scaling factors.
 - Group A/B solves the issues in forward/backward equations, respectively.
 - Everyone submits a report for the forward or backward eqn.
 - Each group representative presents the fixed equations on Wed.

E-step: forward and backward equations

- $\alpha_t(j) = P(X_1 = x_1, \dots, X_t = x_t, S_t = j) = \{ \sum_{i=1}^M \alpha_{t-1}(i) a_{ij} \} \cdot b_j(x_{t+1} | x_t)$ (Group A)
- $\beta_t(i) = P(X_{t+1} = x_{t+1}, \dots, X_T = x_T | S_t = i, X_t) = \sum_{j=1}^M a_{ij} b_j(x_{t+1} | x_t) \beta_{t+1}(j)$ (Group B)
- $\gamma_t(i) = P(S_t = i | X_1 = x_1, \dots, X_T = x_T) = \alpha_t(i) \beta_t(i) / \sum_{i=1}^M \alpha_t(i) \beta_t(i)$

M-step

- MLEs have the form of weighted averages