



Calibration of stochastic biochemical models using single-cell video-microscopy experiments

Andela Davidović

► To cite this version:

Andela Davidović. Calibration of stochastic biochemical models using single-cell video-microscopy experiments. Biohasard, Aug 2019, Rennes, France. pasteur-03402738

HAL Id: pasteur-03402738

<https://pasteur.hal.science/pasteur-03402738>

Submitted on 25 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CALIBRATION OF STOCHASTIC BIOCHEMICAL MODELS

USING SINGLE-CELL VIDEO-MICROSCOPY EXPERIMENTS

Davidović A, Batt G, Ruess J

BioHasard, 29 August 2019

Contents



PART 1

State of the art

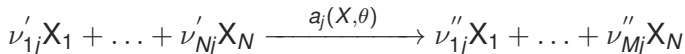
1.1 Biochemical networks

- Biochemical networks perform computational tasks like electronic circuits
- performed by molecules that chemically and physically interact with each other
- These are intrinsically **stochastic** processes where molecular populations are integers that evolve stochastically

1.1 Biochemical networks

Stochastic chemical kinetics

- $\mathcal{S}_1, \dots, \mathcal{S}_N$ - species; $\mathcal{R}_1, \dots, \mathcal{R}_M$ - reactions
- $X_i(t)$ - number of molecules of $\mathcal{S}_i$ at time t



- ν_{ij} - change in X_i due to reaction $\mathcal{R}_j$
- propensity
 $a_j(x, \theta)dt := P(\mathcal{R}_j \text{ occurs in time } [t, t + dt) \mid X(t) = x)$

Time evolution equation - Chemical Master Equation

$$\frac{dP(x, t)}{dt} = \sum_{j=1}^M [a_j(x - \nu_j, \theta)P(x - \nu_j, t) - a_j(x, \theta)P(x, t)]$$

1.2 Learn parameters of the system

Microscopy data

- some of the chemical species are observable (fluorescent)
- take measurements of one cell $Y = \{y_1, \dots, y_S\}$ at times $t_1, \dots, t_S$

$$y_s = C \cdot X(t_s) + \xi_s, \quad \xi_s \sim \mathcal{N}(0, \sigma^2)$$

- Y is the observable part of the full system X

To **learn the parameters of the system** we need to compute $P(Y|\theta)$.

Problem:

CME is usually not solvable and is computationally intractable.

1.2 Learn parameters of the system

Linear noise approximation (LNA) of CME

$$X(t) = \varphi(t) + \xi(t)$$

- $\varphi(t) = (\varphi_1(t), \dots, \varphi_N(t))$ - macroscopic state (ODEs)
- $\xi(t)$ - stochastic process

Full state of the system for all measurement times is

$$\mathcal{X} = (X(t_1), \dots, X(t_S)) \sim \mathcal{N}(\mu(\theta), \Sigma(\theta))$$

where θ - parameters of the system; $\mu(\theta) = (\varphi(t_1), \dots, \varphi(t_S))$
 $\Sigma(\theta)$ is a covariance matrix of size $NS \times NS$, where each element is a solution to another couple of ODEs.

- LNA issues:
 - computationally intensive
 - does not work for complex systems

1.2 Learn parameters of the system

LNA + Kalman Filter

The single cell likelihood can be factorised

$$P(Y | \theta) = P(y_1 | \theta) \cdot P(y_2 | y_1, \theta) \cdots P(y_S | y_1, \dots, y_{S-1}, \theta)$$

Use Kalman Filter at each measurement time:

LNA over $[0, t_1]$

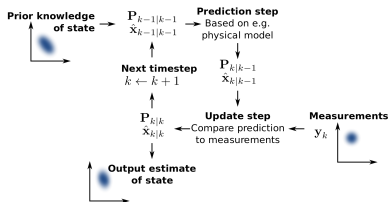
Iterate for $s = 2, \dots, S$

1. KF to obtain

$$P(x_{s-1} | y_{s-1}, \dots, y_1, \theta)$$

2. LNA over $[t_{s-1}, t_s]$ to obtain

$$P(y_s | y_{s-1}, \dots, y_1, \theta)$$



Complexity in $N^2 S$ instead of $N^2 S^2$.



PART 2

In silico testing

2.1 Accuracy and precision

A positive feedback loop

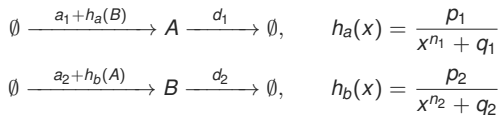
$$\emptyset \xrightarrow{a_0 + a \cdot h(X)} X \xrightarrow{d} \emptyset, \quad h(x) = \frac{x^n}{x^n + K^n}$$

blue - exact, red - LNA+KF, green - open-loop LNA

- Bimodal distribution
- Learning $\theta = a$, using different $t_m = t_{s+1} - t_s$, comparing LNA vs LNA+KF
- Accurate estimate using LNA + KF in the relevant domain - around maximum estimator

2.1 Accuracy and precision

Genetic Toggle Switch



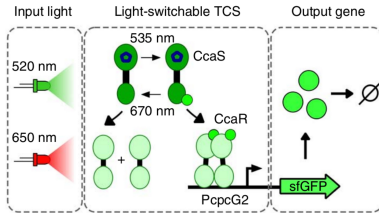
- Model of repression and production of proteins A and B , measuring only B .
- Learning d_2 , using LNA+KF, and different $t_m = t_{s+1} - t_s$
- Overall estimations are very good, especially for small t_m



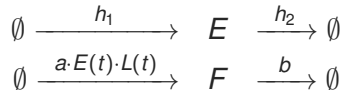
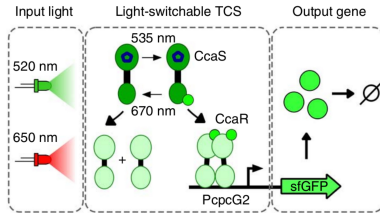
PART 3

Parallelized single-cell experiments

3.1 Model: Green light/Red light



3.1 Model: Green light/Red light



We measure $s \cdot F$

- E - responsiveness of the cell
- F - fluorescence
- $u(t)$ - light control ON/OFF
- $\theta = \{a, b, s, \mu_E, h_2, c, n_H, K\}$

$$L(t) = \frac{(c \cdot H(t))^{n_H}}{(c \cdot H(t))^{n_H} + K^{n_H}}$$

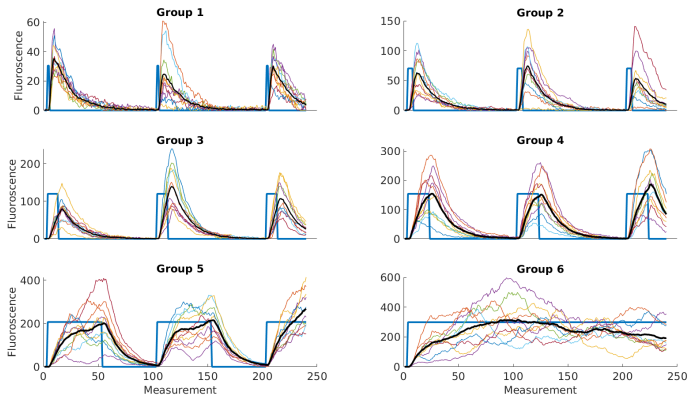
$$\frac{dH(t)}{dt} = u(t) - c \cdot H(t)$$

3.2 Biological questions

- Q1 For a fixed number of cell trajectories what is the benefit of diversifying light stimulation patterns at the single cell level?
- Q2 What is the robustness of the proposed method for parameter inference with respect to a low number of cell trajectories?

3.3 Numerical results

Simulated data



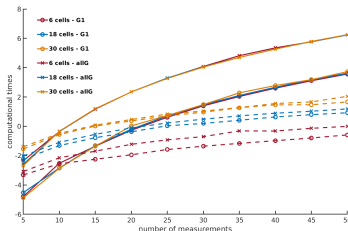
3.3 Numerical results

Computational times LNA vs LNA+KF

- likelihood for exact parameters
- for one group (G1)
- for all groups (allG)
- for different numbers of cells

Observation

- LNA+KF linear growth
- LNA quadratic



3.3 Numerical results

Calibration of the model using LNA+KF

Learning from 60 cells.

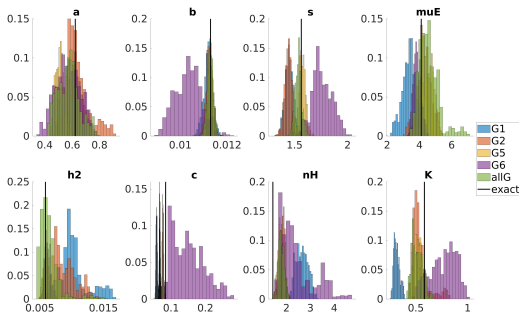
Comparing posteriors of groups after 10 000 runs of MCMC.

3.3 Numerical results

Calibration of the model using LNA+KF

Learning from 60 cells.

Comparing posteriors of groups after 10 000 runs of MCMC.

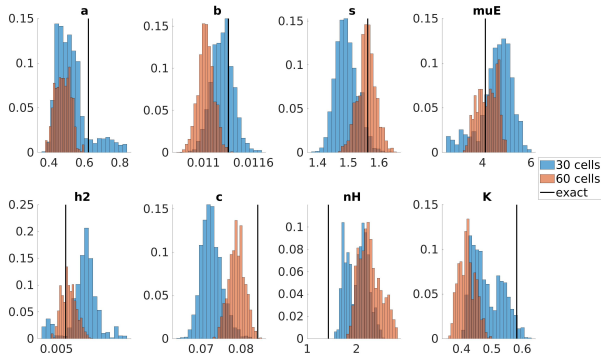


This work used the computational and storage services (TARS) provided by the IT department at Institut Pasteur.

3.3 Numerical results

Posteriors with different numbers of cells

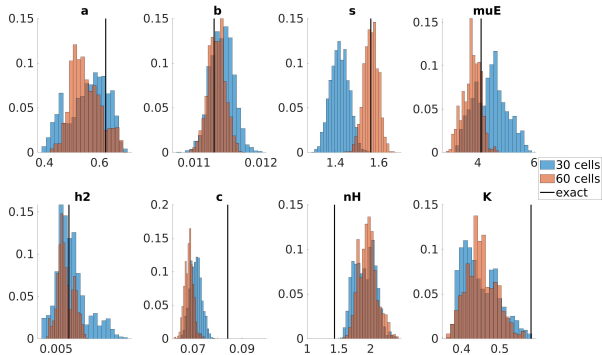
For G5



3.3 Numerical results

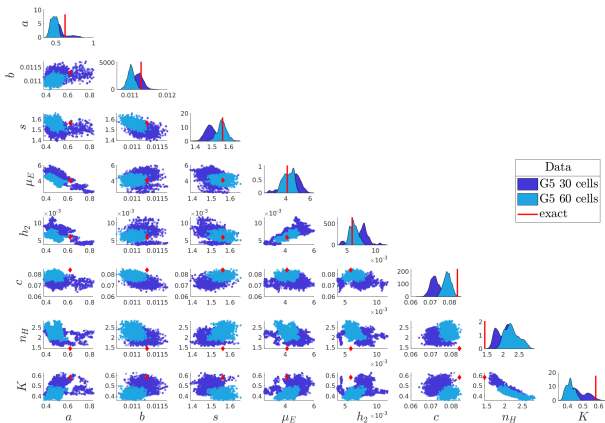
Posteriors with different numbers of cells

For allG



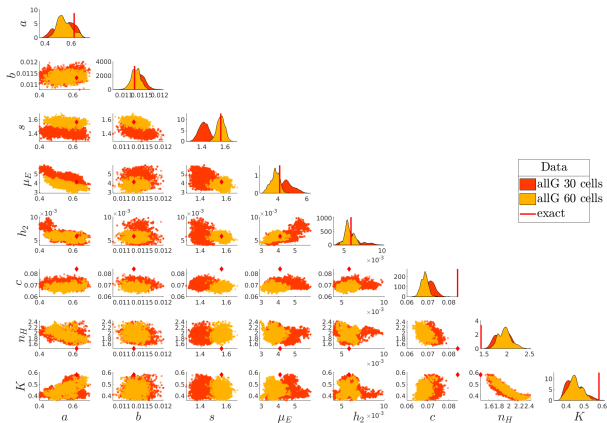
3.3 Numerical results

Scatter plots



3.3 Numerical results

Scatter plots



3.3 Numerical results

What is the point?

- If the time scale of the light signal is well chosen, learning from cells of only this group (G5) or from cells with variable light patterns (allG) is very similar
- small preference for allG
- variable light signal is more **robust!**



PART 4

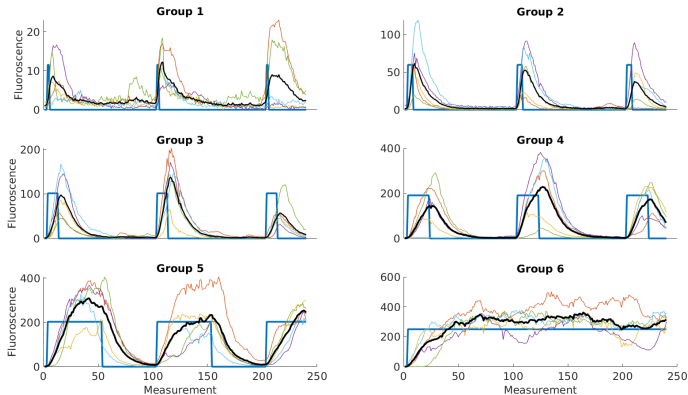
Experimental data

4.1 Calibration with experimental data

Data

30 cells in total, 5 per group.

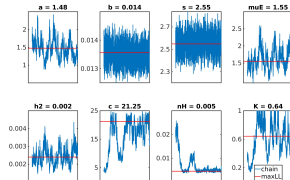
Measurements every 6 min during 24 hours (after cleaning the data).



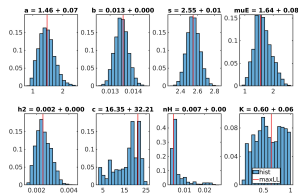
4.1 Calibration with experimental data

- 60000 iterations of MCMC
- MLL estimations
 - $a = 1.5$
 - $b = 0.0135$
 - $s = 2.5$
 - $\mu_E = 1.5$
 - $h_2 = 0.002$ temporal fluctuations of cells responsiveness to light
- Non-identifiable parameters
 - c, n_H, K - hill function parameters

parameter chains and MLL:



histograms, means and vars:



4.1 Calibration with experimental data

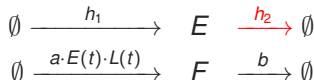
60000 iterations of MCMC

MLL estimations

- $a = 1.5$
- $b = 0.0135$
- $s = 2.5$
- $\mu_E = 1.5$
- $h_2 = 0.002$ temporal fluctuations of cells responsiveness to light

Non-identifiable parameters

- c, n_H, K - hill function parameters



We measure $s \cdot F$

$$L(t) = \frac{(c \cdot H(t))^{n_H}}{(c \cdot H(t))^{n_H} + K^{n_H}}$$

$$\frac{dH(t)}{dt} = u(t) - c \cdot H(t)$$

Thanks for your attention!

Take-home messages:

- LNA with Kalman Filter is a good method for approximating likelihood computation
- Computationally convenient for the systems with a lot of measurement times, precise even for a small number of trajectories.
- Diversifying light patterns in light-inducible cells allows for more precise and more robust learning of parameters of the system.