Identification of biological models from single-cell data

Andres Mauricio Gonzalez Vargas

Joint work with
Giancarlo Ferrari Trecate (UNIPV),
J. Uhlendorf , A. Llamosi, J. Schaul, E. Cinquemani, G. Batt (INRIA)

Università degli Studi di Pavia

Università degli Studi di Pavia

Identification & Control Lab

Giancarlo Ferrari Trecate

Andres M. Gonzalez Vargas

Inria

Contraintes Group

Ibis Group

Gregory Batt

Artemis Llamosi

Eugenio Cinquemani

Previous Collaborators: Jannis Uhlendorf, Joe Schaul

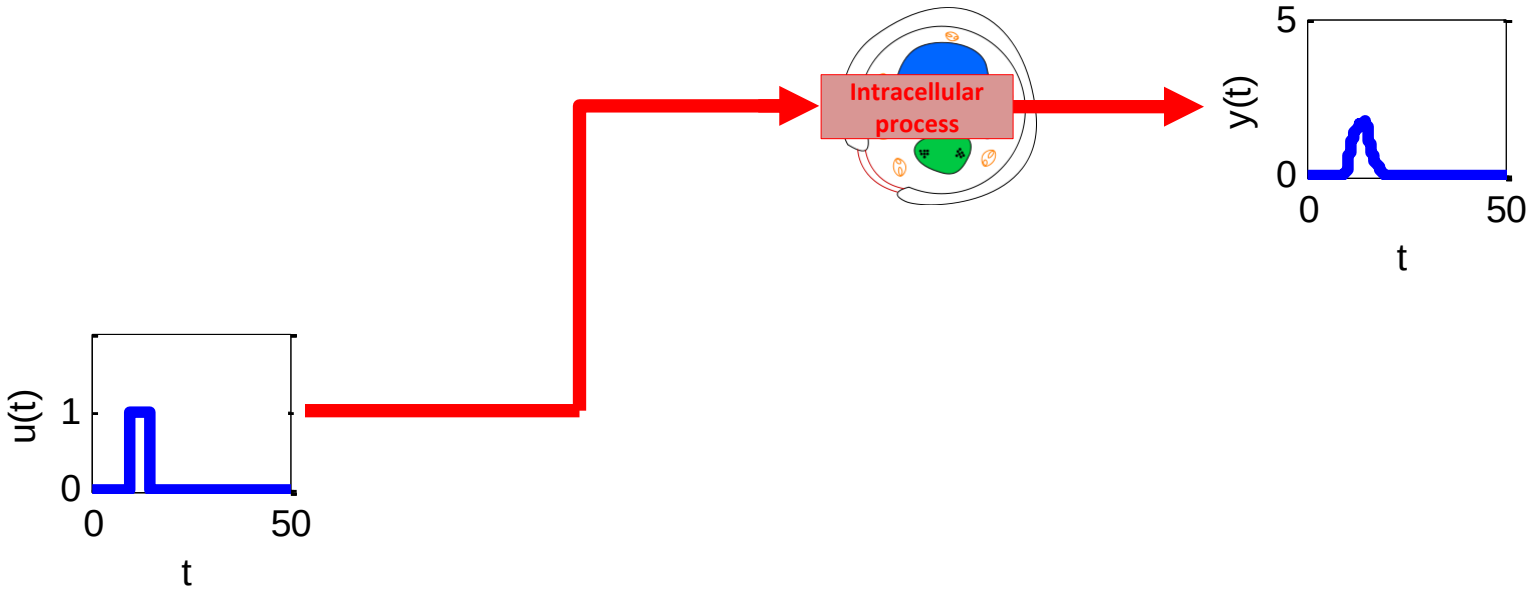
Outline

- Biological models in cell populations
- Case Study: Hyperosmotic gene expression in yeast
- Modeling frameworks
- Previous work: ME and CME models
- Current work: ME and two – stage approach
- Conclusions and future work

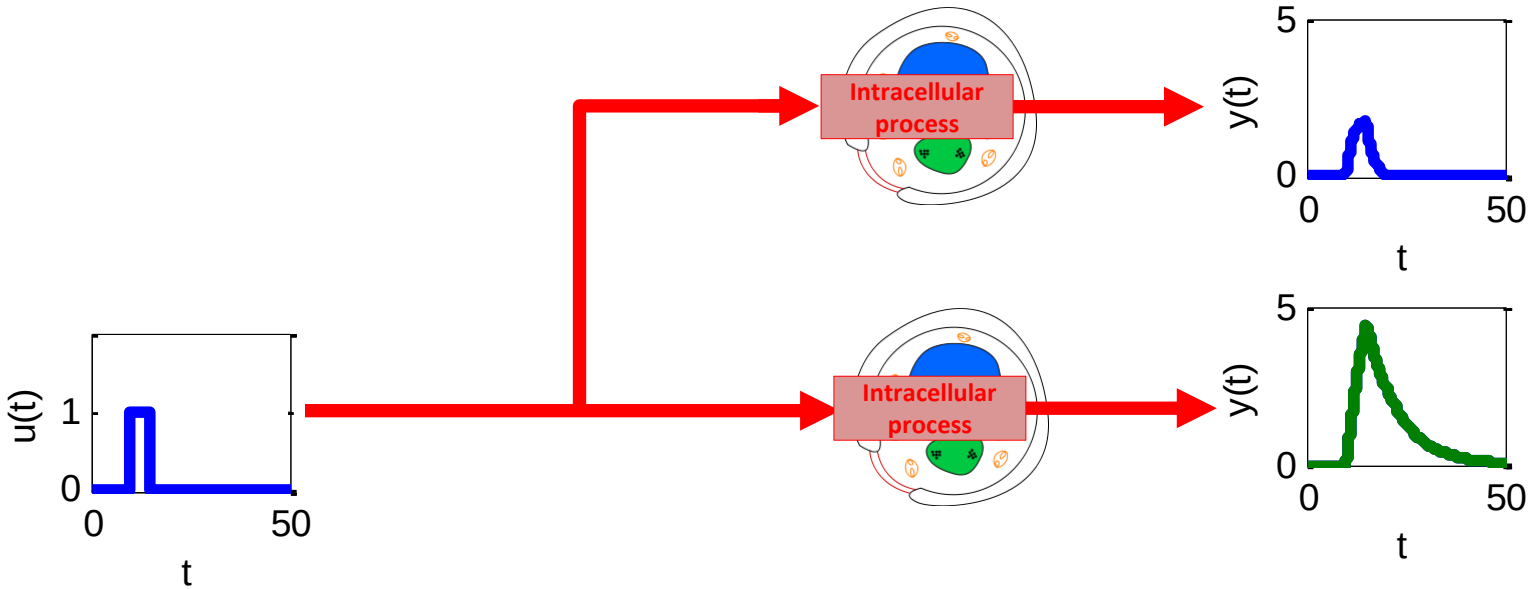
Outline

- Biological models in cell populations
- Case Study: Hyperosmotic gene expression in yeast
- Modeling frameworks
- Previous work: ME and CME models
- Current work: ME and two – stage approach
- Conclusions and future work

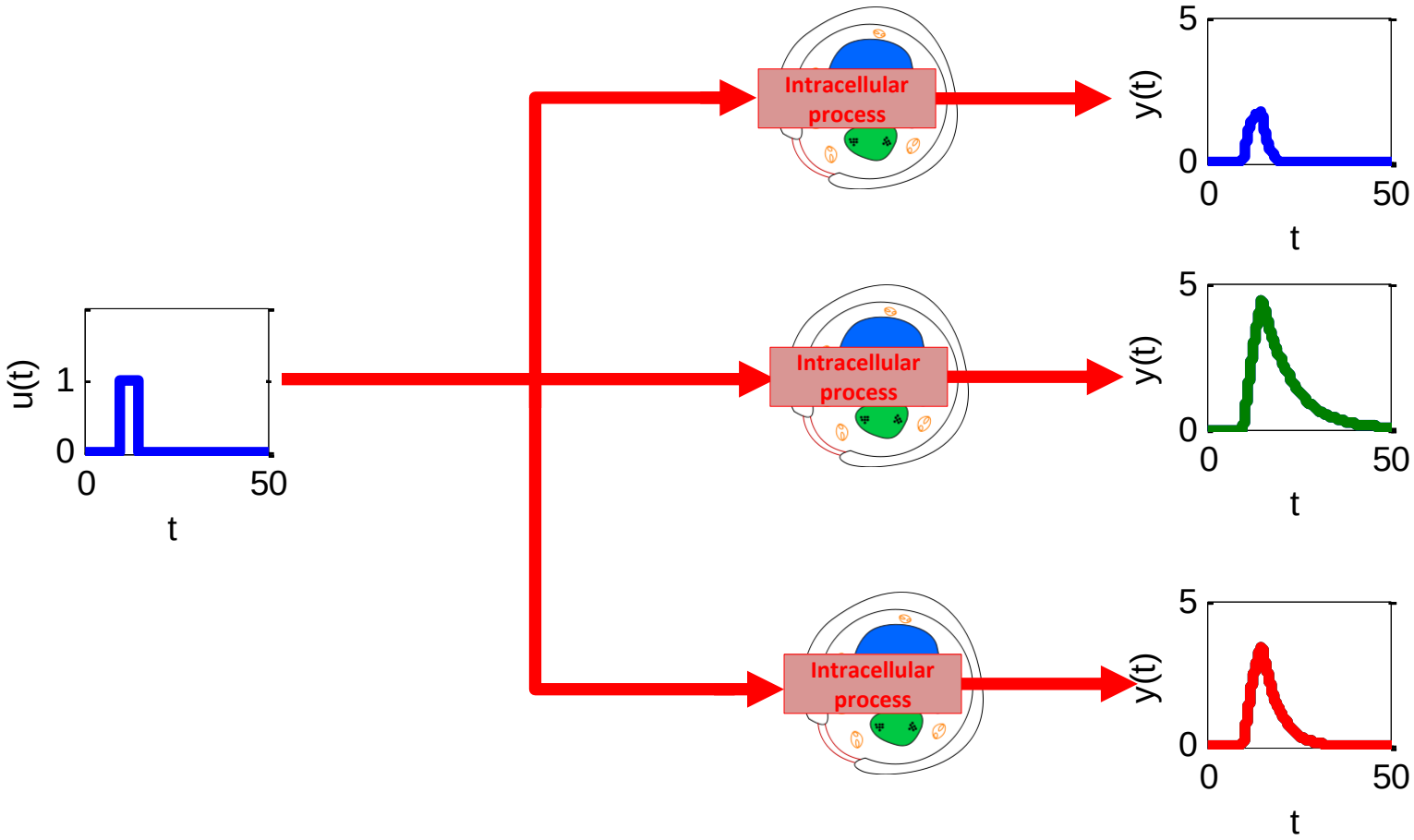
Biological models in cell populations



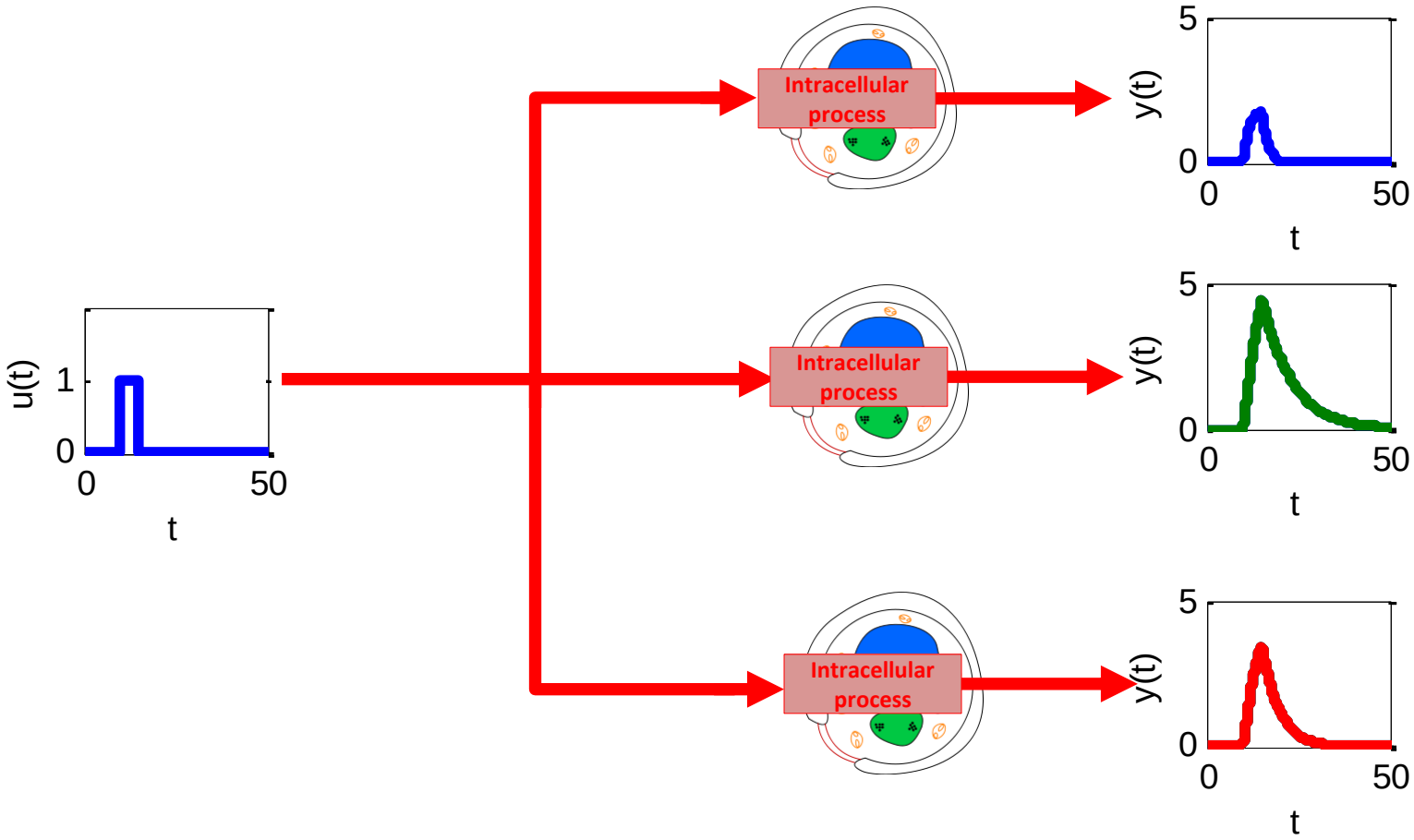
Biological models in cell populations



Biological models in cell populations

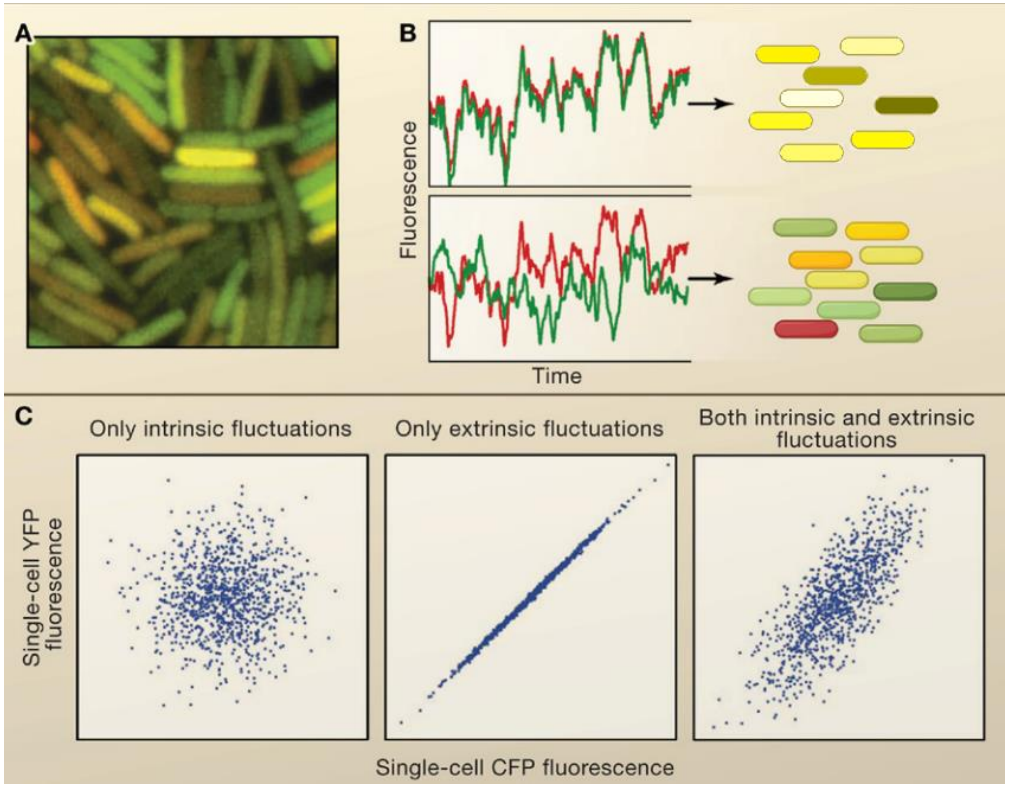


Biological models in cell populations



Objective: Model biochemical networks at single-cell and population level

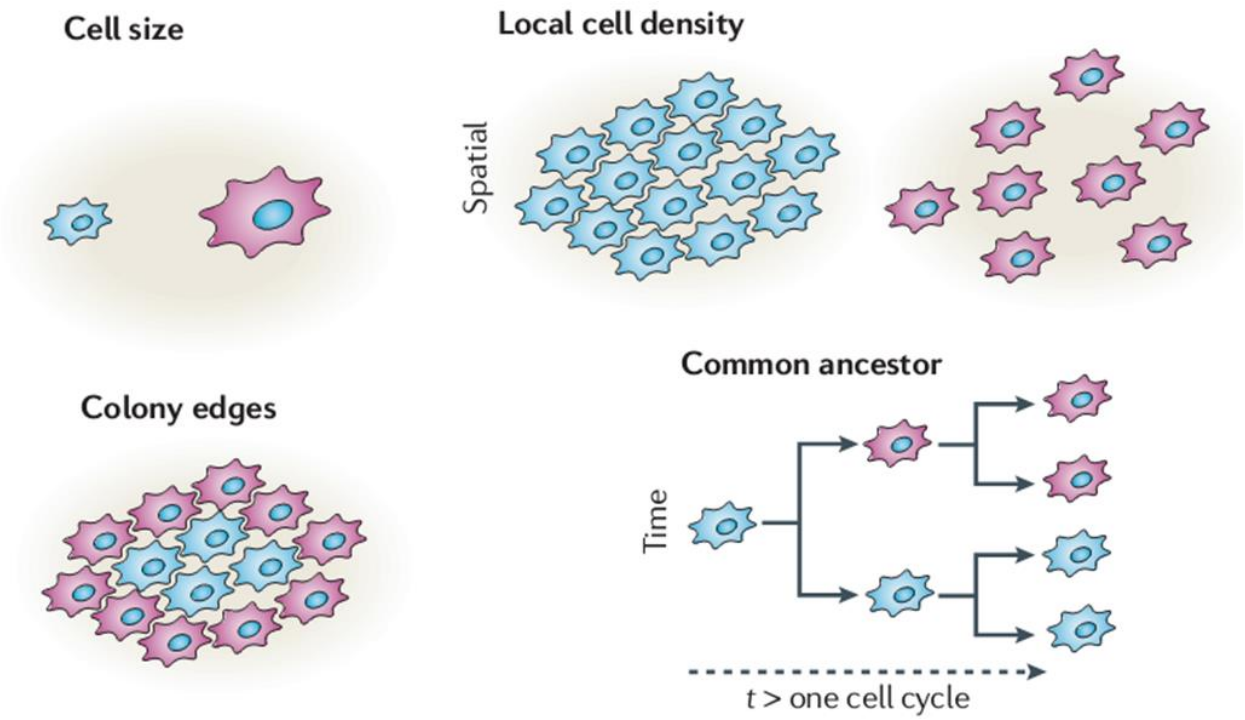
Variability in gene expression



Raj and Oudernardeen, Cell 2008

Extrinsic variability: Variations in expression from cell to cell.
Intrinsic variability: Variations in expression inside the cell.

Causes of extrinsic variability

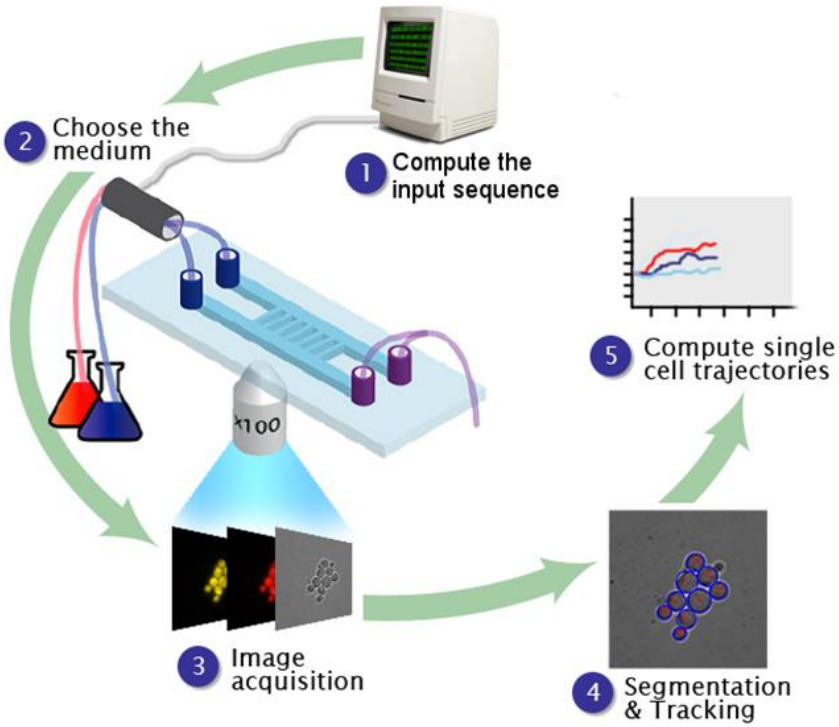


Snijder and Pelkmans. Nature Reviews –Molecular Cell Biology, 2008.

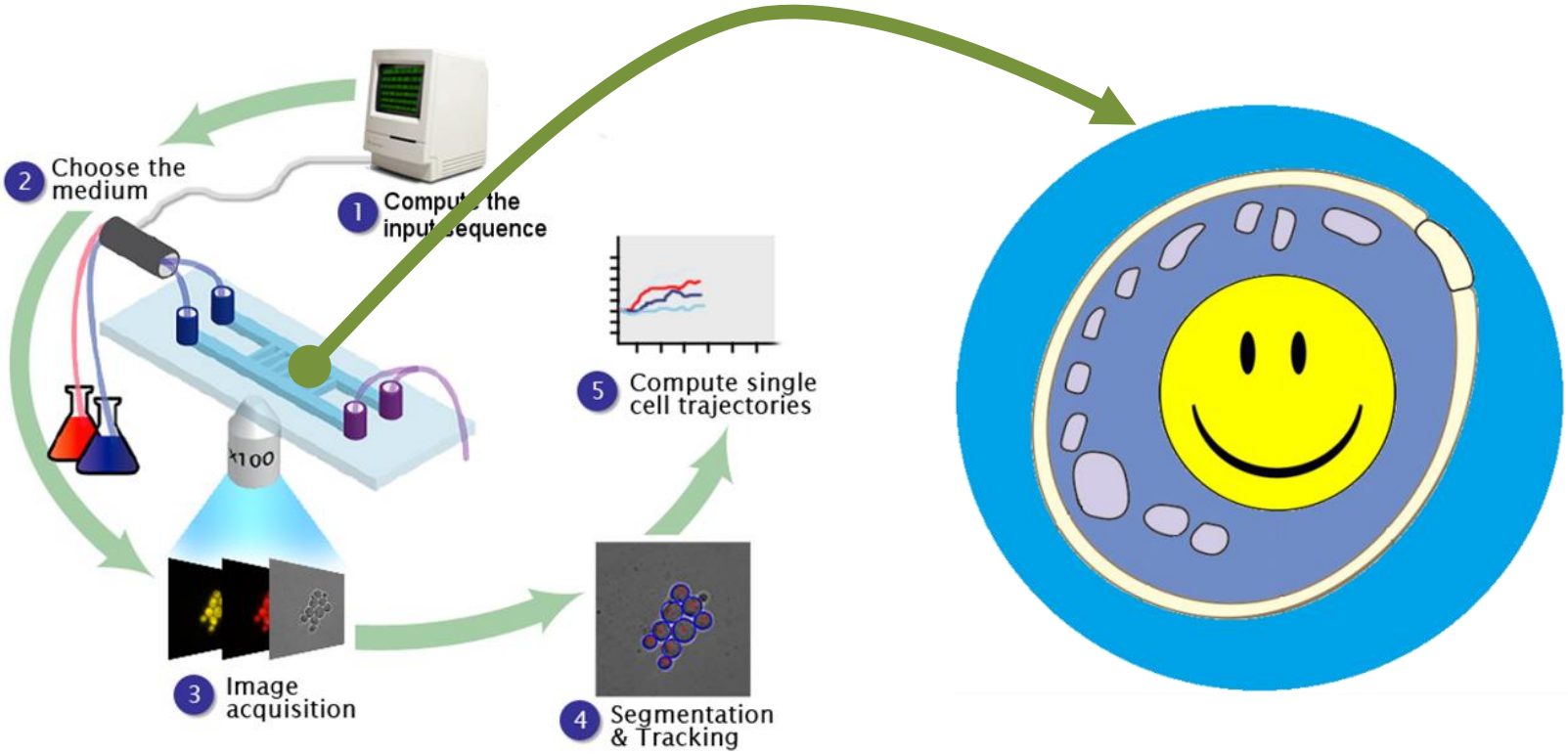
Outline

- Biological models in cell populations
- Case Study: Hyperosmotic gene expression in yeast
- Modeling frameworks
- Previous work: ME and CME models
- Current work: ME and two – stage approach
- Conclusions and future work

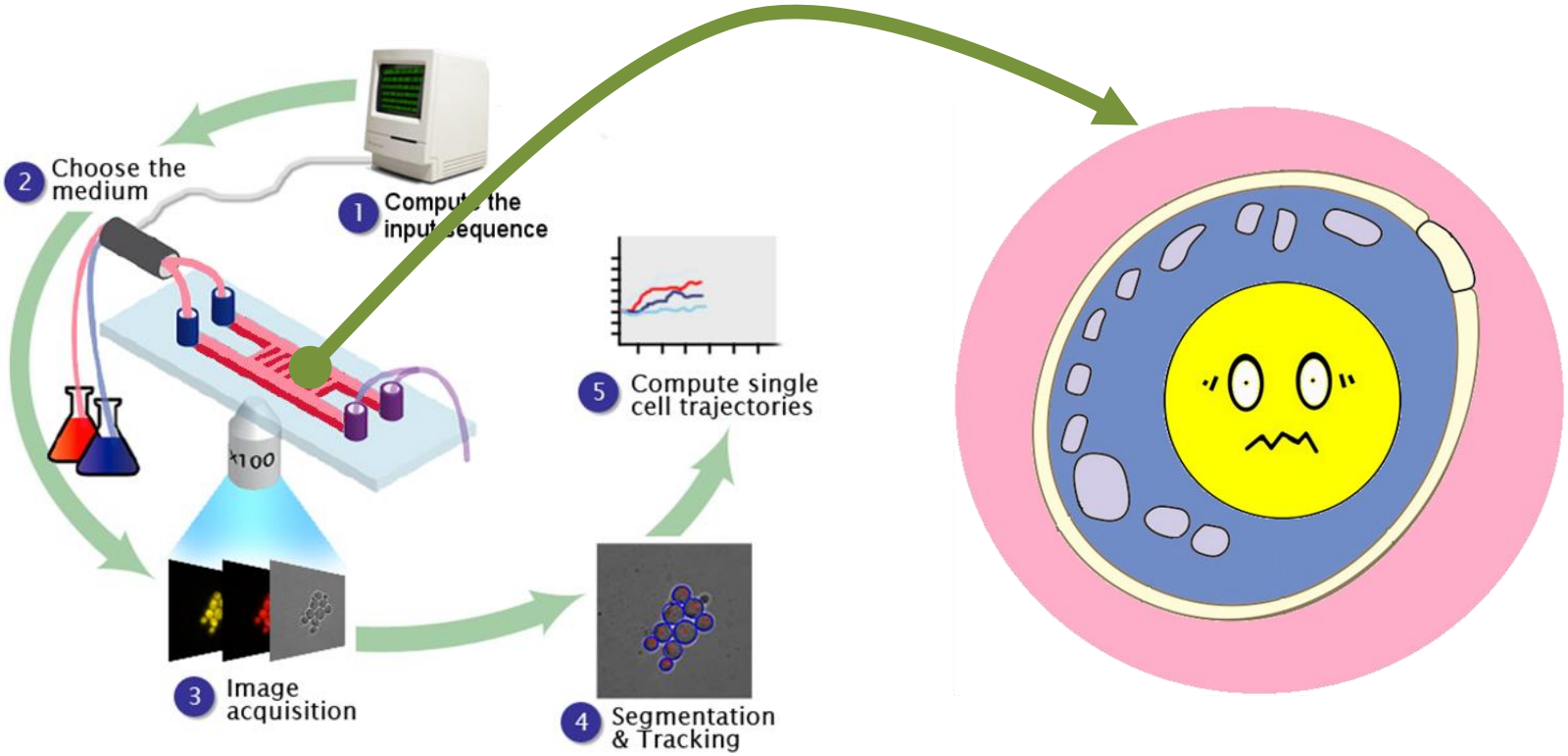
Experimental setup



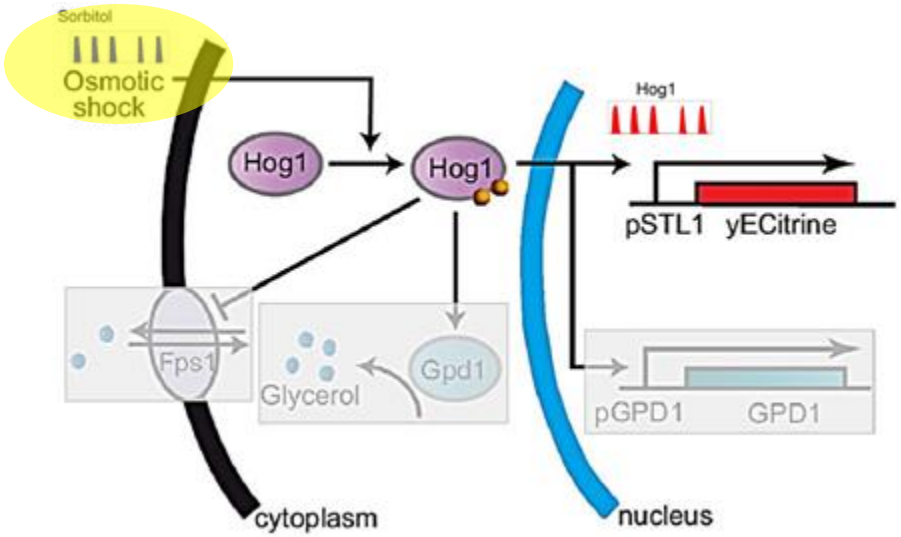
Experimental setup



Experimental setup



Hyperosmotic gene expression in yeast



$$pSTL1^{off} \underset{c2}{\overset{c1u(t-\tau)}{\rightleftharpoons}} pSTL1^{on}$$

$$pSTL1^{on} + CR \underset{c4}{\overset{c3}{\rightleftharpoons}} CR \cdot pSTL1^{on}$$

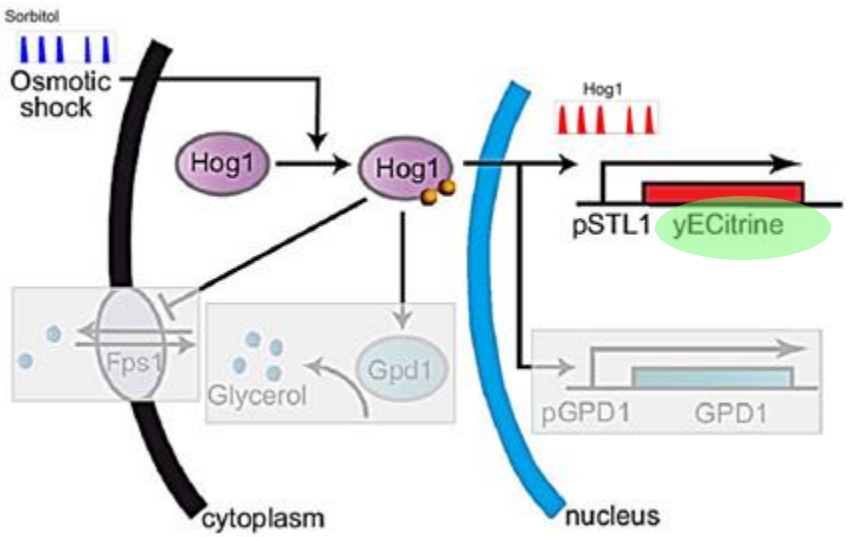
$$CR \cdot pSTL1^{on} \xrightarrow{c5} CR \cdot pSTL1^{on} + mRNA$$

$$mRNA \xrightarrow{c6} mRNA + yECitrine$$

$$yECitrine \xrightarrow{c7} \emptyset$$

$$mRNA \xrightarrow{c8} \emptyset$$

Hyperosmotic gene expression in yeast



$$pSTL1^{off} \underset{c2}{\overset{c1u(t-\tau)}{\rightleftharpoons}} pSTL1^{on}$$

$$pSTL1^{on} + CR \underset{c4}{\overset{c3}{\rightleftharpoons}} CR \cdot pSTL1^{on}$$

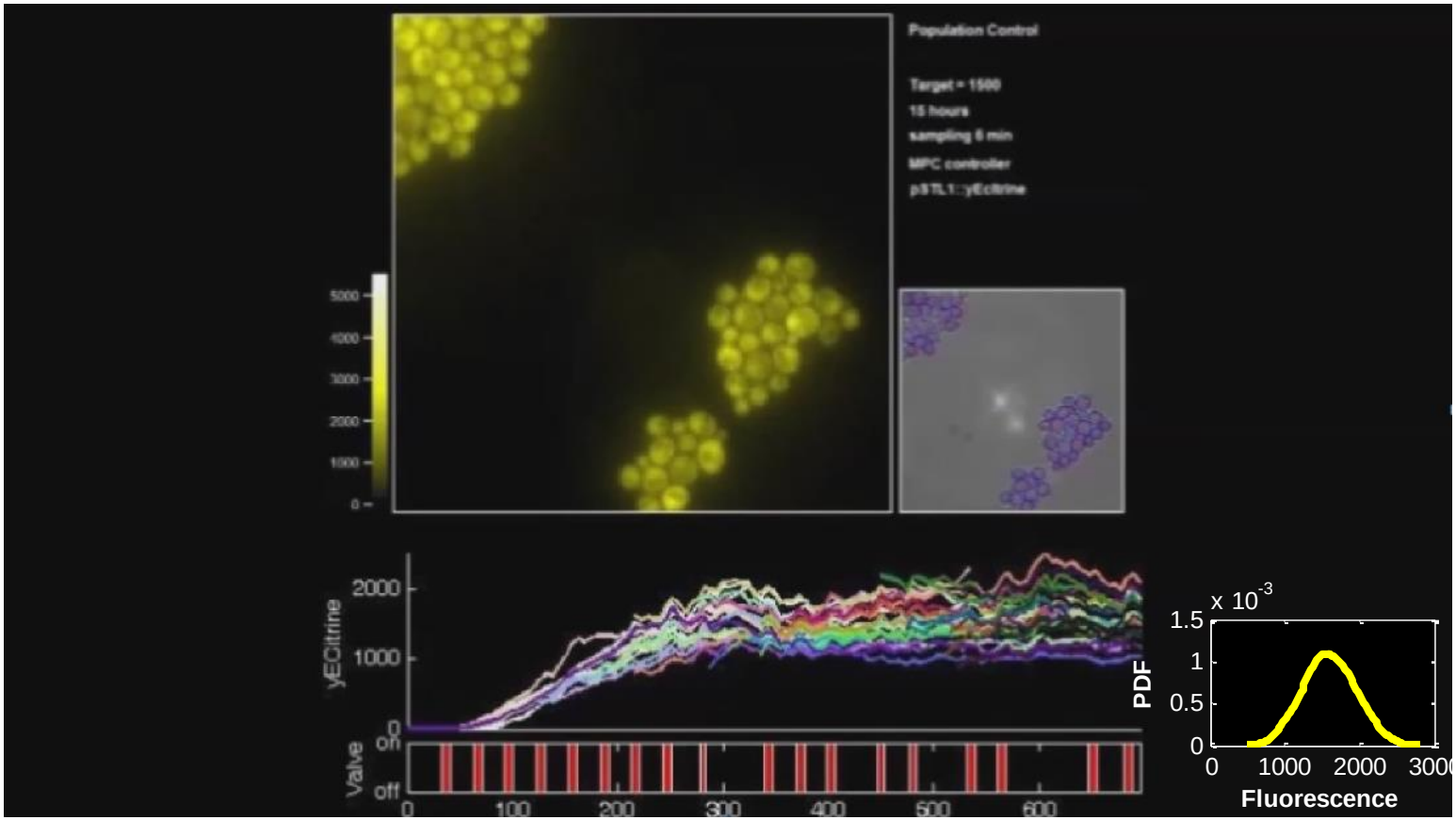
$$CR \cdot pSTL1^{on} \xrightarrow{c5} CR \cdot pSTL1^{on} + mRNA$$

$$mRNA \xrightarrow{c6} mRNA + yECitrine$$

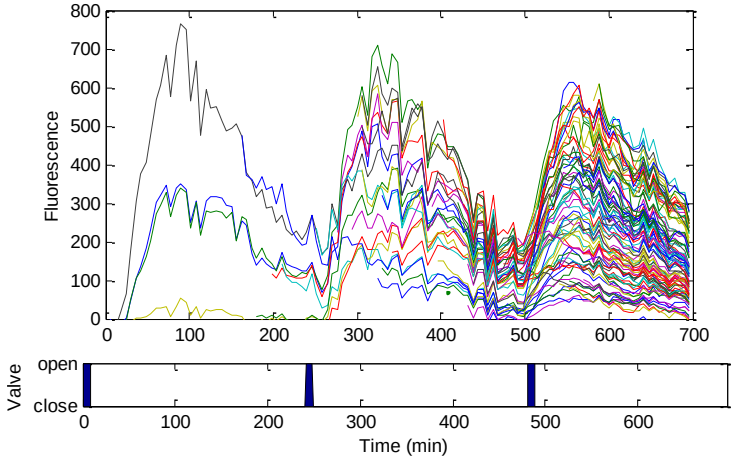
$$yECitrine \xrightarrow{c7} \emptyset$$

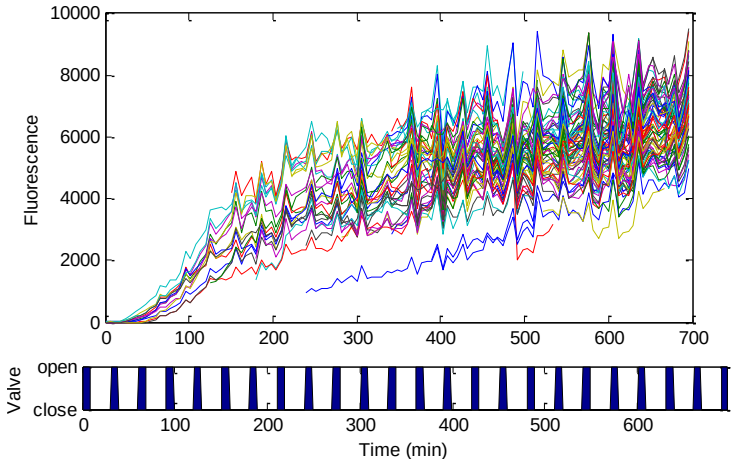
$$mRNA \xrightarrow{c8} \emptyset$$

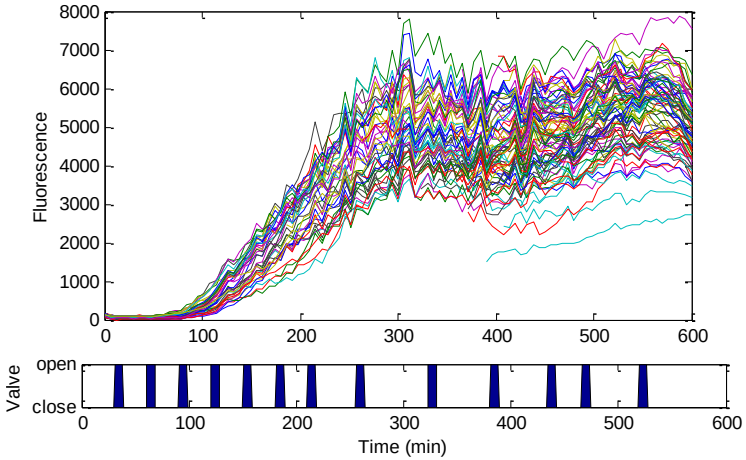
Single cell data

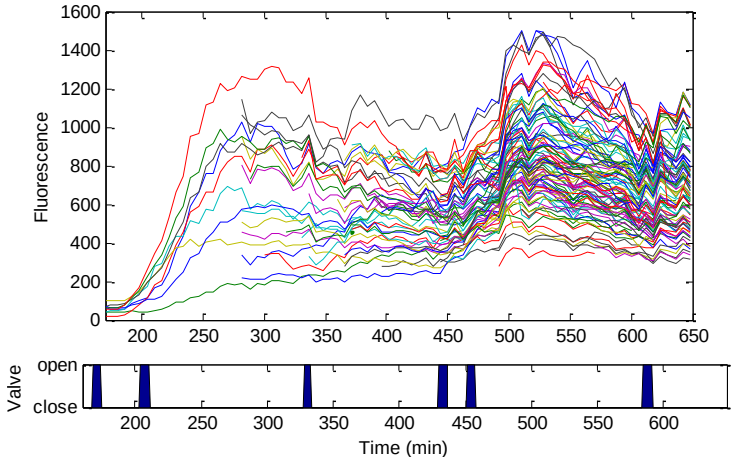


Single cell data









Outline

- Biological models in cell populations
- Case Study: Hyperosmotic gene expression in yeast
- Modeling frameworks
- Previous work: ME and CME models
- Current work: ME and two – stage approach
- Conclusions and future work

Mean cell model (MC)

Population model:

$$y_j = f(Z_j, \beta) + \varepsilon_j \quad , \qquad j = 1, \dots, J$$

Datapoint at time t_j

Regressors (t,u)

Population parameters

Measurement noise

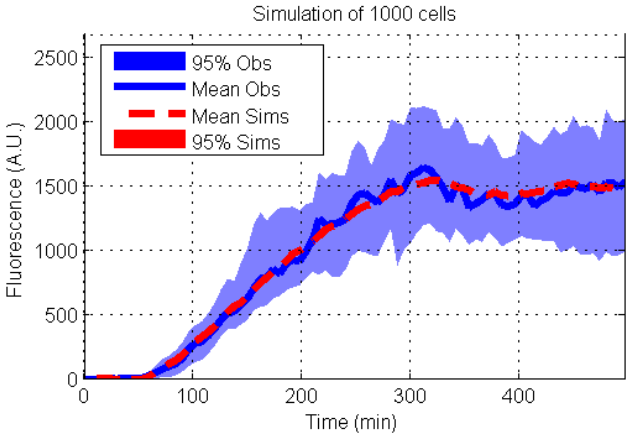
Identification methods

- MSE minimization.

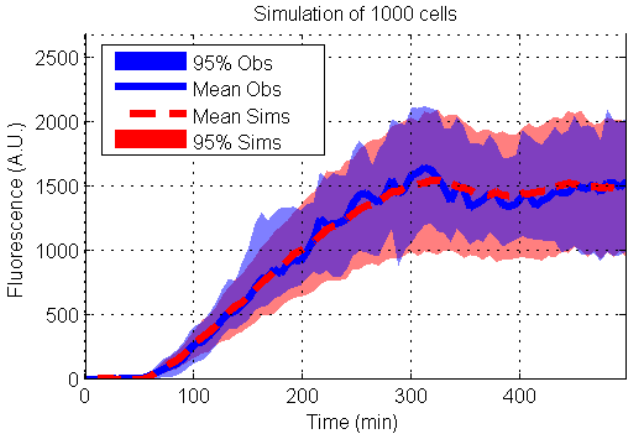
Simulation:

- ODE with one set of parameters, plus gaussian noise

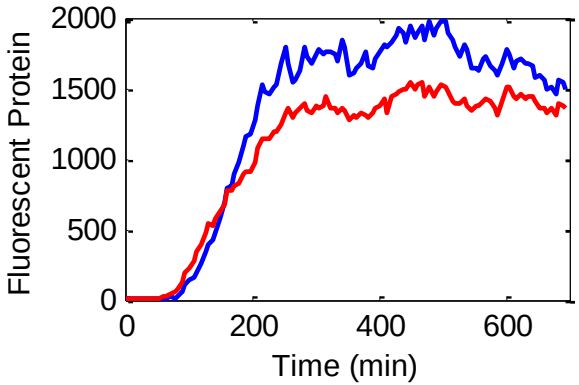
Mean cell model (MC)



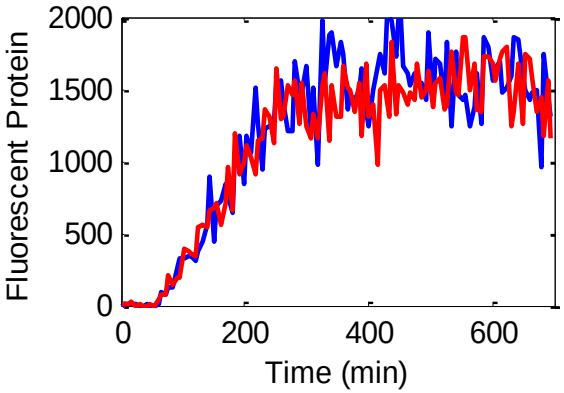
+
Measurement
Noise



Observations



Simulations



Chemical master equation model (CME)

Chemical Master Equation:

$$\frac{d}{dt}p(x,t) = \sum_{r=1}^{R} a_r(x - v_r)p(x - v_r, t) - a_r(x)p(x,t)$$

Probability of being in state x at time t

Propensity function

# of molecules

Change in # of molecules

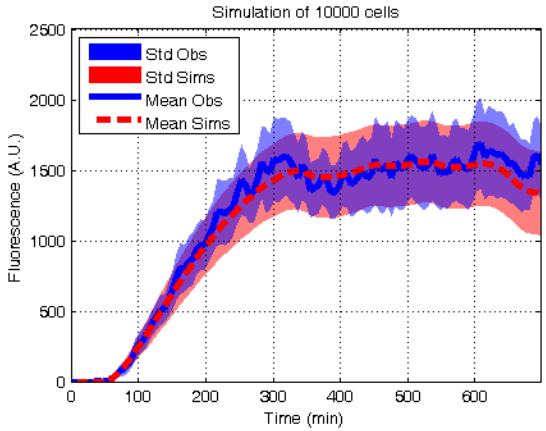
Identification methods:

- Moment-based inference
- Finite state projection
- Langevin equation
- Others

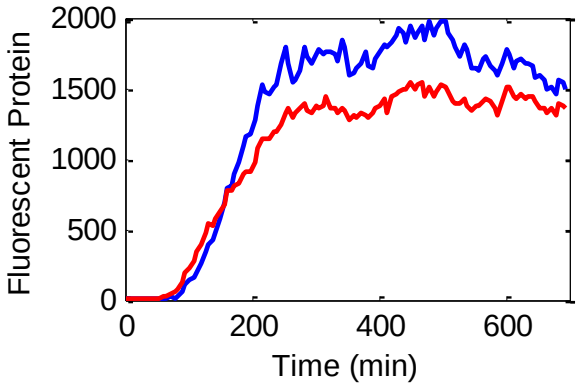
Simulation:

- SSA (Gillespie algorithm) + Gaussian noise

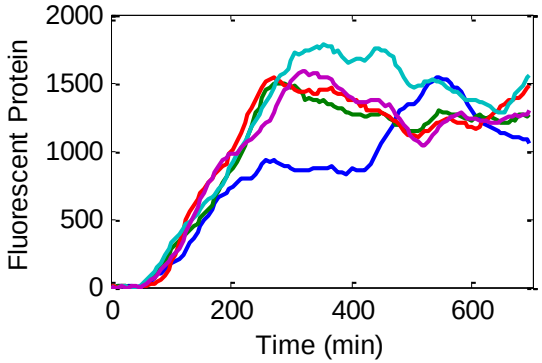
Chemical master equation model (CME)



Observations



Simulations



Mixed effects model (ME)

Individual model:

$$y_{ij} = f(Z_{ij}, \beta_i) + \varepsilon_{ij} \quad , \qquad j = 1, \dots, J_i$$

Datapoint at time t_j

Regressors (t,u)

Individual parameters

Measurement noise

Population model:

$$\beta_i = d(\alpha_i, \beta, b_i) \, , \qquad i = 1, \dots, N \qquad b_i \sim N(0, C)$$

Individual covariates

Random effects

Fixed effects

Extrinsic variability

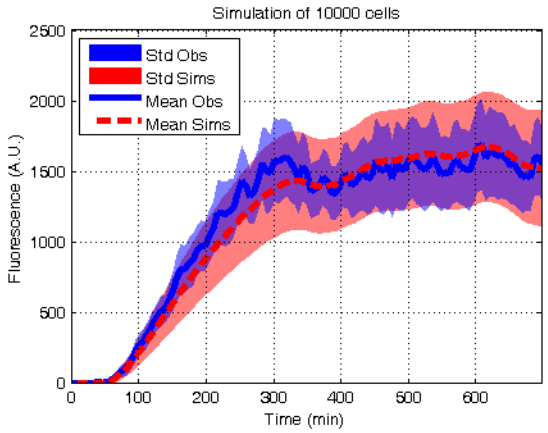
Identification methods

- Marginal likelihood maximization (Expectation-Maximization Algorithm)

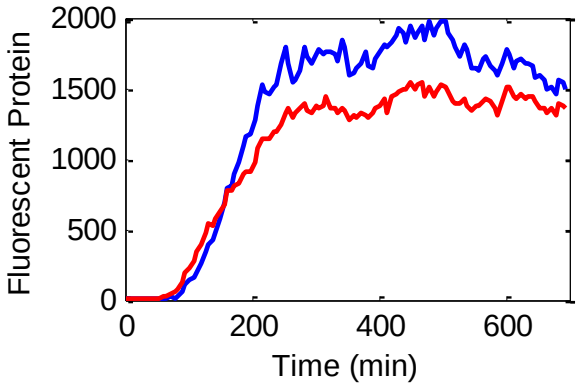
Simulation:

- ODE with random parameters + gaussian noise

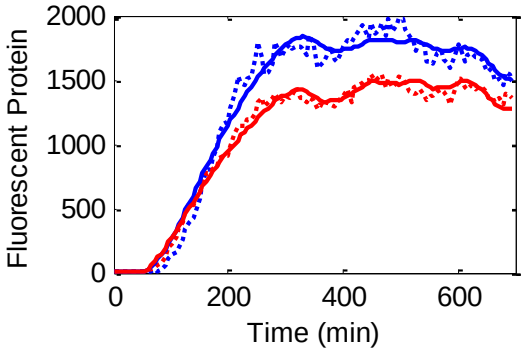
Mixed effects model (ME)



Observations



Simulations



Model summary

	Population model	Single cell model	Measurement Noise	Extrinsic Variability	Intrinsic Variability	Inference	Simulation
MC	Yes	No	Yes	No	No	MSE minimization	ODE
ME	Yes	Deterministic	Yes	Yes	Possible	Likelihood maximization	ODE
CME	Yes	Stochastic	Yes	Possible	Yes	Moment-based, SDE, Finite states	SSA

Outline

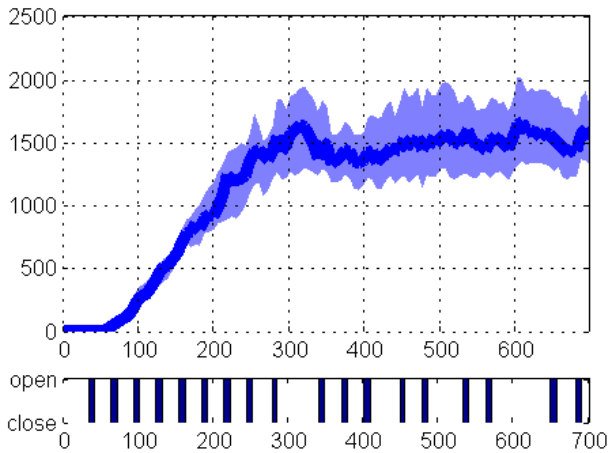
- Biological models in cell populations
- Case Study: Hyperosmotic gene expression in yeast
- Modeling frameworks
- Previous work: ME and CME models
- Current work: ME and two – stage approach
- Conclusions and future work

Previous work

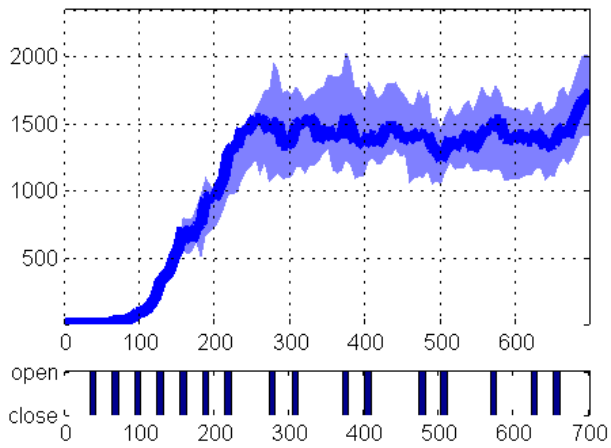
- Compare Mixed Effects models vs Chemical Master Equation models.
  - Use Moment Based Inference for CME

Identification and validation data

Identification Data



Validation Data



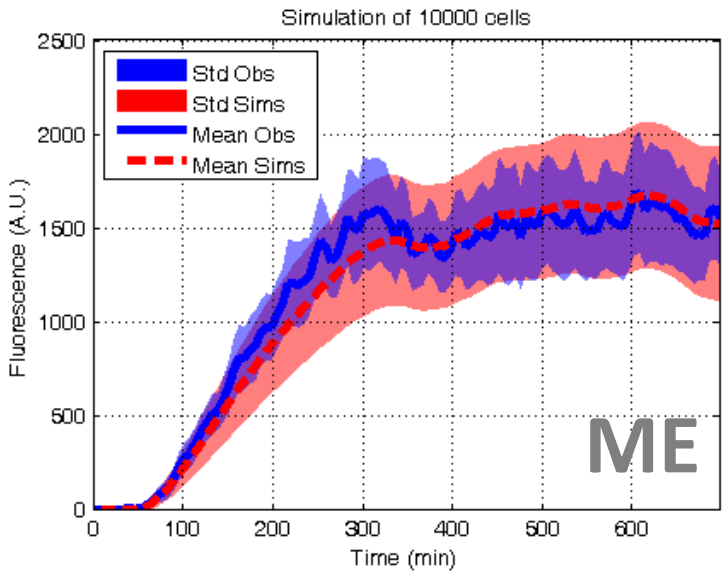
Validation:

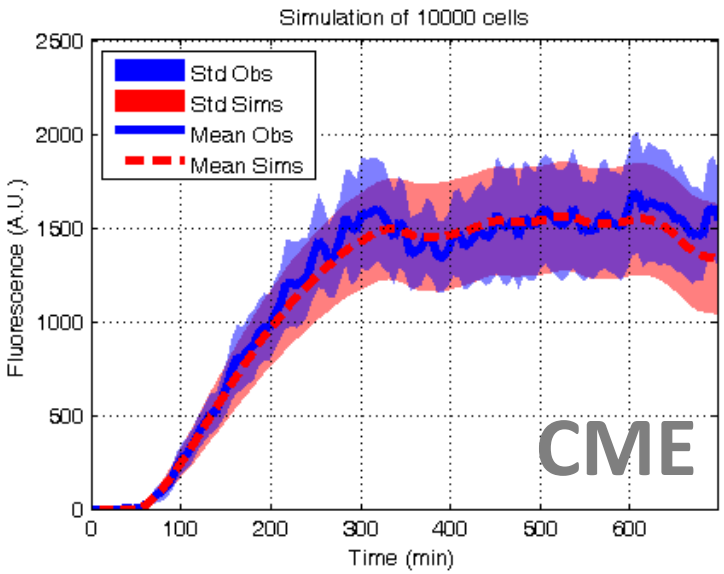
Simulate 10000 cells with the identification input and 10000 cells with the validation input

Compute Normalized RMSE (for mean and stdev)
Compute Kolmogorov-Smirnov (KS) test at each time instant

Total Success rate (h-Kol)
Average significance (p-Kol)

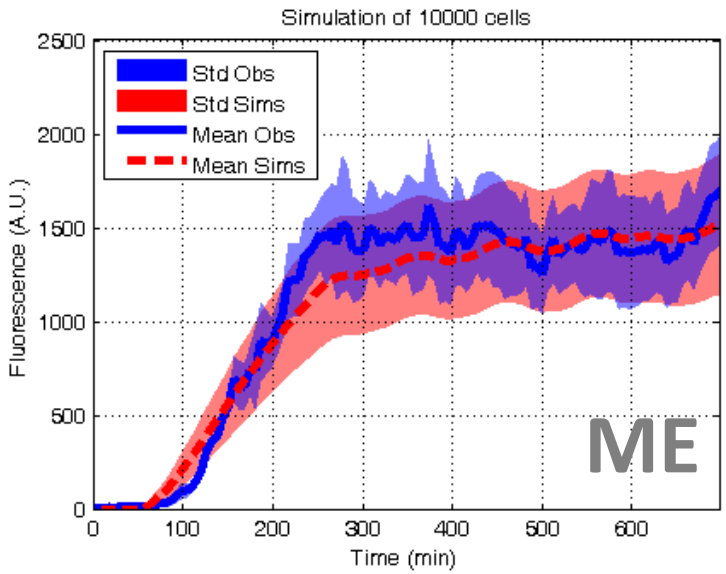
Performance on identification dataset

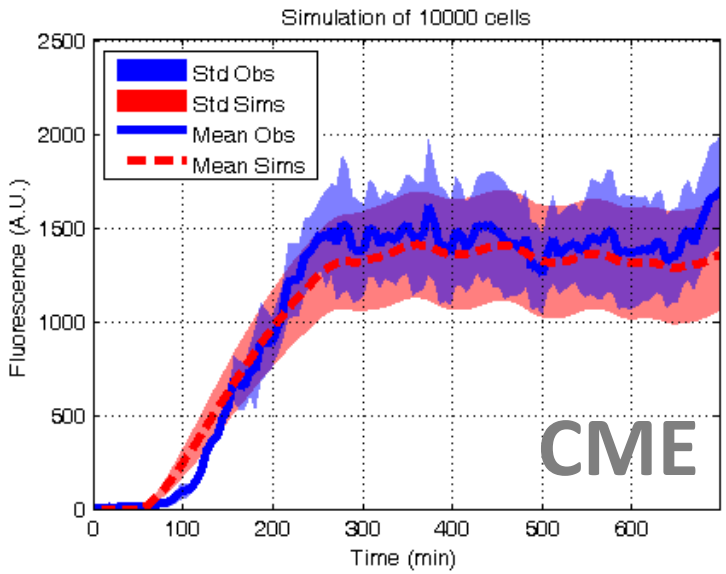




	ME	CME
$NRMSE_M$	0.06	0.04
$NRMSE_S$	0.25	0.11
Avg. p-Kol	0.25	0.49
h-Kol	79%	87%

Performance on validation dataset





	ME	CME
$NRMSE_M$	0.08	0.06
$NRMSE_S$	0.20	0.13
Avg. p-Kol	0.34	0.32
h-Kol	87%	74%

Outline

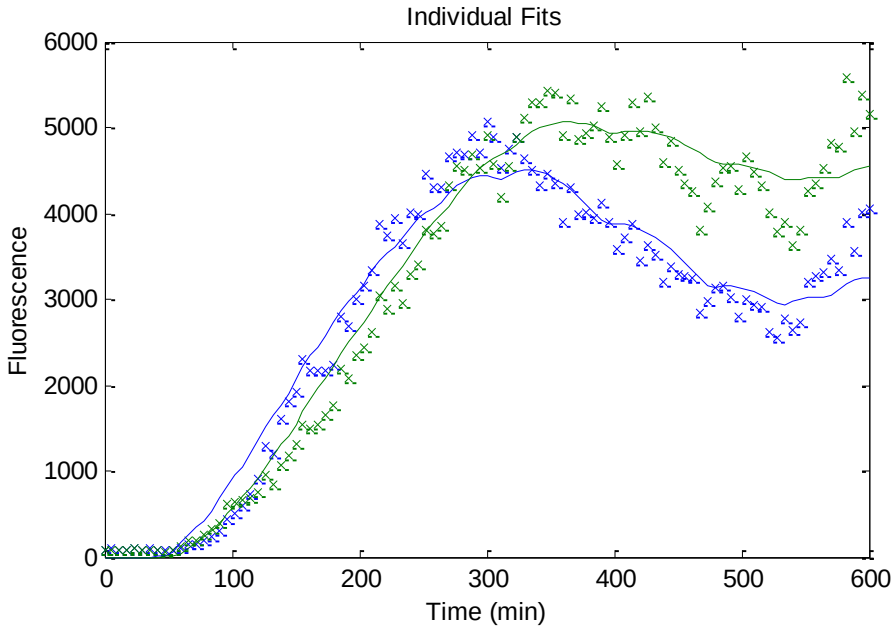
- Biological models in cell populations
- Case Study: Hyperosmotic gene expression in yeast
- Modeling frameworks
- Previous work: ME and CME models
- Current work: ME and two – stage approach
- Conclusions and future work

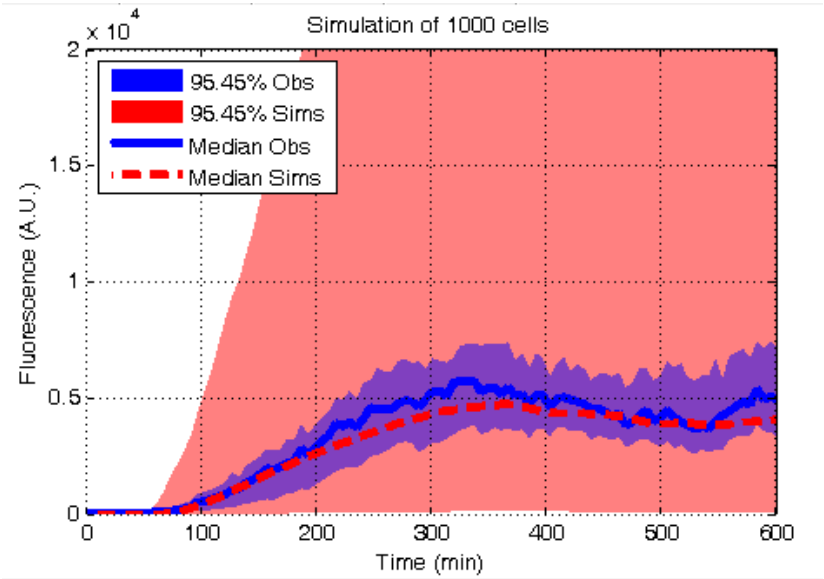
Current work

- Compare two-stage approach vs mixed effects models.

Two-stage estimation

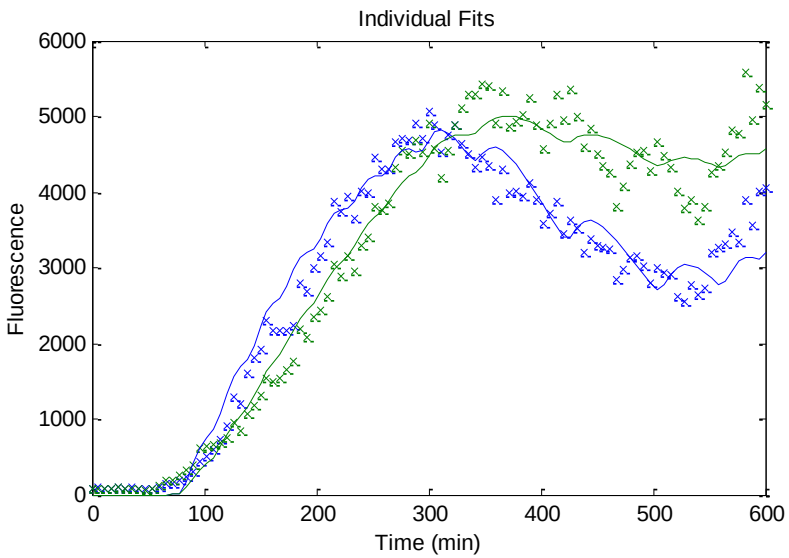
1. Compute individual cell parameters ($E(\beta_i), cov(\beta_i)$) using standard techniques (MSE minimization)
2. Compute parameter distribution based (on individual parameters

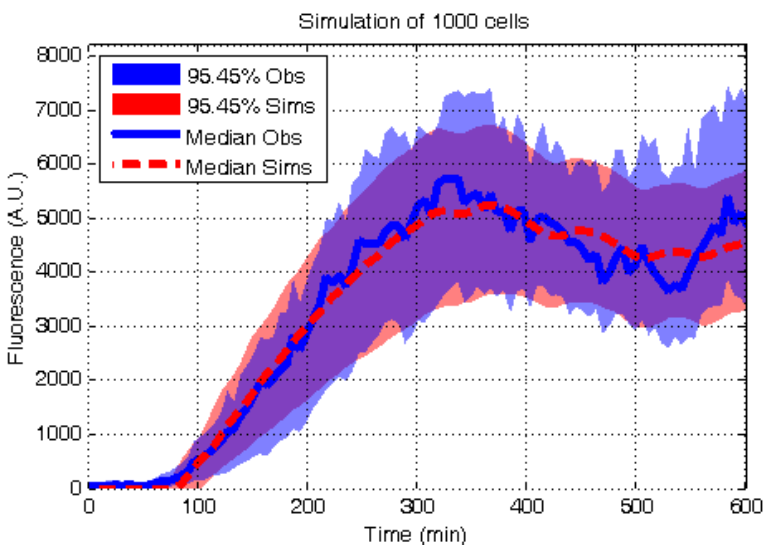




Mixed effects estimation

1. Compute a multivariate parameter distribution that maximizes the likelihood of the data (using stochastic expectation maximization algorithms)





Outline

- Biological models in cell populations
- Case Study: Hyperosmotic gene expression in yeast
- Modeling frameworks
- Previous work: ME and CME models
- Current work: ME and two – stage approach
- Conclusions and future work

Conclusions

- CME and ME models can describe the available population data in with a comparable success.
  - Improvements are needed to separate intrinsic from extrinsic variability.

- CME
  - Accounts for intrinsic variability.
  - Inference is computationally efficient when moment-based methods are used
  - Tailored to flow-cytometry data (time-course of distributions of fluorescence)

- ME
  - Accounts for extrinsic variability
  - Can model specific single cells.
  - Requires time-lapse single cell data (e.g. fluorescence video microscopy)
  - Facilitates the discovery of biological deterministic correlations in the parameters

Future Work

- Include intrinsic variability in M.E. through stochastic differential equations.
- Use lineage information to find correlation between mother-daughter cell parameters.
- Use these models to develop control algorithms.

Thanks for your attention!

Acknowledgements

University of Pavia

Giancarlo Ferrari Trecate

INRIA Rhone-Alpes

Eugenio Cinquemani

INRIA Paris-Rocquencourt

Gregory Batt
Jannis Uhlendorf
Artemis Llamosi
Joe Schaul