Tutorial on modelling and analysis of population systems with applications to molecular biology

Eugenio Cinquemani, Giancarlo Ferrari-Trecate, John Lygeros

ECC Hycon2 Workshop, Paris
Thursday, June 24, 2014

Outline

Part I :

- Introductory concepts
- Gene expression variability
  - Intrinsic vs. extrinsic noise
- Mixed-Effects modelling for extrinsic variability
- Yeast osmotic shock response: Identification and validation results

Part II :

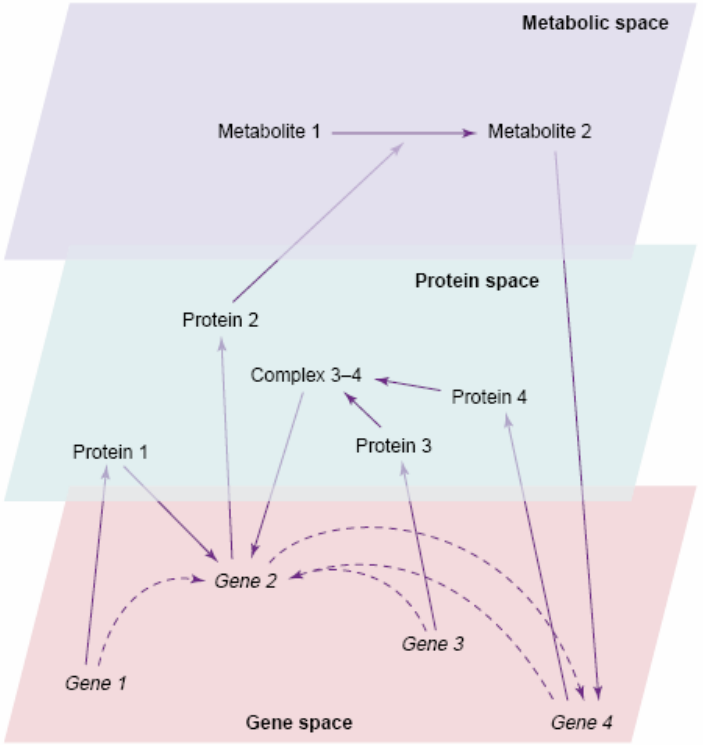
- Chemical Master Equation and intrinsic uncertainty
- System Identification
  - Joint estimation of intrinsic and extrinsic variability
  - Population bimodality
- Concluding remarks and outlook

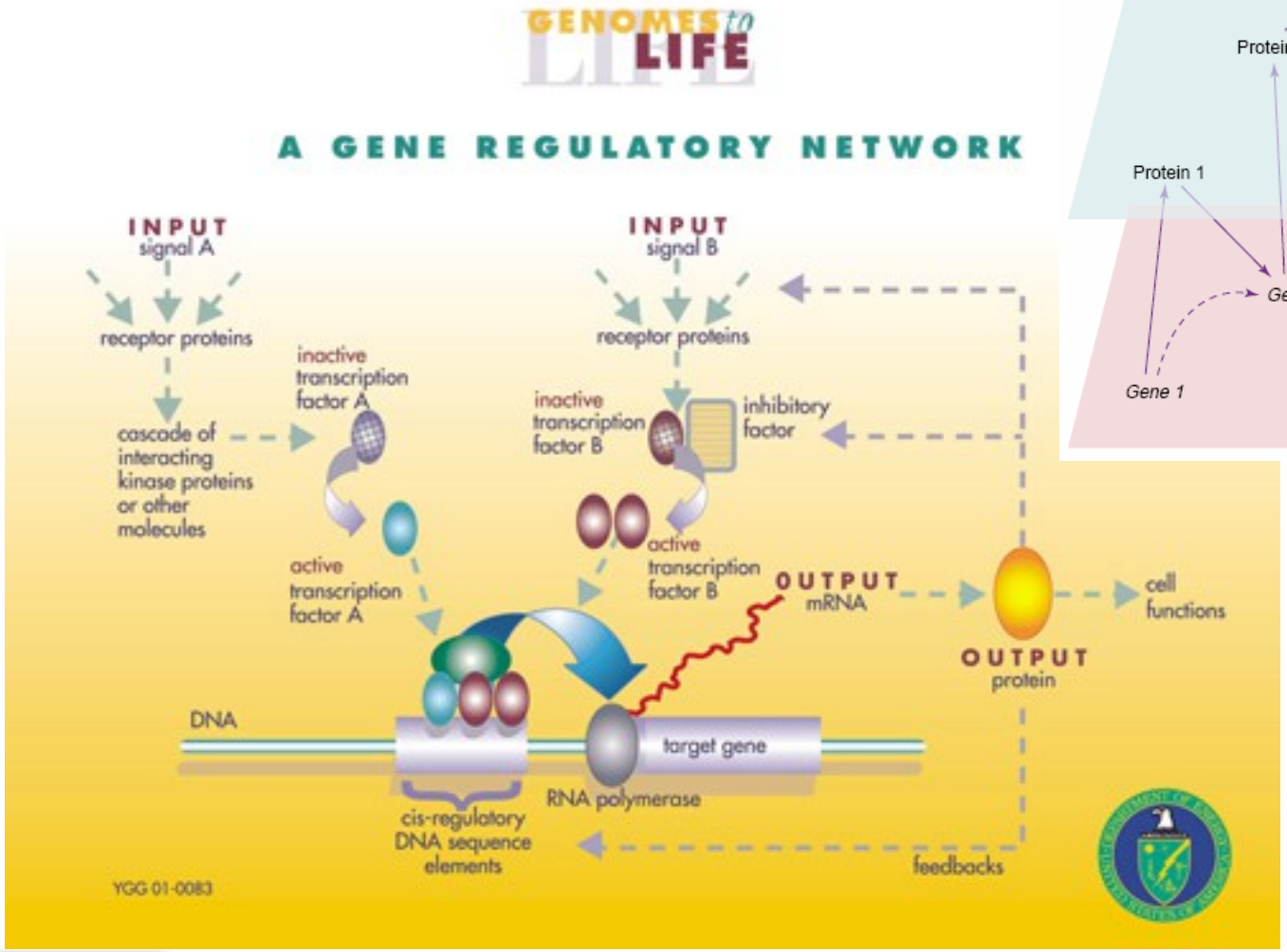
Part I – Mixed-effects modelling of extrinsic noise in cell populations

Work with :

A.Gonzales, G.Ferrari-Trecate (UNIPV)
G.Batt, J.Uhlendorf, J. Schaul, A.Llamosi (LIFEWARE, INRIA Paris Rocquencourt)
P.Hersen (MSC, Université Paris Diderot)

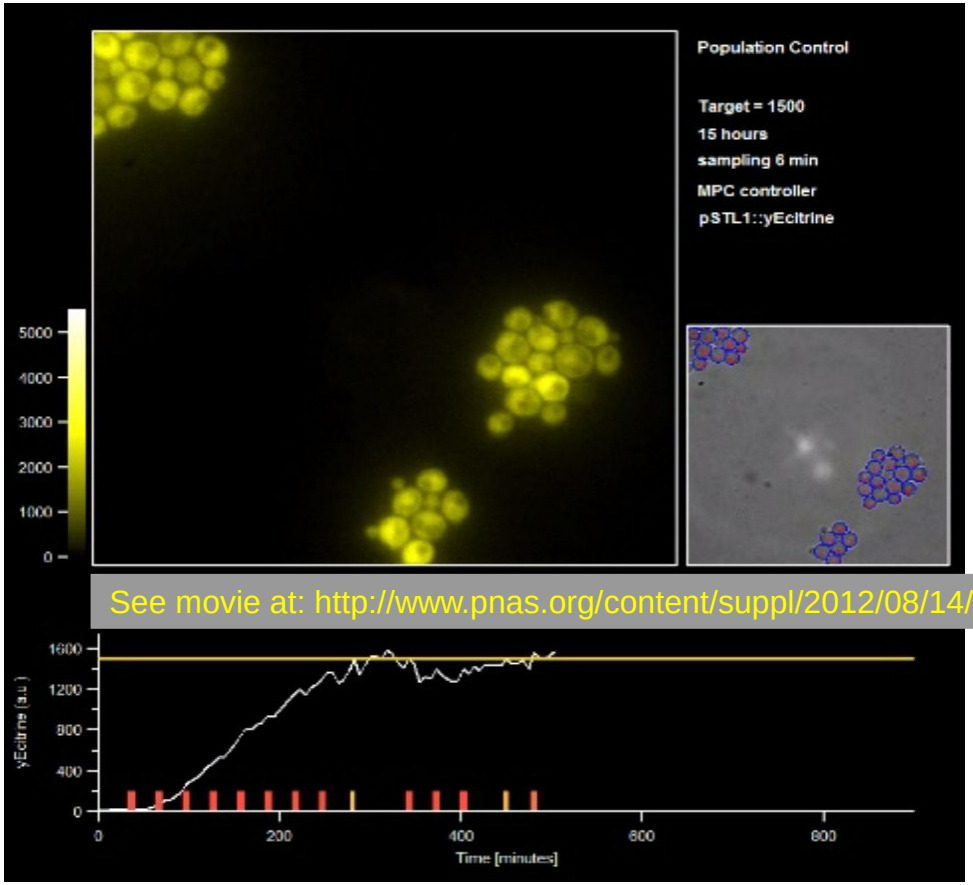
Gene regulatory networks and the central dogma



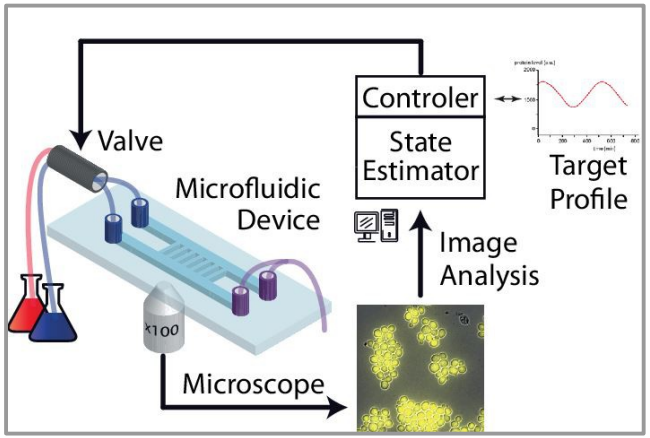


(Wikipedia)

Gene expression in single cells

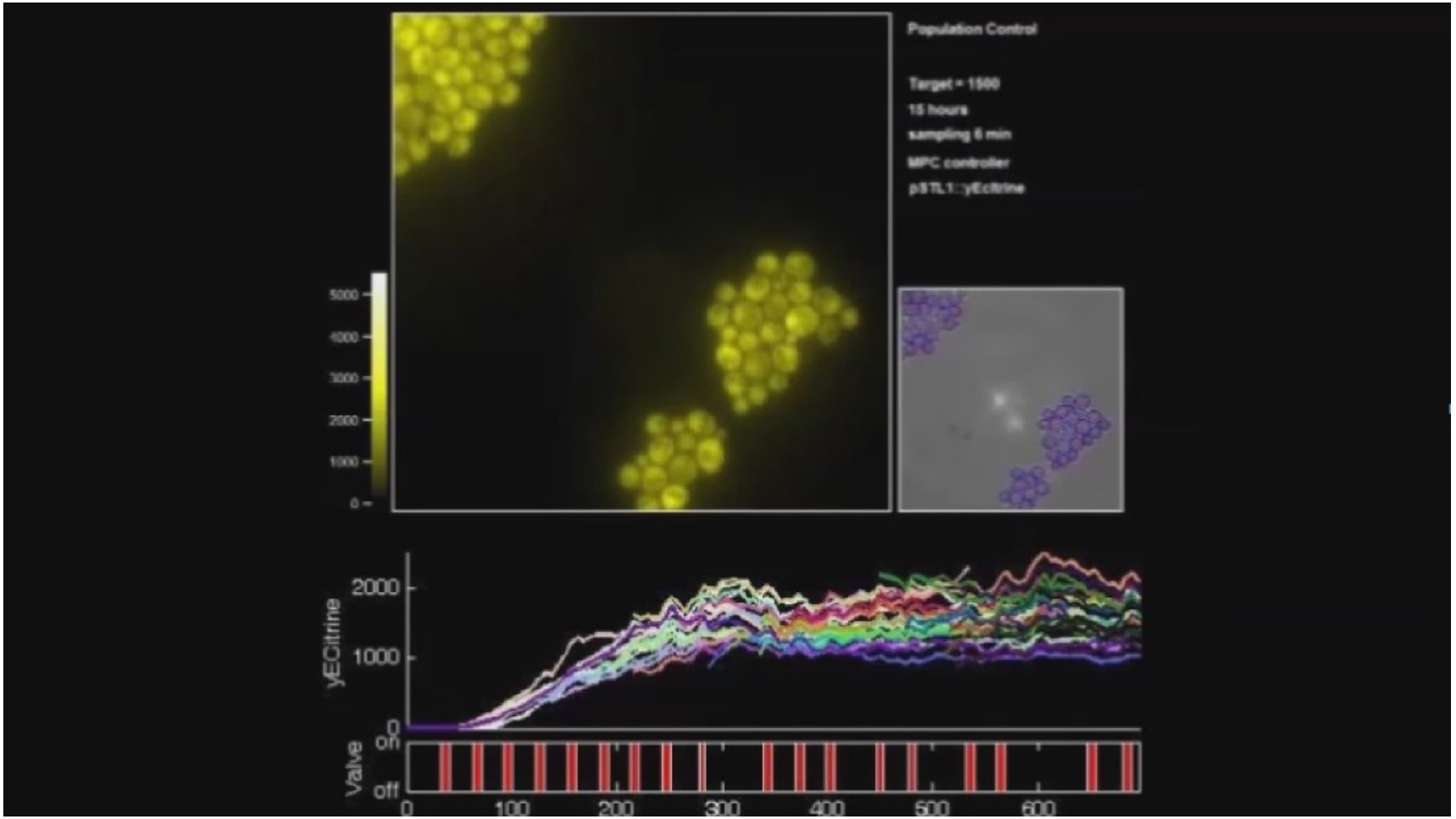


See movie at: http://www.pnas.org/content/suppl/2012/08/14/1206810109.DCSupplemental/sm01.wmv



(Uhlendorf *et al.* PNAS 2012)

Gene expression variability



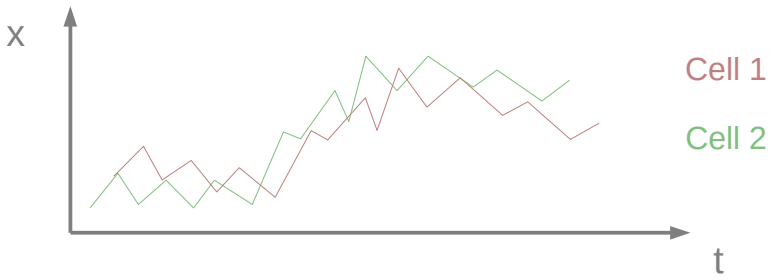
(Adapted from Uhlendorf *et al.* PNAS 2012)

Extrinsic vs. intrinsic noise

- Intrinsic noise : randomness of biochemical events within the gene expression process

$$P[x(t+dt) - x(t) = v_j] = a_j(x(t))\,dt + o(dt)$$

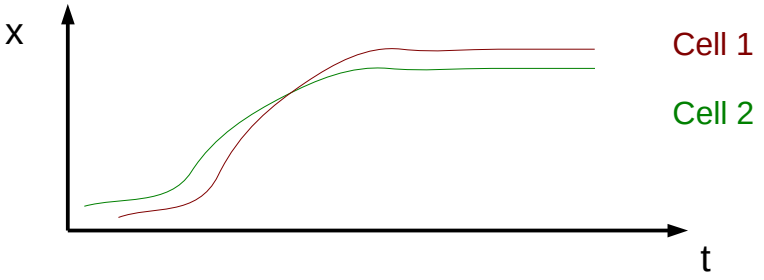
for reactions *j=1,…,m*



- Extrinsic noise : variability of cell properties external to but affecting gene expression

$$dx_\theta(t) = \sum_j v_j a_j(x_\theta(t))\,dt, \quad \theta = \theta^1, \ldots, \theta^N$$

where N is the number of cells



Extrinsic noise : Mixed-Effects (ME) modelling

Individual model: $y_j^i = F(\theta^i, u(\cdot), t_j) + \varepsilon_j^i, \qquad \varepsilon_j^i \sim p_{\varepsilon j}$

"Regressors"

Datapoint at time t_j

Measurement noise

Individual parameters

- solution of ODE model with individual parameters, observed with noise

Extrinsic variability

Population model: $\theta^i = g(\alpha^i, \beta, \beta^i), \qquad \beta_i \sim p(\Gamma)$

Individual Covariates

Population statistics

Random Effects

Fixed effects

Inferring ME models from single-cell data

Problem: Determine population statistics and single cell parameters from data

- Naive approach :

$$\hat{\theta}^i = \arg\max\, p(y^i|\theta^i), \quad i = 1, \ldots, N$$
$$\hat{\Gamma} = \text{statistics}(\hat{\theta}^1, \ldots, \hat{\theta}^N)$$

- ME approach :

$$\hat{\Gamma} = \arg\max\, p(y^1, \ldots, y^N|\Gamma)$$
$$\hat{\theta}^i = \arg\max\, p(\theta^i|y^i, \hat{\Gamma}), \quad i = 1, \ldots, N$$

  based on marginalization :

$$p(y^i|\Gamma) = \int p(y^i|\theta^i)p(\theta^i|\Gamma)d\theta$$

Off-the-shelf ME inference algorithms exist (e.g. SAEM, implemented in Monolix)

Yeast osmotic shock response

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

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

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

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

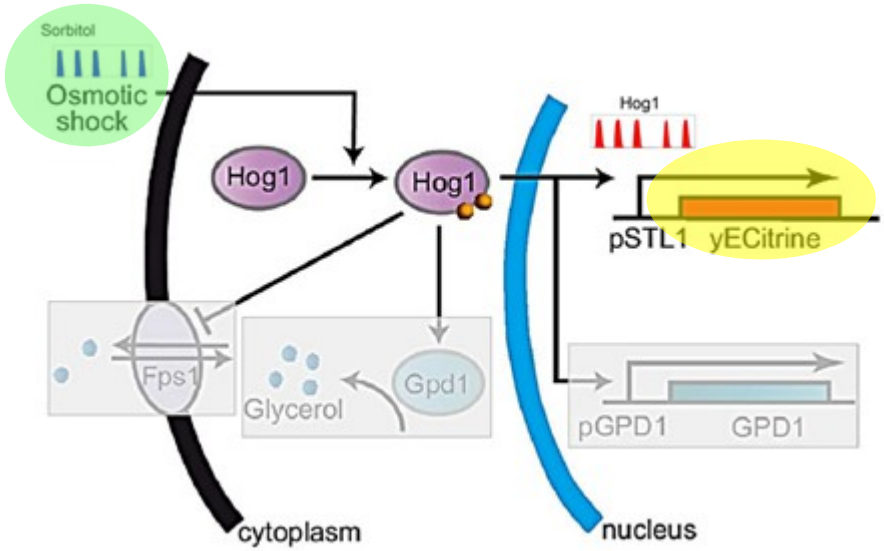
$$yECitrine \xrightarrow{c7} \emptyset$$

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

$$\vdots$$

$$\frac{d}{dt} mRNA(t) = c_5 f\left(u(t-\tau)\right) - c_8 mRNA(t)$$

$$\frac{d}{dt} yECitrine(t) = c_6 mRNA(t) - c_7 yECitrine(t)$$



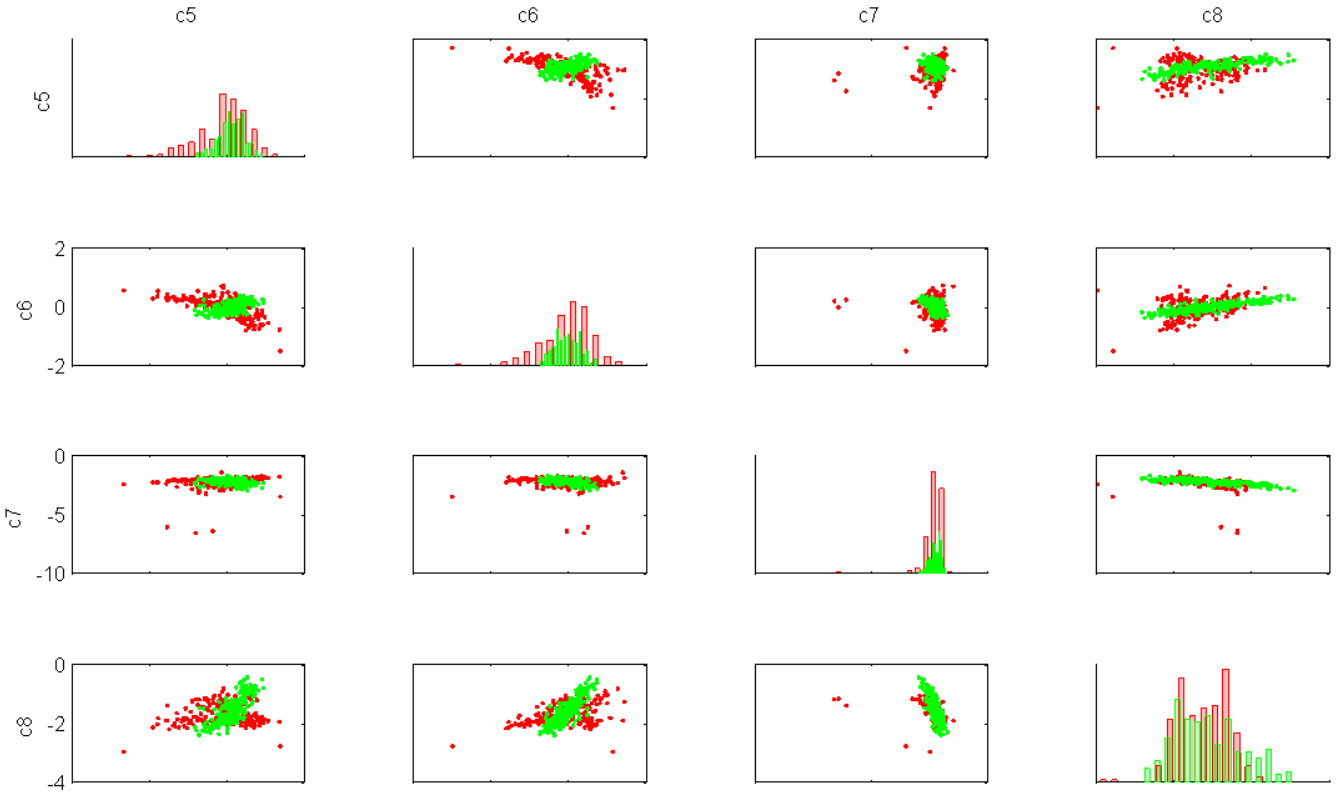
(Adapted from Uhlendorf *et al.* PNAS 2012)

$$y_j^i = yECitrine^i(t_j) + \varepsilon_j^i, \qquad \varepsilon_j^i \sim \mathcal{N}\left(0, \sigma^2\left(yECitrine^i(t_j)\right)\right)$$

Identification results

Identification data generated by a first control experiment on about 100 cells

- Parameter estimates of TS more dispersed than those of ME

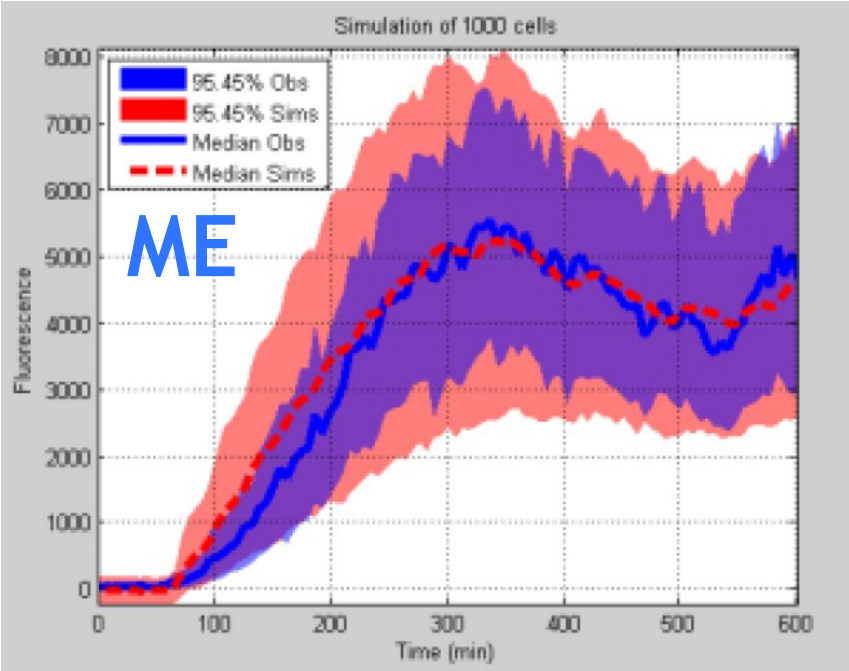


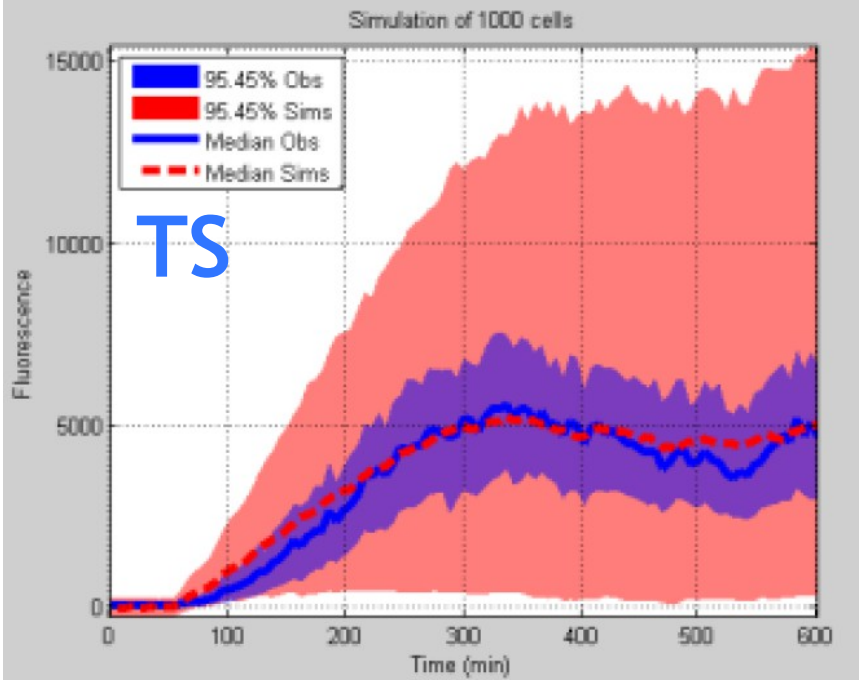
1
2

Validation results

Validation data generated by a new control experiment (with a different input)

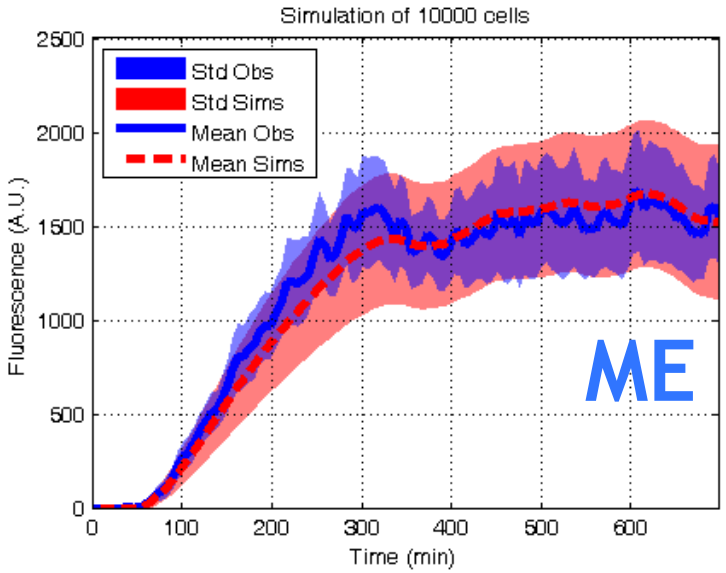
- Simulations with ME population statistics reproduce validation dataset rather accurately
- Simulations with TS population statistics are overly dispersed

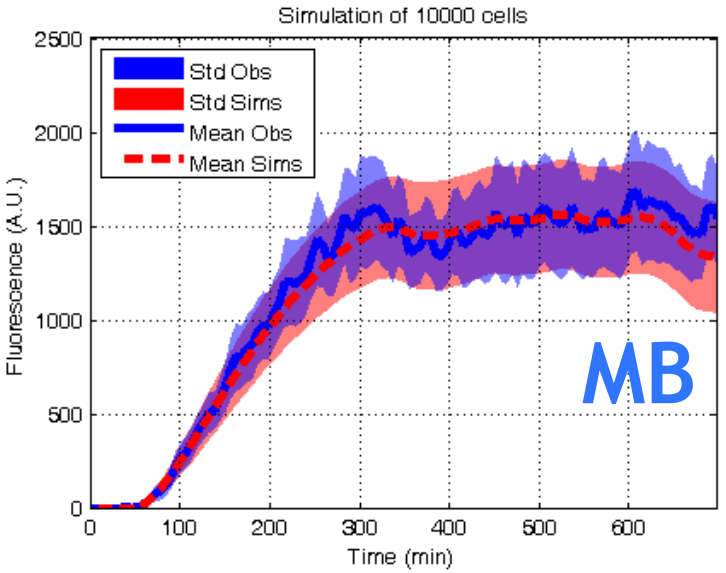




Comparison with intrinsic noise approach

Moment-Based (MB) inference (Zechner *et al.*, PNAS 2012) : Find parameter values of a purely intrinsic noise model that best explain mean/variance of the cell output over time



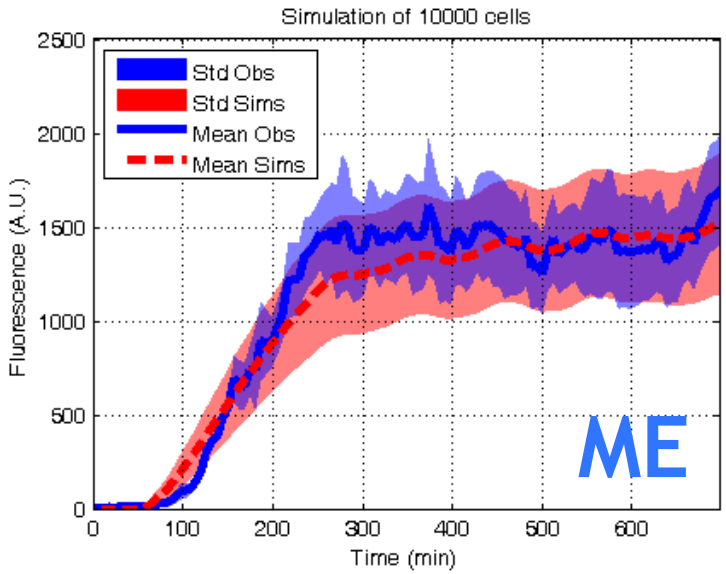


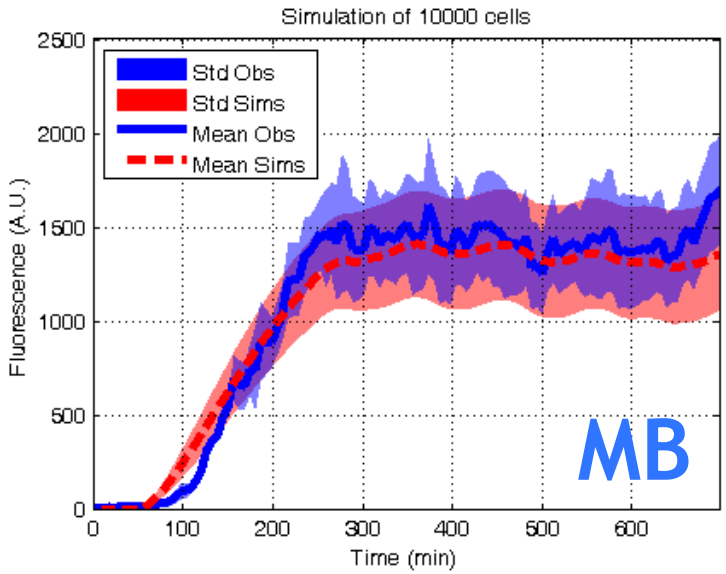
(Gonzalez et al., ECC 2013)

- Reasonable data fitting results
- Moment-Based (MB) identification explains data slightly better than ME

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

Validation results





(Gonzalez et al., ECC 2013)

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

- Reasonable validation results
- ME explains validation data at least as good as moment fitting

- … Intrinsic or extrinsic noise ? Or both ?

Results and perspectives

Results so far (early results in ECC 2013, publications in preparation) :

- Reasonable estimates of population statistics (means, covariances)
- Individual parameter estimates correlate with biological quantities of interest

Perspectives : Many…

- Further methodological advances (modelling/identification tools)
- Analysis of biological system(s)