Inheritance and variability of kinetic gene expression parameters in microbial cells: Modelling and inference from lineage tree data

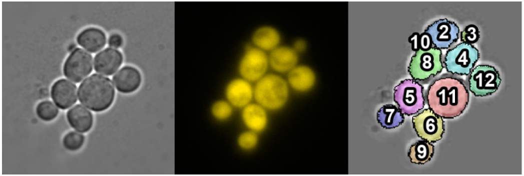
Aline Marguet[1], Marc Lavielle[2], Eugenio Cinquemani[1].

---

[1] Univ. Grenoble Alpes, Inria, 38000 Grenoble, France

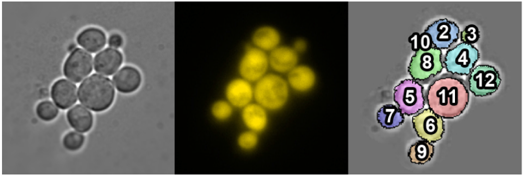
[2] Inria Saclay & Ecole Polytechnique, Palaiseau, France

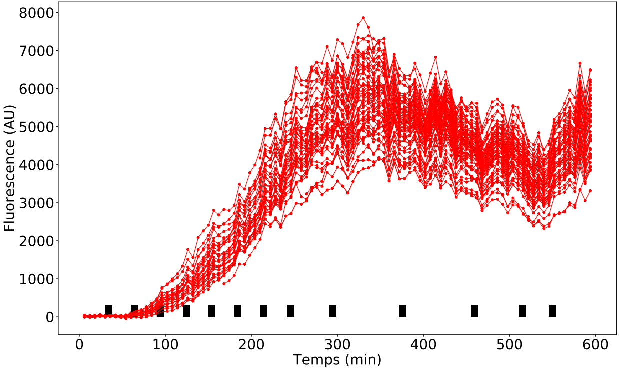
Gene expression variability



Llamosi et *al.*, What population reveals about cell identity: Single-cell parameter estimation of models of gene expression in yeast, 2016, *PLOS Comput. Biol.* 12(2), 2016.

Gene expression variability





Llamosi et *al.*, What population reveals about cell identity: Single-cell parameter estimation of models of gene expression in yeast, 2016, *PLOS Comput. Biol.* 12(2), 2016.

Mathematical modeling of variability

Usually, cells are modeled as independent individuals.

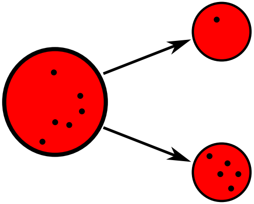
cell division $\Rightarrow$ correlations between mother and daughter cells

Mathematical modeling of variability

Usually, cells are modeled as independent individuals.

cell division $\Rightarrow$ correlations between mother and daughter cells

- random partitioning

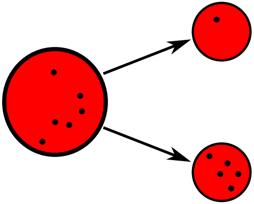


Mathematical modeling of variability

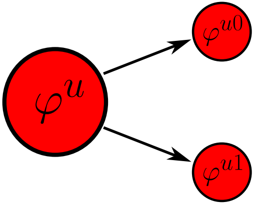
Usually, cells are modeled as independent individuals.

cell division ⇒ correlations between mother and daughter cells

- random partitioning



- inheritance

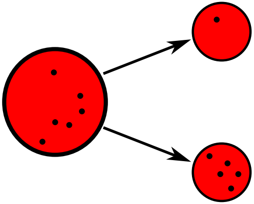


Mathematical modeling of variability

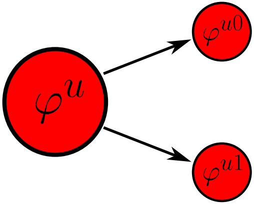
Usually, cells are modeled as independent individuals.

cell division $\Rightarrow$ correlations between mother and daughter cells

- ♦ random partitioning



- ♦ inheritance



Questions: ♦ do parameters usually treated as independent across cells show inheritance? ♦ to what extend are these parameters conserved from one generation to the next?

Gene expression modelling over a lineage tree

Identification from lineage tree data

Validation *in silico* of the ARME algorithm

Application to the study of yeast osmotic shock response

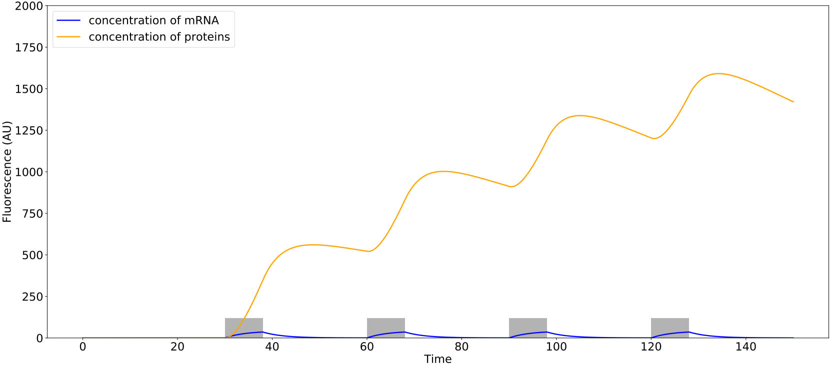
Plan

- Gene expression modelling over a lineage tree
- Identification from lineage tree data
- Validation *in silico* of the ARME algorithm
- Application to the study of yeast osmotic shock response

Dynamical model of gene expression

$$\begin{cases} \dot{m}(t) &= k_m u(t) - g_m m(t) \\ \dot{p}(t) &= k_p m(t) - g_p p(t). \end{cases}$$

- $u(t)$: activity of transcription factors,
- k_m, g_m: production and decay rate of the mRNA,
- k_p, g_p: production and decay rate of the protein.



Individual parameters: mixed-effect modelling

Each cell v in the population has its own parameters

$$\psi_v = (k_m^v, g_m^v, k_p^v, g_p^v)$$

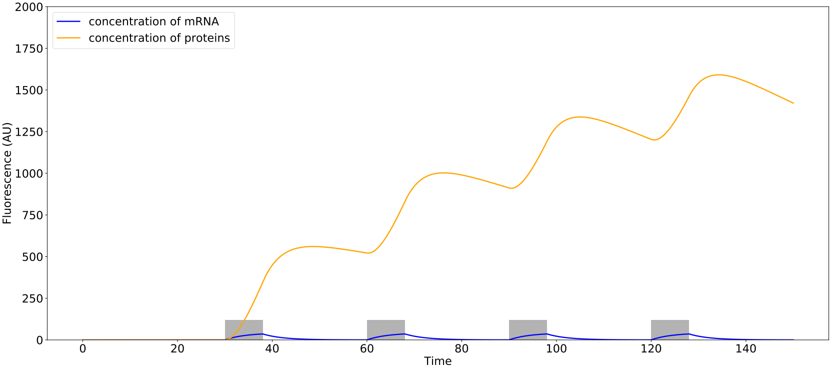
Individual parameters: mixed-effect modelling

Each cell v in the population has its own parameters

$$\psi_v = (k_m^v, g_m^v, k_p^v, g_p^v)$$

- Variability in the response to the same stimulus:

$$\begin{cases} \dot{m}(t) & = k_m^v u(t) - g_m^v m(t) \\ \dot{p}(t) & = k_p^v m(t) - g_p^v p(t). \end{cases}$$



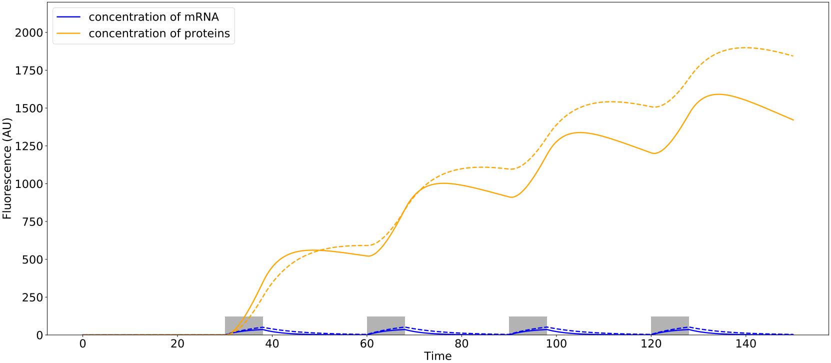
Individual parameters: mixed-effect modelling

Each cell v in the population has its own parameters

$$\psi_v = (k_m^v, g_m^v, k_p^v, g_p^v)$$

- Variability in the response to the same stimulus:

$$\begin{cases} \dot{m}(t) & = k_m^v u(t) - g_m^v m(t) \\ \dot{p}(t) & = k_p^v m(t) - g_p^v p(t). \end{cases}$$



Modelling inheritance

- Individual parameters:

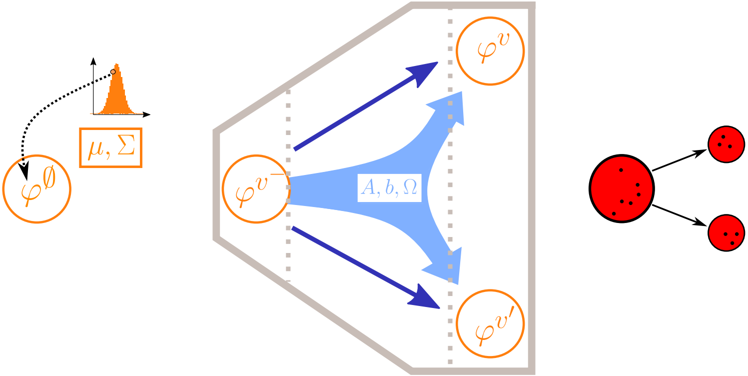
$$\psi^v = (k_m^v, g_m^v, k_p^v, g_p^v) \text{ and } \varphi^v = \log(\psi^v).$$

Modelling inheritance

- Individual parameters:

$$\psi^v = (k_m^v, g_m^v, k_p^v, g_p^v) \text{ and } \varphi^v = \log(\psi^v).$$

- Transmission mechanism:



$$\varphi^{\emptyset} \sim \mathcal{N}(\mu, \Sigma)$$
$$\Sigma = A \Sigma A^T + \Omega$$

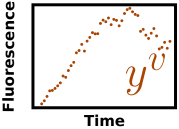
$$\varphi^v = A\varphi^{v^-} + (\mathbf{I} - A)b + \eta^v$$
$$\eta^v \sim \mathcal{N}(0, \Omega)$$

equal sharing
at division.

Identification problem

- Additive noise model for the fluorescence measurements:

$$Y_j^v = p(t_j^v, \psi_v) + h\varepsilon_j^v$$

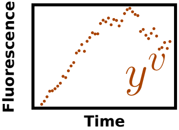


where $h \geq 0$, $(\varepsilon_j^v, j = 1, \ldots, n^v,\ v \in V)$ are i.i.d. $\varepsilon_j^v \sim \mathcal{N}(0,1)$.

Identification problem

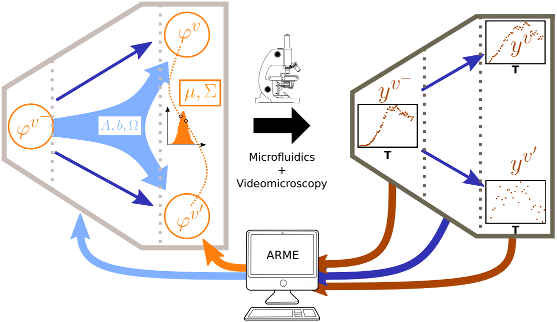
- Additive noise model for the fluorescence measurements:

$$Y_j^v = p(t_j^v, \psi_v) + h\varepsilon_j^v$$



where $h \geq 0$, $(\varepsilon_j^v, j = 1, \ldots, n^v,\ v \in V)$ are i.i.d. $\varepsilon_j^v \sim \mathcal{N}(0,1)$.

- Goal: estimate $\theta = (A, b, \Omega, h)$ from y and lineage informations W.



Plan

Gene expression modelling over a lineage tree

Identification from lineage tree data

Validation *in silico* of the ARME algorithm

Application to the study of yeast osmotic shock response

ARME algorithm : a generalization of the SAEM algorithm

1. Initialisation : $\theta_0 = \vartheta_0$.
2. For k from 0 to N,
   - Simulation step : using MCMC methods, simulate

     $$(\varphi_v^k)_{v\geq 0} \sim p((\varphi_v)_{v\geq 0}|Y, W, \theta_k).$$

   - Stochastic approximation step :

     $$Q_{k+1}(\theta) = Q_k(\theta) + \gamma_k \left(\log(p(Y, (\varphi_v^k)_{v\geq 0}|W, \theta) - Q_k(\theta)\right).$$

   - Maximization step : $\theta_{k+1} = \operatorname{argmax}_\theta \left(Q_{k+1}(\theta)\right)$.

⚠ For the simulation step: dependencies between individuals.

"Convergence of a stochastic approximation version of the EM algorithm." B. Delyon, M. Lavielle, E. Moulines, Ann. Statist. 27 (1999), no. 1, 94–128.

Implementation: detailed simulation step

⋄ Using Metropolis-Hasting algorithm with several proposal law, simulate

$$(\varphi_v^k)_{v\geq 0} \sim p((\varphi_v)_{v\geq 0} | Y, W, \theta_k).$$

Update of $\varphi = (\varphi_v)_{v\geq 0}$

- at the **population level**: takes into account every correlations, very low acceptance rate.

Implementation: detailed simulation step

⋄ Using Metropolis-Hasting algorithm with several proposal law, simulate

$$(\varphi_v^k)_{v\geq 0} \sim p((\varphi_v)_{v\geq 0}|Y, W, \theta_k).$$

Update of $\varphi = (\varphi_v)_{v\geq 0}$

- at the **population level**: takes into account every correlations, very low acceptance rate.

- at the **generation level**: takes into account the correlation with the previous and the next generation, better acceptance rate.

Implementation: detailed simulation step

⋄ Using Metropolis-Hasting algorithm with several proposal law, simulate

$$(\varphi_v^k)_{v\geq 0} \sim p((\varphi_v)_{v\geq 0}|Y, W, \theta_k).$$

Update of $\varphi = (\varphi_v)_{v\geq 0}$

- at the **population level**: takes into account every correlations, very low acceptance rate.

- at the **generation level**: takes into account the correlation with the previous and the next generation, better acceptance rate.

- at the **individual level**: does not take into account any correlation, adaptative acceptance rate.

Plan

Gene expression modelling over a lineage tree

Identification from lineage tree data

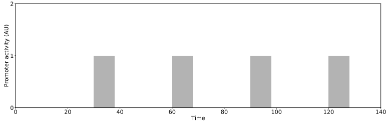
Validation *in silico* of the ARME algorithm

Application to the study of yeast osmotic shock response

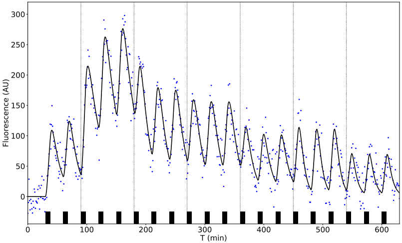
Validation *in silico* of the ARME algorithm

Simulation of 20 datasets with

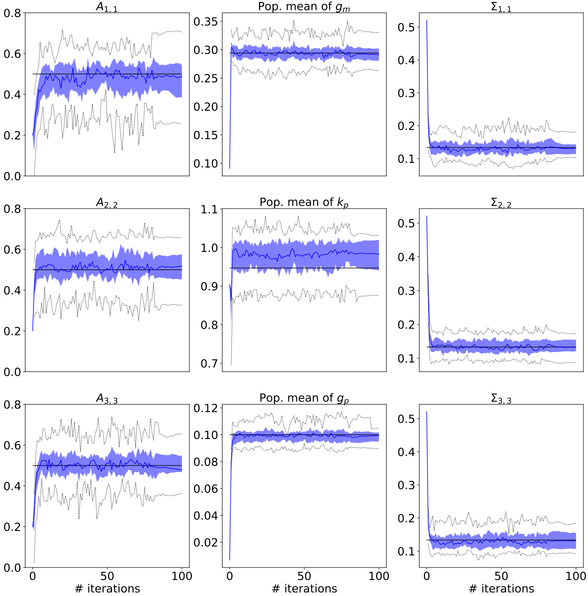
- $u(t) =$



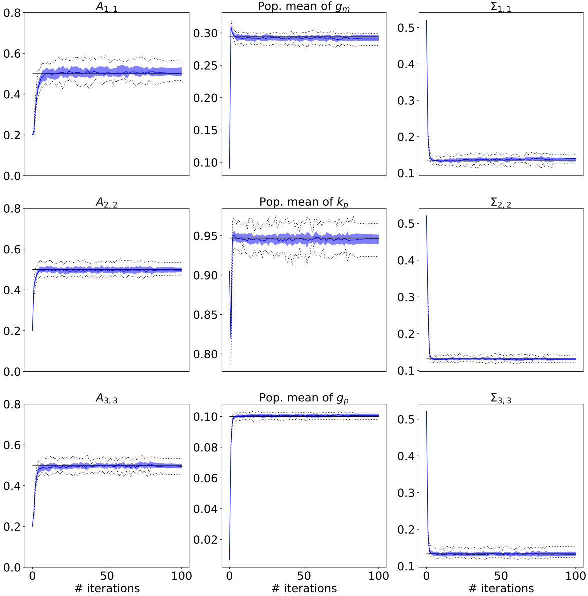
- 128 individuals (7 generations),
- a fixed value of k_m, for identifiability reasons,
- inheritance parameter $A = \mathrm{Diag}(0.5, 0.5, 0.5)$,
- global mean parameter $b = [\log(0.294), \log(0.947), \log(0.1)]^T$,
- global covariance parameter $\Omega = \mathrm{Diag}(0.1, 0.1, 0.1)$,
- noise of measure $h = 20$,



Validation *in silico* of the ARME algorithm



Validation *in silico* of the ARME algorithm



Plan

Gene expression modelling over a lineage tree

Identification from lineage tree data

Validation *in silico* of the ARME algorithm

Application to the study of yeast osmotic shock response

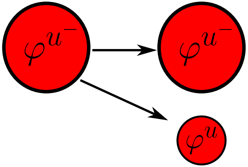
Application to the study of yeast osmotic shock response

- Extended model for gene expression (Llamosi *et al.* (2016)):

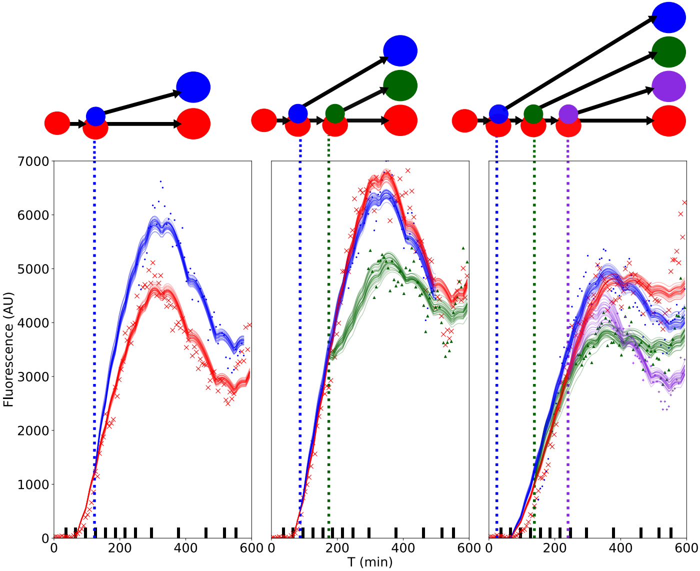
$$\begin{cases} \dot{u}(t) & = & k_h u_c(t) - g_h u(t), \\ \dot{m}(t) & = & k_m u(t) - g_m m(t), \\ \dot{p}(t) & = & k_p m(t) - g_p p(t), \end{cases}$$

with fixed values for k_h and g_h.

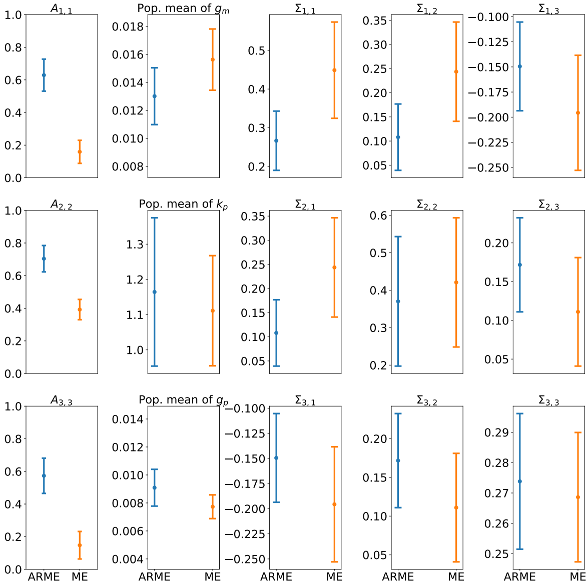
- Maturation time for reporter molecules.
- Budding yeast (*S. cerevisiae*): the mother keeps its own kinetic parameter at division.



Single-cell data fits after ARME identification



Results from identification of a ARME model



Conclusions

- Daughter cell parameters are determined by the mother to an extent as large as 60% (a state-of-the art indirect method assessed this value at $20 - 40\%$).
- Indirect methods underestimate inheritance
- Inheritance is equal for the different parameters: it acts at the level of global regulatory factors (at least for the system and data we examined).

Perspectives

- Consider intrinsic noise, more complex inheritance models, etc.
- Proof of the convergence of the algorithm

Conclusions

- Daughter cell parameters are determined by the mother to an extent as large as 60% (a state-of-the art indirect method assessed this value at $20 - 40\%$).
- Indirect methods underestimate inheritance
- Inheritance is equal for the different parameters: it acts at the level of global regulatory factors (at least for the system and data we examined).

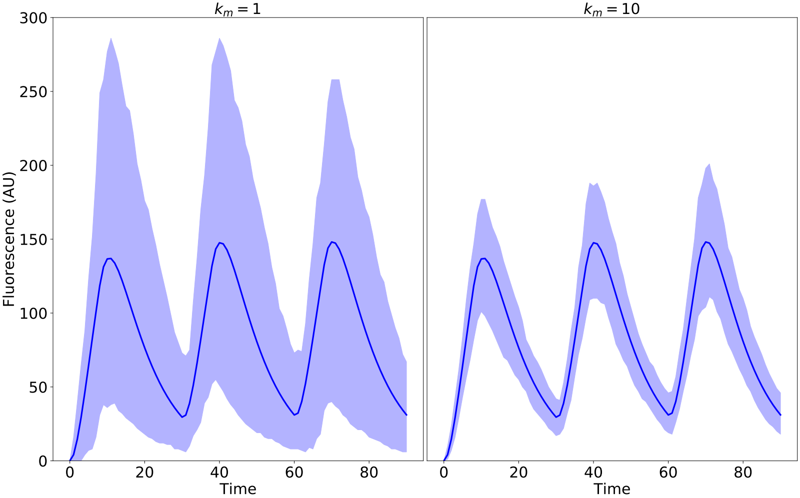
Perspectives

- Consider intrinsic noise, more complex inheritance models, etc.
- Proof of the convergence of the algorithm

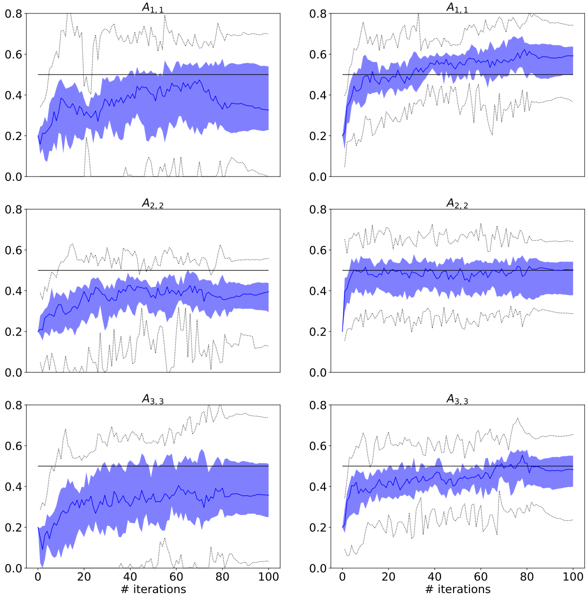
Thank you for your attention!

Performance in presence of intrinsic noise

$$\emptyset \xrightarrow{k_m u(t)} M, \quad M \xrightarrow{g_m} \emptyset, \quad M \xrightarrow{k_p} M + P, \quad P \xrightarrow{g_p} \emptyset,$$



Performance in presence of intrinsic noise



Experimental design