

Bursting in gene expression model

Romain Yvinec¹, Michael C. Mackey², Marta Tyran-Kamińska³, Changjiing Zhuge⁴, Jinzhi Lei⁴

¹BIOS group, INRA Tours, France.

²McGill University, Canada.

³University of Silesia, Poland.

⁴Tsinghua University, China.

Hi Khanh !

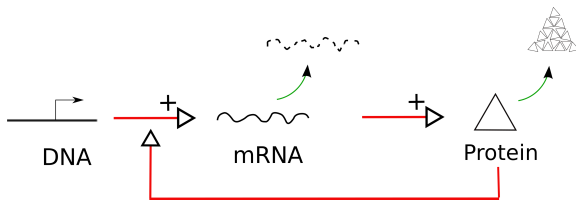


Goodwin's deterministic model

Stochastic gene expression model : forward problem

Stochastic gene expression model : inference

Taking into account division

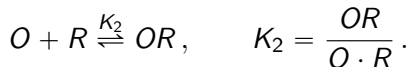
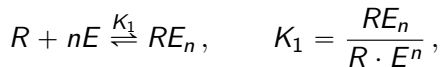


$$\left\{ \begin{array}{l} \frac{dM}{dt} = \lambda_1(E) - \gamma_1 M, \\ \frac{dI}{dt} = \lambda_2 M - \gamma_2 I, \\ \frac{dE}{dt} = \lambda_3 I - \gamma_3 E. \end{array} \right.$$

Goodwin (1965), Griffith (1968), Othmer (1976), Selgrade (1979)...

The transcription rate function λ_1

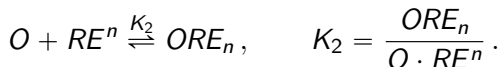
- ▶ **Inducible Operon** : Repressors R interacts with both the Operator O and the Effector E ,



With **QSSA**, and if $O_{tot} \ll R_{tot}$,

$$\lambda_1(E) \sim \frac{O}{O_{tot}} = \frac{1 + K_1 E^n}{1 + K_2 R_{tot} + K_1 E^n}. \quad (1)$$

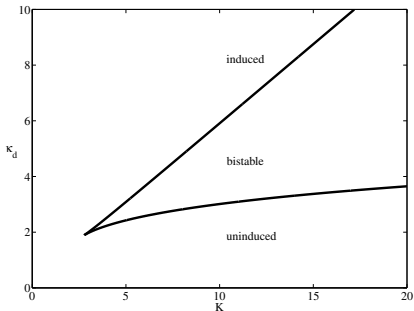
- ▶ **Repressible Operon** : Similar but



and we get

$$\lambda_1(E) \sim \frac{O}{O_{tot}} = \frac{1 + K_1 E^n}{1 + (K_1 + K_2 R_{tot}) E^n}. \quad (2)$$

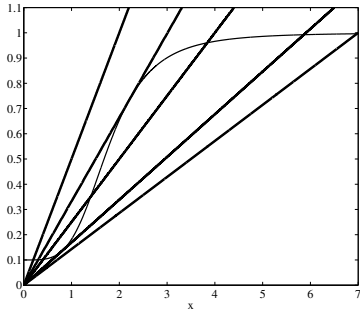
Bifurcation analysis in ODE



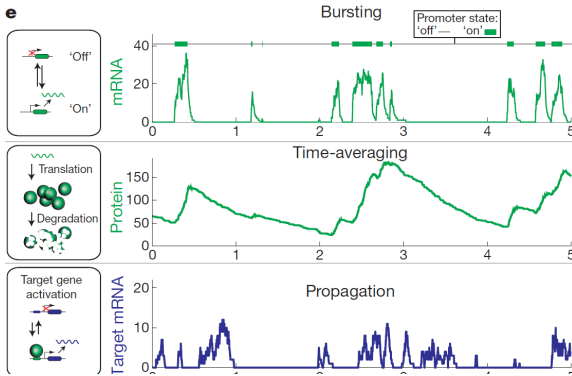
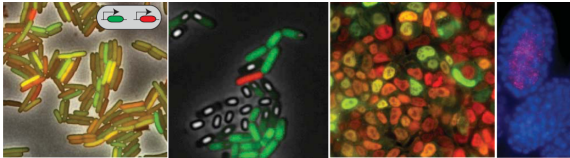
- ▶ **Inducible ($K > 1$) :**
Mono-stability or Bi-stability.
- ▶ **Repressible ($K < 1$) :**
Mono-stability or limit cycle.

$$\begin{cases} \frac{dx_1}{dt} = \gamma_1[\lambda_1(x_3) - x_1], \\ \frac{dx_2}{dt} = \gamma_2(x_1 - x_2), \\ \frac{dx_3}{dt} = \gamma_3(x_2 - x_3). \end{cases}$$

Here $\lambda_1(x) = \kappa_d \frac{1 + x^n}{K + x^n}$.

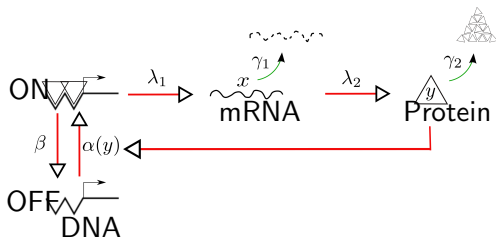


Eldar and Elowitz (Nature 2010)



→ ODEs are forbidden !

'New' Central dogma



$$\left\{ \begin{array}{l} \frac{dx}{dt} = G(t)\lambda_1(y(t)) - \gamma_1 x(t), \\ \frac{dy}{dt} = \lambda_2 x(t) - \gamma_2 y(t), \\ (G = 0) \xrightleftharpoons[\beta(y(t))]{\alpha(y(t))} (G = 1). \end{array} \right.$$

Rigney & Schieve (1977), Berg (1978), Peccoud & Ycart (1995), Thattai & Van Oudenaarden (2001)...

Can we perform a systematic bifurcation theory on such systems ?

- ▶ We are interested in long time behavior.
- ▶ We want to know how many modes has the stationary distribution.
- ▶ This requires in practice 'analytical' solution.

Can we perform a systematic bifurcation theory on such systems ?

- ▶ We are interested in long time behavior.
 - ▶ We want to know how many modes has the stationary distribution.
 - ▶ This requires in practice 'analytical' solution.
- **I don't know how to do it, let's reduce the model !**

A subclass of the 'three-stage' (DNA, mRNA, Protein) model is the 1D-bursting model (Storage model)

$$Lf(x) = \underbrace{-\gamma(x)f'(x)}_{\text{degradation}} + \underbrace{\lambda(x) \int_0^{\infty} (f(x+y) - f(x))h(x,y)dy}_{\text{bursting}}$$

where

- ▶ γ : degradation rate (deterministic drift)
- ▶ λ : bursting rate (jump intensity)
- ▶ h : bursting size distribution (jump size), $\int_0^{\infty} h(x,y)dy = 1$.

Time-dependent probability density $u(x, t)$ satisfies

$$\frac{\partial u(x, t)}{\partial t} - \underbrace{\frac{\partial(\gamma(x)u(x, t))}{\partial x}}_{\text{degradation}} = \underbrace{-\lambda(x)u(x, t)}_{\text{bursting from } x} + \underbrace{\int_0^x \lambda(y)h(y, x-y)u(t, y)dy}_{\text{bursting to } x}$$

$$Lf(x) = \underbrace{-\gamma(x)f'(x)}_{\text{degradation}} + \underbrace{\lambda(x) \int_0^{\infty} (f(x+y) - f(x))h(x,y)dy}_{\text{bursting}}$$

Typical example for h is given by the exponential distribution, with jump size independent of the starting point,

$$h(x,y) = \frac{1}{b} \exp(-y/b), \quad \int_0^{\infty} yh(x,y)dy = b.$$

Any stationary distribution u^* satisfies

$$\int_0^{\infty} Lf(x)u^*(x)dx = 0, \quad \forall \text{ (suitable) test function } f \quad (3)$$

→ it seems difficult to solve (3) for "general" h

$$Lf(x) = \underbrace{-\gamma(x)f'(x)}_{\text{degradation}} + \underbrace{\lambda(x) \int_0^{\infty} (f(x+y) - f(x))h(x,y)dy}_{\text{bursting}}$$

If h is generalized exp. distribution with "x-dependent" burst size,

$$h(x,y) = -\frac{\nu'(x+y)}{\nu(x)}, \nu \searrow 0,$$

$$h(x,y) = \frac{1}{b(x+y)} \exp\left(-\int_x^{x+y} 1/b(z)dz\right), \nu(x) = \exp\left(-\int^x 1/b(z)dz\right)$$

then any stationary distribution u^* satisfies

$$\int_0^{\infty} \left[-\gamma(x)u^*(x) + \nu(x) \int_0^x \frac{\lambda(y)}{\nu(y)} u^*(y) dy \right] f'(x) dx = 0.$$

which leads to

$$u^*(x) = \frac{\nu(x)}{C\gamma(x)} \exp\left(\int^x \frac{\lambda(y)}{\gamma(y)} dy\right)$$

Theorem

If $\gamma > 0$ is continuous, $\int_0^\delta \frac{1}{\gamma} dx = \infty$, $\nu > 0$ abs. C^0 , $\nu \searrow 0$

$$\int_0^\infty \frac{\nu(x)}{\gamma(x)} e^{\int^x \frac{\lambda(y)}{\gamma(y)} dy} dx < \infty, \quad \text{and} \quad \int_0^\infty (-\nu'(x)) e^{\int^x \frac{\lambda(y)}{\gamma(y)} dy} dx < \infty,$$

then $u(t, x)$ converges (for any initial density) towards u^* , e.g.

$$\int_0^\infty |u(t, x) - u^*(x)| dx \rightarrow 0, \text{ as } t \rightarrow \infty.$$

Theorem

If $\gamma > 0$ is continuous, $\int_0^\delta \frac{1}{\gamma} dx = \infty$, $\nu > 0$ abs. C^0 , $\nu \searrow 0$

$$\int_0^\infty \frac{\nu(x)}{\gamma(x)} e^{\int^x \frac{\lambda(y)}{\gamma(y)} dy} dx < \infty, \quad \text{and} \quad \int_0^\infty (-\nu'(x)) e^{\int^x \frac{\lambda(y)}{\gamma(y)} dy} dx < \infty,$$

then $u(t, x)$ converges (for any initial density) towards u^* , e.g.

$$\int_0^\infty |u(t, x) - u^*(x)| dx \rightarrow 0, \text{ as } t \rightarrow \infty.$$

→ The proof is based on Foguel's alternative : roughly, if all states communicate, then either you have either convergence or 'escape'.

Theorem

If $\gamma > 0$ is continuous, $\int_0^\delta \frac{1}{\gamma} dx = \infty$, $\nu > 0$ abs. C^0 , $\nu \searrow 0$

$$\int_0^\infty \frac{\nu(x)}{\gamma(x)} e^{\int^x \frac{\lambda(y)}{\gamma(y)} dy} dx < \infty, \quad \text{and} \quad \int_0^\infty (-\nu'(x)) e^{\int^x \frac{\lambda(y)}{\gamma(y)} dy} dx < \infty,$$

then $u(t, x)$ converges (for any initial density) towards u^* , e.g.

$$\int_0^\infty |u(t, x) - u^*(x)| dx \rightarrow 0, \text{ as } t \rightarrow \infty.$$

Remark

If the first integral is infinite, then the process 'escape' towards 0 or ∞ (in infinite time). If the second integral is infinite, there is explosion in finite time.

Theorem

If $\gamma > 0$ is continuous, $\int_0^\delta \frac{1}{\gamma} dx = \infty$, $\nu > 0$ abs. C^0 , $\nu \searrow 0$

$$\int_0^\infty \frac{\nu(x)}{\gamma(x)} e^{\int^x \frac{\lambda(y)}{\gamma(y)} dy} dx < \infty, \quad \text{and} \quad \int_0^\infty (-\nu'(x)) e^{\int^x \frac{\lambda(y)}{\gamma(y)} dy} dx < \infty,$$

then $u(t, x)$ converges (for any initial density) towards u^* , e.g.

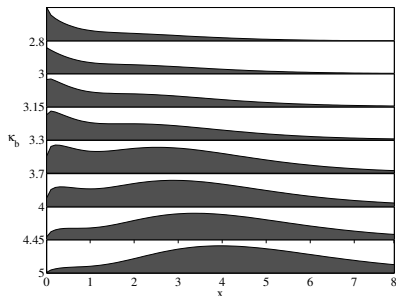
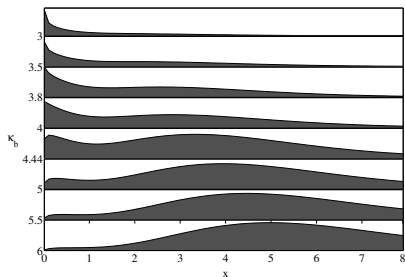
$$\int_0^\infty |u(t, x) - u^*(x)| dx \rightarrow 0, \text{ as } t \rightarrow \infty.$$

Remark

Under different conditions, one may prove exponential convergence (even for 'general' h) without explicit u^* (cf Meyn-Tweedie approach).

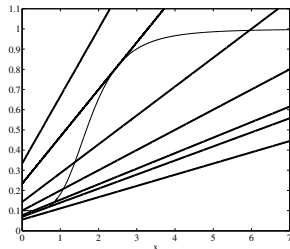
Stationary distribution $(\lambda, \gamma, \nu) \Rightarrow (u^*)$

$$u^*(x) = \frac{\nu(x)}{C\gamma(x)} \exp\left(\int^x \frac{\lambda(y)}{\gamma(y)} dy\right)$$



Here $\lambda(x) = \kappa_b \frac{1 + x^n}{K + x^n}$ and $\nu(x) = \exp(-x/b)$.

Bifurcation analysis in SDE

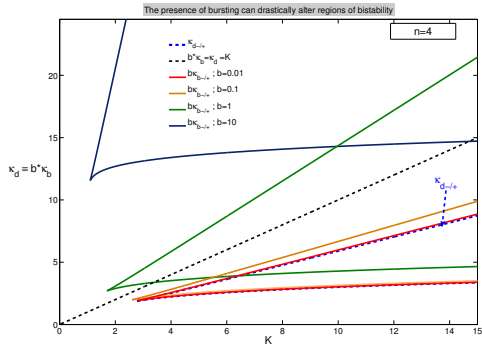


- ▶ **Inducible :**
Uni-modal or Bi-modal.
- ▶ **Repressible :**
Uni-modal.

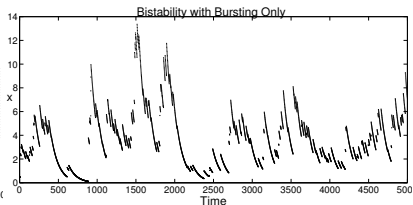
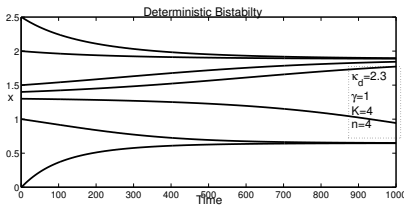
$$u^*(x) = \frac{\nu(x)}{C\gamma(x)} \exp\left(\int^x \frac{\lambda(y)}{\gamma(y)} dy\right)$$

$$\frac{du^*}{dx} = \left[\lambda(x) - \gamma\left(1 + \frac{x}{b}\right)\right] \frac{u^*(x)}{\gamma x}.$$

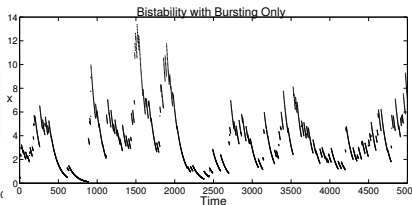
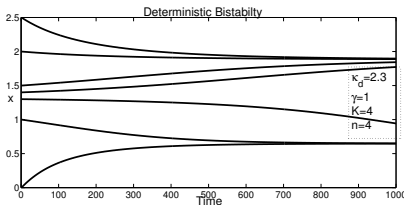
with $\nu(x) = \exp(-x/b)$ and $\gamma(x) = \gamma x$.



- ▶ Careful! The two notions of deterministic bistability and 'stochastic bistability' (bimodality) are in fact *quiet different*



- ▶ Careful! The two notions of deterministic bistability and 'stochastic bistability' (bimodality) are in fact *quiet different*



- ▶ (mean) Switching time : can quantify the 'stability' of each state. MFPT are also explicit if

$$h(x, y) = \frac{1}{b(x+y)} \exp\left(-\int_x^{x+y} 1/b(z)\right)$$

For $x < x_0$, the MFPT to reach a **lower** level x from x_0 is

$$V_{l,x}(x_0) = \int_x^{x_0} \frac{1}{\gamma(y)} dy + \int_x^{x_0} \frac{\lambda(y)}{\gamma(y)\nu(y)} \int_y^{\infty} \frac{\nu(w)}{\gamma(w)} e^{Q(y)-Q(w)} dw dy$$

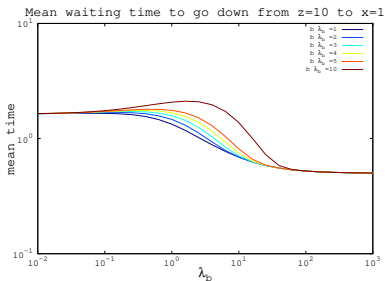
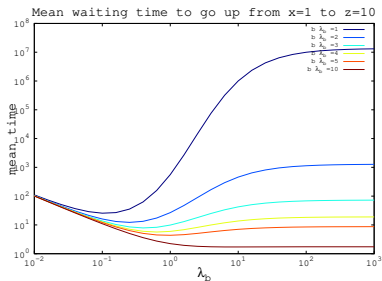
For $z > x_0$. The MFPT to reach a **higher** level z from x_0 is

$$V_{u,z}(x_0) = \int_0^z \frac{\nu(y)}{\gamma(y)\nu(z)} e^{Q(z)-Q(y)} dy - \int_{x_0}^z \frac{1}{\gamma(y)} dy \\ + \int_{x_0}^z \frac{\lambda(y)}{\gamma(y)\nu(y)} \int_0^y \frac{\nu(w)}{\gamma(w)} e^{Q(y)-Q(w)} dw dy$$

Remark

FPT are **strongly asymmetric** here.

The mean waiting time is non-monotonic with respect to the bursting rate (even for fixed mean).



From the 3-stage to the bursting model ?

$$\left\{ \begin{array}{l} \frac{dx}{dt} = G(t)\lambda_1(y(t)) - \gamma_1 x(t), \\ \frac{dy}{dt} = \lambda_2 x(t) - \gamma_2 y(t), \\ (G=0) \xrightleftharpoons[\beta(y(t))]{\alpha(y(t))} (G=1). \end{array} \right. \quad (4)$$

- ▶ If the mRNA lifetime is short ($\gamma_1 \rightarrow \infty$), we can perform an **adiabatic** reduction ($x(t) \approx G(t) \frac{\lambda_1}{\gamma_1}(y(t))$) :

$$\left\{ \begin{array}{l} \frac{dy}{dt} = G(t) \frac{\lambda_2 \lambda_1}{\gamma_1}(y(t)) - \gamma_2 y(t), \\ (G=0) \xrightleftharpoons[\beta(y(t))]{\alpha(y(t))} (G=1). \end{array} \right. \quad (5)$$

$$\left\{ \begin{array}{l} \frac{dy}{dt} = G(t) \frac{\lambda_2 \lambda_1}{\gamma_1} (y(t)) - \gamma_2 y(t), \\ (G = 0) \xrightleftharpoons[\beta(y(t))]{\alpha(y(t))} (G = 1). \end{array} \right. \quad (6)$$

- ▶ If the Gene active periods are short ($\beta \rightarrow \infty$), we obtain the **bursting** model

$$\frac{dy}{dt} = Z(t) - \gamma_2 y(t), \quad (7)$$

where $Z = \sum_i Z_i \delta_{T_i}$ is a jump process, of jump rate $\alpha(y(t))$ and jump size cumulative distribution **of separated form**

$$\mathbb{P}\{y(T_i^+) \geq z \mid y(T_i^-) = y\} = \exp\left(-\int_y^z \frac{\gamma_1 \beta}{\lambda_1 \lambda_2}(w) dw\right).$$

$$\left\{ \begin{array}{l} \frac{dy}{dt} = G(t) \frac{\lambda_2 \lambda_1}{\gamma_1} (y(t)) - \gamma_2 y(t), \\ (G = 0) \xrightleftharpoons[\beta(y(t))]{\alpha(y(t))} (G = 1). \end{array} \right. \quad (6)$$

- ▶ If the Gene active periods are short ($\beta \rightarrow \infty$), we obtain the **bursting** model

$$\frac{dy}{dt} = Z(t) - \gamma_2 y(t), \quad (7)$$

where $Z = \sum_i Z_i \delta_{T_i}$ is a jump process, of jump rate $\alpha(y(t))$ and jump size cumulative distribution **of separated form**

$$\mathbb{P}\{y(T_i^+) \geq z \mid y(T_i^-) = y\} = \exp\left(-\int_y^z \frac{\gamma_1 \beta}{\lambda_1 \lambda_2}(w) dw\right).$$

Remark

For a constitutive gene, $b := \frac{\lambda_1 \lambda_2}{\gamma_1 \beta}$ is the average number of proteins produced per Gene activation event.

$$\left\{ \begin{array}{l} \frac{dy}{dt} = G(t) \frac{\lambda_2 \lambda_1}{\gamma_1} (y(t)) - \gamma_2 y(t), \\ (G = 0) \xrightleftharpoons[\beta(y(t))]{\alpha(y(t))} (G = 1). \end{array} \right. \quad (6)$$

- ▶ If the Gene active periods are short ($\beta \rightarrow \infty$), we obtain the **bursting** model

$$\frac{dy}{dt} = Z(t) - \gamma_2 y(t), \quad (7)$$

where $Z = \sum_i Z_i \delta_{T_i}$ is a jump process, of jump rate $\alpha(y(t))$ and jump size cumulative distribution **of separated form**

$$\mathbb{P}\{y(T_i^+) \geq z \mid y(T_i^-) = y\} = \exp\left(-\int_y^z \frac{\gamma_1 \beta}{\lambda_1 \lambda_2}(w) dw\right).$$

Remark

The bursting limit is a way to introduce 'complicated' post-transcriptional regulation without too much cost in terms of 'mathematical complications'....

Small disgression

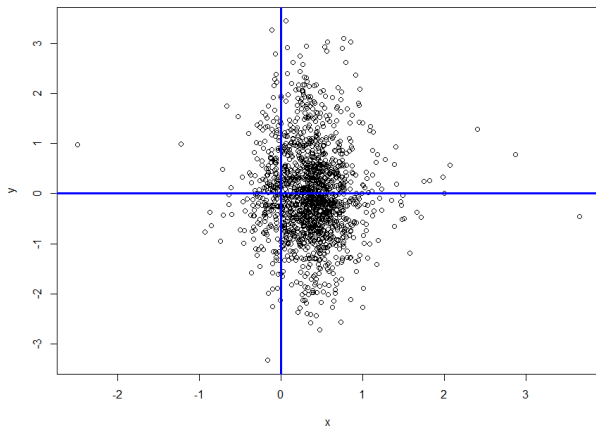
Adiabatic reduction like $x(t) \approx G(t) \frac{\lambda_1}{\gamma_1}(y(t))$ are not really rigorous...

→ x is a fast variable and does not converge in pointwise manner but 'only' in $L^1(0, t)$.

→ The slow variable y only 'sees' **time averaged** variation

Small digression

This can be (one!) explanation of general disagreement between (bulk) RNA-Seq and Proteomics...



Small digression 2

Remark

For the '2-stage' model (Telegraph process),

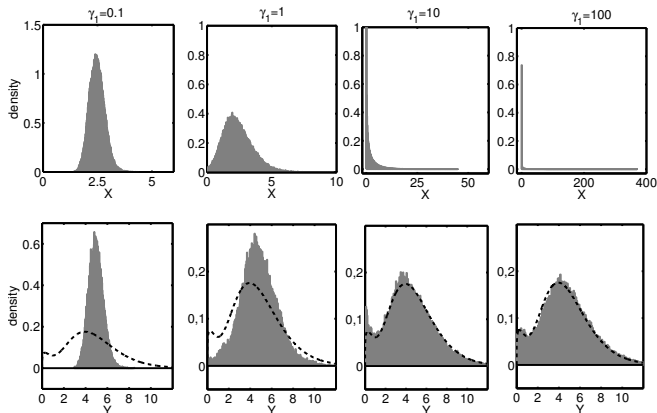
$$\left\{ \begin{array}{l} \frac{dx}{dt} = G(t)\lambda(x(t)) - \gamma x(t), \\ (G = 0) \xrightleftharpoons[\beta(x(t))]{\alpha(x(t))} (G = 1). \end{array} \right. \quad (8)$$

'general' analytical solution are **available** (See Boxma et al. 2005)

$$\frac{du^*}{dx} = \left[\frac{\alpha(x)}{\gamma x} - \frac{\beta(x)}{\lambda(x) - \gamma x} - \frac{\lambda(x)/x - \gamma + \gamma x(\lambda'(x) - \gamma)/(\lambda(x) - \gamma x)}{\lambda(x)} \right] u^* \quad (9)$$

Small recap'

- ▶ We performed an adiabatic reduction to make the problem analytical tractable.
- ▶ We solved the reduced problem for arbitrary coefficients at equilibrium.
- ▶ We performed an (deterministic-analogous) bifurcation study.



Inverse Problem : $(u^*) \Rightarrow (\lambda, \gamma, b)$

For a constitutive gene, we can infer the burst rate (in protein lifetime unit) $\frac{\lambda}{\gamma}$ and the mean burst size b from the first two (stationary) moments

$$\frac{b\lambda}{\gamma} = \mathbb{E}[X],$$

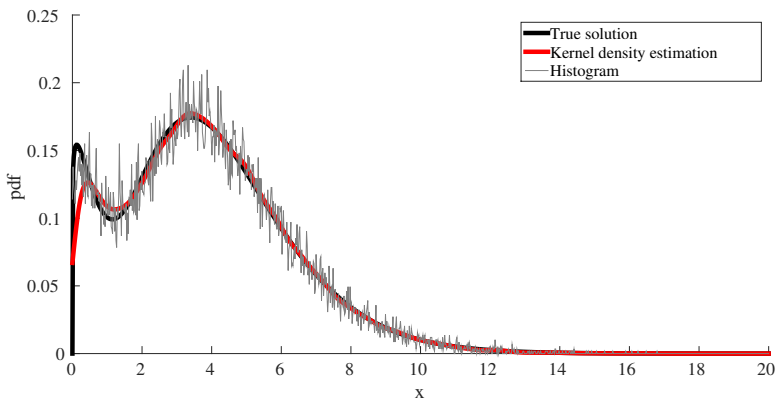
$$b = \frac{\text{Var}(X)}{\mathbb{E}[X]}.$$

For an auto-regulated gene, we can inverse the formula for the stationary pdf :

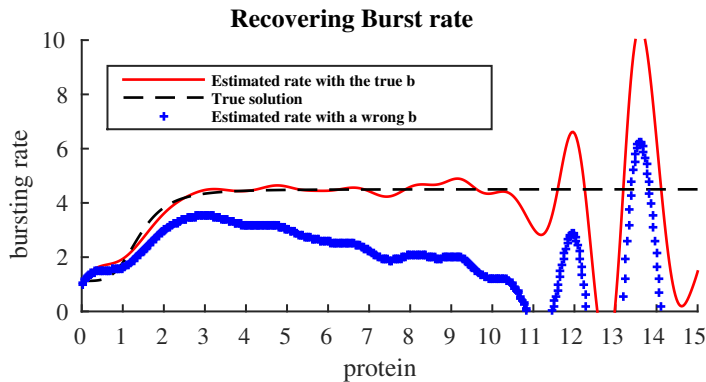
$$\frac{(xu^*(x))'}{u^*(x)} = \frac{\lambda(x)}{\gamma} - \frac{x}{b}.$$

Simulated data

First step : Density reconstruction by Kernel Density Estimate

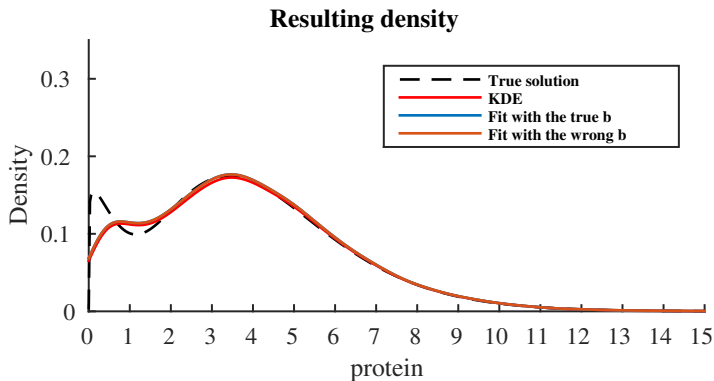


Inferred bursting rate

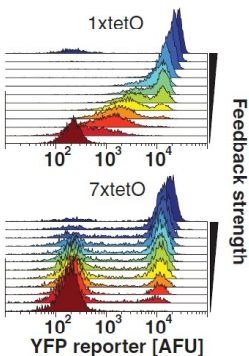
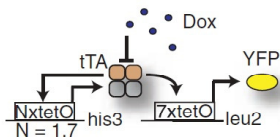


→ 'semi-parametric' inference : highly sensitive to the burst mean b

Resulting Probability Density Function



Single cell data on self-regulating gene



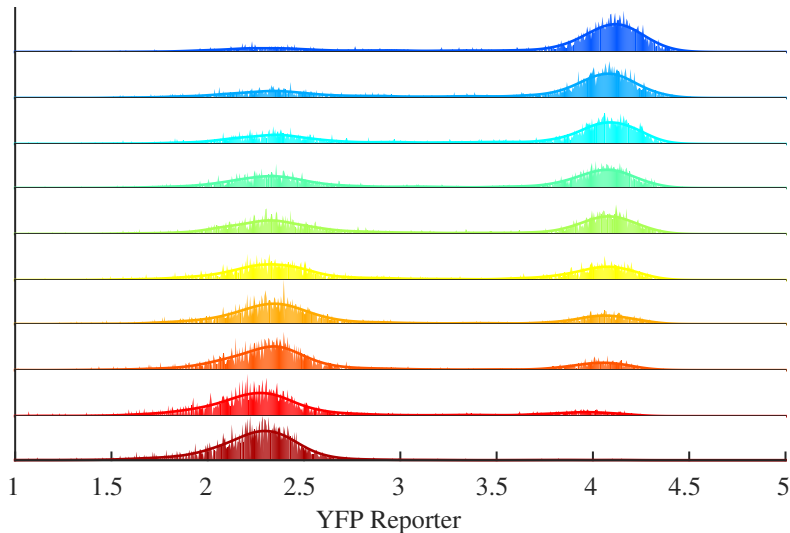
Noise Can Induce Bimodality in Positive Transcriptional Feedback Loops Without Bistability

Tsz-Leung To, *et al.*
Science 327, 1142 (2010);

DOI: 10.1126/science.1178962

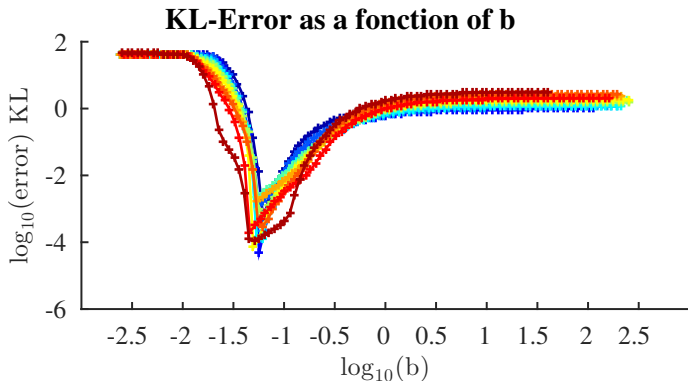
- ▶ Synthetic Tet-Off in budding yeast.
- ▶ Feedback modulated by an external parameter (doxycycline)

1) Kernel Density Estimation

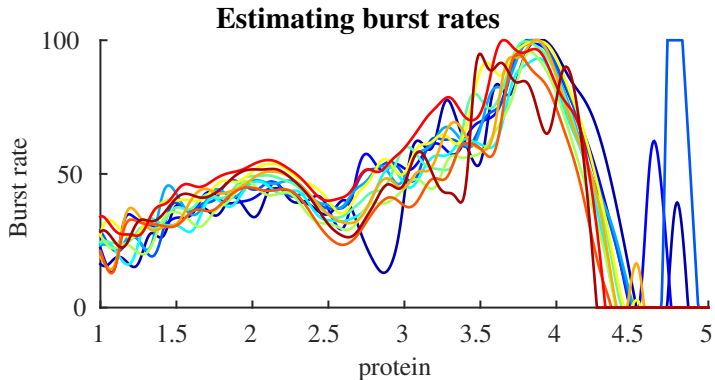


2) Finding the 'best' mean burst size (KL distance)

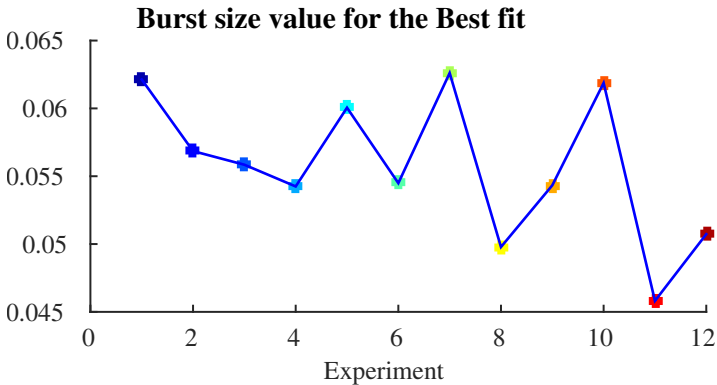
Fix b , infer λ , compute KL distance, modify b , etc...



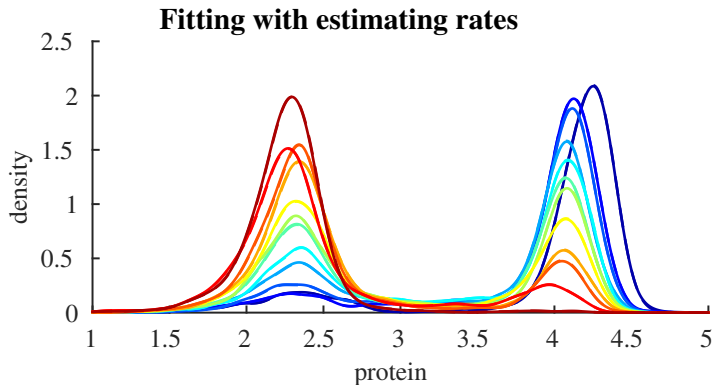
3) Inferred burst rate



3) Inferred mean burst size



4) Resulting Probability Density Function



Small recap' on the inverse problem

- ▶ With the help of the full solution, we obtained a formula to find the parameter functions from the stationary density.
- ▶ We applied this on simulated and real data.
- ▶ The inverse problem is generally ill-posed (cannot find burst size b and burst rate λ at the same time).
- ▶ Although the resulting pdf does usually 'fit' the data.
- ▶ Work still on progress...

Can we deal with cell population ?

Similar results may be obtained for a 'bursting-division' model.

$$Lf(x) = d(x) \int_0^x (f(y) - f(x))\kappa(x, y)dy + \lambda(x) \int_0^\infty (f(x+y) - f(x))h(x, y)dy$$

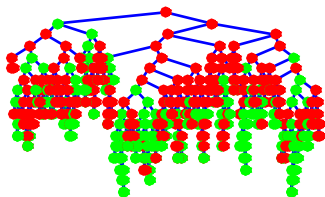
For instance, with uniform repartition kernel ($\kappa(x, y) = 1/x$), constant division rate d and constant exponential burst size ($h(x, y) = \exp(-y/b)$),

$$\frac{d}{dy} u^* = \left[-\frac{\lambda'(y) + d}{\lambda(y) + d} + \frac{\lambda(y)}{\lambda(y) + d} \left(\frac{1}{x} + \frac{1}{b} \right) - \frac{xb^2}{bx + 1} - \frac{1}{x} \right] u^*(y)$$

This may be used to predict the long time behavior of a dividing cell population

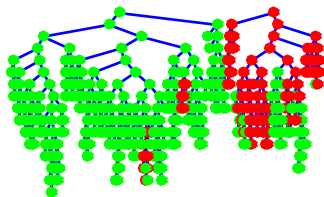
scenario 1

A cell tree

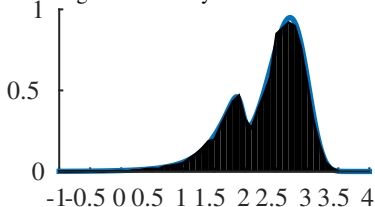


scenario 2

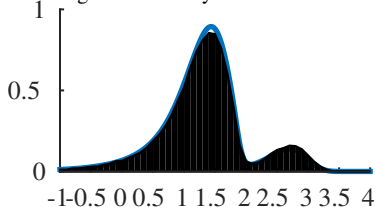
A cell tree



Histogram and analytical solution



Histogram and analytical solution



Thanks for your attention !

- ▶ *Molecular distributions in gene regulatory dynamics*, M.C Mackey, M. Tyran-Kamińska and R.Y., *Journal of Theoretical Biology* (2011) 274 :84-96
- ▶ *Dynamic Behavior of Stochastic Gene Expression Models in the Presence of Bursting*, M.C Mackey, M. Tyran-Kamińska and R.Y., *SIAM Journal on Applied Mathematics* (2013) 73 :1830-1852
- ▶ *Adiabatic reduction of a model of stochastic gene expression with jump Markov process*, R.Y., C. Zhuge, J. Lei, M.C Mackey, *Journal of Mathematical Biology* (2014) 68 :1051-1070