Direct and Inverse Problems in Biological Dynamics

Marie Doumic

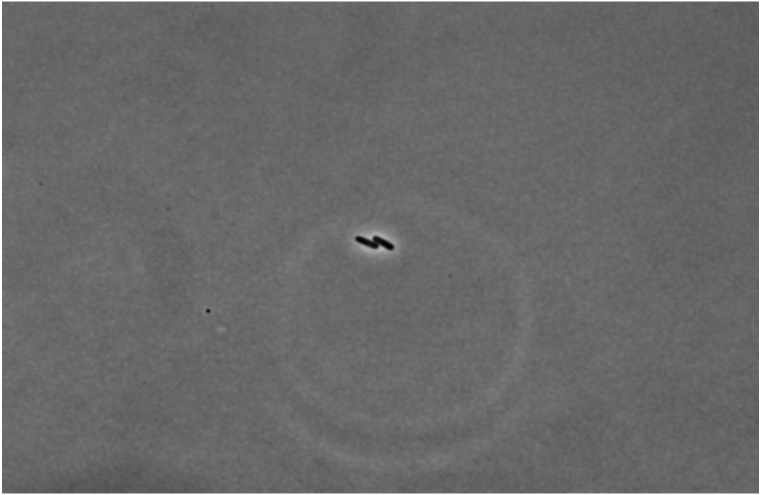
INRIA and UPMC, Paris
Wolfgang Pauli Institute, Vienna

Mauritius, December 6-9, 2016

European Research Council

INVENTORS FOR THE DIGITAL WORLD

Bacterial growth (E. coli here)



From E. J. Stewart, R. Madden, G. Paul, F. Taddei, Plos Biol, 2005

What triggers bacterial division?

OR

Different ways of investigation:

- details the intracellular mechanisms
  many studies (e.g. E Harry, L Monahan, L Thompson, Int. Rev. Cytol., 2006.)
- Observe and understand the population dynamics

What triggers bacterial division?

OR

Different ways of investigation:

- details the intracellular mechanisms
  many studies (e.g. E Harry, L Monahan, L Thompson, Int. Rev. Cytol., 2006.)
- Observe and understand the population dynamics

Question: Can we deduce laws from our observations?

Steps towards such ”laws”

Steps towards such "laws"

1. Make the most of direct observations
   Methods: statistical analysis, density estimation...

Steps towards such ”laws”

1. Make the most of direct observations
   Methods: statistical analysis, density estimation...

2. Make assumptions or simplifications

Steps towards such "laws"

1. Make the most of direct observations
   Methods: statistical analysis, density estimation...
2. Make assumptions or simplifications
3. Build model(s)
   Methods: probabilistic processes / ODE or PDE ...

Steps towards such "laws"

1. Make the most of direct observations
   Methods: statistical analysis, density estimation...
2. Make assumptions or simplifications
3. Build model(s)
   Methods: probabilistic processes / ODE or PDE ...
4. Calibrate the model(s): estimation of unobserved parameters
   Methods: inverse problems, statistics

Steps towards such "laws"

1. Make the most of direct observations
   Methods: statistical analysis, density estimation...
2. Make assumptions or simplifications
3. Build model(s)
   Methods: probabilistic processes / ODE or PDE ...
4. Calibrate the model(s): estimation of unobserved parameters
   Methods: inverse problems, statistics
5. Back to the data to (in)validate the model(s)

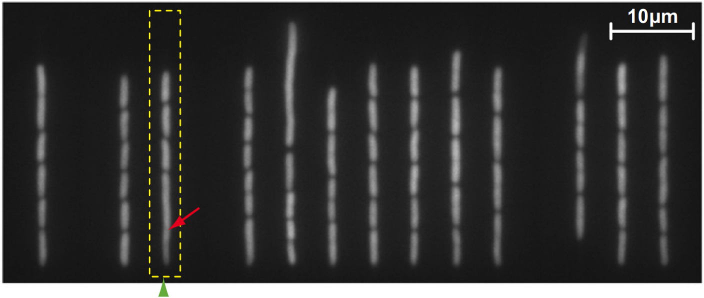
First step: take the most of our data

(before writing down a math model)

1. Direct observations

2 types of data:

- initial video: all descendants till a certain time, several microcolonies (Stewart et al, Plos Biol, 2005)
- 1 daughter cell kept at each generation, till a certain time, several lineages (Wang, Robert et al, Current Biology, 2010)



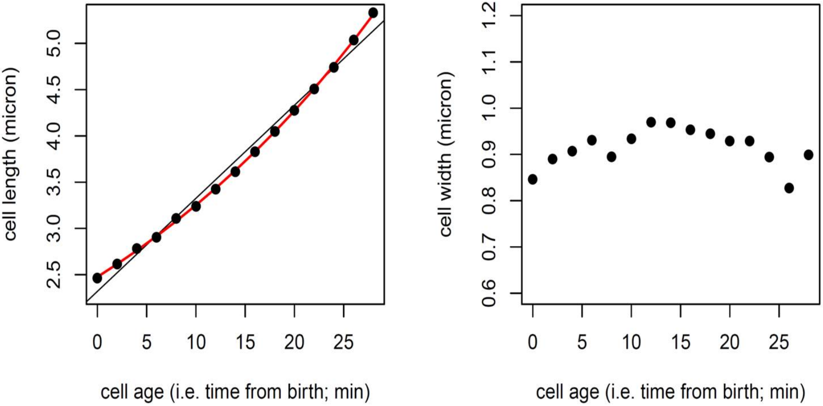
The way we observe the data influence the math modeling.

Direct observations: individual growth

commonly accepted after much debate: exponential growth:

$$\frac{dx}{dt} = \kappa x.$$

(Stewart et al, Plos Biol, 2005)



1. Direct observation: individual growth

variability of the exponential rate κ among cells

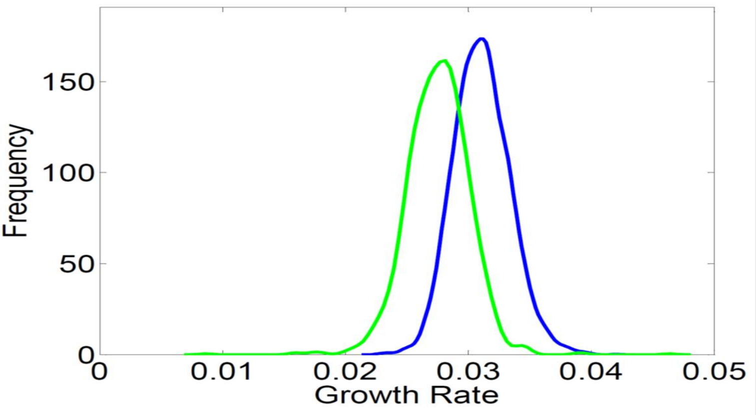


Figure : growth rate distrib. (min^{-1})

1. Direct observations: population growth

Growth of the population: exponential with Malthus parameter λ (almost) equal to the (average) individual growth rate κ.
Doubling time ($= Log(2)/\kappa$) of approx. 20 min.

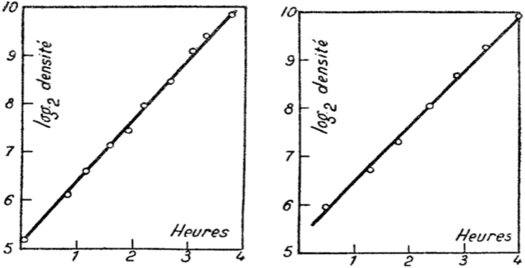


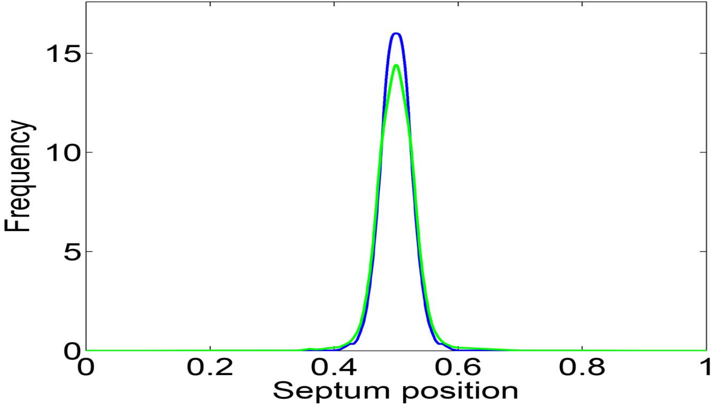
Fig. 10. — Phase exponentielle de la croissance d'une culture de *B. coli* en milieu synthétique, avec 300 mgr. par l. de glucose. Coordonnées semi-logarithmiques.

Fig. 11. — Phase exponentielle de la croissance d'une culture de *B. subtilis* en milieu synthétique, avec 500 mgr. par l. de saccharose. Coordonnées semi-logarithmiques.

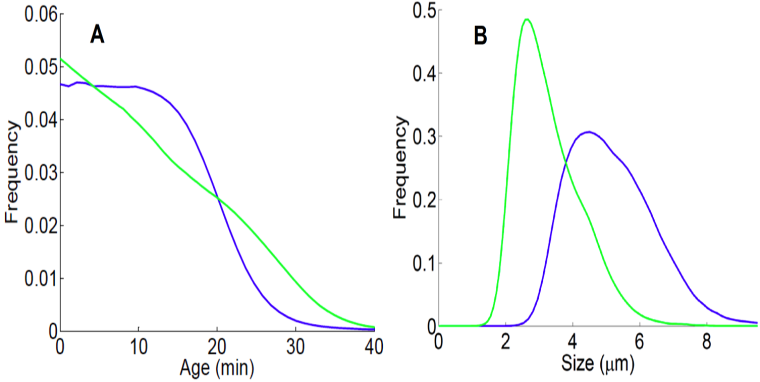
Figure : Monod's 1942 thesis on *E. Coli* culture cells.

1. Direct observation: division

Distribution of the ratio (size of daughter/size of mother)



1. Direct observation: distributions



Blue: 1 branch/genealogical data Green: whole tree data till a certain time

Second step: making assumptions

(before writing down a math model)

2. Assumptions: some simplification

based on direct observations:

- daughter cell size= half of mother cell size
- growth rate = constant among cells (neglect variability)

$$\frac{dx}{dt} = \kappa x$$

- infinite nutrient and space
- first cell selected at random

2. Assumptions: modeling

- no memory
- a particle of size x may divide with a division rate B depending on age
  OR
- a particle of size x may divide with a division rate B depending on size
  OR
- a particle of size x may divide with a division rate B depending on size AND age AND/OR something else...

Third step: models

(that we will analyse and calibrate)

3. Models

2 main ways of translating mathematically the previous assumptions:

1. probability: model each cell
2. PDE: model the population of cells, considered either as large or in expectancy

3. Models: Branching processes modeling see Meyn & Tweedie, 1993 and M.H.A. Davis, 1993

Piecewise Deterministic Markov Processes (PDMP):

- start: a singe cell of size x_0.
- cell's growth: deterministic.
- at each time, it has an instantaneous probabillity rate B to divide (jump); B depends on size x or age a of the cell.
- At division, two offspring of age 0 and initial size $x_1/2$, where x_1 is the size of the mother at division.
- The two offspring start independent growth (Markov property) according to the (deterministic) rate κ and divide according to the (probabilistic) rate B.

3. The probabilistic model
see Meyn & Tweedie, 1993 and M.H.A. Davis, 1993

Genealogical tree: infinite random marked tree

$$\mathcal{U} = \bigcup_{n=0}^{\infty} \{0,1\}^n \quad \text{with} \quad \{0,1\}^0 := \emptyset.$$

To each node $u \in \mathcal{U}$, we associate a cell with size at birth ξ_u and lifetime ζ_u.
If u^- denotes the parent of u then

$$\boxed{\xi_u = \frac{\xi_{u^-}}{2} \exp\left(\kappa \zeta_{u^-}\right).}$$

3. Models: From probability back to PDE...

Equivalent view: Piecewise Deterministic Markov Process (PDMP):
To each cell labeled by $u \in \mathcal{U}$, we associate a birth time b_u.
$X(t) = \big(X_1(t), X_2(t), \dots\big)$ process of the sizes of the population at time t, or $A(t) = \big(A_1(t), A_2(t), \dots\big)$ of ages at time t.
$X(t)$ has values in the space of finite point measures on $\mathbb{R}_+ \setminus \{0\}$ via

$$\mathcal{M}_{X(t)} = \sum_{i=1}^{\sharp X(t)} \delta_{X_i(t)}, \qquad \mathcal{M}_{A(t)} = \sum_{i=1}^{\sharp A(t)} \delta_{A_i(t)}$$

Branch tree case: always 1 and only 1 Dirac mass $\delta_{X_i(t)}$, with $i =$ number of divisions till time t.

3. Age model: renewal equation

Set, for (regular compactly supported) f

$$\langle n(t,\cdot), f\rangle := \mathbb{E}\Big[\sum_{i=1}^{\infty} f\big(A_i(t)\big)\Big].$$

In a weak sense:

$$\partial_t n(t,a) + \partial_a n(t,a) = -B(a)n(t,a),$$

$$n(t,0) = 2\int_0^{\infty} B(a)n(t,a)da \quad OR \quad n(t,0) = \int_0^{\infty} B(a)n(t,a)da$$

So the mean empirical distribution of $A(t)$ satisfies the deterministic renewal equation.

3. Size model: growth-fragmentation equation

Set, for (regular compactly supported) f

$$\langle n(t,\cdot), f\rangle := \mathbb{E}\left[\sum_{i=1}^{\infty} f\left(X_i(t)\right)\right].$$

Proof: tagged fragment approach (Bertoin, Haas, ...), many-to-one formula (Bansaye et al, 2009, Cloez, 2011, ...)

We have (in a weak sense) IF we keep the 2 daughters at each generation:

$$\boxed{\partial_t n(t,x) + \partial_x\left(\kappa x\, n(t,x)\right) + B(x)n(t,x) = 4B(2x)n(t,2x).}$$

So the mean empirical distribution of $X(t)$ satisfies the deterministic growth-fragmentation / size-structured / cell division equation (with binary fission and equal mitosis).

3. Size model: growth-fragmentation equation

Set, for (regular compactly supported) f

$$\langle n(t,\cdot), f\rangle := \mathbb{E}\Big[\sum_{i=1}^{\infty} f\big(X_i(t)\big)\Big].$$

Proof: tagged fragment approach (Bertoin, Haas, ...), many-to-one formula (Bansaye et al, 2009, Cloez, 2011, ...)

We have (in a weak sense) IF we keep 1 daughter at each generation:

$$\boxed{\partial_t n(t,x) + \partial_x\big(\kappa x\, n(t,x)\big) + B(x)n(t,x) = 2B(2x)n(t,2x).}$$

So the mean empirical distribution of $X(t)$ satisfies a deterministic conservative growth-fragmentation equation (also encountered e.g. for TCP/IP protocol)

3. Age and Size model: PDE

$n(t,a,x)$ density of cells of size x and age a.
PDE obtained from the PDMP (as previously) or by a mass balance:

$$\frac{\partial}{\partial t}n + \frac{\partial}{\partial a}n + \frac{\partial}{\partial x}\big(\kappa x n\big) = -B(a,x)n(t,a,x),$$

$$n(t,a=0,x) = 4\int\limits_0^\infty B(a,2x)n(t,a,2x)da$$

with $n(0,a,x) = n^{(0)}(a,x)$, $x \geq 0$.
IF $B = B(x)$: back to growth-fragmentation equation
IF $B = B(a)$: back to renewal equation
IF we keep only 1 daughter at each generation: the boundary condition becomes:

$$n(t,a=0,x) = 2\int\limits_0^\infty B(a,2x)n(t,a,2x)da$$

Fourth step: model calibration

(which first needs analysis)

4. Model calibration

Only unobserved parameter: the division rate B.
Estimation procedure:

- mathematical analysis: asymptotic regime (PDMP or PDE)
- estimation methods
- comparison of calibrated model results and data

Long-time asymptotics: PDE - Age model

historically the first structured-population model to be studied (Kermack and Mc Kendrick, 1927 ; Metz and Diekmann, 1981)
$n(t,a) \to e^{\lambda t} N(a)$, with λ and N uniquely determined by

$$\frac{\partial}{\partial a} N + \lambda N = -B(a)N, \qquad N(0) = 2\int_0^\infty B(a)N(a)da.$$

Explicit solution: $N(a) = N(0)e^{-\lambda a - \int_0^a B(s)ds}$,
λ uniquely determined by the boundary condition:
either $\lambda = 0$ (1 branch case) or

$$2\int_0^\infty B(a)e^{-\lambda a - \int_0^s B(s)ds} da = 1$$

Long-time asymptotics: PDE - Size model

Looking for solutions $n(t,x) = e^{\lambda t} N(x)$

Under proper assumptions $\exists!$ $(\lambda > 0, N \geq 0)$ solution of

$$\begin{cases} \frac{\partial}{\partial x}(\kappa x N(x)) + \lambda N(x) = -B(x)N(x) + 4B(2x)N(2x)dx, \\ N(x) \geq 0, \qquad \int_0^\infty N(x)dx = 1. \end{cases} \tag{1}$$

Here it stands that $\kappa = \lambda$ and by the "General Relative Entropy" method

$$\boxed{\int_{\mathbb{R}_+} \left| n(t,x)e^{-\lambda t} - \langle n^{(0)}, x\rangle N(x) \right| x dx \to 0 \quad \text{as} \quad t \to \infty}$$

Reference book: B. Perthame, *Transport Equations in Biology*, 2007

Fourth step: estimation procedure

Fourth step: estimation procedure

4. Estimation methods

3 methods:

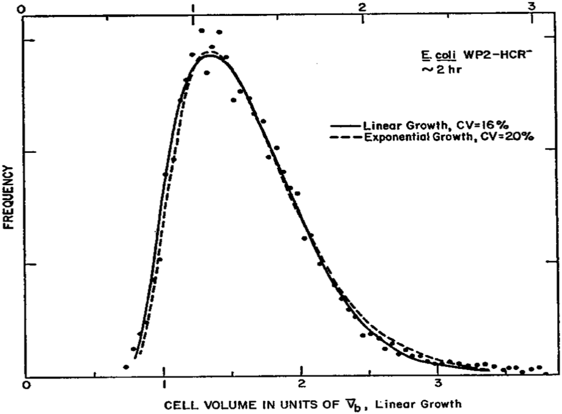
- use the "all cells" distributions: "indirect/inverse" approach, based on $N(x)$ or $N(a)$
- use the "at division" distributions: "direct" approach: PDMP or $B(x)N(x)/\int BNdx$
- use both ! "direct" approach: measure of both $B(x)N(x)/\int BNdx$, and $N(x)$

With E. coli: choose any of the 3 schemes and select the most accurate

Preliminaries: How to estimate these densities?

4. First method, preliminaries: estimation of $N(x)$

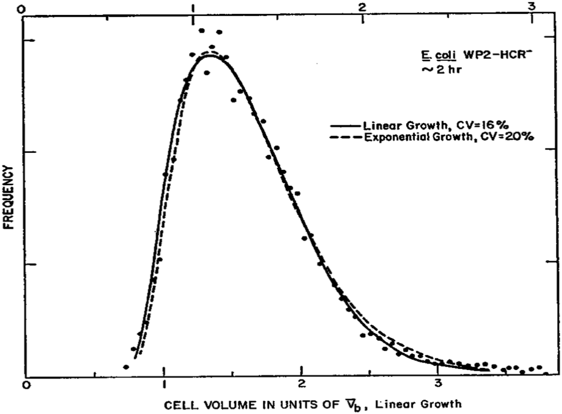
1st historical observations, the simplest and often the only possible ones, and confirm the asymptotic behavior:



Observation (from Kubitschek, 1969): DOUBLING TIME and STEADY SIZE DISTRIBUTION.

4.1. First method: an indirect approach

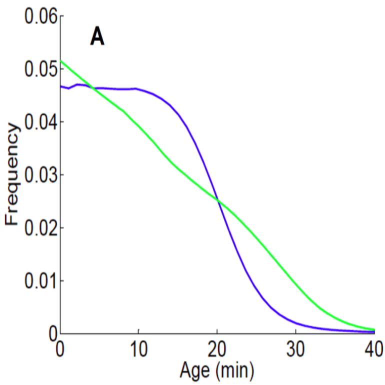
1st historical observations, the simplest and often the only possible ones, and confirm the asymptotic behavior:



Observation (from Kubitschek, 1969): DOUBLING TIME and STEADY SIZE DISTRIBUTION.

4.1. First method: an indirect approach

Any cell at any time put together in this asymptotic distribution



cf. video at the beginning: around 30.000 to 60.000 observations
(Blue: 1 branch, Green: whole tree)

4.1. Inverse Problem for the age model

(see also M. Gyllenberg, A. Osipov and L. Pivrinta, 2002 & 2003)

From a (noisy) measure of $N(a)$ and λ, we look for $B(a)$.
Since we have the explicit relation

$$N(a) = N(0)e^{-\lambda a - \int_0^a B(s)ds},$$

we get

$$B(a) = -\lambda - \frac{\partial_a N(a)}{N(a)}.$$

From a noisy version of N: regularization is needed.

4.1. Inverse Problem for the size model

Inverse Problem: estimating the division rate $B(x)$

From: measurements of (κ, N) with

$$\frac{\partial}{\partial x}(\kappa x N(x)) + \lambda N(x) = -B(x)N(x) + 4B(2x)N(2x)dx.$$

Choice of a **Hilbert space:** $L^2(\mathbb{R}_+, x^p dx)$
(Engl, Hanke, Neubauer, *Regularization of Inverse Problems*, 1995)

Similar to the age problem: the equation implies a derivative for N

4.1. Inverse Problem for the Size Model

Estimate B through

$$L(N) = G(BN), \qquad \text{with}$$

$$G(f)(x) = 4f(2x) - f(x), \tag{2}$$
$$L(N)(x) = \kappa\partial_x\big(xN(x)\big) + \kappa N(x), \tag{3}$$

2 main steps:

- Solve $G(f) = L$ for f, L in suitable weighted L^2 spaces: PDE part. the problem $N \to f = BN$ is now linear.
- Find an estimate for $L(N)$ in this L^2 space: PDE or statistical part

4.1. Inverse Problem for the Size Model

Step 1: solve a dilation equation

Defining

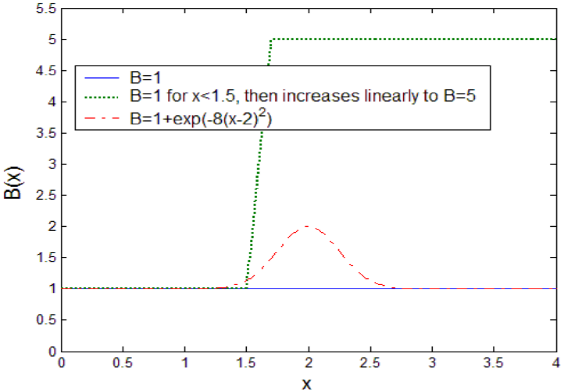
$$G : f \to G(f) = 4f(2x) - f(x)$$

We want to inverse G in a weighted L^2 space.: knowing $L \in L^2$, find $f \in L^2$ solution of

$$L(x) = 4f(2x) - f(x)$$

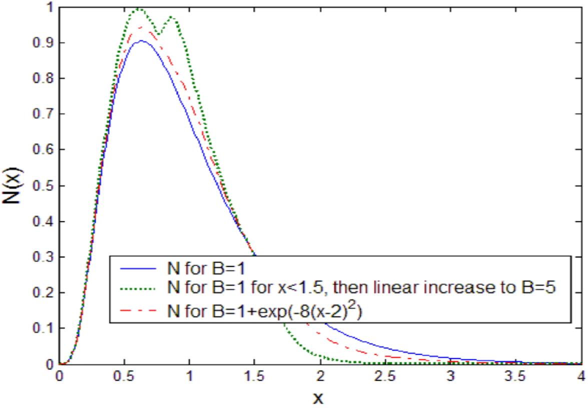
Key point: possibly several solutions

Numerical Results - Size Structured



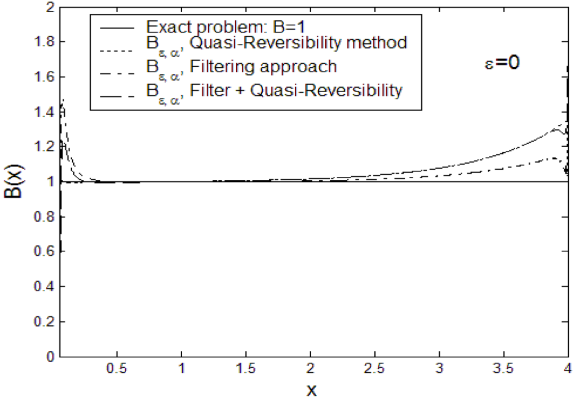
Three tested division rates *B*

Numerical Results - Size Structured



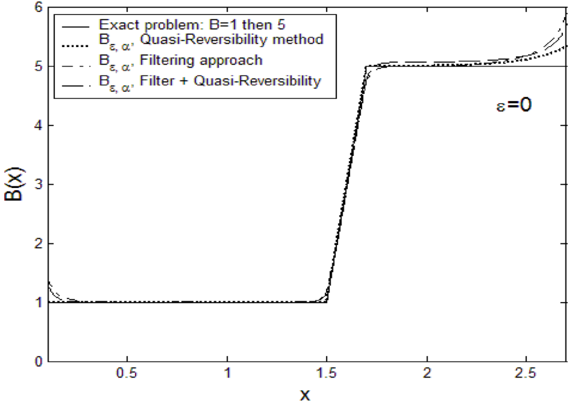
Three related asymptotic distributions N

Numerical Results - - Size Structured



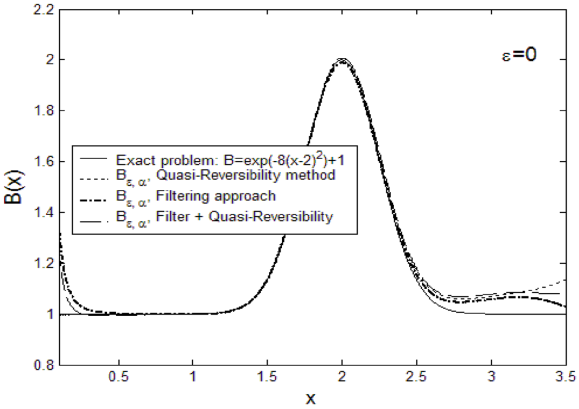
Results with no noise - constant B

Numerical Results - - Size Structured



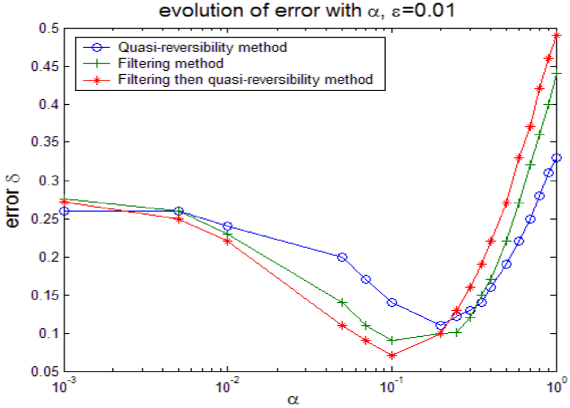
Results with no noise - step B

Numerical Results - - Size Structured



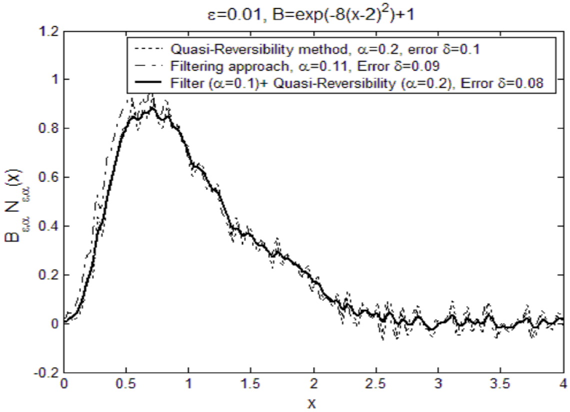
Results with no noise - varying *B*

Numerical Results - - Size Structured



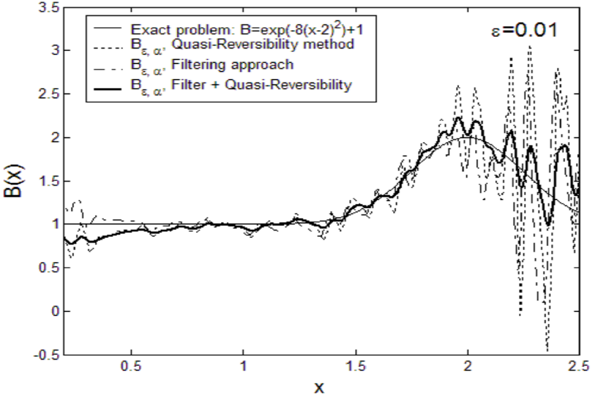
Results with noise $\varepsilon = 0.01$ - Error with respect to the regularization parameter α

Numerical Results - - Size Structured



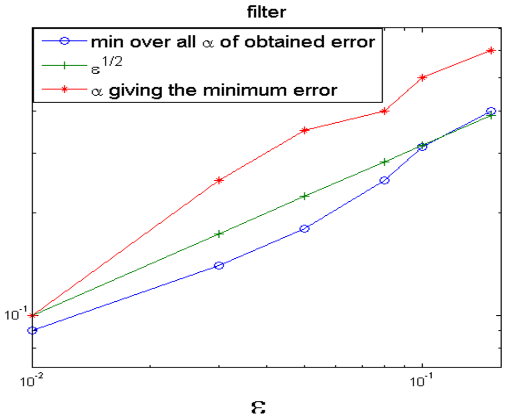
Results with noise $\varepsilon = 0.01$ - BN

Numerical Results - - Size Structured



Results with noise $\varepsilon = 0.01$ - B

Numerical Results - - Size Structured



Optimal α with respect to ε, compared to $\sqrt{\varepsilon}$ and the optimal error

4.2. Second method: direct and full observation

Statistical reconstruction
(MD, M. Hoffmann, N. Krell, L. Robert, Bernoulli, 2014)
Observation scheme

$$\{(\xi_u, \zeta_u), \quad u \in \mathcal{U}_n\},$$

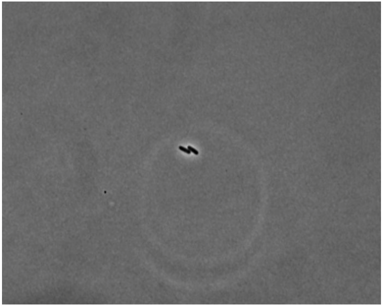
with $\mathcal{U}_n \subset \mathcal{U}$ a set of n nodes having the property

$$\text{If } u \in \mathcal{U}_n \text{ then } u^- \in \mathcal{U}_n.$$

Asymptotics taken as $n \to \infty$.

We use the link between $f(t)$ the density of the lifetime and the division rate B.

Step 5: Finally back to the data...



OR

Will we be able to select or reject our models ?

5. Back to the data

(M.D., M. Hoffmann, N. Krell, L. Robert, BMC Biology, 2014)

To test a model:

- calibrate it (previously seen methods and data)
- simulate the **age-size** PDE model:

$$\frac{\partial}{\partial t} n + \frac{\partial}{\partial a} n + \frac{\partial}{\partial x}\big(\kappa x n\big) = -B(a,x) n(t,a,x),$$

$$n(t, a=0, x) = 4 \int_0^\infty B(a,2x) n(t,a,2x) da$$

  till its asymptotic steady behaviour $n(t,a,x) = e^{\lambda t} N(a,x)$

- compare quantitatively data and simulations
- conclude !

5. Back to the data

(M.D., M. Hoffmann, N. Krell, L. Robert, BMC Biology, 2014)

To test a model:

- calibrate it (previously seen methods and data)
- simulate the **age-size** PDE model:

$$\frac{\partial}{\partial t} n + \frac{\partial}{\partial a} n + \frac{\partial}{\partial x}\big(\kappa x n\big) = -B(a,x)n(t,a,x),$$

$$n(t,a=0,x) = 4\int\limits_0^\infty B(a,2x)n(t,a,2x)da$$

  till its asymptotic steady behaviour $n(t,a,x) = e^{\lambda t} N(a,x)$
- compare quantitatively data and simulations
- conclude ! If possible...

5. Back to the data
experimental age/size data - whole tree till a certain time

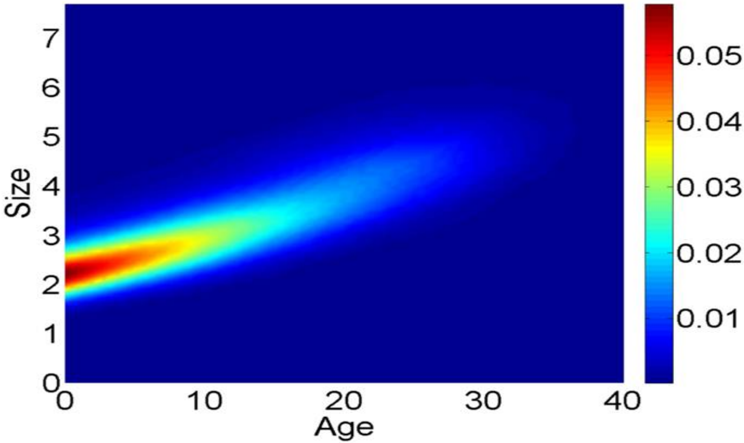


Figure : Age Size Distribution for all cells - whole tree data

5. Back to the data
experimental age/size data - 1 branch data

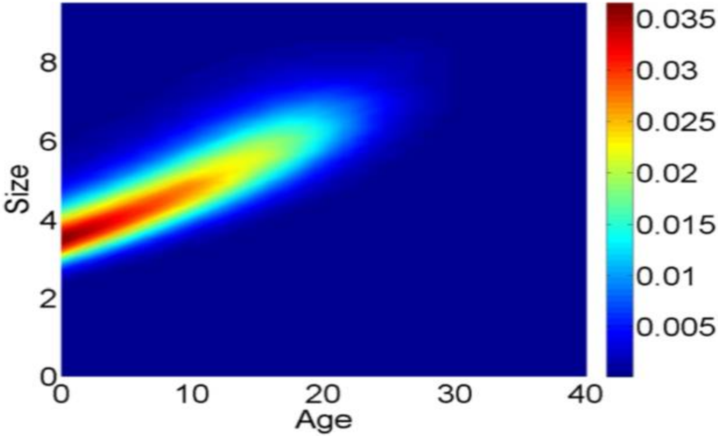


Figure : Age Size Distribution for all cells - tree branches data

Testing the Age Model

5. Back to the data: testing the Age Model

(M.D., M. Hoffmann, N. Krell, L. Robert, BMC Biology, 2014)

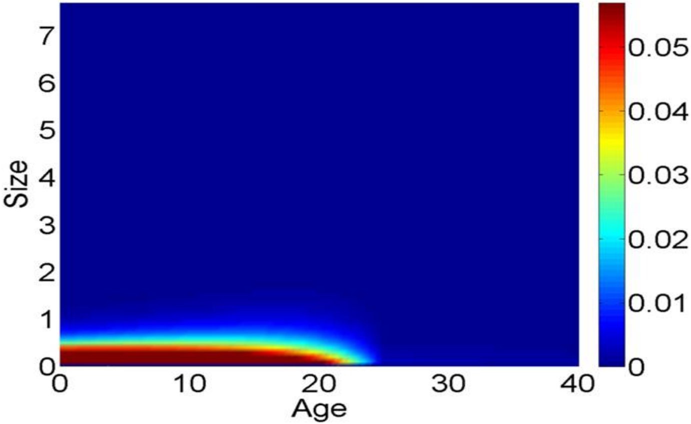


Figure : Age Size simulation for the Age Model - whole tree data

5. Back to the data: testing the Age Model with a corrected growth rate

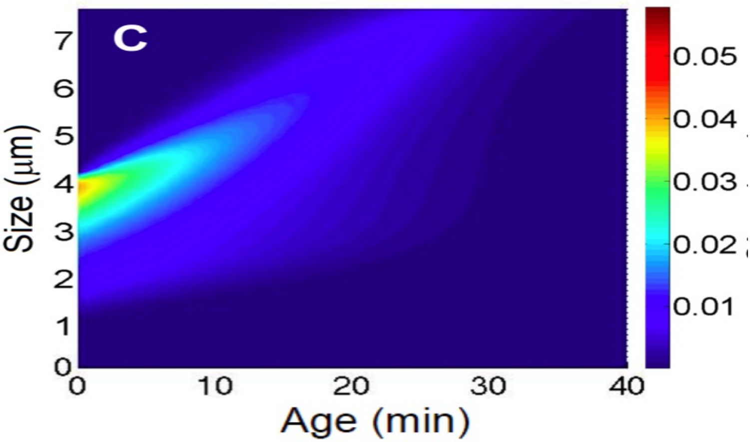


Figure : Age Size simulation for the Age Model - whole tree data

5. Back to the data: testing the Age Model

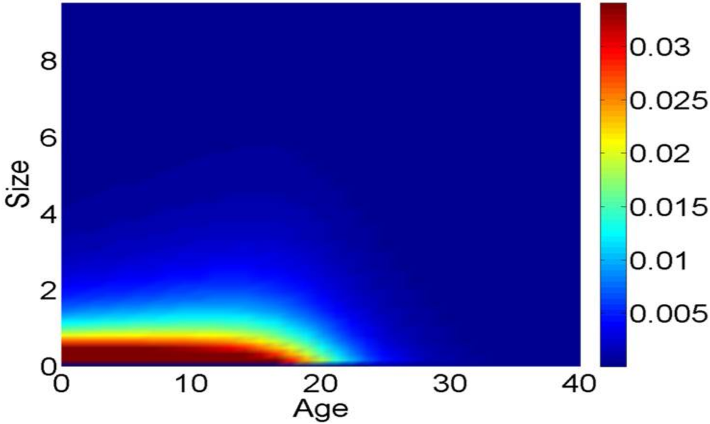


Figure : Age Size simulation for the Age Model - branch tree data

5. Back to the data: testing the Age Model with a corrected growth rate

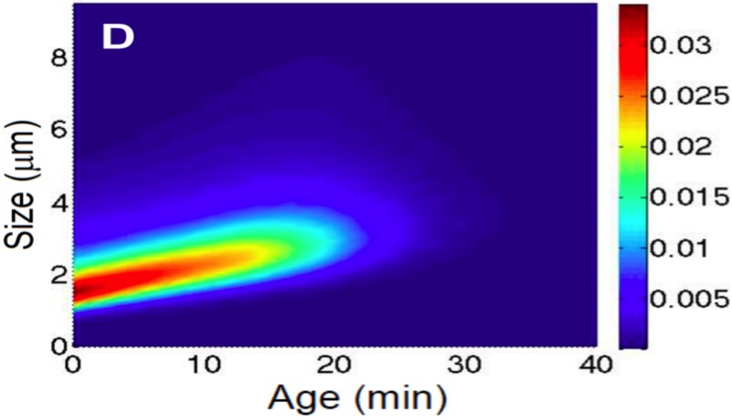


Figure : Age Size simulation for the Age Model - branch tree data

5. Age Model: conclusion

- As it is, this model is rejected
- Theoretical reason: exponential growth + age-dependent division rate lead to accumulation towards 0.
- Refer to theoretical results for the asymptotic regime: we need $\frac{B(x)}{x} \in L^1_0$ (M.D., P. Gabriel, 2010)
- This theory is not sufficient: corrected growth rate dependence on these corrections is too important

Testing the Size Model

5. Back to the data: testing the Size Model

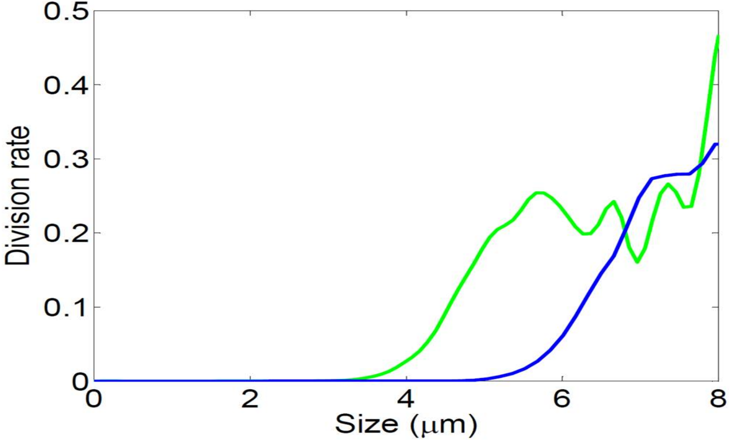
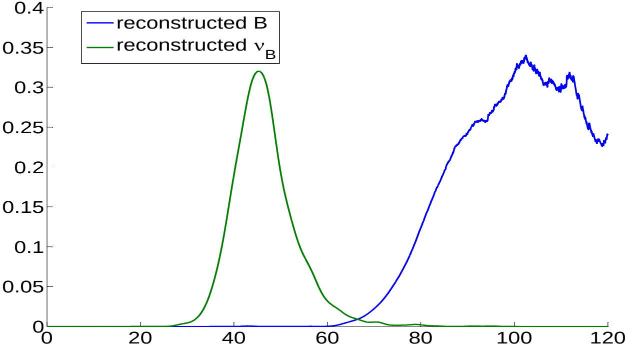


Figure : Reconstruction of the division rate - green: whole tree, blue: branches data

5. Size Model: reconstruction for B



5. Back to the data: testing the Size Model

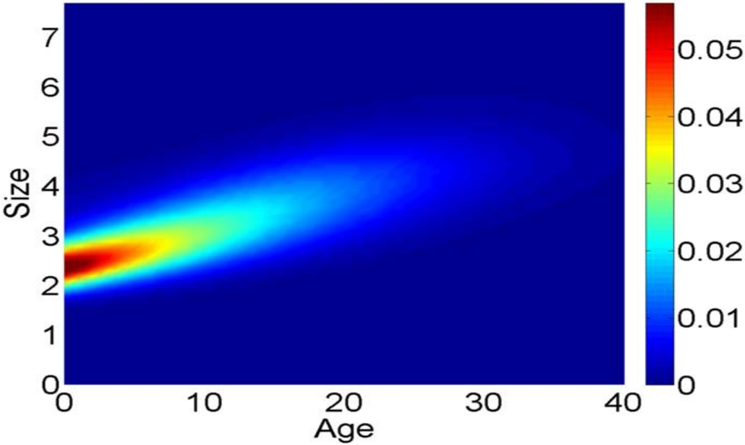


Figure : Age Size simulation for the Size Model - whole tree data

5. Back to the data: testing the Size Model

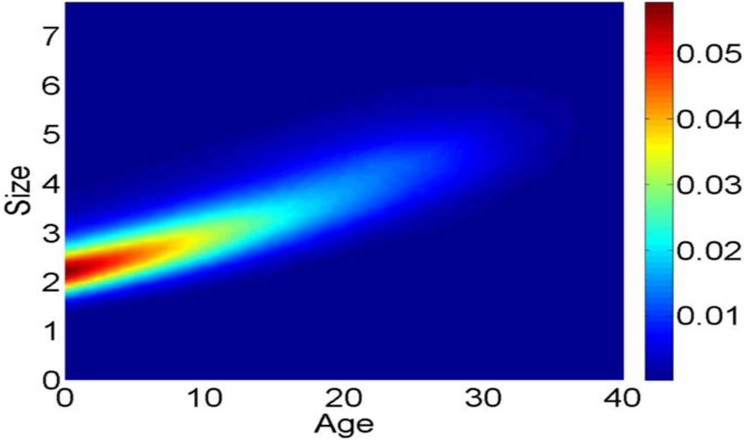


Figure : Age Size experimental data - whole tree data

5. Back to the data: testing the Size Model

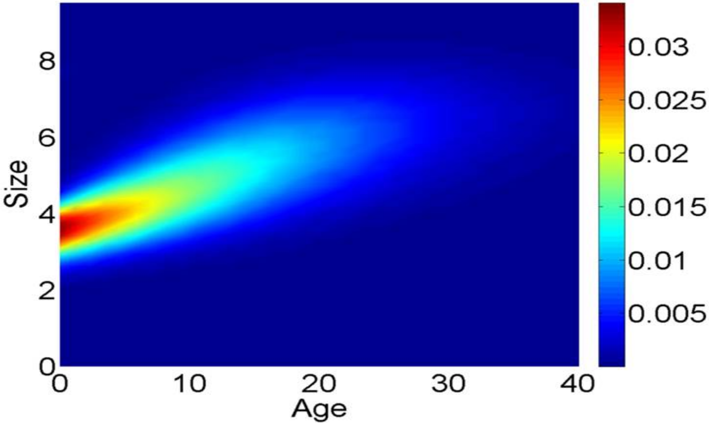


Figure : Age Size simulation for the Size Model - branch tree data

5. Back to the data: testing the Size Model

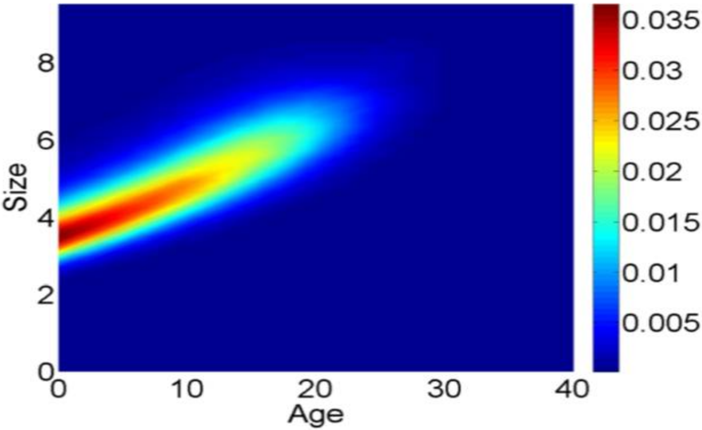


Figure : Age Size experimental data - branch tree data

Extensions of the model

Variability:

$$\frac{\partial}{\partial t} n(t,x,v) + \frac{\partial}{\partial x}\big(v x n(t,x,v)\big) =$$

$$-B(x)n(t,x,v) + 2\int\limits_x^\infty \int\limits_0^\infty B(y)k(y,x)\rho(v',v)n(t,y,v')dy, dv'$$

with $\int_0^\infty \rho(v',v)dv = 1$

Extensions of the model

Variability:

$$\frac{\partial}{\partial t} n(t,x,v) + \frac{\partial}{\partial x}\big(vxn(t,x,v)\big) =$$

$$-B(x)n(t,x,v) + 2\int\limits_x^\infty \int\limits_0^\infty B(y)k(y,x)\rho(v',v)n(t,y,v')dy, dv'$$

with $\int_0^\infty \rho(v',v)dv = 1$

Age + variability:

$$\frac{\partial}{\partial t} n(t,a,x,v) + \frac{\partial}{\partial x}\big(vxn(t,a,x,v)\big) = -B(a,x)n(t,a,x,v),$$
$$n(t,a=0,x,v) = 2\int\limits_x^\infty \int\limits_0^\infty B(a,y)k(y,x)\rho(v',v)n(t,a,y,v')dy\,dv'\,da$$

(related (maturity) models: Lebowitz, Rubinow, 1977 - Rotenberg, 1983 - Mischler, Perthame, Ryzhik, 2002,...)

5. Incorporating variability

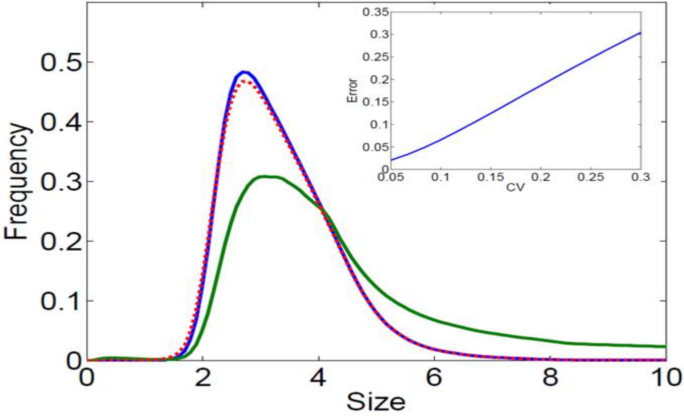


Figure : Effect on the distribution of growth rate variability

5. Incorporating variability

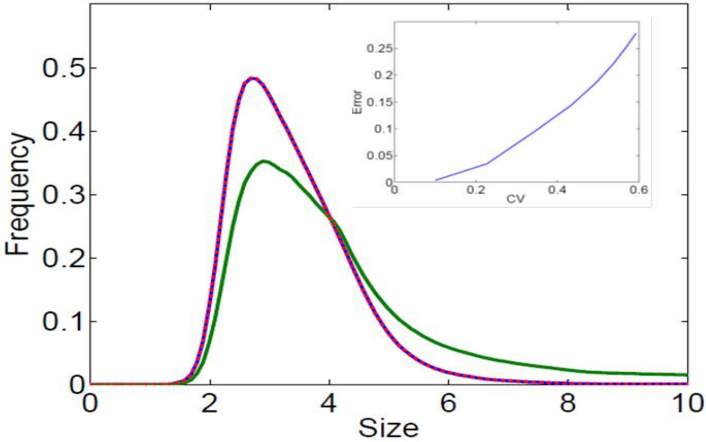


Figure : Effect on the distribution of variability in daughter sizes

5. What about an Age-Size Model ?

To test it, we would need an extra variable:

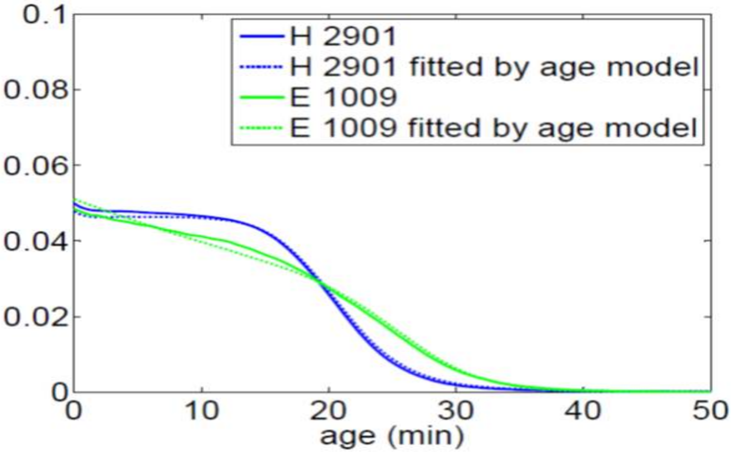


Figure : Age distribution: data and fit by the age model

5. What about an Age-Size Model ?

To test it, we would need an extra variable:

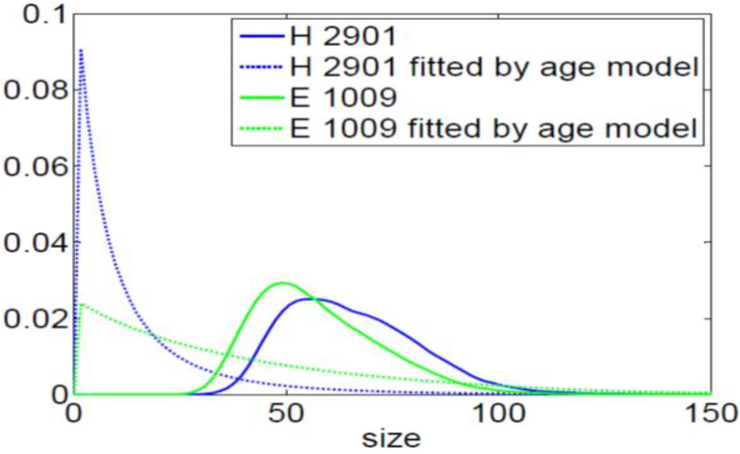


Figure : Size distribution: data and fit by the age model

Conclusion

- Method may be adapted to other cases and models
- Strong coherence and complementarity between PDE and statistical approaches, and still many open mathematical problems
- a basis for new biological questions: coordination between growth and division, influence of variability...

And many huge thanks to...

Pierre Gabriel, Thibault Bourgeron, Miguel Escobedo, Magali Tournus, Benoit Perthame, Jorge Zubelli, Pedro Maia, Marc Hoffmann, Patricia Reynaud-Bouret, Lydia Robert, Vincent Rivoirard, Nathalie Krell, Adélaïde Olivier

to be continued!