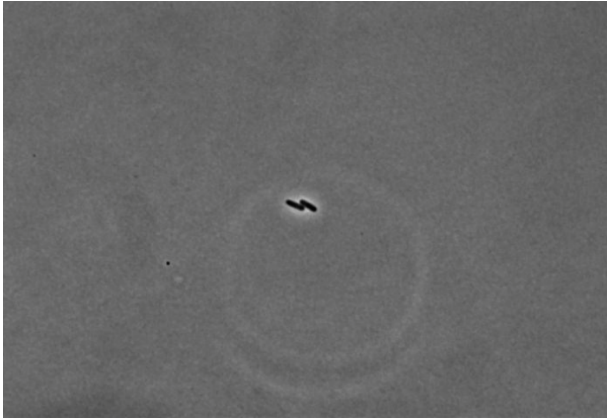# Nonparametric estimation of growth models: combining PDE, PDMP and statistics

Marie Doumic, Marc Hoffmann, Nathalie Krell, Lydia Robert

Rennes, May the 15th, 2013

Growing and dividing populations: what triggers the growth?

What triggers the bacterial growth?

Different ways of investigation:

- details the intracellular mechanisms
  still poorly understood
- Observe and understand the population dynamics

HERE: we focus on this second way: a phenomenological approach?

Population dynamics observations

Different characteristics of the cells may be observed. In the previous movie:

Population dynamics observations

Different characteristics of the cells may be observed. In the
previous movie:

- the age distribution
- the size distribution
- the 2 daughters sizes
- the growth rate distribution
- the age-at-division distribution
- the size-at-division distribution
- the genealogical influence (inheritance of some traits)...

Population dynamics observations

Different characteristics of the cells may be observed. In the previous movie:

- the age distribution
- the size distribution
- the 2 daughters sizes
- the growth rate distribution
- the age-at-division distribution
- the size-at-division distribution
- the genealogical influence (inheritance of some traits)...

Question: Can we deduce laws from our observations?

Steps towards such "laws" (and outline of the talk)

1. Make the most of direct observations

2. Make assumptions or simplifications

3. Build model(s)

4. Calibrate the model(s): estimation of unobserved parameters

5. Back to the data to (in)validate the model(s)

1. Direct observations

In the previous movie: 1 photo is taken and analysed each minute or each 2 minutes.

We have 2 types of data::

- ▶ Data as in the initial movie: all descendants till a certain time, several microcolonies (Stewart et al, Plos Biol., 2005)
- ▶ 1 daugther cell kept at each generation, till a certain time, several lineages (Wang, Robert et al, Current Biology, 2010) microfluidic device

THE WAY WE OBSERVE THE DATA INFLUENCE THE MATHEMATICAL MODELLING.

1. Direct observations

Division of the cells: distribution of the ratio (size of daughter/size of mother)
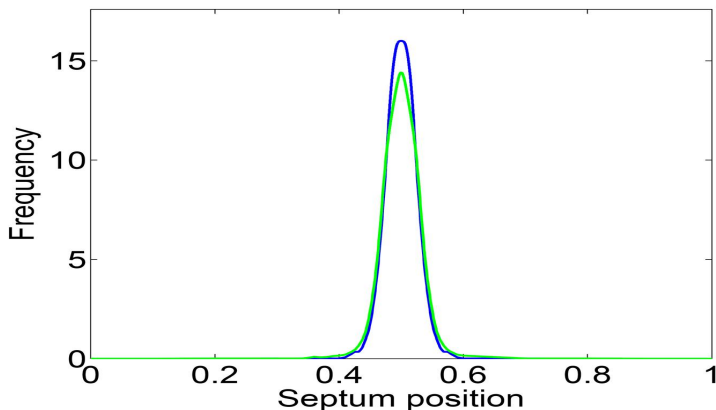
1. Direct observations

Bacterial growth: commonly admitted after much debate: exponential growth:
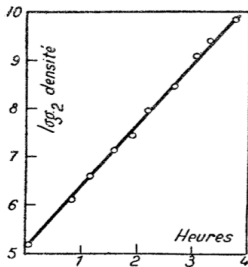
$$\frac{dx}{dt} = \kappa x.$$

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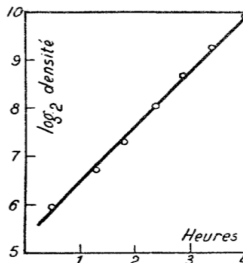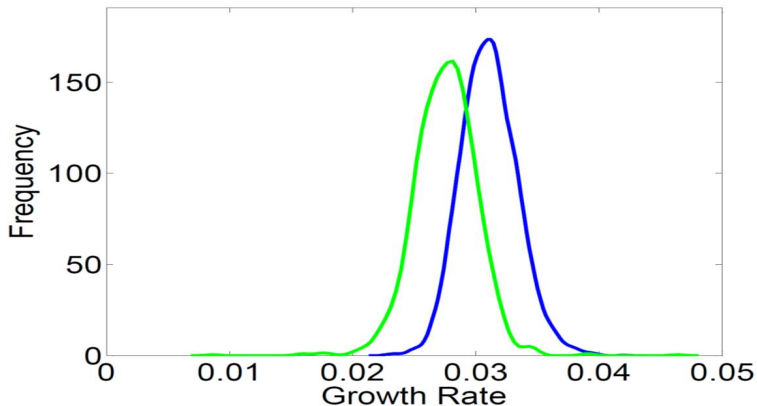
Fig. 11. — Phase exponentielle de la croissance d'une culture de *B. subtilis* en milieu synthétique, avec 500 mgr. par l. de saccharose. Coordonnées semi-logarithmiques.

Figure: Monod's 1942 thesis on *B. Coli* culture cells.
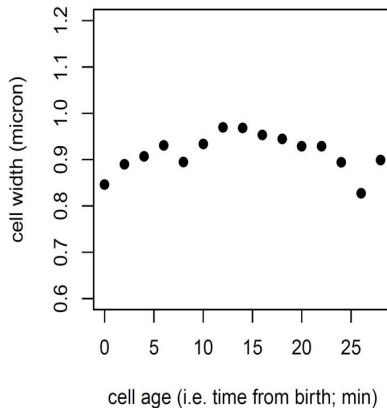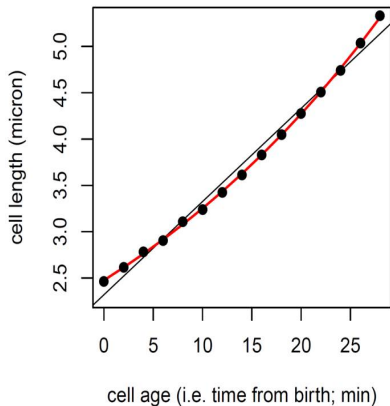
1. Direct observations

Growth of the cells: variability among exponential growth rates
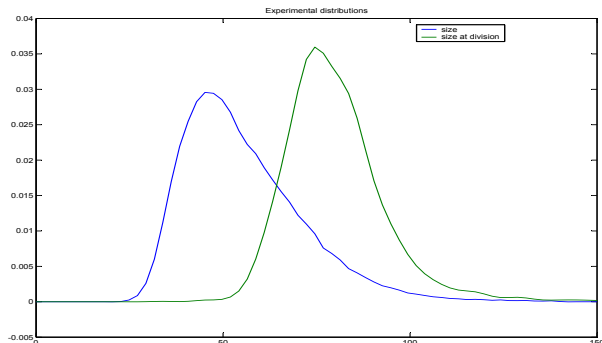
Direct observations

Recent data (Stewart et al, Plos Biol, 2005)

1. Direct observations

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. Well-known fact by biologists.

Blue: size distribution of ALL cells, green: of dividing cells

2. Assumptions

Approximation assumptions 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

Models assumption:

- 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...

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 modelling

Piecewise Deterministic Markov Processes (PDMP):

- We start with a singe cell of size x_0.

- The cell size grows exponentially according to a constant rate κ, , in a deterministic way.

- at each size, a cell has an instantaneous probabillity rate B to divide (jump); B depend on either its size x or its age a (time elapsed since the time of its birth).

- At division, the mother cell gives rize to two offsprings of age 0 and initial size $x_1/2$, where x_1 is the size of the mother at division.

- The two offsprings start independent growth (Markov property) according to the (deterministic) rate κ and divide according to the (probabilistic) rate B.

3. The probabilistic model
close to models of V. Bansaye, B. Cloez, ...

- The population evolution is associated with an infinite random marked tree. Let

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

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

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

3. Models: Back to PDE...

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

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

- If we keep only 1 daughter cell at each jump: always 1 and only 1 Dirac mass $\delta_{X_i(t)}$, with i = number of division till time t.

3. Age model: renewal equation (PhD of A. Olivier)

- Set, for (regular compactly supported) f

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

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

- We have (in a weak sense):

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

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

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

3. Size model: growth-fragmentation equation

▶ Set, for (regular compactly supported) f

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

▶ Methods: 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 daugthers at each generation:

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

▶ Therefore the mean empirical distribution of $X(t)$ satisfies the deterministic transport-fragmentation equation.

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:

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

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

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:

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

4. Model calibration

To test the models, we first need to calibrate them.
Only non measured 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

Classical problem, solved for a long time: see e.g. the textbooks
Metz and Diekmann, 1981 and B. Perthame, 2007.
if we look at a solution under the form $n(t, a) = e^{\lambda t} N(a)$, $a \geq 0$

$$\frac{\partial}{\partial a} N + \lambda N = -B(a)N, \qquad N(0) = 2\int\limits_0^\infty B(a)N(a)da.$$

Explicit solution: $N(a) = N(0)e^{-\lambda a - \int\limits_0^a B(s)ds}$, and λ is uniquely
determined by the boundary condition: either $\lambda = 0$ (1 branch) or

$$2\int\limits_0^\infty B(a)e^{-\lambda a - \int\limits_0^s B(s)ds} da = 1$$

Long-time asymptotics: PDE - Size model

if we look at a solution under the form $n(t,x) = e^{\lambda t} N(x)$, $x \geq 0$

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

Here it stands that $\kappa = \lambda$ and (Michel, 2007 or MD, Gabriel, 2010)

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

Very active field of research: Pakdaman, Perthame, Salort ;
Balagué, Canizo, Gabriel... **and for PDMP: ask people of WS2!**

# 4. Estimation methods (Size Model)

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 $N(x)$

Here we can use both schemes and select the most accurate
The indirect scheme needs less data and is the only possible one in many situations

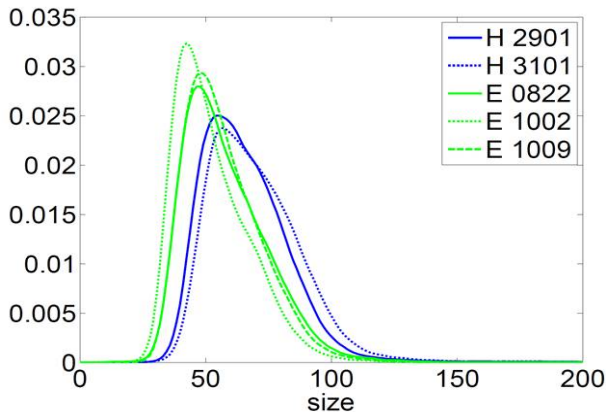# 4.1. First method: an indirect approach

1st historical observations, the simplest and often the only possible ones, and confirms the asymptotic behaviour:
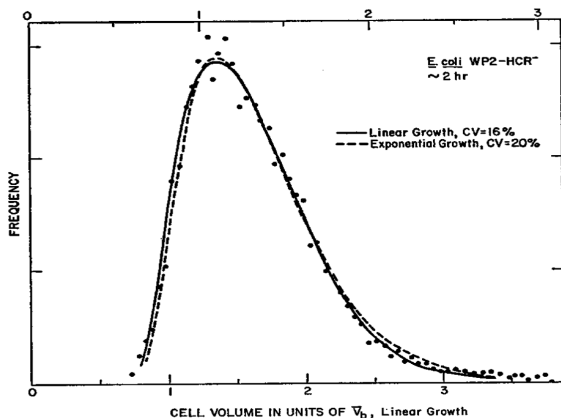
Observation (from Kubitschek, 1969): DOUBLING TIME and STEADY SIZE DISTRIBUTION.
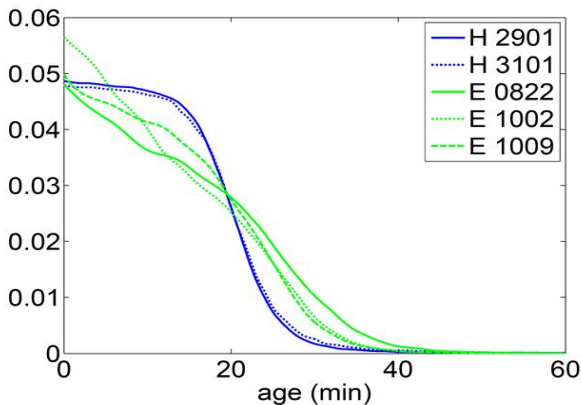
4.1. First method: an indirect approach

Size distribution

cf. movie at the beginning: 33000 observations
(Blue: 1 branch, Green: whole tree)

4.1. First method: all cell distribution

Age distribution

cf. movie at the beginning: 33000 observations
(Blue: 1 branch, Green: whole tree)

4.1. Inverse Problem for the age model

(known from the 1980's. See Gyllenberg et al., 1982)

Quite simple: you have a (noisy) measure of $N(a)$, you look for $B(a)$, and the explicit relation

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

so by direct calculation

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

From a noisy version of N: you need to regularization is needed.

4.1. Inverse Problem for the age model: statistical treatment

We observe a sample of n cells, of ages $a_1, \cdots, a_n$ realizations of $A_1, \ldots, A_n$, *i.i.d.* random variables with density N,
That is, your measure of $N(a)$ is

$$N_\varepsilon(a) = \frac{1}{n} \sum_{i=1}^n \delta_{a=a_i}$$

Regularization: kernel method for instance: mollifier ρ_α

$$N_{\varepsilon,\alpha}(a) = \rho_\alpha * \left(\frac{1}{n} \sum_{i=1}^n \delta_{a=a_i} \right)$$

and define

$$B_{\varepsilon,\alpha}(a) = -\lambda - \frac{\partial_a N_{\varepsilon,\alpha}(a)}{N_{\varepsilon,\alpha}(a)}.$$

4.1. Inverse Problem for the size model

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

From: measures 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 an **Hilbert space:** $L^2(\mathbb{R}_+)$ (Engl, Hanke, Neubauer, *Regularization of Inverse Problems*, 1995)

Theorem (MD, L.M. Tine, J. Math. Biol., 2012)
... If $B \in H^s$, then $N \in H^{s+1}$.

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

4.1. Size model

Estimate B through

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

$$\mathfrak{L}(f)(x) = 4f(2x) - f(x), \tag{2}$$

$$L(N)(x) = \kappa \partial_x\big(xN(x)\big) + \kappa N(x), \tag{3}$$

2 main steps:

- Solve $\mathcal{L}(f) = L$ for f, L in adequate spaces: PDE part the problem $N \to H = BN$ is now linear.
- Find an estimate for $L(N)$ in this adequate space: PDE or statistical part

For the first step: solve

$$L(x) = 4H(2x) - H(x),$$

see [MD, B. Perthame, J.P. Zubelli, Inv. Prob., 2009]
and [MD, L.M. Tine, JMB, 2012] (for general fragmentation)
and soon: work in progress with T. Bourgeron and M. Escobedo...
(for self-similar fragmentation)

4.1 size model: second step

From a measure $N_\varepsilon \in \mathbf{L^2}$, estimate

$$L(N) = \partial_x\big(\kappa x N(x)\big) + \kappa N(x).$$

degree of ill-posedness 1 for a L^2-noise: regularization method to treat the derivative & gain 1 degree of regularity.

A simple method: Filter L with $\rho_\alpha(x) = \frac{1}{\alpha}\rho(\frac{x}{\alpha})$, $\alpha > 0$:

$$L_{\alpha,\varepsilon} = \rho_\alpha\star\bigg(\kappa N_\varepsilon + \kappa\frac{\partial}{\partial x}\big(N_\varepsilon(\frac{y}{2})\big)\bigg), \qquad y > 0,$$

Proposition. *For $B \in H^s$, and appropriate $\rho \in \mathcal{C}_c^\infty(\mathbb{R})$,* $\int_0^\infty \rho(x)\,dx = 1$ *for $\alpha = O(\varepsilon^{\frac{1}{s+1}})$*

$$||B_{\varepsilon,\alpha}N_{\varepsilon,\alpha} - B\,N||_{L^2(dx)} \lesssim \frac{1}{\alpha}||N_\varepsilon - N||_{L^2(dx)} + \alpha^s||N||_{H^s(\mathbb{R}_+)} \lesssim \varepsilon^{\frac{s}{s+1}}$$

(see also Perthame, Zubelli, 2007 - Groh, Krebs, Wagner, 2011)

Indirect Observation Scheme
Step 2: statistical approach

Joint work with M. Hoffmann, P. Reynaud-Bouret & V. Rivoirard
Till now: we have supposed

$$||N - N_\varepsilon||_{L^2} \leq \varepsilon$$

But why an L^2 norm ? What about real data ?

Indirect Observation Scheme
Step 2: statistical approach

Joint work with M. Hoffmann, P. Reynaud-Bouret & V. Rivoirard

Till now: we have supposed

$$||N - N_\varepsilon||_{L^2} \le \varepsilon$$

But why an L^2 norm ? What about real data ?

We observe a sample of n cells, of sizes $x_1, \cdots, x_n$ realizations of $X_1, \ldots, X_n$, $i.i.d.$ random variables with density N, $i.e.$ where

$$\mathbb{P}(X_1 \in dx_1, \ldots, X_n \in dx_n) := \prod_{i=1}^{n} N(x_i) dx_i.$$

Estimation of N
Goldenschluger-Lepski kernel estimation

Let $: \mathbb{R} \to \mathbb{R}_+$ continuous function $\quad / \quad \int K = 1$ and $\int K^2 < \infty$.

$$\hat{N}_h(x) := \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) = K_h \star \left(\frac{1}{n} \sum_{i=1}^n \delta_{x=X_i} \right),$$

$K_h = \frac{1}{h} K(./h)$ replaces the notation ρ_α...

How to adaptively select h ? Goldensshluger and Lepski (2009, 2010)

We thus have a statistical estimator $\widehat{L}_n = K_h \star L(\hat{\lambda}, \frac{1}{n} \sum \delta_{X_i})$,

we plug the first PDE step to inverse $\mathcal{L}$ and its discrete numerical approximation $\mathcal{L}^k$

which gives us an estimate $\hat{H}$ of BN via $\mathcal{L}_{k_n}^{-1}(\widehat{L}_n)$.

We finally set $\hat{B} = \hat{H}/\hat{N}$ and $\tilde{B} = \max(\min(\hat{B}, \sqrt{n}), -\sqrt{n})$.

Rate of convergence for the estimation of B

Theorem (MD, Hoffmann, Reynaud-Bouret, Rivoirard, SIAM Num. Anal., 2012)

If $B \in H^s$ ($s > 1/2$), then (under suitable assumptions)

$$\mathbb{E}\left[\left\|(\tilde{B} - B)1_{[a,b]}\right\|_2\right] = O\left(n^{-\frac{s}{2s+3}}\right).$$

This rate is optimal and is to be compared with the (provably optimal) deterministic rate $\varepsilon^{s/(s+1)}$.

4.2. Second method: direct and full observation

Statistical reconstruction (with M. Hoffmann, N. Krell, L. Robert)
Observation scheme

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

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

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

3 fundamental cases:

- sparse tree case: a line of descendants $(\emptyset, u_1, \cdots, u_n)$
- full tree case: $n = 2^{k_n}$, k_n first generations
- measures stop at a given time (independent of the number of generations)

Asymptotics taken as $n \to \infty$.

4.2. Statistical analysis

We have

$$\mathbb{P}(\zeta_u \in [t, t+dt]\,|\,\zeta_u \geq t, \xi_u = x) = B(xe^{\kappa t})dt$$

for the size model, or

$$\mathbb{P}(\zeta_u \in [t, t+dt]\,|\,\zeta_u \geq t, \xi_u = x) = B(t)dt$$

for the age model from which we obtain the density of the lifetime $\zeta_u = t$ conditional on the size at birth $= x$:

$$f(t, x) = B(xe^{\kappa t}) \exp\left(-\int_0^t B(xe^{\kappa s})ds\right).$$

for the size model or

$$f(t) = B(t) \exp\left(-\int_0^t B(s)ds\right).$$

for the age model

Estimation of B - age model (PhD of A. Olivier - in progress...)

PDE setting: we observe a sample of n cells, of sizes $a_1, \cdots, a_n$ realizations of $A_1, \ldots, A_n$, *i.i.d.* random variables with density $f(a) = B(a)N(a)/\int BNda$, and it is well-known that (branch tree)

$$B(a) = \frac{f(a)}{\int\limits_{a}^{\infty} f(s)ds}.$$

For the whole tree data: "bias" term: f is replaced by $f(a)e^{\lambda a}$.

Estimation of B - size model

- We obtain a simple and explicit representation for the transition kernel $\mathcal{P}_B$ (which links the daughter size (age) law to its mother size (age) law) on $\mathcal{S} = [0, \infty)$, reminiscent of conditional survival function estimation.

- Under appropriate condition on B close to the conditions for the eigenvalue PDE problem, the Markov chain on $\mathcal{S} = [0, \infty)$ is geometrically ergodic. (but not reversible.)

- Under some assumptions, we have existence (and uniqueness) of an invariant measure on $\mathcal{S}$

$$\nu_B(dx) = \nu_B(x)dx$$

  i.e. such that $\nu_B \mathcal{P}_B = \nu_B$.

- we have a contraction property which proves the convergence to the invariant measure uniformly in $B \in \mathcal{F}^\lambda(\mathfrak{c})$, $\rho \in \mathcal{M}(\rho_{\min})$, for an appropriate Lyapunov function V.

Key representation

- We conclude

$$B(y) = \frac{\kappa y}{2} \frac{\nu_B(y/2)}{\mathbb{E}_{\nu_B} \left[\mathbf{1}_{\{\xi_u^- \leq y, \, \xi_u \geq y/2\}} \right]}.$$

- Statistical inference: introduce a kernel function to estimate ν_B...

Final estimator

- Final estimator

$$\widehat{B}_n(y) \quad = \frac{y}{2} \frac{n^{-1}\sum_{u\in\mathcal{U}_n} K_h(\xi_u - y/2)}{n^{-1}\sum_{u\in\mathcal{U}_n} \frac{1}{\tau_{u^-}}\mathbf{1}\{\xi_{u^-}\leq y, \xi_u \geq y/2\} \bigvee \varpi},$$

- The estimator $\widehat{B}_n(y)$ is specified by K, the bandwidth h and the threshold ϖ.

- Error estimates If $B \in H^s$, for appropriate bandwidths, we have

$$\boxed{\mathbb{E}_\mu\left[\|\widehat{B}_n - B\|^2_{L^2(\mathcal{D})}\right]^{1/2} \lesssim (\log n) n^{-s/(2s+1)}}$$

- This rate is almost optimal and is to be compared with the indirect approach.

Quid of a deterministic transcription of our "final estimator" ?

Remember the key representation:

$$B(y) = \frac{\kappa y}{2} \frac{\nu_B(y/2)}{\mathbb{E}_{\nu_B}\left[\mathbf{1}_{\{\xi_u^- \leq y,\, \xi_u \geq y/2\}}\right]} = \frac{\kappa y}{2} \frac{\nu_B(y/2)}{\int\limits_{\frac{y}{2}}^{y} \nu_B(x)dx}.$$

Genealogical scheme: the conservative equation replaces the growth-fragmentation equation:

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

Asymptotic behaviour: steady state:

$$\partial_x\big(\kappa x\, N(x)\big) + B(x)N(x) = 2B(2x)N(2x).$$

Quid of a deterministic transcription of our "final estimator" ?

We measure ξ_u distributed along $\nu_B(x) \equiv 2B(2x)N(2x)$
Write the PDE in terms of BN :

$$\partial_x \big(\kappa x\, N(x) \big) = 2B(2x)N(2x) - B(x)N(x).$$

Integrate:

$$\kappa y N(y) = \int_0^y (2B(2x)N(2x) - B(x)N(x))dx = \int_y^{2y} BN(x)dx$$

So that if we measure BN (or $2BN(2x)$) we estimate B through

$$B(y) = \frac{BN(y)}{N(y)} = \kappa y \frac{BN(y)}{\int_y^{2y} BN(x)dx} = \frac{\nu_B(\frac{y}{2})}{2\kappa y \int_{\frac{y}{2}}^{y} \nu_B(x)dx}$$

Comparison of the convergence rates

- **Deterministic problem**: well-posed! Degree of ill-posedness $a = 0$ - estimate in $O(\varepsilon)$

- **Statistical viewpoint**: density estimate, $H^{-1/2}$ to L^2 so that $a = 1/2$

$$\varepsilon^{s/(s+1/2)} = n^{-s/(2s+1)}$$

- to be compared to the indirect method: error in the order of $\varepsilon^{s/(s+3/2)} = n^{-s/(2s+3)}$.

5. Back to the data

To test a model:

- calibrate it (aboveseen method)
- simulate the age-size PDE model:

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

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

  till it reaches its asymptotic behaviour

- compare data and simulations
- if possible: conclude...

5. Back to the data...
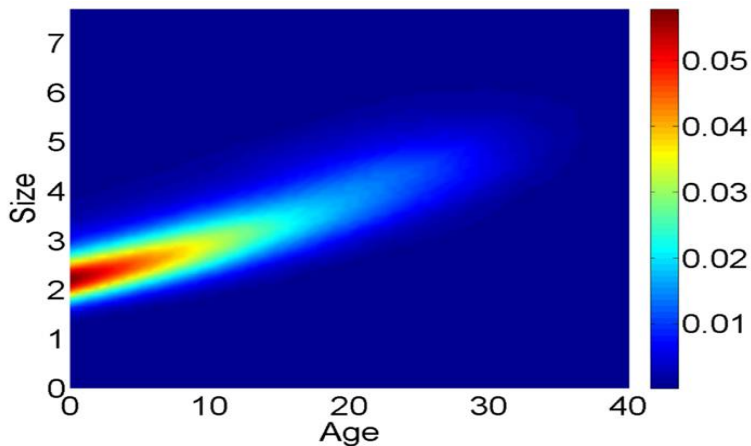
Figure: Age Size Distribution for all cells - whole tree data

5. Back to the data...
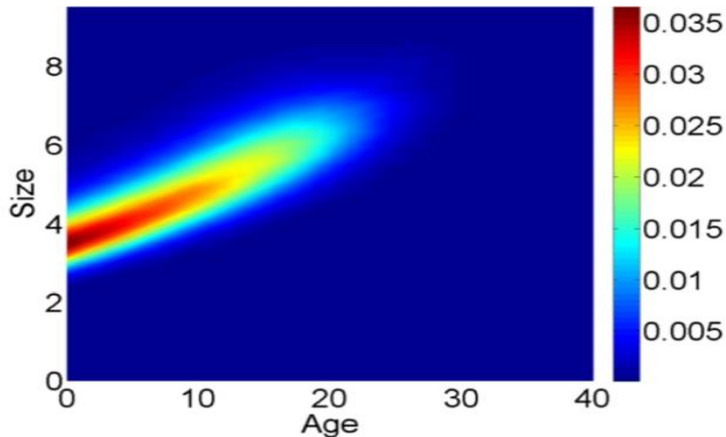
Figure: Age Size Distribution for all cells - tree branches data

5. Back to the data: testing the Age Model
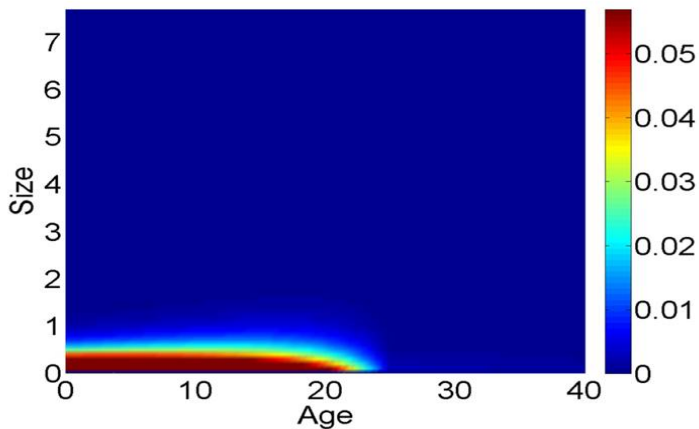
Figure: Age Size simulation for the Age Model - whole tree data

5. Back to the data: testing the Age Model
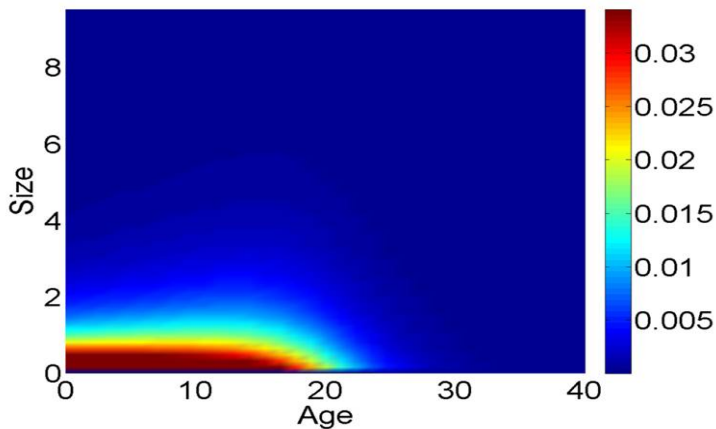
Figure: Age Size simulation for the Age Model - branch tree data
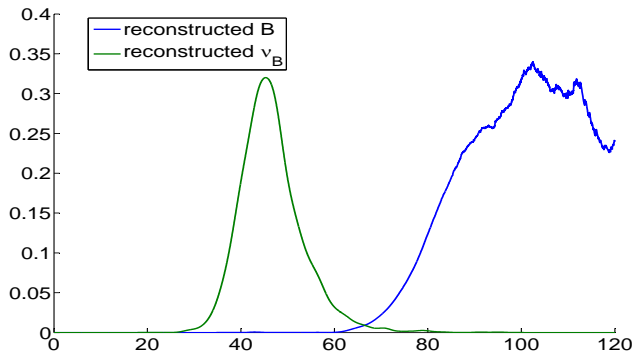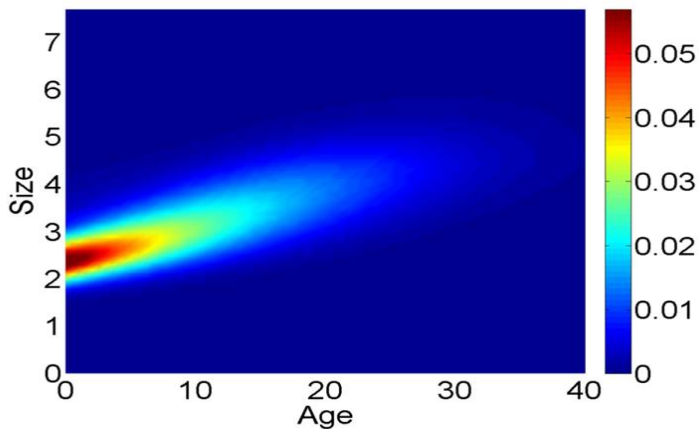
5. Back to the data: testing the Age Model

- This model as it is is rejected
- theoretical reason: exponential growth + age-dependent division rate leads to accumulation towards 0.
- Refer to theoretical results for the asymptotic regime: we need $\frac{B(x)}{x} \in L^1_0$...
- Numerical tests even with constant growth rate not satisfactory.

5. Back to the data: testing the Size Model
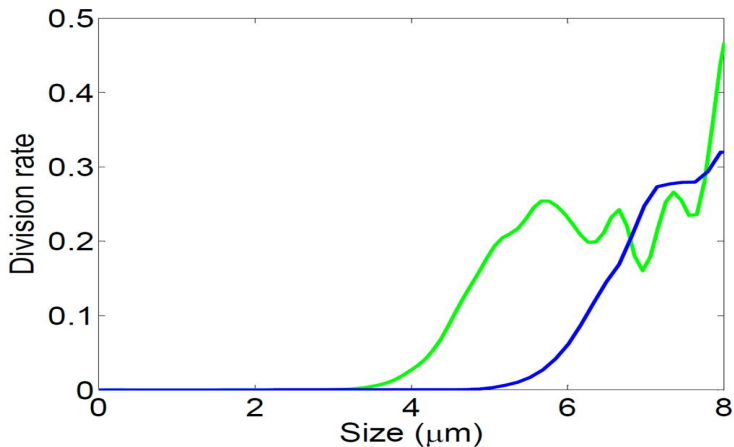
Figure: Reconstruction of the division rate - green: whole tree, blue: branches data

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
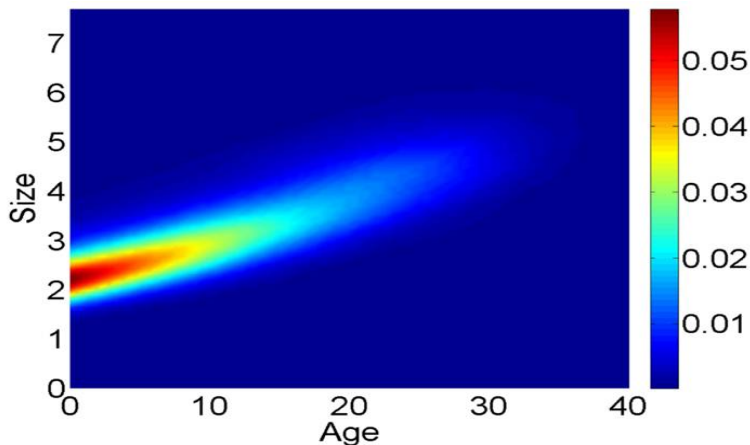
5. Back to the data: testing the Size Model

Figure: Age Size experimental data - whole tree data
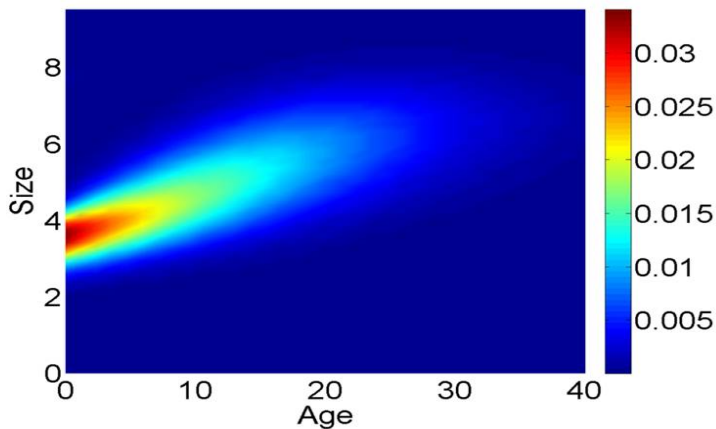
5. Back to the data: testing the Size Model

Figure: Age Size simulation for the Size Model - branch tree data
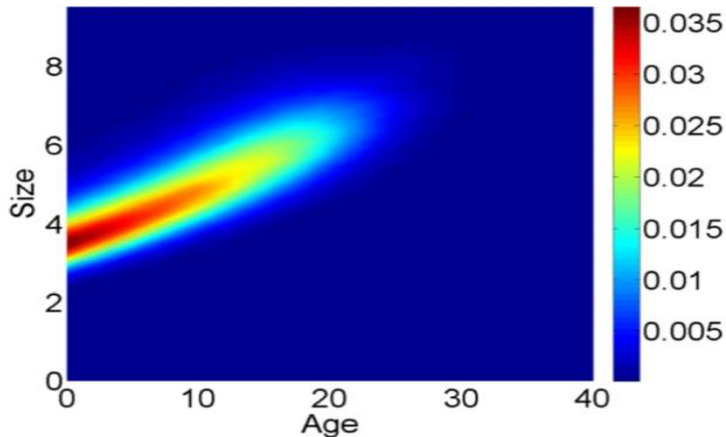
5. Back to the data: testing the Size Model

Figure: Age Size experimental data - branch tree data

5. Back to the data: testing the Size Model

- This model as it is is... reasonable
- errors in L^2 distance are in the order of what is expected: Min of $O(n_1^{-\frac{s}{2s+1}})$ and $O(n_2^{-\frac{s}{2s+3}})$, with n_1 : number of data at division, n_2 : total number of data.

Conclusion

- In bacterial growth, size is a structuring variable, age is not
- Method may be adapted to other cases
- Indirect observation scheme: ill-posed problem.
- Full observation scheme: "well-posed" problem (rejoins density estimation).
- Strong coherence and complementarity between PDE and statistical approaches
- Many open problems: improve the model, investigate the influence of variability...

Conclusion

- Indirect observation scheme: ill-posed problem.
- Full observation scheme: "well-posed" problem (rejoins density estimation).
- Link between stochastic and deterministic modelling via many-to-one formulas for transport-fragmentation processes.
- Strong coherence and complementarity between PDE and statistical approaches
- Variability encompassed into richer stochastic models, with deterministic counterparts if we enlarge the state space
- Other issues: adaptivity, relative size of two offsprings, age dependence.

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

Extensions of the model

Variability:

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

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

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

Age + variability:

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

$$n(t, a = 0, x, v) = 2 \int\limits_x^\infty \int\limits_0^\infty B(a, y) k(y, x) \rho(v', v) n(t, a, y, v') dy \, dv' da$$

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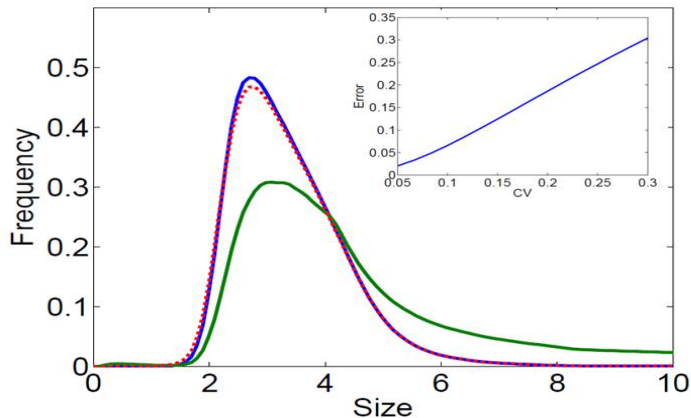
5. Incorporating variability

Figure: Effect on the distribution of growth rate variability
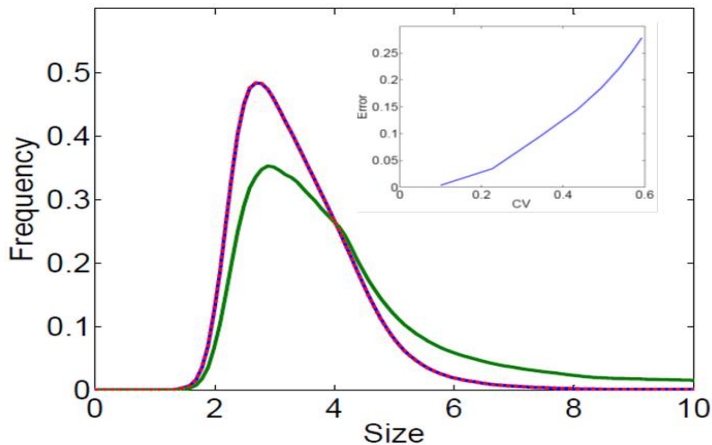
5. Incorporating variability

Figure: Effect on the distribution of variability in daughter sizes

5. What about an Age-Size Model ?
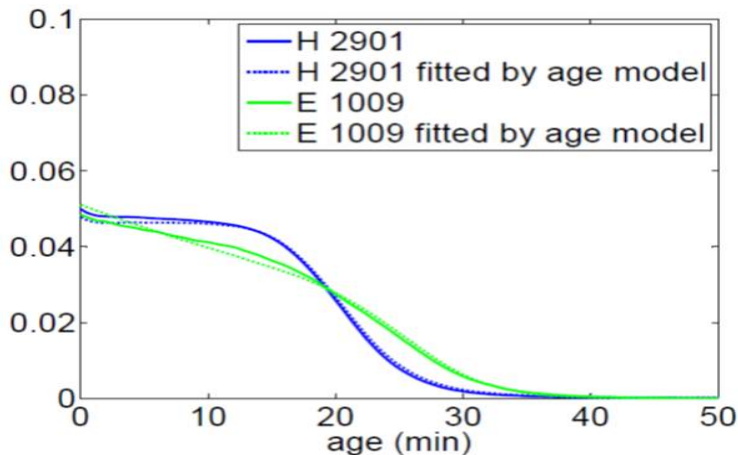
To test it, we would need an extra variable:

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

5. What about an Age-Size Model ?
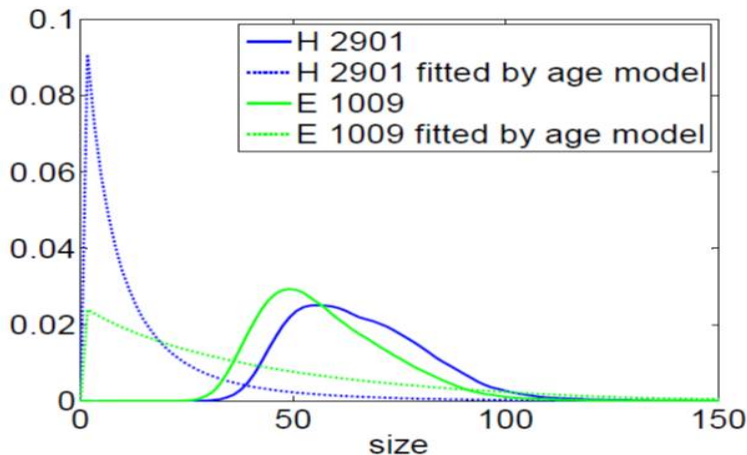
To test it, we would need an extra variable:

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