Are there limits to inference in the simplest ecological models?

Some thoughts and speculations

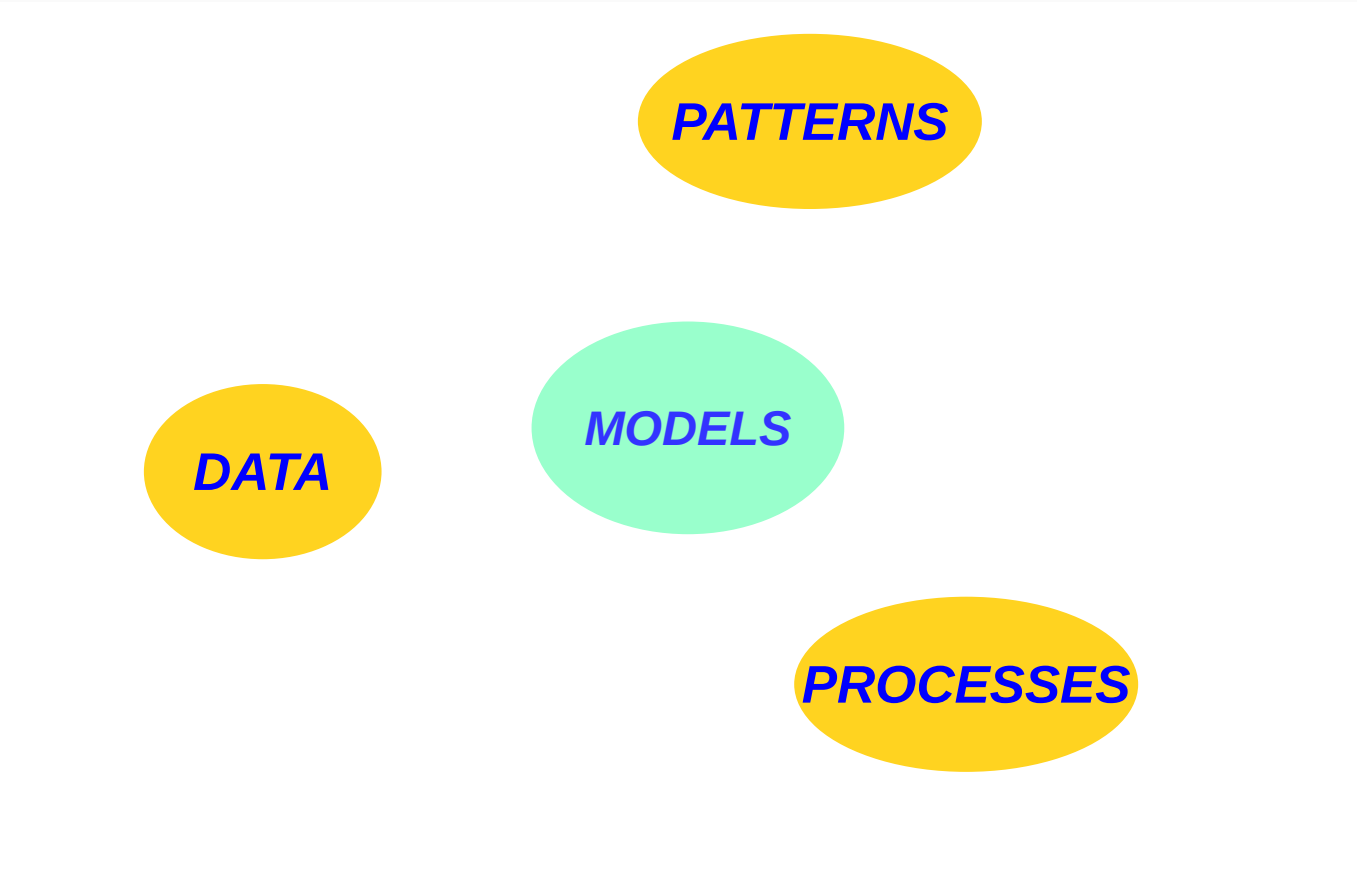
Sylvain Billiard

Université de Lille - France - Lab. Evo-Eco-Paléo - UMR CNRS 8198

After many discussions with probabilists and statisticians

- Vincent Bansaye (CMAP, École Polytechnique)
- Éliza Vergu (INRA Jouy-en-Josas)
- Jean-René Chazottes (CPHT, École Polytechnique)
- Maud Delattre (AgroParisTech)
- Chi Tran (Université Gustave Eiffel)
- Nicolas Champagnat (Institut E. Cartan, INRIA)
- Peter Czuppon (University of Muenster)
- Sylvie Méléard (CMAP, École Polytechnique)
- Pierre Collet (CPHT, École Polytechnique)
- ...

The question: Inference in (population) biology

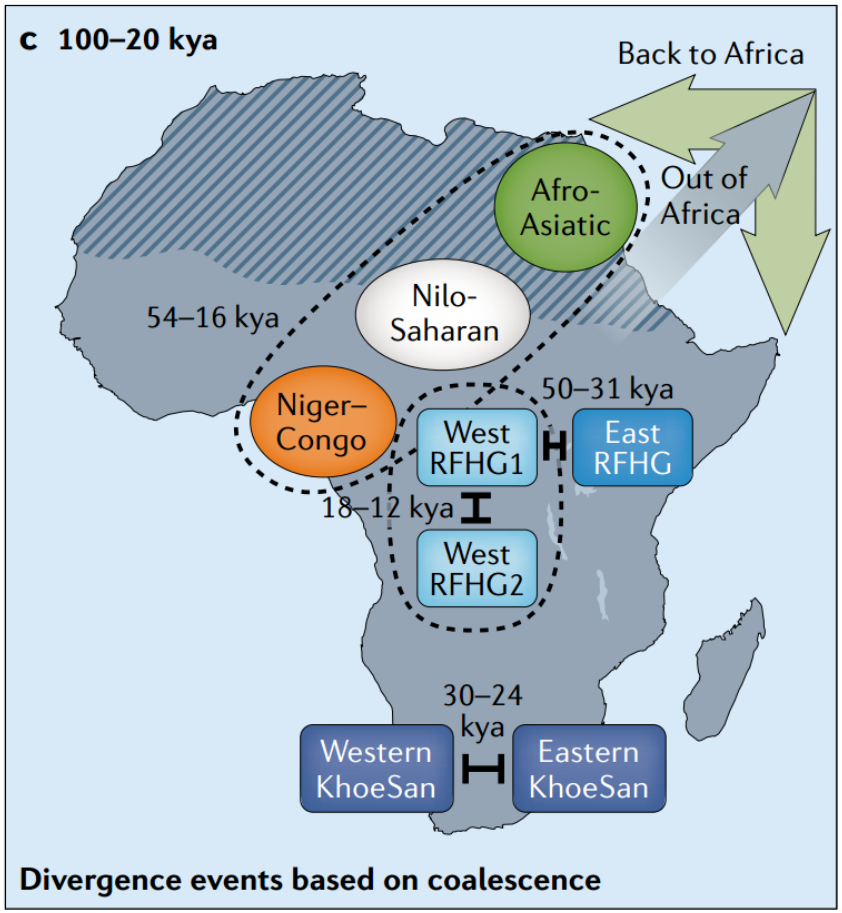


Information about the world from data?

Examples: Genetic diversity and demography reconstruction

African genetic diversity and adaptation inform a precision medicine agenda

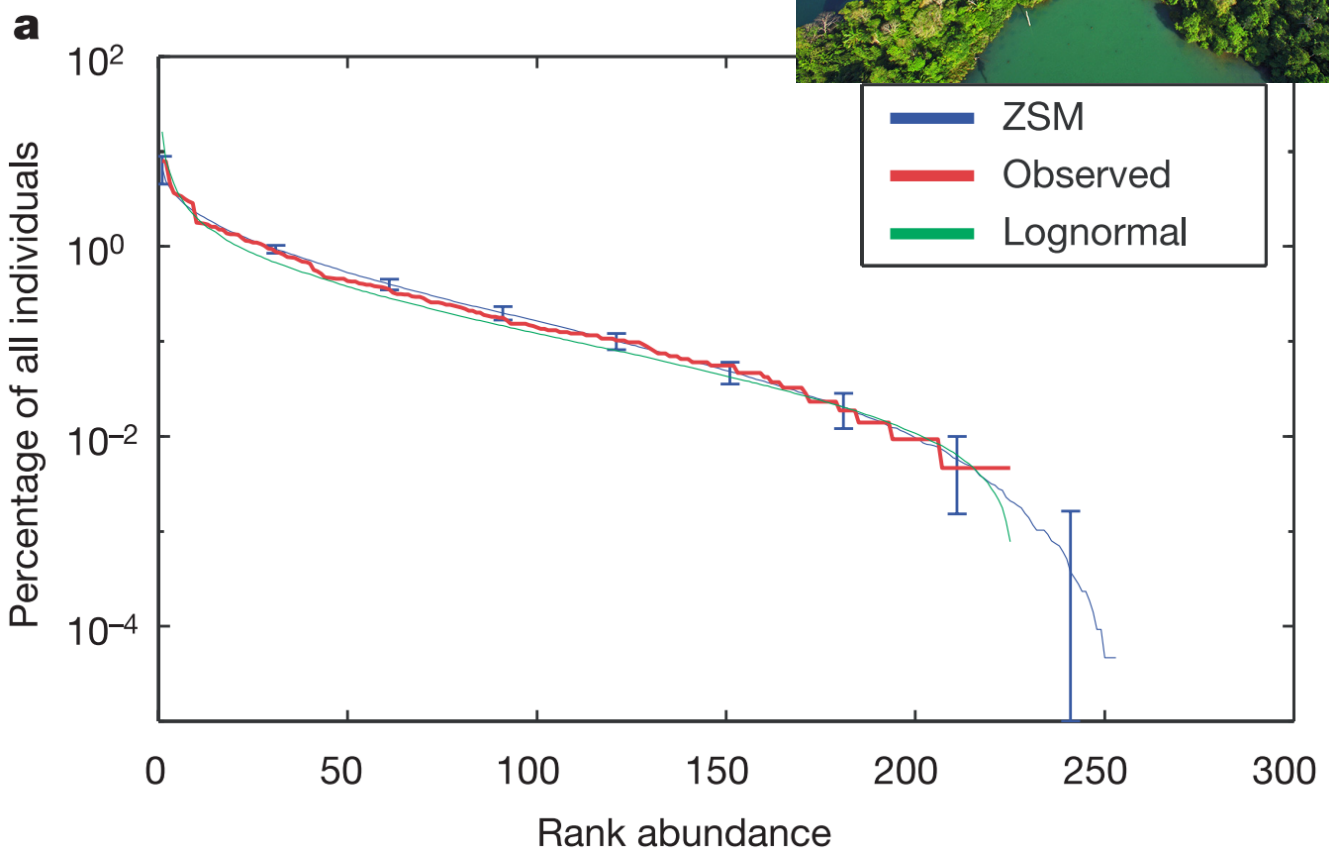
Luisa Pereira[1,2], Leon Mutesa[3], Paulina Tindana[4] and Michèle Ramsay[5]✉



Examples: Species diversity and processes

A test of the unified neutral theory of biodiversity

Brian J. McGill



Limits to inference

- Models much simpler than natural populations
- Models very complex (many parameters)
- Several processes can give similar patterns
- No clues from the past (most often)

  (no idea of initial conditions)

- No independent data to validate inference

  (except from computer simulations but ... circularity issues)
- Big data but little information (autocorrelation, non-independence)

  ...

What can we infer from the simplest population system?

- bacteria (clonal, relatively simple)
- individual count (exact dynamics data)
- homogeneous environment
- few individuals (no competition)
- small timescale (no mutation, no trait evolution)
- controlled experiment (repeatability)
- Direct observations are possible

  ...

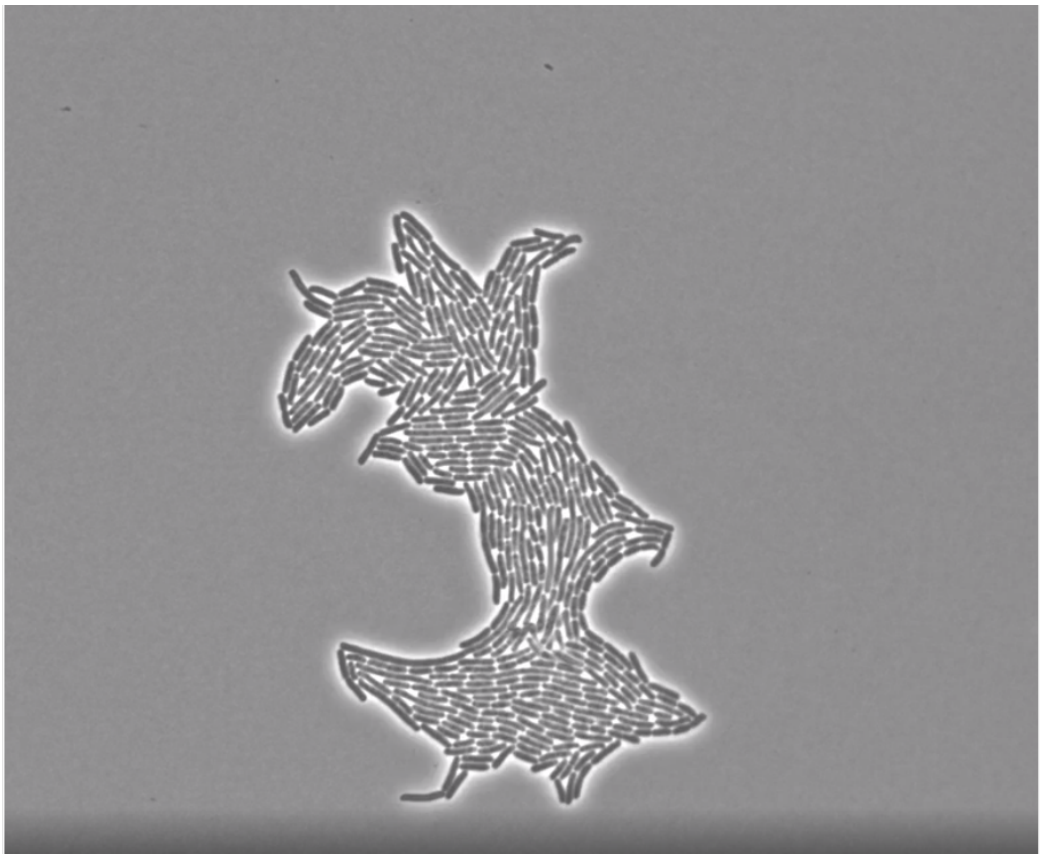
$\rightarrow$ A simple birth-death model.

eLIFE
elifesciences.org

RESEARCH ARTICLE

Antibiotic-induced population fluctuations and stochastic clearance of bacteria

Jessica Coates[1†], Bo Ryoung Park[2†], Dai Le[2], Emrah Şimşek[2], Waqas Chaudhry[2], Minsu Kim[1,2,3*]



Coates et al. 2018's Experiment

Protocol

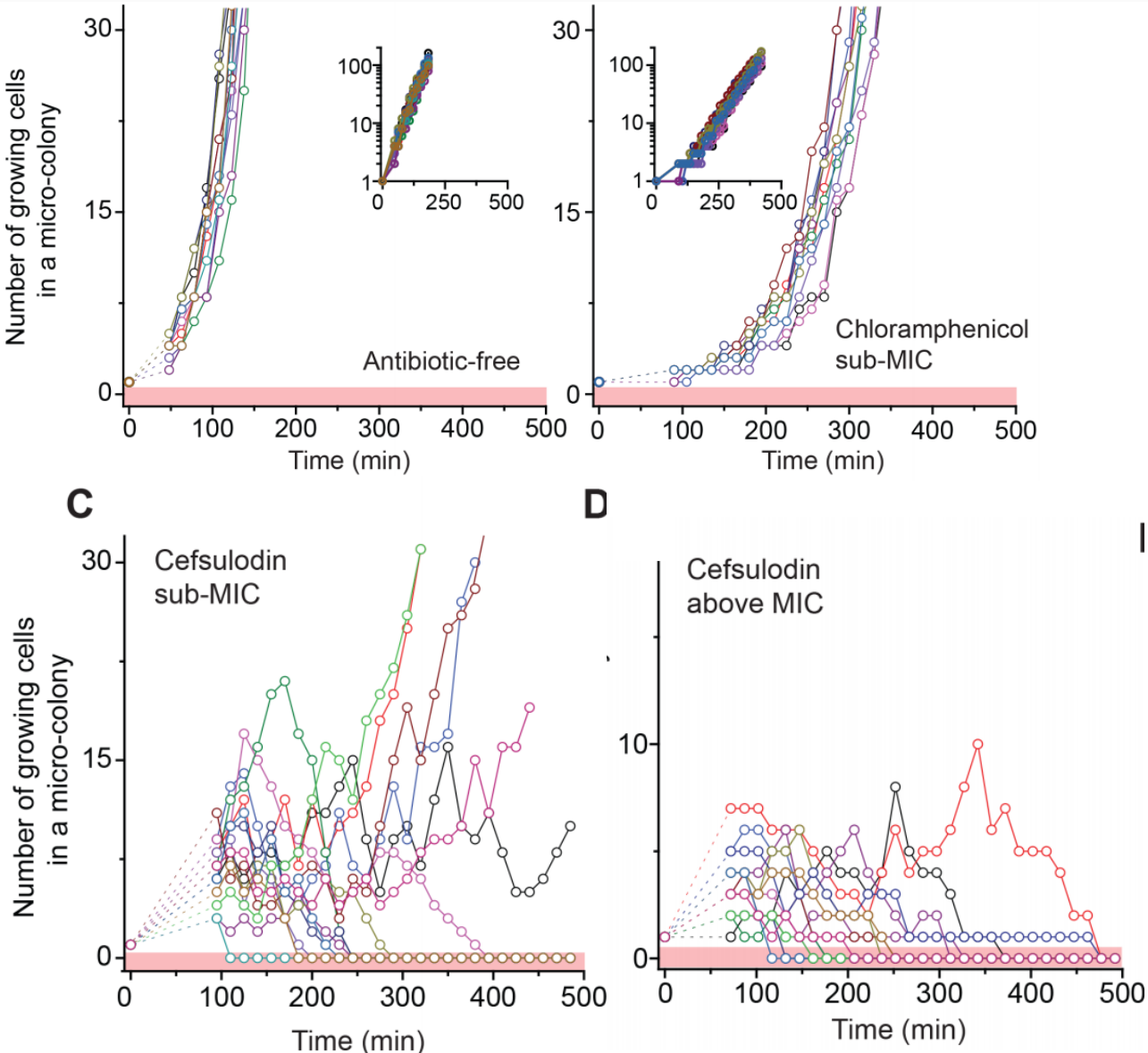
- Inoculation of colonies on plates with a single cell
- Picture every ~ 10 mins
- Rich medium
- Automatized cell count
- Three types of medium:
  1. Antibiotics-free (control)
  2. Bacteriostatic (slows down or stops bacteria growth and division)
  3. Bactericidal (kills bacteria)

What we want to know:
Can we recover that antibiotics differently affect

- the individual birth rate
- the individual death rate

Movies

Data: population dynamics



Inference from a pure birth model (Yule's process)

The model

- X_t: the number of cells at time t
- b: cell birth rate
- Δt: the time frame between observations

Distribution: Negative binomial

$$\mathbb{P}(X_{t+\Delta t} - X_t = k | X_t = n) = \binom{n+k-1}{n-1} p^n (1-p)^k$$

with $p = e^{-b\Delta t}$

Method: Exact Likelihood

$$\mathcal{L}(b) = \sum_i \log \mathbb{P}(X_{t_i} - X_{t_{i-1}} = k_i | X_{t_{i-1}} = k_{i-1}, b)$$

with

- k_i the number of cells in observation i
- t_i the time of observation i

Inference from a pure birth model (Yule's process)

Results

- Control (no antibiotics): $\widehat{b} = 0.023$
  (AIC: 801.8, BIC: 804.7)
- Chloramphenicol (bacteriostatic): $\widehat{b} = 0.014$
  (AIC: 1082.8, BIC: 1085.7)
- Cefsulodin (bactericidal, supercritical): $\widehat{b} = ?$
  (Likelihood not defined)
- Cefsulodin (bactericidal, subcritical): $\widehat{b} = ?$
  (Likelihood not defined)

Direct observations
Cell division rate: 0.025 min.$^{-1}$ (Minsu Kim, pers. comm.)

Inference from a linear birth-death model

The model

- X_t: the number of cells at time t
- b: cell birth rate
- d: cell death rate
- $\gamma = b + d$: an allometry parameter scaling demographic rates
- Δt: the time frame between observations
- Assumption: b and d, and the population size, scale with $K \to \infty$

SDE Approximation (e.g. Bansaye & Méléard 2015

$$dX_t = (b - d)X_t dt + \sqrt{2\gamma X_t} dB_t$$

Inference from a linear birth-death model

Discretization with an Euler's scheme

$$X_{t+\Delta t} - X_t = (b-d)X_t\Delta t + \sqrt{2(b+d)X_t}\,(B_{t+\Delta t} - B_t)$$

Method: Likelihood aproximation in an Euler's Scheme

$$\mathcal{L}(b,d) = -\frac{1}{2}\left(\sum_i \frac{\left(X_{t_i} - X_{t_{i-1}} - (b-d)X_{t_{i-1}}\Delta t_i\right)^2}{2(b+d)X_{t_{i-1}}\Delta t_i}\right) + \sum_i \log(4\pi(b+d)X_{t_i}\Delta t_i)$$

Algorithm

- R package MsdeParEst from Delattre et al 2016
- Mixed effect in the drift part
- Fixed effect in the diffusion part

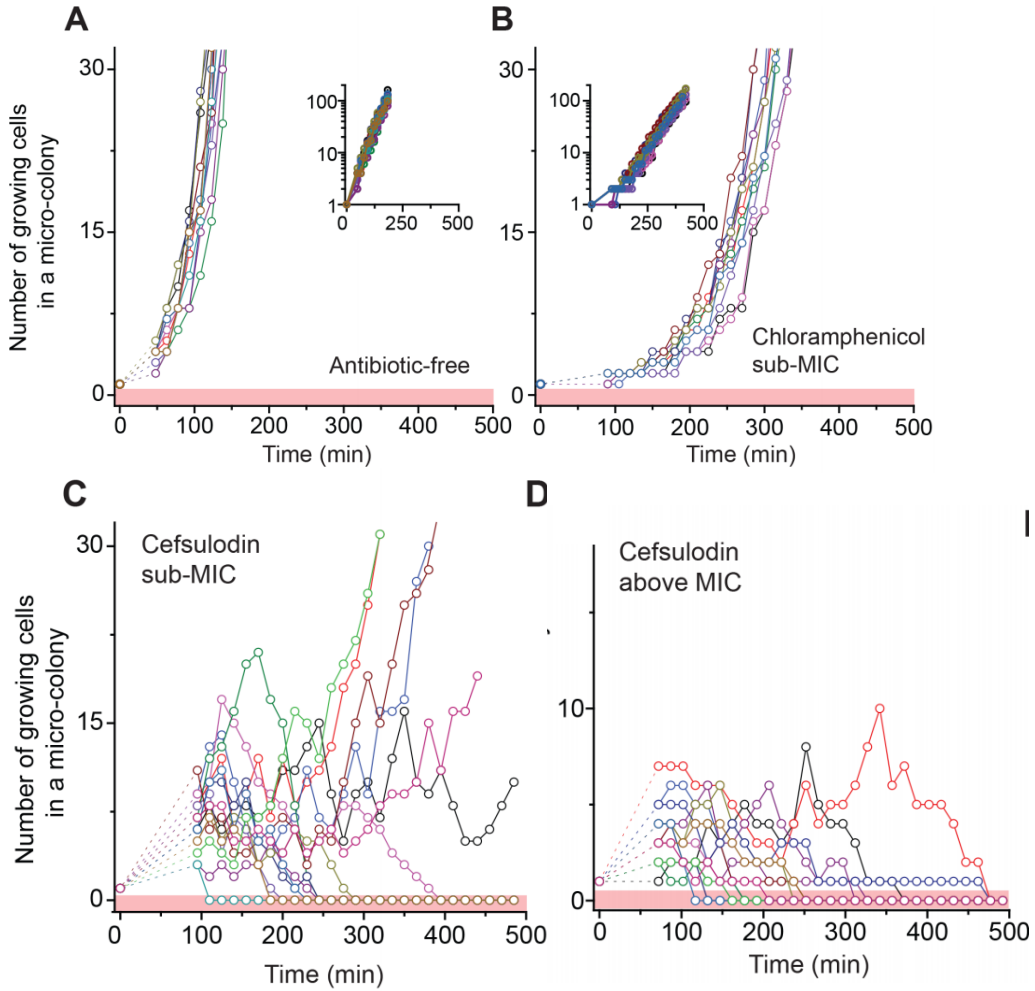
Results

- Control (no antibiotics): $\widehat{b} = 0.039$ and $\widehat{d} = 0.016$
  (AIC: 864.9, BIC: 866.1)
- Chloramphenicol (bacteriostatic): $\widehat{b} = 0.016$ and $\widehat{d} = 0.002$
  (AIC: 902.7, BIC: 902.6)
- Cefsulodin (bactericidal, supercritical): $\widehat{b} = 0.023$ and $\widehat{d} = 0.017$
  (AIC: 297.3, BIC: 299.9)
- Cefsulodin (bactericidal, subcritical): $\widehat{b} = 0.007$ and $\widehat{d} = 0.008$
  (AIC: 170.9, BIC: 173.5)

Issues from the models

Model	Medium	Rates estimate	Growth rate
Pure Birth			
(AIC: 801)	Control	$\widehat{b} = 0.023$	0.023
(AIC: 1083)	Bacteriostatic	$\widehat{b} = 0.014$	0.014
—	Bactericidal (supercritical)	—	—
—	Bactericidal (subcritical)	—	—
Birth-Death			
(AIC: 865)	Control	$\widehat{b} = 0.039$, $\widehat{d} = 0.016$	0.023
(AIC: 903)	Bacteriostatic	$\widehat{b} = 0.016$, $\widehat{d} = 0.002$	0.014
(AIC: 297)	Bactericidal (supercritical)	$\widehat{b} = 0.023$, $\widehat{d} = 0.017$	0.006
(AIC: 171)	Bactericidal (subcritical)	$\widehat{b} = 0.007$, $\widehat{d} = 0.008$	-0.001

Issues from the data



- No extinction before first observation

  $\rightarrow$ Observations conditional on survival

  $\rightarrow$ Observation bias

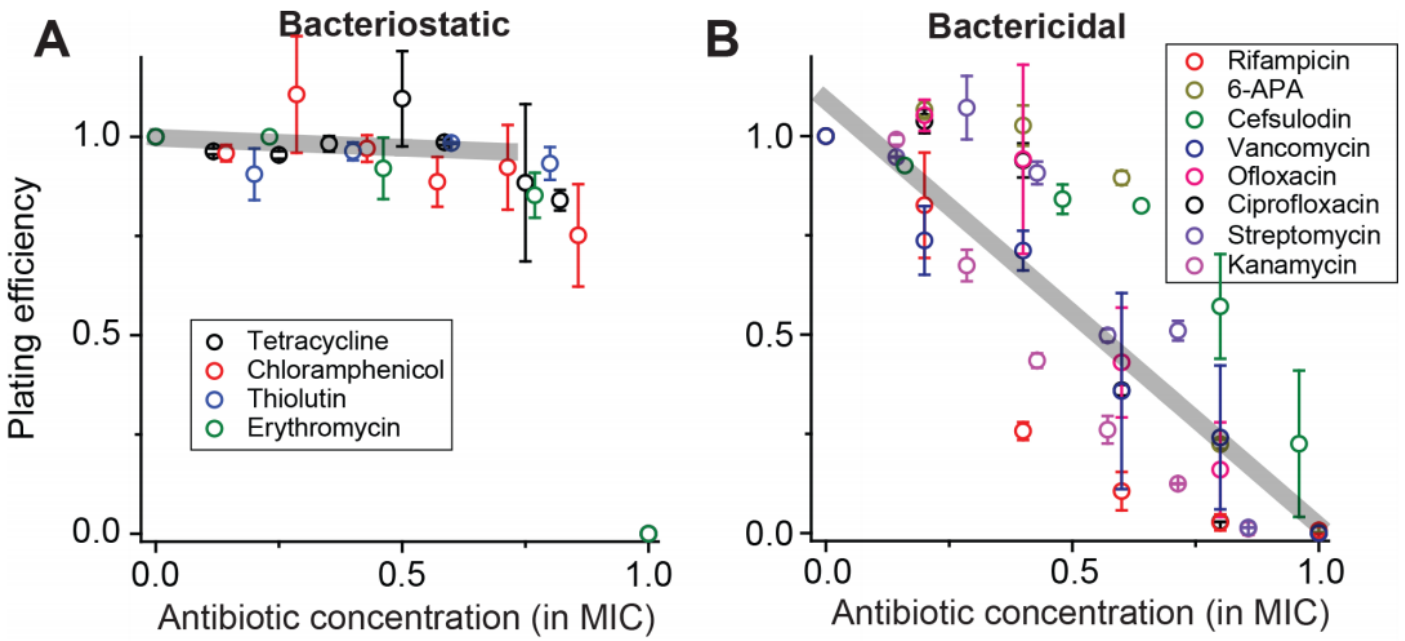
Issues from the interpretation

- Population clearly spatialized
- Competition between individuals
- Medium certainly not homogeneous, not constant
- Cells growth before division: multiscale models?

  $\rightarrow$ interpretation of $\widehat{b}$ and $\widehat{d}$?

Probability of extinction: unexplained

- Pure birth model: $P(\text{extinction}|X_0 = 1) = 0$
- (supercritical) Birth-Death model : $P(\text{extinction}|X_0 = 1) = \frac{d}{b}$



Perspectives

- Even simpler biological systems?
- Include conditional on survival?
- Include observations protocol and apparels into models?
- More informative statistics?
- (*a priori*) Choice of the scale of observation?

  (why not focusing on a single cell?)
- (*a priori*) Choice of the relevant processes?

  (why not bacteria movement?)
- Time is exponential in models, can we test this?
- Inherent and unsurpassable limits?

The remaining question

Most relevant model to estimate b and d on such data?

Acknowledgements

Minsu Kim (for sharing data)
Maud Delattre (for her help with parameter estimations from the SDE)
The Organizers