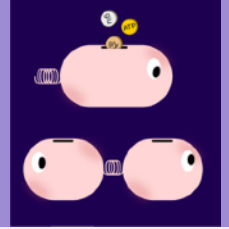
Economic Principles in Cell Biology

Paris, July 10-14, 2023



Metabolic diversity

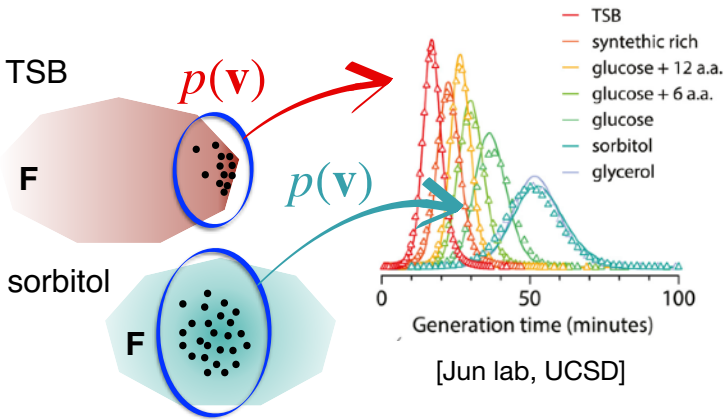
Andrea De Martino, Politecnico di Torino [Part 1]
Daniele De Martino, Instituto Biofisika [Part 2]

- Part 1: (Optimal) Probability densities on the flux polytope
- Part 2: Inference of single-cell quantities

INRAE Département MathNum EURIP LEARNING PLANET INSTITUTE

- Constraint-based models are mostly calibrated for *population averages*
- But cells within a population differ in both `observable' (e.g. growth rate) and `internal' properties (e.g. metabolic phenotype, like fermentative vs respiratory)
- In some cases, diversity is a plus (see e.g. bacterial persistence)
- **Question:** can we capture single-cell properties within the frame of CBMs?

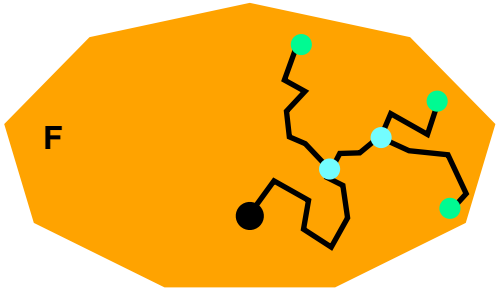
For instance



[Jun lab, UCSD]

- **Feasible space (F)**: defined by mass balance conditions (**Sv**=**0**) and ranges of variability for each v_i ; dim(**F**)=$\mathcal{O}(10^2)$
- **Basic idea:** empirical distributions represent marginals of an unknown high-dimensional distribution $p(\mathbf{v})$ on **F**
- Two ways to understand $p(\mathbf{v})$:
  - **Dynamics**: $p(\mathbf{v},t) \rightarrow p(\mathbf{v})$
  - **Statics (variational)**: "$p(\mathbf{v})$ is optimal"

A minimal model of population dynamics in F



- Same feasible space **F** for all cells
- $n(\mathbf{v},t)$ = nr of cells with flux vector **v** at time t
- Time evolution of n due to (i) replication (rate $\lambda(\mathbf{v})$), (ii) **diffusion** in **F** (small random changes in the flux vector), and (iii) **advection** (cells adjusting **v** to maximize $\lambda(\mathbf{v})$)
  - *The dynamics is sensitive to the growth-rate landscape*
- Finite carrying capacity
- Steady state: balance of diffusion and advection

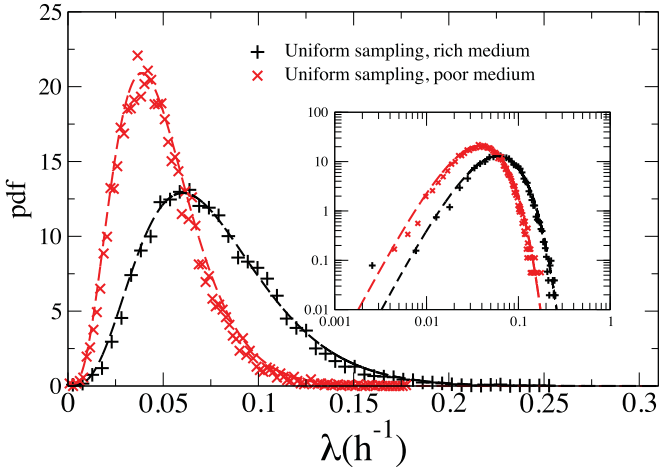
1d case (simple) : $J_{\text{diff}} = -D\frac{\partial n}{\partial v}$, $J_{\text{adv}} = \chi n \frac{\partial \lambda}{\partial v}$

[De Martino et al 2016]

$$J_{\text{diff}} + J_{\text{adv}} = 0 \quad \rightarrow \quad \frac{\partial n}{\partial v} = \beta n \frac{\partial \lambda}{\partial v} \quad \rightarrow \quad n(v) \sim e^{\beta\lambda(v)}$$

$$\beta = \chi / D$$

- *The dynamics is sensitive to the growth-rate landscape*



$$q(\lambda) \propto \lambda^b(\lambda_{\max} - \lambda)^a \quad , \quad a \gg b \quad , \quad a \gg 1$$

- Small random changes to v are overwhelmingly more likely to reduce the growth rate than increase it

"FBA" (max λ)

$\chi \gg D$

$n(v) \sim e^{\beta\lambda(v)}$

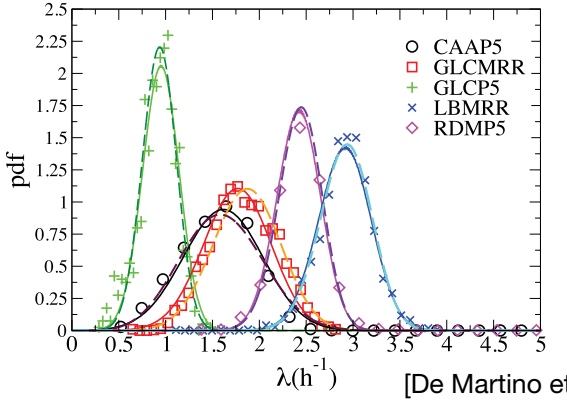
$\beta = \chi/D$

$D \gg \chi$

Uniformly distrib. over F

Comparisons

- Compare marginals of $p(\mathbf{v})$ for the growth rate with data (fitting parameter: β)



[De Martino et al 2016]

Table 1. Inferred maximum growth rates, level of optimization and rate of metabolic change for the experimental data [19, 20] fitted with the stationary distributions retrieved by the MaxEnt framework of section 3 and from the dynamical model of section 4. Growth rates are measured in h^{-1}, while σ and $\beta\lambda_{max}$ are adimensional.

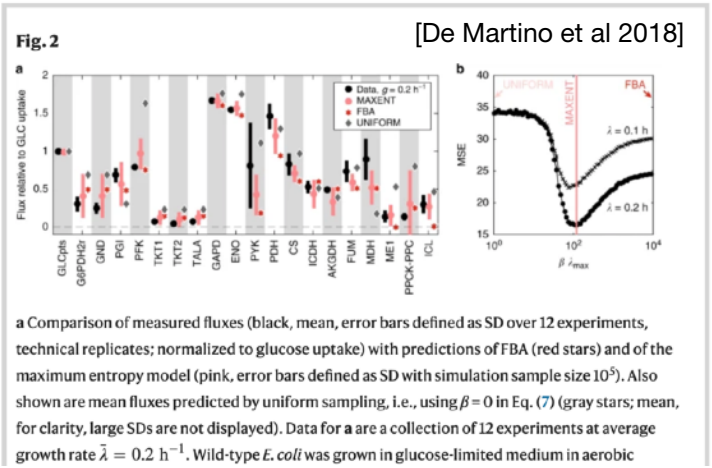
Data set	MaxEnt		Dynamical	
	λ_{max} (h^{-1})	$\beta\lambda_{max}$ (adim.)	λ_{max} (h^{-1})	σ (adim.)
[19] rich medium	5.9	220	7.2	10^{-5}
[19] poor medium	3.2	220	3.8	10^{-5}
[20] GLCP5	3.5	220	4.3	10^{-5}
[20] GLCMRR	7	220	8	10^{-5}
[20] CAAP5	8.6	190	9	1.2×10^{-5}
[20] RDMP5	5.5	300	6.4	5×10^{-6}
[20] LBMRR	6.6	300	7.7	5×10^{-6}

$$p(\mathbf{v}) = \frac{e^{\beta\lambda(\mathbf{v})}}{Z(\beta)}$$

(data: Kennard et al 2016)

- Compare marginals of $p(\mathbf{v})$ vs MS fluxes
  - Assume $p(\boldsymbol{v})$ with empirical β
  - Sample $p(\boldsymbol{v})$ (e.g. MC)
  - Compute marginals
  - Compare vs experiments

[De Martino et al 2018]



a Comparison of measured fluxes (black, mean, error bars defined as SD over 12 experiments, technical replicates; normalized to glucose uptake) with predictions of FBA (red stars) and of the maximum entropy model (pink, error bars defined as SD with simulation sample size 10^5). Also shown are mean fluxes predicted by uniform sampling, i.e., using $\beta = 0$ in Eq. (7) (gray stars; mean, for clarity, large SDs are not displayed). Data for **a** are a collection of 12 experiments at average growth rate $\bar{\lambda} = 0.2\ h^{-1}$. Wild-type *E. coli* was grown in glucose-limited medium in aerobic

Variational route to *n*(v)~exp[βλ(v)]

- Mean growth rate <λ> ~ population fitness
- **An `energy-entropy' tradeoff:** distributions *p*(**v**) with large <λ> have small entropy and v.v.

$$\max_{p(\mathbf{v})} H[p] \quad \text{s.t.} \quad \langle\lambda\rangle \quad \rightarrow \quad p(\mathbf{v}) = \frac{e^{\beta\lambda(\mathbf{v})}}{Z(\beta)}$$

Entropy of *p* :

$$H[p] = -\int_F p(\mathbf{v}) \ln p(\mathbf{v})\, d^N v$$

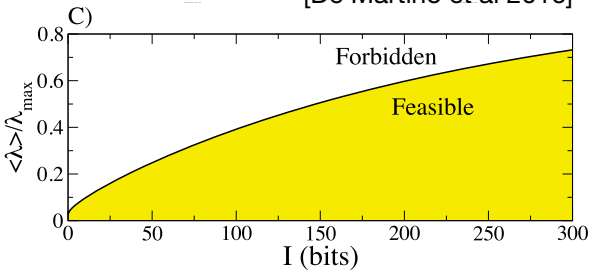
- ***Lesson (2016):*** *at the metabolic level (CBMs), cells within a population appear to have maximal growth-rate heterogeneity for the population's fitness (!)*
- ***To go more in depth:*** *some more theory + inference...*

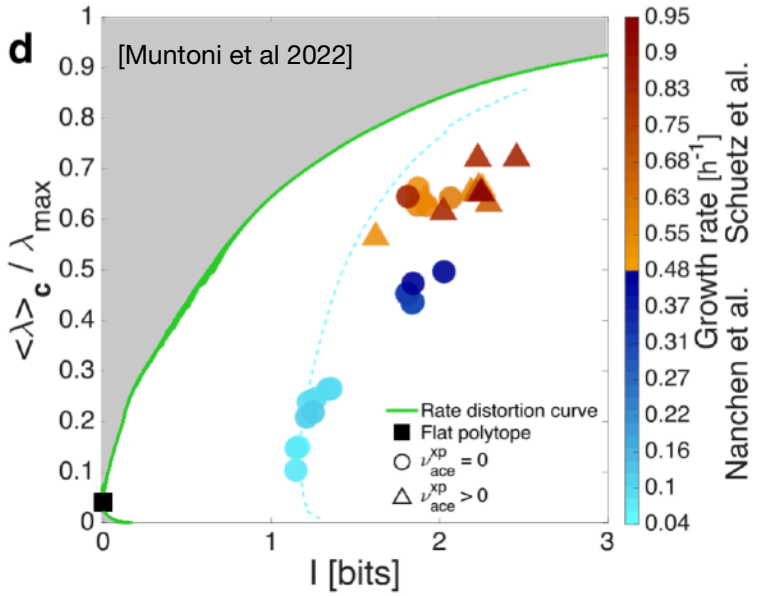
- **Relationship between <λ> and *H***

$$H(0) - H(\beta) \equiv I \ln 2 = \beta\langle\lambda\rangle - \int_0^\beta \langle\lambda\rangle\, d\beta'$$

- **Re-phrasing:** what is the minimum number of bits (*I*) to be encoded in *p(**v**)* in order to achieve a given "fitness" (mean growth rate)?

[De Martino et al 2016]

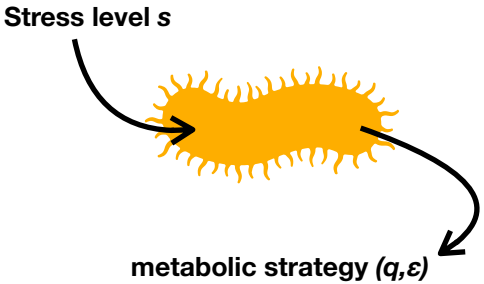



d
[Muntoni et al 2022]
$\langle\lambda\rangle_c / \lambda_{max}$
I [bits]
Growth rate [h^{-1}]
Nanchen et al.
Schuetz et al.
Rate distortion curve
Flat polytope
$\nu^{xp}_{ace} = 0$
$\nu^{xp}_{ace} > 0$

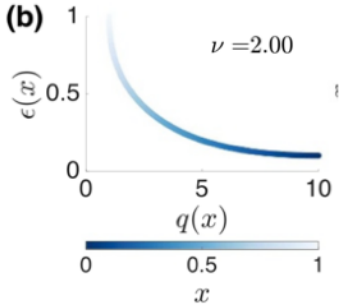
Heterogeneity as an optimal response

Best metabolic strategy in a fluctuating medium

[Muntoni et al 2023]



q-ε trade-off



- *q(x)* = specific intake
- ε(x) = specific proteome cost

$$\lambda(x,s) = \frac{\phi}{w + sq(x) + \epsilon(x)}$$

- What if *s* fluctuates?
- *P(s)* (distrib of stress levels)
- Fast fluctuations: maximize <λ> (avg over *s*)

$$\langle \lambda \rangle = \int ds\, P(s) \int dx\, P(x \mid s) \lambda(x,s)$$

- Optimize over conditional response *P(x|s)* ...
- ... subject to mutual information of *x* and *s*

$$I(x;s) = \int ds\, P(s) \int dx\, P(x \mid s) \log_2 \frac{P(x \mid s)}{P(x)}$$

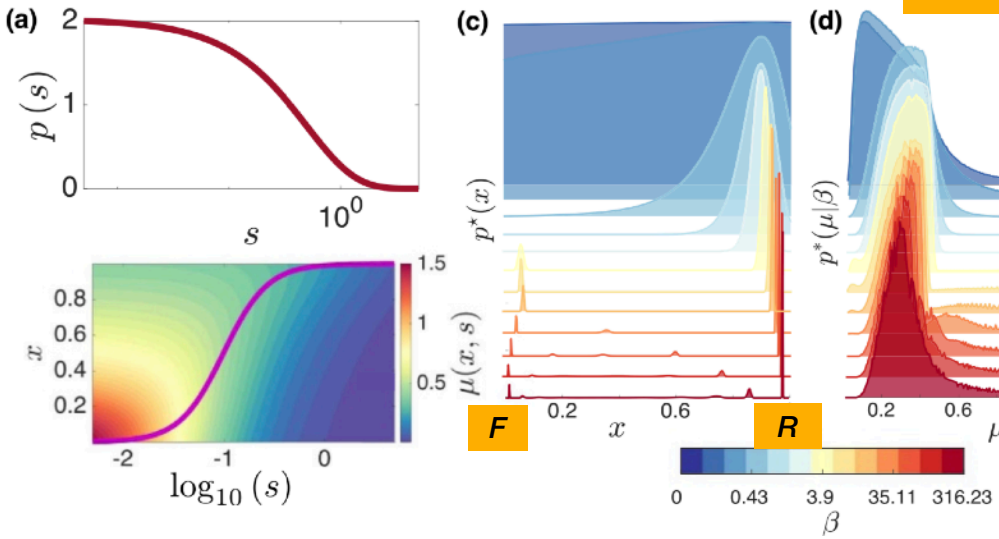
- Solution:

$$P(x \mid s) = \frac{P(x)}{Z(s,\beta)}\, e^{\beta \lambda(x,s)}$$

P(s)

P(x) (distrib of metabolic strategies, diff. *β*)

Distrib. of growth rates (diff. *β*)

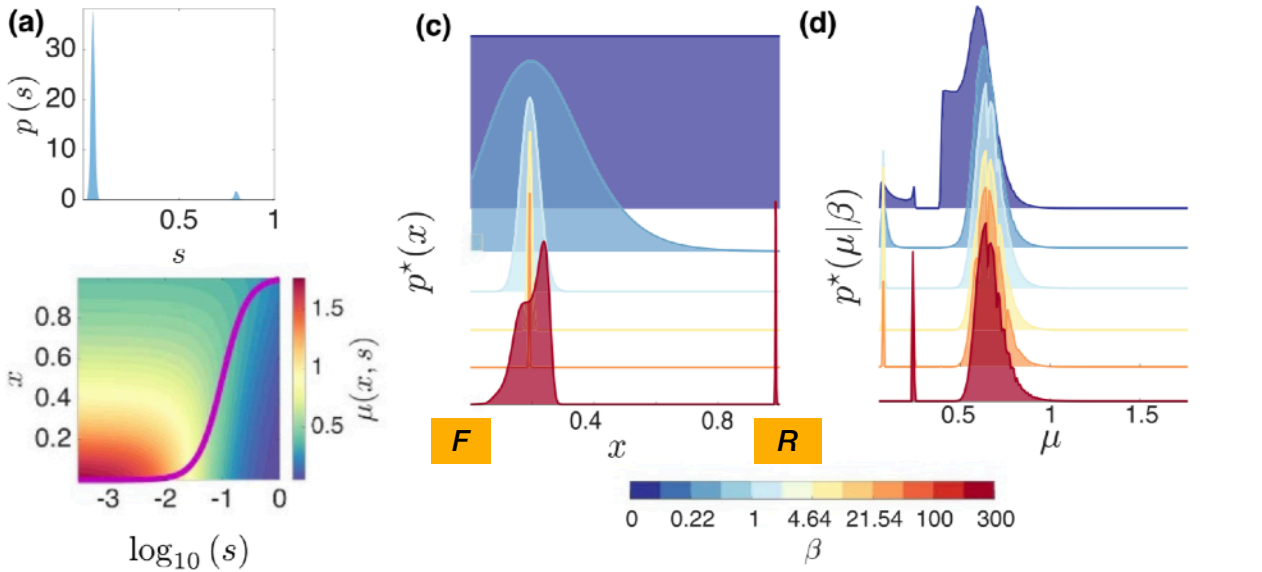


E.g. exponential *P(s)*

P(s)

P(x) (distrib of metabolic strategies, diff. *β*)

Distrib. of growth rates (diff. *β*)



E.g. bimodal *P(s)*

Summary

- Metabolic diversity from dynamics
- Beyond bulk properties: probability densities on the flux polytope (with some simplifying assumptions)
- Bacterial populations close to maximizing diversity at given fitness
- Diversity as optimal response (e.g. in fluctuating media)
- **This half:** distributions that are "optimal" (in some sense)
- **Next half:** learning distributions from data (and see how far they are from optima)

POLITECNICO DI TORINO

MARIE CURIE ACTIONS

Italian Institute for Genomic Medicine

IIGM

*joint work with **D De Martino (BFI Bilbao)**, **AP Muntoni (IIGM)**,*
A Braunstein, A Pagnani, T Gueudrè, M Miotto, F Capuani
&
many more

Inferring
single cell metabolic fluxes

Daniele De Martino
Ikerbasque Fellow
Biofisika Institute FBB/IBF

Paris School
Principles of cell Physiology

ikerbasque
Basque Foundation for Science

BIOFISIKA
Basque Centre for Biophysics

daniele.demartino@ehu.eus

Synopsis

1)Introduction

Bayes Theorem with example

2) Inferring single cell fluxes

- Growth rates from imaging
- Uptakes/turnover from nano-sims

3)Perspectives: inverse modeling

- Intercellular exchange from Nanofibers sensing

The Monty Hall problem

1) There are three closed doors and only one hides a prize

2) You can choose one (that will not be opened yet)

3) The game-show host (that knows where the prize is)
opens one of the other two in such a way as not to reveal the prize and offers you to swap your door for the remaining closed door

Is it convenient to swap?

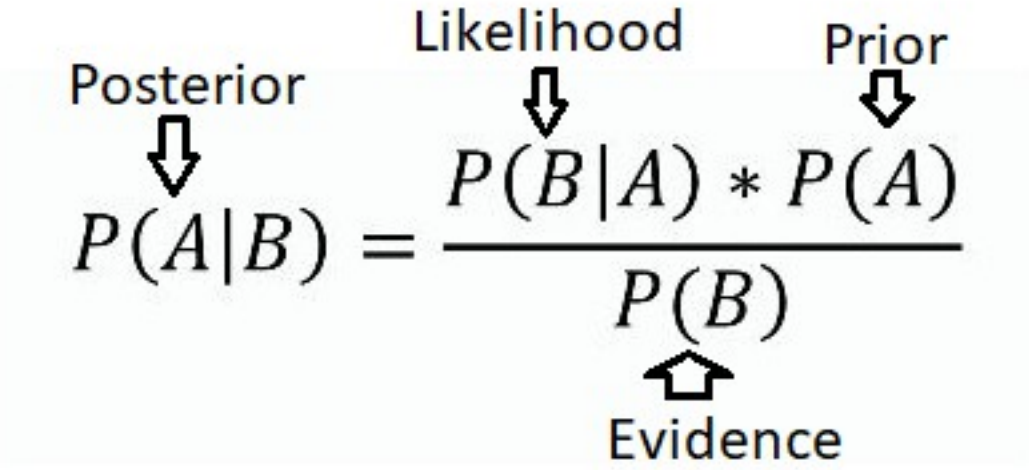
Eg you choose door 1, the host opens 3 showing it is empty, will you swap to 2?

Bayes Theorem

$$P(A|B)P(B) = P(B|A)P(A)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Demonstration: symmetry of conditional probability by definition



Solution

Solution to exercise 3.8 (p.57). Let $\mathcal{H}_i$ denote the hypothesis that the prize is behind door i. We make the following assumptions: the three hypotheses $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$ are equiprobable *a priori*, i.e.,

$$P(\mathcal{H}_1) = P(\mathcal{H}_2) = P(\mathcal{H}_3) = \frac{1}{3}. \tag{3.36}$$

The datum we receive, after choosing door 1, is one of $D=3$ and $D=2$ (meaning door 3 or 2 is opened, respectively). We assume that these two possible outcomes have the following probabilities. If the prize is behind door 1 then the host has a free choice; in this case we assume that the host selects at random between $D=2$ and $D=3$. Otherwise the choice of the host is forced and the probabilities are 0 and 1.

$$\begin{array}{|c|c|c|} P(D=2\,|\,\mathcal{H}_1)=1/2 & P(D=2\,|\,\mathcal{H}_2)=0 & P(D=2\,|\,\mathcal{H}_3)=1 \\ P(D=3\,|\,\mathcal{H}_1)=1/2 & P(D=3\,|\,\mathcal{H}_2)=1 & P(D=3\,|\,\mathcal{H}_3)=0 \end{array} \tag{3.37}$$

Now, using Bayes' theorem, we evaluate the posterior probabilities of the hypotheses:

$$P(\mathcal{H}_i\,|\,D=3) = \frac{P(D=3\,|\,\mathcal{H}_i)P(\mathcal{H}_i)}{P(D=3)} \tag{3.38}$$

$$\left|\; P(\mathcal{H}_1\,|\,D=3) = \frac{(1/2)(1/3)}{P(D=3)} \;\right|\; P(\mathcal{H}_2\,|\,D=3) = \frac{(1)(1/3)}{P(D=3)} \;\left|\; P(\mathcal{H}_3\,|\,D=3) = \frac{(0)(1/3)}{P(D=3)} \;\right| \tag{3.39}$$

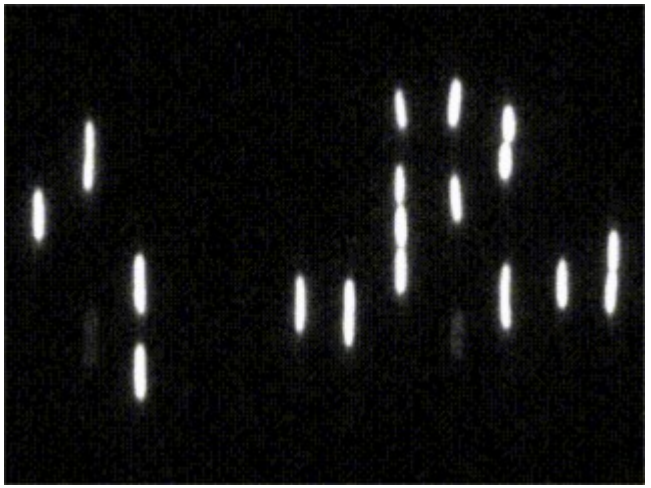
The denominator $P(D=3)$ is $(1/2)$ because it is the normalizing constant for this posterior distribution. So

$$\left|\; P(\mathcal{H}_1\,|\,D=3) \;=\; 1/3 \;\right|\; P(\mathcal{H}_2\,|\,D=3) \;=\; 2/3 \;\left|\; P(\mathcal{H}_3\,|\,D=3) \;=\; 0. \;\right| \tag{3.40}$$

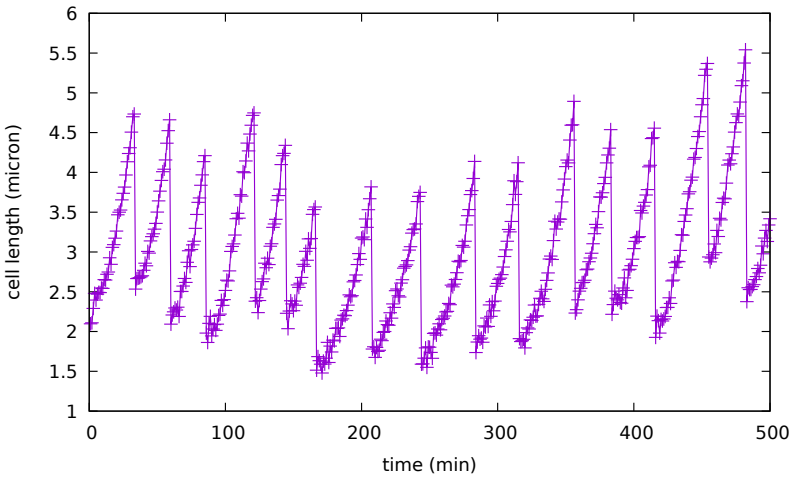
So the contestant should switch to door 2 in order to have the biggest chance of getting the prize.

MacKay, D. J. (2003). Information theory, inference and learning algorithms.

Inferring single cell growth rates



Mother machine movie



Time traces length vs time

Data from:
Tanouchi, Y., Pai, A., Park, H., Huang, S., Buchler, N. E., & You, L. (2017).
Long-term growth data of Escherichia coli at a single-cell level. Scientific data, 4(1), 1-5.

The Linear fit, revisited (I)

$$l(t) = l_0 e^{\lambda t}$$

Hypothesis 1: exponential growth!

$$\log l(t_i)/l_0 = \lambda t_i + \mathrm{noise}(t_i)$$

Let us linearize taking the log!
i runs over data points

Hyopthesis 2: noise terms are Gaussian random variables independent and identically distributed with stdv sigma

$$p(data|\lambda)p(\lambda) = p(\lambda|data)p(data)$$

Bayes theorem

$$p(\lambda) \sim \mathrm{const.}$$

Hypothesis 3: approx. uniform prior

$$p(data) = \int p(data|\lambda)p(\lambda)d\lambda$$

The “evidence” is just a constant

Linear fit revisited (II)

$$P(\lambda|data) \propto P(data|\lambda) \propto \prod_i e^{-\frac{(\log l(t_i)/l_0 - \lambda t_i)^2}{2\sigma^2}}$$

We have the full posterior!

$$\mathcal{L} = -\frac{1}{2\sigma^2}\sum_i (\log l(t_i)/l_0 - \lambda t_i)^2 - N \log\sqrt{2\pi\sigma^2}$$

Log-likelihood

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0$$

$$\lambda^* = \frac{\sum_i t_i (\log l(t_i)/l_0)}{\sum_i t_i^2}$$

Max likelihood sol. coincides with Chi^2 min..

Constant rate or constant speed?

$$l(t) = l_0 + vt$$

Alternative model: constant elongation speed!

$$l(t_i) = l_0 + vt_i + \text{noise}(t_i) \qquad p(v) \sim \text{const.}$$

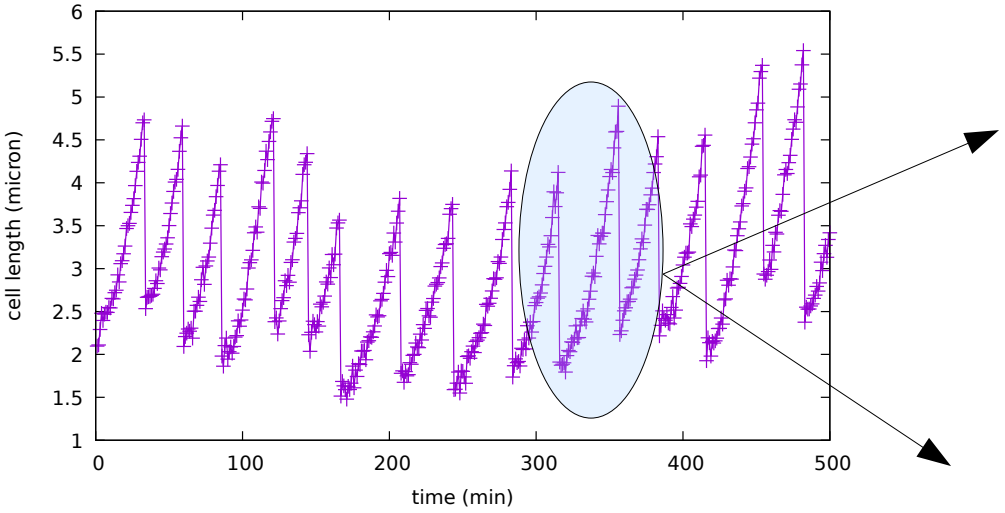
$$P(v|data) \propto P(data|v) \propto \prod_i e^{-\frac{(l(t_i)-l_0-vt_i)^2}{2\sigma^2}}$$

$$\mathcal{L} = -\frac{1}{2\sigma^2}\sum_i (l(t_i) - l_0 - vt_i)^2 - N\log\sqrt{2\pi\sigma^2}$$

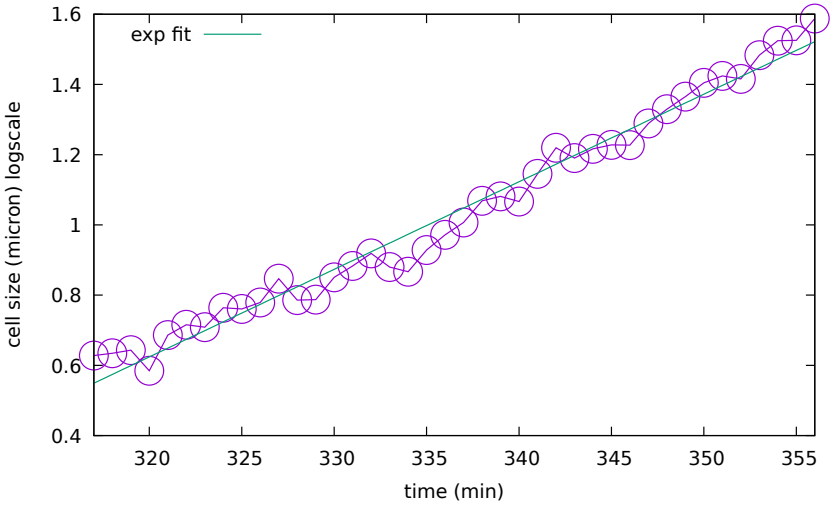
$$\frac{\partial \mathcal{L}}{\partial v} = 0$$

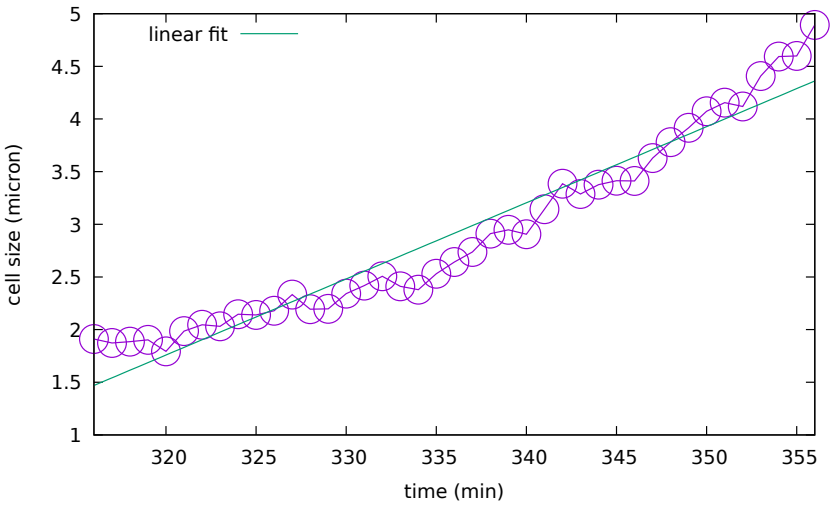
$$v^* = \frac{\sum_i t_i(l(t_i) - l_0)}{\sum_i t_i^2}$$

Let us test them!



Exponential growth





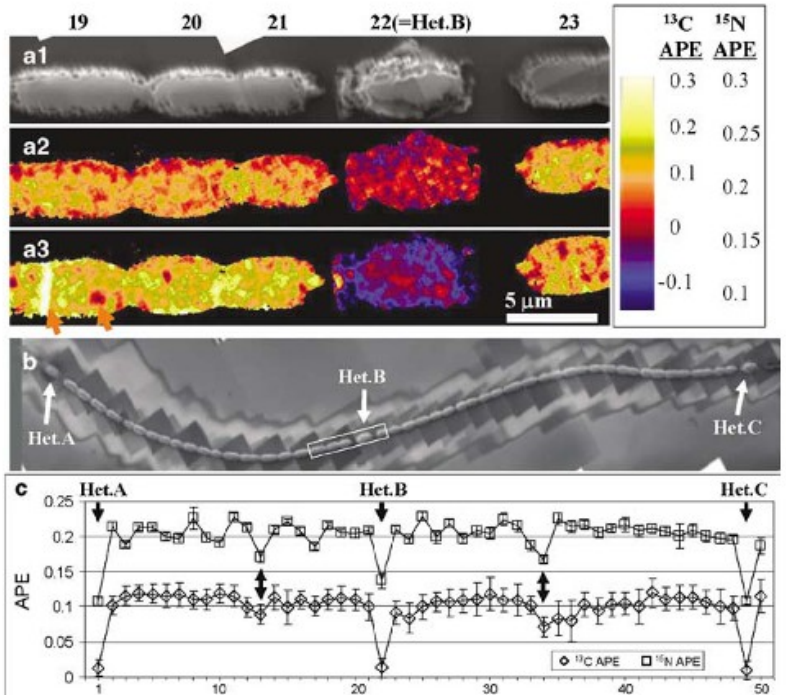
Linear growth

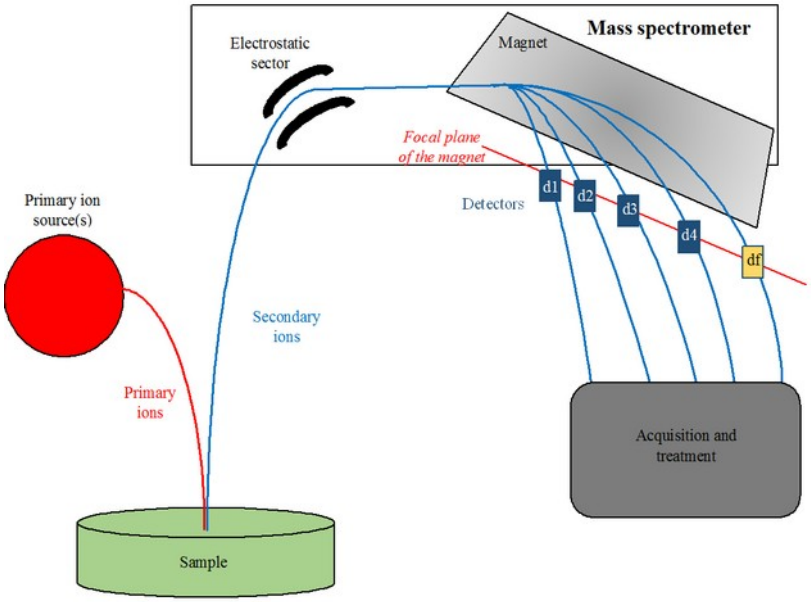
log-likelihood(exp)/log-likelihood(lin)>20
Single cell exponential growth!
(this is non-trivial no?)

Inferring single cells uptakes/turnover/growth from nanosims data

Mass spectrometry + isotope labeling is the most widespread technique for flux analysis → extension to single cells?

“Nanoscale secondary ion mass spectrometry”
+ stable isotope labeling





https://en.wikipedia.org/

Radu et al. "Carbon and nitrogen fixation and metabolite exchange in and between individual cells of Anabaena oscillarioides." The ISME journal 1.4 (2007): 354-360.

From ratios to rates

$$C(t) = C_0 + (C_1 - C_0)(1 - e^{-rt})$$

Main hypothesis:
linear kinetics for concentrations

$$x_0 = \frac{C_0^H}{C_0^L} \qquad x(t) = \frac{C^H}{C^L} = \frac{C_0^H + (C_1^H - C_0^H)(1 - e^{-rt})}{C_0^L + (C_1^L - C_0^L)(1 - e^{-rt})}$$

$$x_1 = \frac{C_1^H}{C_1^L} \qquad C_0^H + C_0^L = C_1^H + C_1^L$$

$$x(t) = \frac{x_1(x_0 + 1) + (x_0 - x_1)e^{-rt}}{1 + x_0 + (x_1 - x_0)e^{-rt}}$$

$$r = -\frac{1}{T} \log \left(\frac{(x_1 - x)(x_0 + 1)}{(x_1 - x_0)(x + 1)} \right)$$

T: incubation time
x:observed ratio
x_1: ratio, labeled
x_0: ratio, natural
r: rate

But what is really r?

A simple model for the linear kinetics: a growing rod fed by diffusion

$$\dot{N} = \alpha(c_e - c)S$$

Simple diffusion through surface S

$$\dot{c} = \frac{\dot{N}V - \dot{V}N}{V^2}$$

Derivative of a ratio

$$\lambda = \dot{V}/V$$

Growth rate

$$S/V = \theta$$

Rod hypothesis

$$\dot{c} = u - (\gamma + \lambda)c$$

$$u = \alpha\theta c_e$$

$$\gamma = \alpha\theta$$

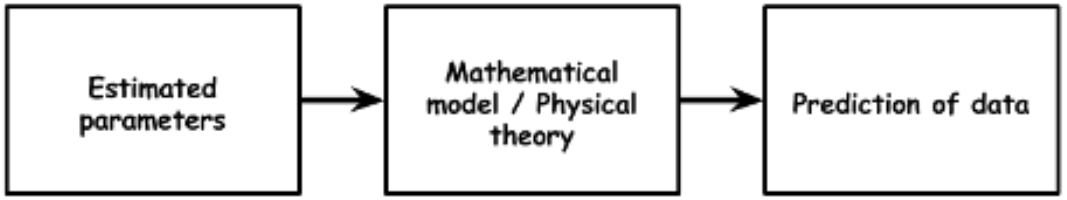
Linear kinetics!

$$r = \gamma + \lambda$$

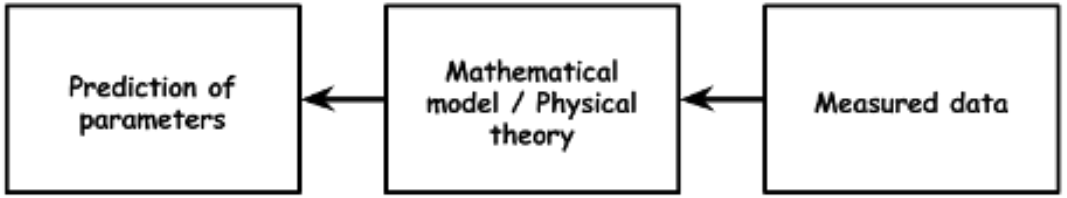
Incorporation rate = growth rate+turnover
r is an upper bound for the growth rate!

More in general: Inverse modeling

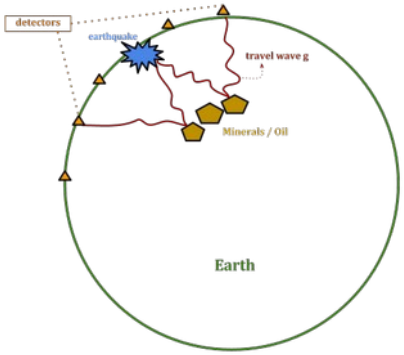
The forward problem



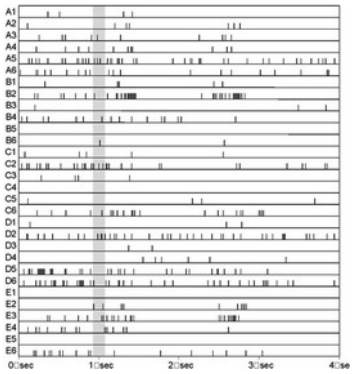
The inverse problem



e.g. geophysics
Inverting wave equations
to reconstruct earth density profiles

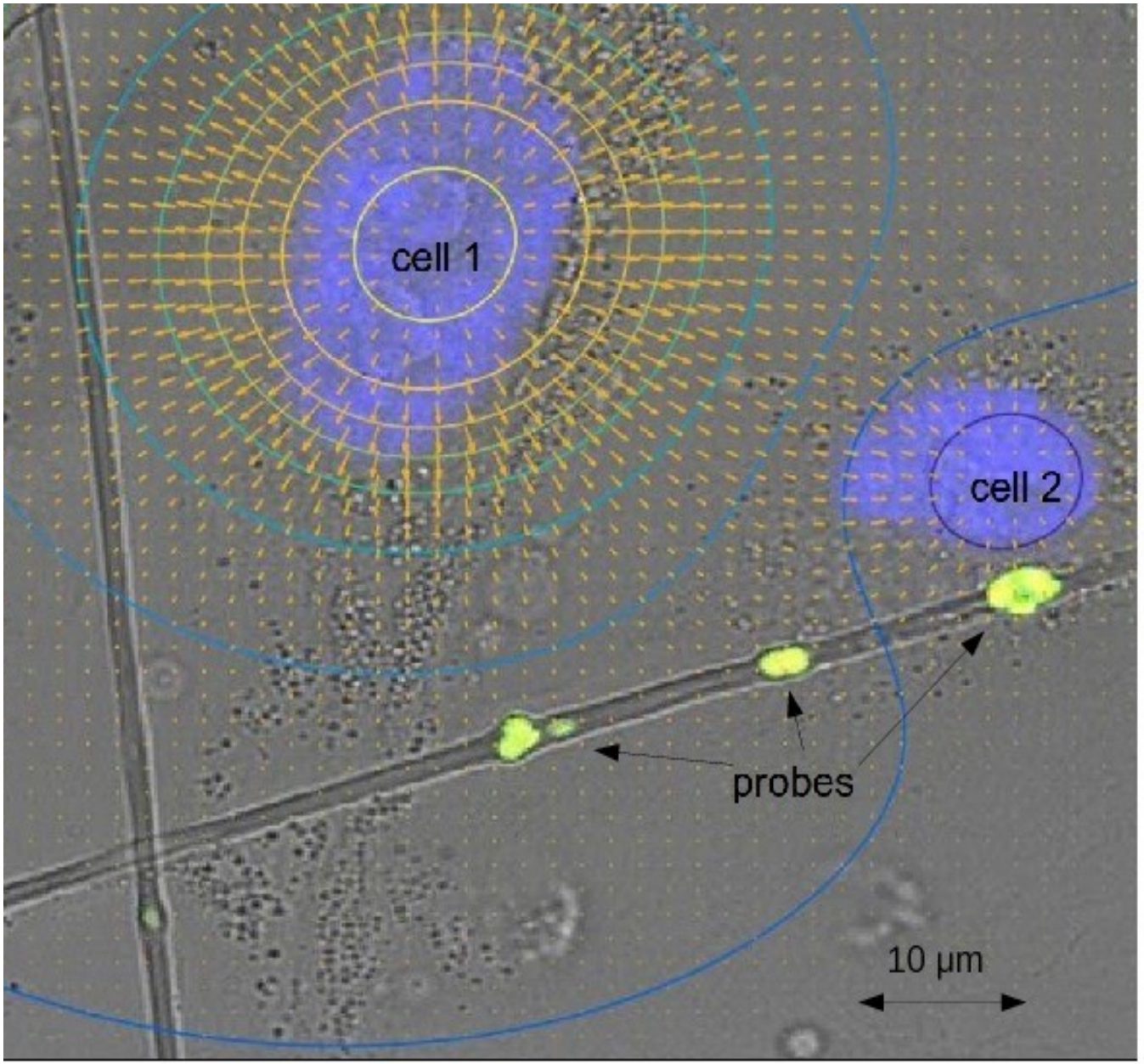


Neuroscience: Inferring neural interactions from spike data
Schneidman et al (2006), Nguyen, Zecchina & Berg (2017)



?

Flux inference & gradient reconstruction

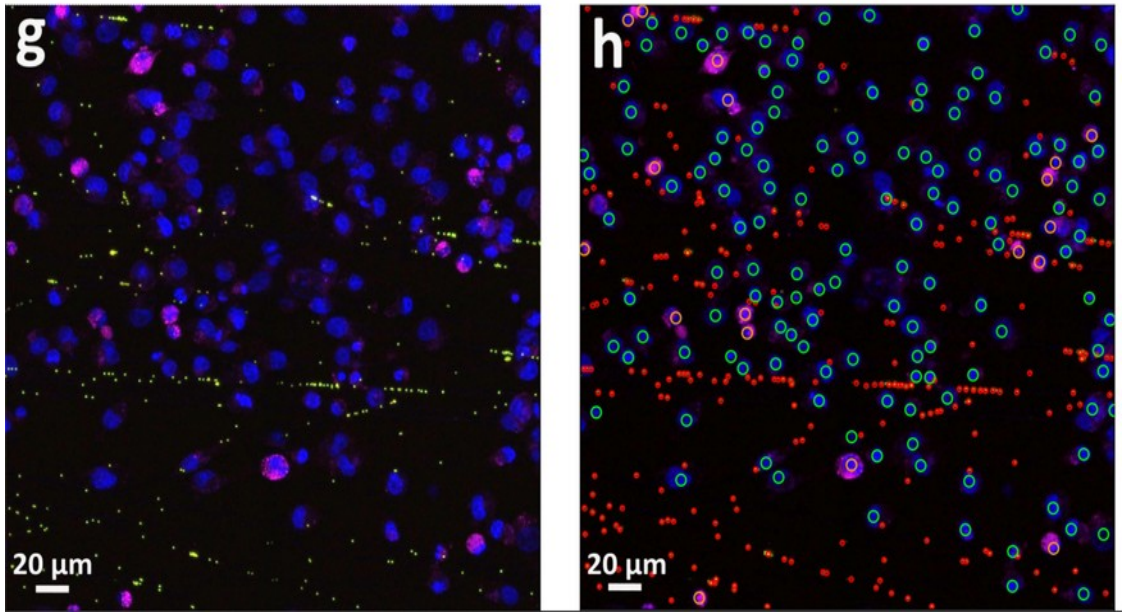


Inverse modeling of the Laplace equation

$$\nabla^2 c(\mathbf{r}) = 0$$

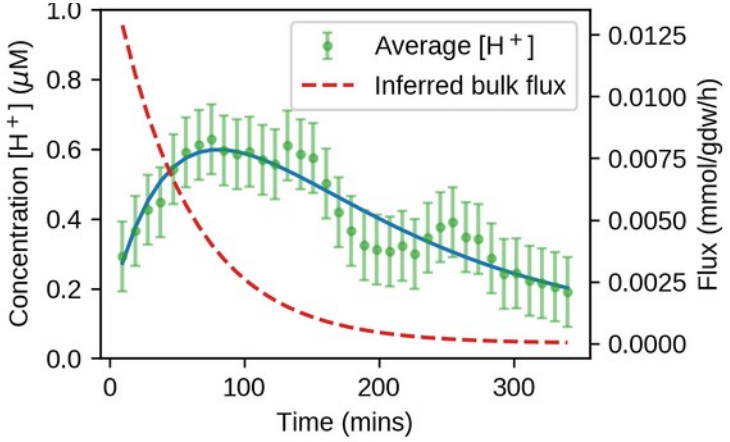
Given (noisy) measurements of C find the (complex) Boundary conditions

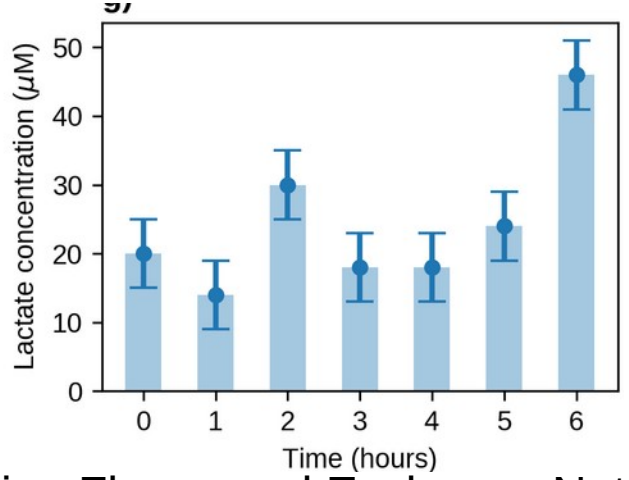
A single cell picture of tumor acidification (Warburg)



?

Non-trivial segmentation and tracking problem





Onesto et al. "Probing Single-Cell Fermentation Fluxes and Exchange Networks via pH-Sensing Hybrid Nanofibers." ACS nano (2022).

Some equations..

$$c(\mathbf{r}) = \sum_{i=1}^{N} \frac{u_i}{D|\mathbf{r} - \mathbf{r}_i|} + U \int_B \frac{ds}{D|\mathbf{r} - \mathbf{r}(s)|}$$

Multipolar expansion truncated to the first term

$$\chi^2_L(\mathbf{u}) = \sum_{\mu} \frac{(c_\mu - \sum_i A_{\mu i} u_i)^2}{2c_\mu^2 \sigma_\mu^2}$$

Find parameters from maximum likelihood
+ sampling the full posterior
(to estimate the errors!)

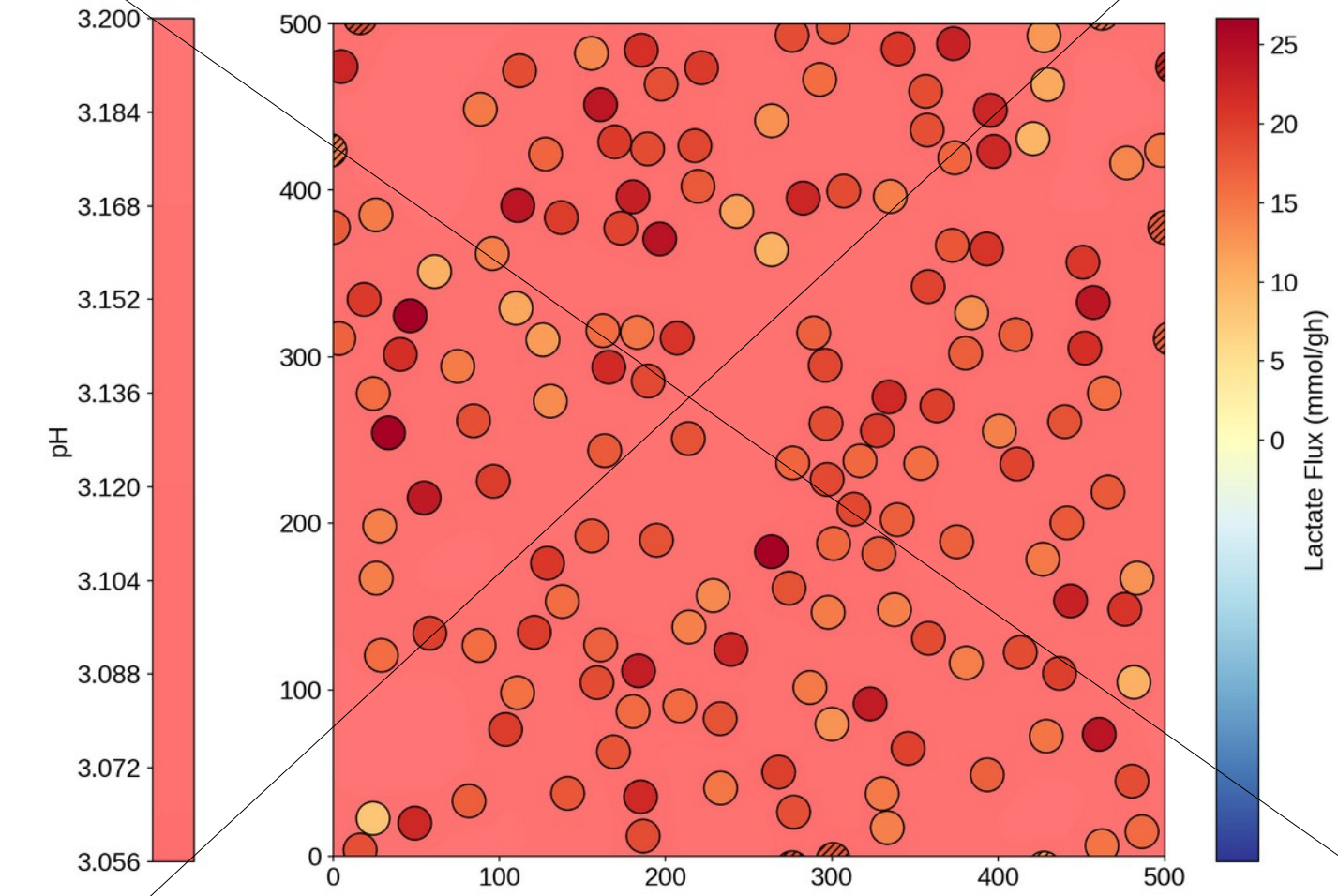
$$u_j^* = \sum_i (B_{ij})^{-1} b_i$$

$$b_i = \sum_{\mu} \frac{A_{\mu i}}{c_\mu \sigma_\mu^2}$$

$$B_{ij} = \sum_{\mu} \frac{A_{\mu i} A_{\mu j}}{c_\mu^2 \sigma_\mu^2}.$$

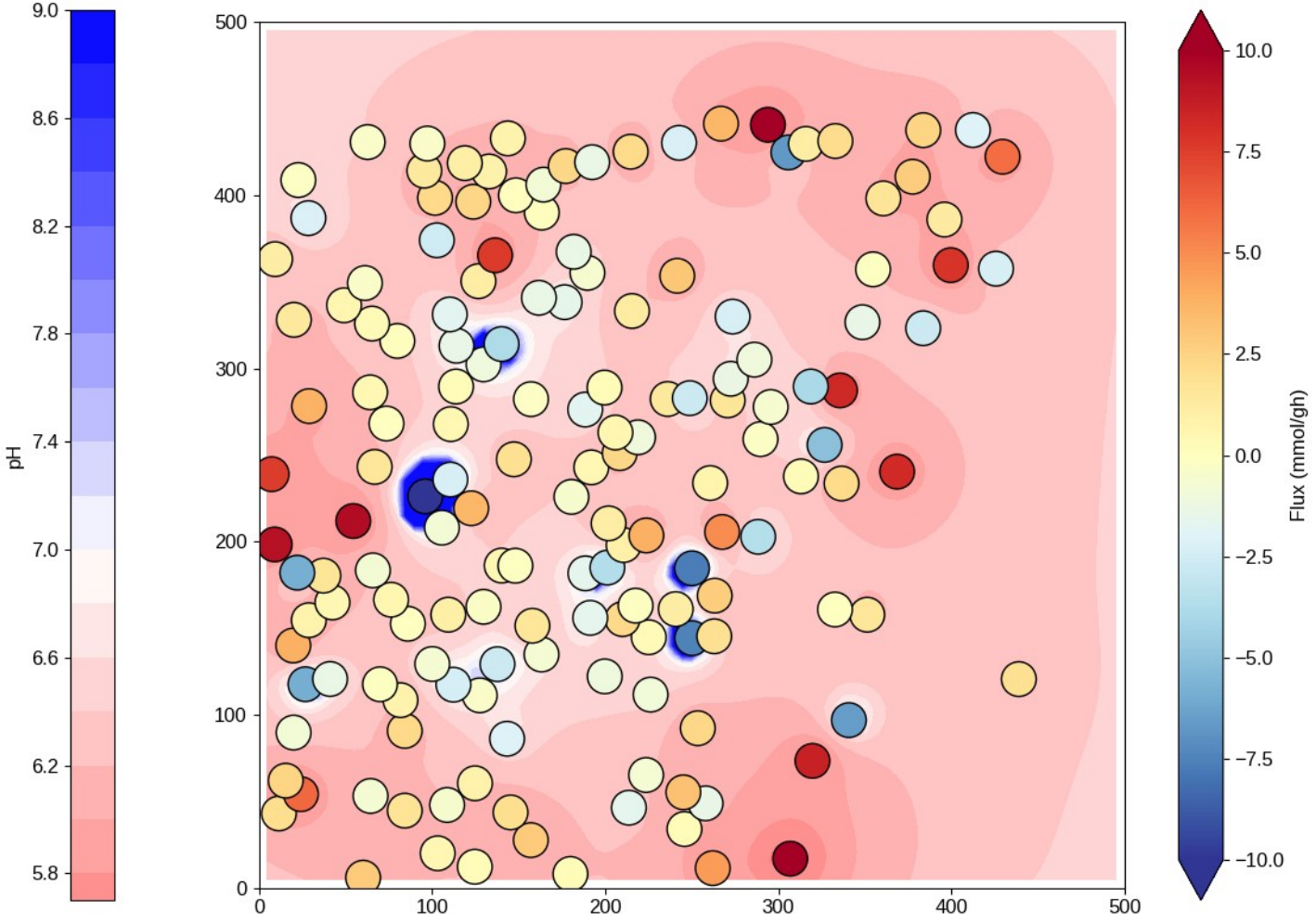
Many assumptions:
stationarity, spherical cows, etc

We do not obtain this!



FBA Simulations of cells maximizing growth/ATP

We obtain this!



Tumor acidification as a spillover from an unbalanced exchange network

Summary & references

you can't do inference – or data compression – without making assumptions.

MacKay, D. J. (2003). Information theory, inference and learning algorithms.

Tanouchi, Y., Pai, A., Park, H., Huang, S., Buchler, N. E., & You, L. (2017). Long-term growth data of Escherichia coli at a single-cell level. Scientific data, 4(1), 1-5.

Radu et al. "Carbon and nitrogen fixation and metabolite exchange in and between individual cells of Anabaena oscillarioides." The ISME journal 1.4 (2007): 354-360

Onesto et al. "Probing Single-Cell Fermentation Fluxes and Exchange Networks via pH-Sensing Hybrid Nanofibers." ACS nano (2022).