

# Structured Models for Cell Populations: Direct and Inverse Problems

Vinicius Albani<sup>2,a</sup> and Jorge P. Zubelli<sup>1,2,b</sup>

<sup>1</sup>IMPA, Estr. Dona Castorina, 110, 22460-320 Rio de Janeiro, Brazil

<sup>2</sup>Computational Science Center, University of Vienna, Oskar Morgenstern-Platz 1, 1090 Vienna, Austria

**Abstract.** Structured population models in biology lead to integro-differential equations that describe the evolution in time of the population density taking into account a given feature such as the age, the size, or the volume. These models possess interesting analytic properties and have been used extensively in a number of areas. In this article, we consider a size-structured model for cell division and revisit the question of determining the division (birth) rate from the measured stable size distribution of the population. We formulate such question as an inverse problem for an integro-differential equation posed on the half line. The focus is to compare from a computational view point the performance of different algorithms. Finally, we will discuss real data reconstructions with *E. coli* data.

## 1 Structured Population Models

Let us consider a population density  $n = n(t, x)$  where  $t \in \mathbb{R}_+$  is time and  $x \in \mathbb{R}_+$  is a parameter, e.g., the mass of a unicellular organism, the DNA content of a cell, the cell age, or its volume. A more specific example is the cell volume in a population of *E. coli*. This unicellular organism is commonly found in the lower intestine of warm-blooded organisms. The cells are typically rod-shaped and are about  $2\mu\text{m}$  long, with  $0.5\mu\text{m}$  diameter. Then, the average volume falls between  $0.6 - 0.7\mu\text{m}^3$ . These unicellular organisms are extensively studied *in vitro* and *in vivo*.

Let  $n(t, x)$  denote the population density of cells of "size"  $x$  at time  $t$ . Then,  $n$  satisfies the integro-differential equation

$$\partial_t n(t, x) + \partial_x [g(x)n(t, x)] = \int_0^\infty k(x, x')n(t, x')dx', \quad (1)$$

where  $g(x)$  is the microscopic growth rate of individuals at size  $x$ , and  $k(x, x')$  is the proportion of cells of size  $x'$  that divide into two cells, one of size  $x$ , and the other of size  $x' - x$ . Under this generality, the model is hard to calibrate and to make predictions.

In one spatial dimension, the amount of data needed to estimate the division kernel  $k(\cdot, \cdot)$  would be unrealistic in practical applications without further hypothesis. See [1–3] for variants of the division

---

<sup>a</sup>e-mail: vvla@impa.br

<sup>b</sup>e-mail: zubelli@impa.br

kernel estimation problem. Thus, in what follows, we shall consider the following simplified version of the model given by Equation (1):

$$\begin{cases} \partial_t n(t, x) + \partial_x n(t, x) + B(x)n(t, x) = 4B(2x)n(t, 2x), & x, t \geq 0, \\ n(t, x = 0) = 0, & t > 0, \quad n(0, x) = n^0(x) \geq 0, & x \geq 0. \end{cases} \quad (2)$$

We remark that the choice of  $g \equiv 1$  was made and that a natural alternative would be that of an affine function.

It is well-known [4–6] that there exist unique  $\lambda_0$  and  $N = N(x)$ , forming an eigenpair, such that, after a suitable time re-normalization, the solutions of (2) satisfy the limit

$$n(t, x)e^{-\lambda_0 t} \longrightarrow \rho N(x), \quad \text{as } t \rightarrow \infty, \quad (3)$$

where the limit is considered under weighted  $L^p$  topologies, and the pair  $(\lambda_0, N)$  is a solution for

$$\begin{cases} \partial_x N(x) + (\lambda_0 + B(x))N(x) = 4B(2x)N(2x), & x \geq 0, \\ N(x = 0) = 0, & N(x) > 0, \text{ for } x > 0, \int_0^\infty N(x)dx = 1. \end{cases} \quad (4)$$

Moreover, exponential rates have been proved to hold for fairly general rates  $B$ . See [6, 7]. Such  $N$  is the so-called stable-size distribution.

Let us assume that the birth rate  $B$  is a measurable function and such that, there exist positive constants  $B_m$ , and  $B_M$  satisfying

$$0 < B_m \leq B(x) \leq B_M < \infty. \quad (5)$$

Then, we can define the *direct problem* as, given a birth rate  $B$  satisfying such conditions, finding the eigenpair  $(\lambda_0, N)$  of Problem (4).

The corresponding *inverse problem* is to recover in a stable way the birth rate  $B$  from noisy data  $N$  and the rate  $\lambda_0$ . The first step is to study the *so-called* parameter-to-solution map. It can be shown that the following result holds.

**Theorem 1 (Perthame-Zubelli [8])** *Under assumption (5), the map  $B \mapsto (\lambda_0, N)$ , from  $L^\infty(\mathbb{R}_+)$  into  $[B_m, B_M] \times L^1 \cap L^\infty(\mathbb{R}_+)$  is continuous under the weak-\* topology of  $L^\infty(\mathbb{R}_+)$ , locally Lipschitz continuous under the strong topology of  $L^2(\mathbb{R}_+)$  into  $L^2(\mathbb{R}_+)$ , and of class  $C^1$  in  $L^2(\mathbb{R}_+)$ .*

## 2 The Inverse Problem

If the measurement  $N$  were very smooth, one could directly consider to solve for  $B$ , the cell-division equation (4) written with  $y = 2x$ ,

$$4B(y)N(y) = B(y/2)N(y/2) + \lambda_0 N(y/2) + 2\partial_y N(y/2), \quad y > 0. \quad (6)$$

This is a well-posed equation on  $B$  as long as  $N$  satisfies regularity properties such as  $\partial_y N(y/2)$  is in  $L^p$ , for some  $p \geq 1$ . However, this is not the case for reasonable practical data.

One possible regularization strategy is the following

$$\alpha \partial_y (B_\alpha N) + 4B_\alpha(y)N(y) = B_\alpha(y/2)N(y/2) + \lambda_0 N(y/2) + 2\partial_y N(y/2), \quad y > 0, \quad (7)$$

with  $B_\alpha N(0) = 0$ , where  $0 < \alpha < 1$  is a small parameter.

In [8], it was shown that Equation (7), when solved for  $B$  has desired properties such as strong stability with respect to perturbations. As a consequence, given a noise level  $\varepsilon$ , choosing  $\alpha = \alpha(\varepsilon)$ , we have  $\|B_{\varepsilon, \alpha} - B\|_{L^2(N_\varepsilon^2 dx)} \leq K \sqrt{\varepsilon}$ .

In this note we take a more pedestrian approach to tackle the inverse problem than that used in [8–10]. More specifically, we want to find minimizers for the following Tikhonov-type functional

$$\mathcal{F}(B) = \|N(B) - N^{\text{obs}}\|_{L^2(\mathbb{R}_+)}^2 + \alpha f_{B_0}(B), \quad (8)$$

with  $B \in L^2(\mathbb{R}_+)$ , satisfying (5), and  $\alpha = 0.05$ . The penalization functional used are:

Smoothing:  $f_{B_0}(B) = 0.01\|B - B_0\|_{L^2(\mathbb{R}_+)}^2 + \|\partial_x B\|_{L^2(\mathbb{R}_+)}^2$ , and  
 Kullback-Leibler:  $f_{B_0}(B) = \int_0^\infty B(x) \log(B_0(x)/B(x)) - (B_0(x) - B(x))dx$ ,

where with Kullback-Leibler,  $B_0(x) = 0.01$  with *E. coli* data and  $B(x) = 1$  with synthetic data, and with the smoothing functional,  $B_0$  equals to the division-rate obtained in [9] with *E. coli* data and also  $B(x) = 1$  with synthetic data.

### 3 Numerical Results

We now present some numerical examples of the calibration of the division-rate  $B$  from the stable distribution  $N$  using Tikhonov-type regularization and statistical inverse-problem techniques, with synthetic as well as real *E. coli* data. The latter data is the same used in [9]. The numerical solution of the PDE problem (4) is obtained with the numerical scheme presented in [10].

By describing the direct problem under a discrete Bayesian model [11], we build a posterior density, which is proportional to the exponential of the functional in Equation (8). The reconstructions are obtained with Maximum a Posteriori (MAP) and Conditional Mean (CM) point estimators. The MAP estimate as well as the Tikhonov-type reconstructions are obtained with the MATLAB’s function LSQNONLIN. To evaluate the CM, we use the Metropolis-Hastings algorithm, which is initialized with the MAP estimate and was generated with  $10^4$  samples. See [11]. With synthetic data, the minimization starts with  $B(x) = 1$ , and with *E. coli* data it starts with the division rates obtained in [9].

The synthetic data is generated with smooth and non-smooth division-rates  $B$ , in a mesh two times finer than the one used to solve the inverse problem. The data is corrupted by a multiplicative Gaussian noise with mean one and covariance matrix  $\delta^2 Id$ , where  $\delta = 0.05$ .

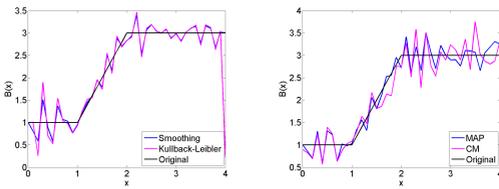


Figure 1: Reconstructions of a non-smooth  $B$  using Tikhonov regularization (left), and statistical techniques (right).

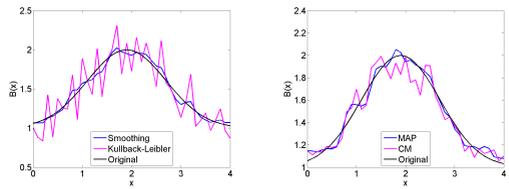


Figure 2: Reconstructions of a smooth  $B$  using Tikhonov regularization (left), and statistical techniques (right).

Figures 1 and 2 present the reconstructions of  $B$  with synthetic data. In both cases, reconstructions are compatible with the true  $B$ . Note that, for the non-smooth  $B$ , we also used a  $l^1$  prior density on the first derivative of  $B$  in the statistical approach. For the smooth case, instead, we used the  $l^2$  prior density on the second derivative of  $B$ .

Figures 3 presents the reconstructions of the division-rates using Tikhonov regularization and the statistical approach interpolated by cubic spline, both compared with  $B$  obtained in [9]. Figure 4

presents the corresponding stable density functions for both cases. Note that, we used the *E. coli* data without interpolation.

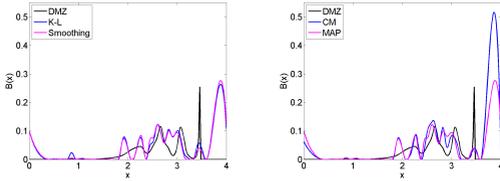


Figure 3: Reconstructions of  $B$  using Tikhonov-type (Smoothing and Kullback-Leibler) regularization (left), and statistical techniques (right).

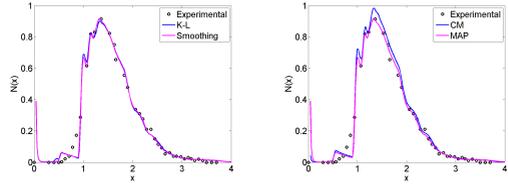


Figure 4: The density  $N$  corresponding to the reconstructions of  $B$  using Tikhonov-type (Smoothing and Kullback-Leibler) regularization (left), and statistical techniques (right).

The numerical tests were performed with a computer with processor Intel(R) Core(TM) i7-4510U CPU @ 2.00GHz, 8.00 GB of RAM, running Windows 7 Ultimate 64-bit Operating System (Service Pack 1) and MATLAB R2013b 64-bit. The computational time using *E. coli* data for each method were:  $3.05 \times 10^3$  seconds (Tikhonov with smoothing functional),  $3.67 \times 10^2$  seconds (Tikhonov with Kullback-Leibler),  $3.45 \times 10^3$  seconds (MAP) and  $7.65 \times 10^3$  seconds (CM). The normalized  $L^2$ -distance between  $N^{\text{obs}}$  and  $N(B)$  where  $B$  was obtained with the different techniques above were:  $7.98 \times 10^{-2}$  (Tikhonov with smoothing functional),  $8.44 \times 10^{-2}$  (Tikhonov with Kullback-Leibler),  $7.98 \times 10^{-2}$  (MAP),  $1.00 \times 10^{-1}$  (CM) and  $7.20 \times 10^{-2}$  (in the paper [9]).

	Nonsmooth Division-rate			Smooth Division-rate		
	Residual	Error	Time (sec)	Residual	Error	Time (sec)
Smoothing	$1.97 \times 10^{-2}$	$8.91 \times 10^{-2}$	6.19	$1.01 \times 10^{-2}$	$3.11 \times 10^{-2}$	8.77
K-L	$2.39 \times 10^{-2}$	$2.16 \times 10^{-1}$	76.39	$2.18 \times 10^{-2}$	$1.43 \times 10^{-1}$	17.20
MAP	$1.98 \times 10^{-2}$	$9.85 \times 10^{-2}$	5.15	$7.00 \times 10^{-3}$	$3.38 \times 10^{-2}$	8.66
CM	$5.86 \times 10^{-1}$	$1.34 \times 10^{-1}$	$1.09 \times 10^2$	$5.26 \times 10^{-1}$	$6.62 \times 10^{-2}$	$1.19 \times 10^2$

Table 1: Results with synthetic data. Note that,  $Residual = \|N(B) - N^{\text{obs}}\|/\|N^{\text{obs}}\|$ ,  $Error = \|B - B^{\text{true}}\|/\|B^{\text{true}}\|$ , Time is the computational time, *Smoothing* stands for Tikhonov with the smoothing functional, *K-L* for Kullback-Leibler Tikhonov.

Visually, our reconstructions present a similar shape as the one obtained in [9]. Moreover, the corresponding stable size-distributions presented a good adherence to the experimental data.

## 4 Discussion

The solution of the inverse problem relies heavily on the existence of a limiting behavior for population structure. Indeed, by looking at the long time behavior, we do not need to assume a particular cell distribution. This limits the applicability of the methodology developed in [8] and justifies the investigation presented in this work since it is a more general technique (with its corresponding advantages and disadvantages). Thus, we made a point of solving the calibration problem of the equal division rate by different techniques including Tikhonov-type regularization and Statistical methods.

The reconstructions obtained with Tikhonov-type regularization and the MAP estimator presented accurate results with smaller computational time compared to the CM estimate, since the latter explores the posterior density, in a sample space with a large number of dimensions. Among the advantages of the Tikhonov regularization approach we cite its stability and its versatility.

## References

- [1] M. Doumic, M. Hoffmann, P. Reynaud-Bouret, V. Rivoirard, *SIAM J. Numer. Anal.* **50**, 925 (2012)
- [2] M. Doumic, L.M. Tine, *J. Math. Biol.* **67**, 69 (2013)
- [3] T. Bourgeron, M. Doumic, M. Escobedo, *Inverse Problems* **30**, 025007, 28 (2014)
- [4] P. Michel, *Math. Models Methods Appl. Sci.* **16**, 1125 (2006)
- [5] P. Michel, S. Mischler, B. Perthame, *J. Math. Pures Appl* **84**, 1235 (2005)
- [6] B. Perthame, L. Ryzhik, *J. Differential Equations* **210**, 155 (2005)
- [7] B. Perthame, *Transport Equations in Biology* (Birkhäuser Verlag, 2007)
- [8] B. Perthame, J.P. Zubelli, *Inverse Problems* **23**, 1037 (2007)
- [9] M. Doumic, P. Maia, J. Zubelli, *Acta Biotheoretica* **58**, 405 (2010)
- [10] M. Doumic, B. Perthame, J.P. Zubelli, *Inverse Problems* **25**, 045008 (2009)
- [11] E. Somersalo, J. Kapiro, *Statistical and Computational Inverse Problems*, Vol. 160 of *Applied Mathematical Sciences* (Springer, 2004)