

On the Calibration of a Size-Structured Population Model from Experimental Data

Marie Doumic, Pedro Maia and Jorge P. Zubelli

Abstract. The aim of this work is twofold. First, we survey the techniques developed in [1, 2] to reconstruct the division (birth) rate from the cell volume distribution data in certain structured population models. Secondly, we implement such techniques on experimental cell volume distributions available in the literature so as to validate the theoretical and numerical results. As a proof of concept, we use the data reported in the *classical* work of Kubitschek [3] concerning *Escherichia coli in vitro* experiments measured by means of a Coulter transducer-multichannel analyzer system (Coulter Electronics, Inc., Hialeah, Fla, USA.) Despite the rather old measurement technology, the reconstructed division rates still display potentially useful biological features.

Mathematics Subject Classification (2000). Primary 35R30; Secondary 35F15 92D25.

Keywords. structured populations, inverse problems, experimental data, biological applications.

1. Introduction

Structured growth models have been the subject of much attention for the last decades [4, 5]. Amongst such models, equal division ones play a crucial role for their relative simplicity. In equal division models, each cell of volume $2x$ subdivides into two cells of size x according to a certain rate $B(2x)$. In order to understand better such models and related microbiological mechanisms, it would be important to calibrate the birth rate $B(\cdot)$ by means of observed cell densities $n(t, x)$, where t is the time parameter. Thus, in this article, we consider the problem of calibrating from measured data the following size-structured model for cell division:

$$\begin{cases} \frac{\partial}{\partial t}n(t, x) + \frac{\partial}{\partial x}[g(x)n(t, x)] + B(x)n(t, x) = 4B(2x)n(t, 2x), & x, t \geq 0, \\ g(x = 0)n(t, x = 0) = 0, t > 0, \\ n(0, x) = n^0(x) \geq 0. \end{cases} \quad (1.1)$$

The function g describes the microscopic growth behaviour of individual cells.

Usually two different models have been considered. The first one consists of constant g , which without loss of generality we take to be one. This is usually called *linear growth model* and leads to characteristic lines of the form $x(t) = x_0 + t$. The second one is $g(x) = \kappa x$, and leads to characteristic lines of the form $x(t) = x_0 \exp(\kappa t)$ where κ is a constant. It is called *exponential growth model*. Obviously both of them subsume some kind of unlimited growth of individual organisms. There has been substantial amount of debate in the literature concerning the correct way of modeling the function g . See for example [6, 7, 8] and references therein.

Two macroscopic quantities of biological interest are naturally computed from the model (1.1). The total cell quantity $\mathcal{N}(t) = \int_0^\infty n(t, x)dx$ and the total biomass $\mathcal{M}(t) = \int_0^\infty xn(t, x)dx$. Integrating equation (1.1) yields $\dot{\mathcal{N}}(t) = \int_0^\infty B(x)n(t, x)dx$. This means that number of cells increase only by division. Integrating the identity

function times (1.1) yields $\dot{\mathcal{M}}(t) = \int_0^\infty g(x)n(t,x)dx$. This means that the biomass increases only by nutrient uptake.

Under reasonable assumptions on B , the long time behavior of $n(t,x)$ is of the form $n(t,x) \approx m_0 N(x)e^{\lambda_0 t}$. More precisely, it is proved in [9] (see also [10, 11] for the case $g(x) = 1$) that, under fairly general conditions on the coefficients, there is a unique solution (N, λ_0) to the following eigenvalue problem

$$\begin{cases} \frac{\partial}{\partial x}(gN) + (\lambda_0 + B(x))N = 4B(2x)N(2x), & x > 0, \\ g(x=0)N(x=0) = 0, \\ N(x) > 0 \text{ for } x > 0, & \int_0^\infty N(x)dx = 1, \end{cases} \quad (1.2)$$

where $\lambda_0 > 0$ and $x^\alpha gN \in L^\infty \cap L^1$ for all $\alpha \geq 0$. Furthermore, it was also shown in [9, 10, 12] that

$$n(t,x)e^{-\lambda_0 t} \xrightarrow[t \rightarrow \infty]{} m_0 N(x), \quad \text{in } L^1(\mathbb{R}_+, \phi(x)dx), \quad (1.3)$$

where the weight ϕ is the unique solution to the adjoint problem to Equation (1.2). In other words, λ_0 is the growth rate of such a system and is usually called ‘‘Malthus parameter’’ in population biology. By integrating the equation and integrating it against the weight x , we also know that λ_0 is related to N by the relation

$$\lambda_0 = \int_0^\infty g(x)N(x)dx \Big/ \int_0^\infty xN(x)dx. \quad (1.4)$$

The plan for the article is the following: In Section 2, we discuss some biological preliminaries of the model under consideration. In Section 3 we describe the theoretical aspects of the inverse problem under consideration. This will give us the underlying theoretical basis for the calibration techniques. In Section 4, we present the calibration results. In Section 5 we conclude with a discussion of the results and suggestions for further research.

2. Biological Preliminaries

In many situations it is important to understand when an organism is mature and ready to reproduce. The concept of size at maturation was developed as a theoretical tool to express how individuals deal with environmental variability within growth rates. In addition, individuals may have unequal fitness and might respond differently to the same environmental conditions. In this context, we can interpret the bacterial-growth experiments of Kubitschek [3] as a very special situation: (1) the cells grow in a chemostat and all individuals have the same amount of nutrients, in all generations of the population. (2) the cells are clones from each other and have the same fitness, responding equally to the same environmental effects. (3) Kubitschek [3] also did a synchronic experiment, using temperature control techniques, so that all the cells would divide at the same time.

One might argue that in this situation, with no difference in the fitness of the individuals and in the environmental conditions, there should be almost no variability in the preferred size of division. However, the reconstructions of $B(x)$ detect an intrinsic variability that is inherent from the division process. This justifies the importance of the reconstructed division rate. Moreover, *E. coli* is commonly used as a model for more complex organisms and the present work can be considered a first step toward the modelling of more complex cell division phenomena.

We now address the biological suitability of the mathematical model. The specific size-structured model under consideration in this work is suitable to microorganisms with the following characteristics:

- (a) All the cells of a given class of microorganisms grow according to some deterministic rule.
- (b) All cells, after attaining a sufficient size, divide into two, and only two, identical daughter cells.
- (c) The cell’s age (generation) does not influence the growth law and the preferable size of division.
- (d) The total number of individuals of the population increases exponentially.
- (e) There is a negligible number of deaths during normal growth.

We will now argue that the statements above are reasonable for the bacteria *E. Coli* growing under certain conditions:

Assumption (a) is in the spirit of [6], where the author argues for the existence of a growth law of cells that can be found and understood.

Perhaps the most delicate assumption is (b). The cell division process may result in daughters with unequal sizes. Equal division does not occur surely for all individuals of a microbiological population, but we may quantify how much they deviate from a perfect division by a statistical description, as explained in the work of Koch [7].

The bacteria *E. Coli* is known for its quite small variation in its subdivisions [7, 13, 14]. It is usually assumed (see [7]) that the fluctuations in the critical size-at-division of individual cells is random and not correlated with other cell cycle events in the current cell generation or in earlier generations. This justifies (c) for the bacteria *E. Coli*.

It is well known that *E. coli* cells in normal growth conditions achieves an exponential phase [15] thus confirming claim (d) in agreement with Equation 1.3.

Finally, for statement (e), cell death during normal growth is considered by Koch [7] as a minor source of variability. It is true that some cells may be dead or moribund and may distort the distribution if they are clustered at some cell size range. But these effects can be neglected if there is plenty of nutrient, which occurs in the exponential phase of the bacterial growth. This hypothesis must be carefully revised when the population is submitted to starvation and other stressful types of experiments. The present model would not apply for instance, to the budding yeast *S. cerevisiae* [16]. However, it could easily be adapted to take into account a known death rate.

3. Methodology: Inverse problems and theoretical results

Based on the limiting behavior of the distribution $n(t, x)$ as t grows, the problem of calibration of B assuming $g \equiv 1$ reduces to the following: How can we estimate the division rate B from the knowledge of the steady dynamics N and λ_0 ? This corresponds to finding B a solution to

$$4B(2x)N(2x) - B(x)N(x) = L(x) := \frac{\partial}{\partial x}N(x) + \lambda_0 N(x), \quad x > 0, \quad (3.1)$$

assuming that (N, λ_0) is known, or, thanks to (1.4), that N is known. However, in practical applications we have only an *approximate* knowledge of (N, λ_0) , given by noisy data $(N_\epsilon, \lambda_\epsilon)$, with $N_\epsilon \in L^2_+(\mathbb{R}_+)$ for instance. This means that we have no way of controlling $\frac{\partial}{\partial x}N_\epsilon$, so we cannot control the precision of a solution B_ϵ to problem (3.1) when a perturbed N_ϵ replaces N . Thus, in the context of noisy data, the inverse problem under consideration is ill-posed and regularization is needed.

As alluded above, N_ϵ denotes the measured stable distribution. The subindex refers to the difference in the appropriate norm to the value N associated to the true B of the model. The precise value of ϵ is obviously unknown, but it depends on a number of factors such as accuracy of the measurements, quality of the model and proximity of the interpolation leading from a finite amount of data to the function with domain \mathbb{R}_+ .

Following [2] three methods will be used, namely: quasi-reversibility, filtering, and a hybrid of these.

Regularization by Quasi-Reversibility. To regularize the problem by the quasi-reversibility proposed in [1] we work with

$$\begin{cases} \alpha \frac{\partial}{\partial y}(B_{\epsilon, \alpha} N_\epsilon) + 4B_{\epsilon, \alpha}(y)N_\epsilon(y) = B_{\epsilon, \alpha}(\frac{y}{2})N_\epsilon(\frac{y}{2}) + \lambda_{\epsilon, \alpha}N_\epsilon(\frac{y}{2}) + 2\frac{\partial}{\partial y}\left(N_\epsilon(\frac{y}{2})\right), & y > 0, \\ (B_{\epsilon, \alpha} N_\epsilon)(0) = 0. \end{cases} \quad (3.2)$$

According to the eigenvalue theory [2], we have to choose

$$\lambda_{\epsilon, \alpha} = \left(\int_0^\infty N_\epsilon dx \right) / \left(\int_0^\infty x N_\epsilon dx + \frac{\alpha}{4} \int_0^\infty N_\epsilon dx \right). \quad (3.3)$$

The method thus consists in solving numerically the Equation (3.2). This was analyzed in detail in Section 3 of [2].

Regularization by Filtering. In this method we basically filter the data so as to smooth out the noisy component of the measurements. For $\alpha > 0$, we introduce

$$\rho_\alpha(x) = \frac{1}{\alpha} \rho\left(\frac{x}{\alpha}\right), \quad \rho \in C_c^\infty(\mathbb{R}), \quad \int_0^\infty \rho(x) dx = 1, \quad \rho \geq 0, \quad \text{Supp}(\rho) \subset [0, 1],$$

and we replace in (3.1) the term $\frac{\partial}{\partial x} N_\epsilon + \lambda_0 N_\epsilon$ by

$$\left(\frac{\partial}{\partial x} N_\epsilon + \lambda_{\epsilon, \alpha} N_\epsilon\right) * \rho_\alpha(x) = N_\epsilon * \left(\frac{\partial}{\partial x} \rho_\alpha + \lambda_{\epsilon, \alpha} \rho_\alpha\right)(x) = \int_0^\infty N_\epsilon(x') \left(\frac{\partial}{\partial x} \rho_\alpha + \lambda_{\epsilon, \alpha} \rho_\alpha\right)(x - x') dx'.$$

We set $N_{\epsilon, \alpha} := N_\epsilon * \rho_\alpha$. In this way, we obtain a smooth term in $L^2(\mathbb{R}_+)$. Furthermore, $N_{\epsilon, \alpha}$ converges to N_ϵ in $L^2(\mathbb{R}_+)$ when α tends to zero. We now have to find $B_{\epsilon, \alpha}$ solution of

$$4B_{\epsilon, \alpha}(2x)N_{\epsilon, \alpha}(2x) - B_{\epsilon, \alpha}(x)N_{\epsilon, \alpha}(x) = \frac{\partial}{\partial x} N_{\epsilon, \alpha} + \lambda_{\epsilon, \alpha} N_{\epsilon, \alpha}(x), \quad x \geq 0. \quad (3.4)$$

Once again, the numerical aspects of the solution of such problem are addressed in Section 3 of [2].

Convergence and the Hybrid Method. In [1, 2] estimates are obtained that guarantee that for the choice of $\alpha = \mathcal{O}(\sqrt{\epsilon})$ the computed value of $B_{\epsilon, \alpha}$ converges to B in the appropriate norms, with a speed of convergence in the order of $\sqrt{\epsilon}$.

In a number of numerical experiments with simulated data we found it convenient to use a hybrid method where we perform filtering and apply the quasi-reversibility technique. In the case of $g(x) = \kappa x$, at present we do not have the theoretical and numerical results of [1, 2]. However, a simple adaptation of our codes led to implementations that yield comparable results. We shall report on both cases.

4. Reconstruction Results

In this section we report the results of the reconstructions in *two* sets of data from [3]. We actually performed the numerical inversion in four sets, but we chose two of them where the experimental set up seemed to fit better our model. As expected with real data, a number of difficulties have to be overcome. To cite a few: a) The data is obtained on a rather scarce set of discrete points. b) It clearly displays a substantial amount of noise. Yet, such noise is very hard to quantify. c) We are not sure whether the conditions for the validity of the model were indeed satisfied, and whether enough time elapsed so as to guarantee a good approximation to stationary limiting distribution.

We performed the reconstructions both for the linear and the exponential models. For each of the data sets, we followed the following steps (both for the linear model as well as for the exponential model):

1. Transcribed the reported measurements from [3] to an array.
2. Completed the boundary data for the volume x close to zero and large enough by setting the boundary conditions to zero (as required by the theory).
3. Interpolated to a uniform grid on the region under consideration using splines.
4. Deduced from (1.4) and from the knowledge of both N_ϵ and λ_0 (given in the article [3] by the doubling times $T_0 = \frac{\text{Log}(2)}{\lambda_0}$) the constant value $g_0 = \lambda_0 \frac{\int x N dx}{\int N dx}$ in the definition of $g(x) = g_0$ (linear growth) or $g(x) = \lambda_0 x$ (exponential growth).
5. Performed a search for a good regularization parameter α that would give a sufficiently small residual in the quasi-reversibility method of [1] and in the hybrid methods. The result was also validated by other heuristical criteria such as the use of the minimum of the ratio $\frac{\text{residual}}{\sqrt{\alpha}}$ [17], see Figure 3.
6. Studied the behavior of the solutions to the inverse problem by varying the regularization parameter.
7. Studied the consistency of the computed limiting distribution for the reconstructed B and the input data by computing the direct problem for some variants of the reconstructed B .

We performed a number of extensive tests and examples, some of which that were not reported here could be found in [18]. In all that follows we take the filter parameter to be $\alpha = 0.0001$. We remark that this choice is very arbitrary, but it is based on the standard deviation of the volumes once we assume the instrumental resolution of the Coulton counter has a standard deviation of the order $\sigma = 0.03\mu\text{m}$ in the diameter [3]. Thus, the standard deviation of the volume, assuming a normal distribution, is of the order of $\sigma_V = (\pi/6)\sqrt{15}\sigma^3 \approx 10^{-4}(\mu\text{m})^3$.

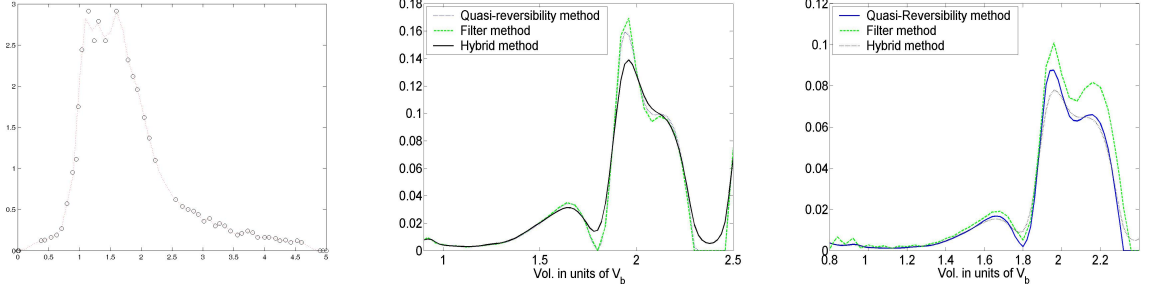


FIGURE 1. Input data from [3] for doubling time of 20 min (left), reconstructed B for the linear model (center), reconstructed B for the exponential model (right). $V_b = 1.36(\mu\text{m})^3$. The value of $\alpha = 0.1$ (larger α give lower peaks).

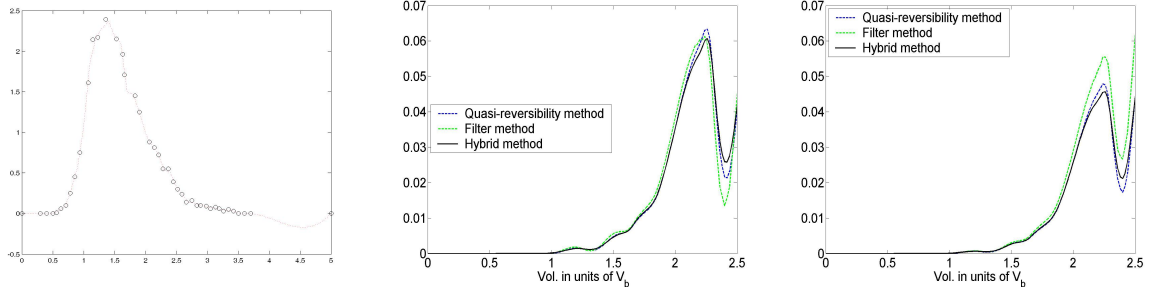


FIGURE 2. Input data from [3] for doubling time of 54 min (left), reconstructed B for the linear model, reconstructed B for the exponential model (right). $V_b = 0.61(\mu\text{m})^3$. The value of $\alpha = 0.2$ for the linear case, $\alpha = 0.1$ for the exponential case.

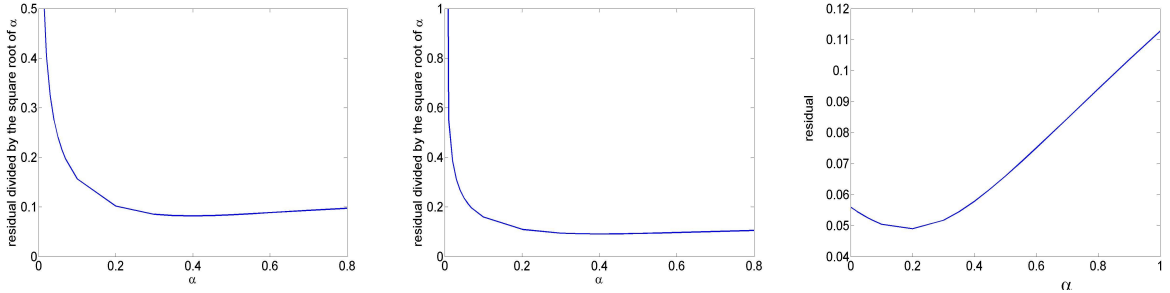


FIGURE 3. Example of criteria for the choice of α . Left: for a doubling time of 20 min, curve of $\frac{\text{residual}}{\sqrt{\alpha}}$ with respect to α , linear case. The minimum is attained for $\alpha = 0.4$ but the curve is very flat for $\alpha \geq 0.2$. Center: same curve for doubling time of 54 min, linear case, the minimum is still for $\alpha = 0.4$. Right: as in Center, doubling time of 54 min and linear case, the curve of the residual exhibits a minimum for $\alpha = 0.2$.

5. Discussion and Perspectives

Most of the reconstructed birth rates we have obtained seem to display a global maximum close to twice the average cell volume. Furthermore, in all cases we have at least a local maximum in that neighborhood. The precise location, however, varies substantially with the doubling time of the *E. coli* experiments, which in turn is a function of the experimental conditions. The peak locations also vary a little with the inversion methods and regularization parameters.

The presence of doublets and triplets in the experimental data may give rise to artifacts for values of x higher than 3. Indeed, it is not clear, due to the highly nonlinear characteristic of the reconstruction whether an incorrect increase in the measurements near a certain x_0 would lead to an increase in the reconstructed value of $B(x_0)$. This is a point that might be investigated further both theoretically and numerically.

We cannot infer any conclusion on the discussion of whether individual cells undergo exponential or linear volume growth. Still, a superficial observation of the results seems to indicate that exponential growth gives smaller residuals. This calls for a more extensive testing on better quality experimental data.

The optimal choice of the regularization parameter α should be investigated further. In this work we took a rather naive choice due to lack of a good estimates on the experimental noise. Indeed, we tried to locate the optimal choice by looking at the curve of the residuals as a function of α . A natural extension of this work would be a more detailed investigation of *a priori* and *a posteriori* choices for α , such as the quasi-optimality criterion [19] or the L-curve method [20].

References

- [1] Benoît Perthame and Jorge P. Zubelli. On the inverse problem for a size-structured population model. *Inverse Problems*, 23(3):1037–1052, 2007.
- [2] Marie Doumic, Benoît Perthame, and Jorge P. Zubelli. Numerical solution of an inverse problem in size-structured population dynamics. *Inverse Problems*, 25, 2009.
- [3] H. E. Kubitschek. Growth during the bacterial cell cycle: Analysis of cell size distribution. *Biophysical Journal*, 9(6):792–809, 1969.
- [4] J. A. J. Metz and O. Diekmann. Formulating models for structured populations. In *The dynamics of physiologically structured populations (Amsterdam, 1983)*, volume 68 of *Lecture Notes in Biomath.*, pages 78–135. Springer, Berlin, 1986.
- [5] Benoît Perthame. Transport equations arising in biology. In *Frontiers in Mathematics*, Frontiers in Mathematics. Birkhauser, 2007.
- [6] Stephen Cooper. Distinguishing between linear and exponential cell growth during the division cycle: Single-cell studies, cell-culture studies, and the object of cell-cycle research. *Theoretical Biology and Medical Modelling*, 3:10, 2006.
- [7] Arthur L. Koch. Biomass growth rate during the prokaryote cell cycle. *Critical Reviews in Microbiology*, 19(1):17–42, 1993.
- [8] J. Mitchison. Single cell studies of the cell cycle and some models. *Theoretical Biology and Medical Modelling*, 2(1):4, 2005.
- [9] M. Doumic and P. Gabriel. Eigenelements of a general aggregation-fragmentation model. submitted, 2009.
- [10] Benoît Perthame and Lenya Ryzhik. Exponential decay for the fragmentation or cell-division equation. *J. Differential Equations*, 210(1):155–177, 2005.
- [11] Philippe Michel. Existence of a solution to the cell division eigenproblem. *Model. Math. Meth. Appl. Sci.*, 16(suppl. issue 1):1125–1153, 2006.
- [12] Philippe Michel, Stéphane Mischler, and Benoît Perthame. General relative entropy inequality: an illustration on growth models. *J. Math. Pures Appl. (9)*, 84(9):1235–1260, 2005.

- [13] F. Trueba. *A Morphometric Analysis of Escherichia Coli and other Rod-Shaped Bacteria*. PhD thesis, University of Amsterdam, 1981.
- [14] R. J. Harvey, A. G. Marr, and P.R. Painter. Kinetics of growth of individual cells of *Escherichia coli* and *Azobacter agilis*. *Journal of Bacteriology*, 93:605–617, 1967.
- [15] Klein Prescott, Harley. *Microbiology*. McGraw-Hill, 2002.
- [16] C. Hatzis and D. Porro. Morphologically-structured models of growing budding yeast populations. *Journal of Biotechnology*, 124:420–438, 2006.
- [17] Heinz W. Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.
- [18] Pedro Maia. Tópicos em teoria da homogeneização e equações de populações estruturadas. Master’s thesis, UFRJ, Brazil, 2009.
- [19] Frank Bauer and Stefan Kindermann. The quasi-optimality criterion for classical inverse problems. *Inverse Problems*, 24, 2008.
- [20] Johann Baumeister and Antonio Leitão. *Topics in inverse problems*. Publicações Matemáticas do IMPA. [IMPA Mathematical Publications]. Instituto Nacional de Matemática Pura e Aplicada (IMPA), Rio de Janeiro, 2005. 25° Colóquio Brasileiro de Matemática. [25th Brazilian Mathematics Colloquium].

Acknowledgments

The authors were supported by the CNPq-INRIA agreement INVEBIO. JPZ was supported by CNPq under grants 302161/2003-1 and 474085/2003-1. JPZ is thankful to the RICAM special semester and to the International Cooperation Agreement Brazil-France. A substantial part of this work was developed during a 3-month international internship of PM at INRIA Rocquencourt during the Spring of 2008 and supported by INRIA. PM and MD thank S. Boatto (UFRJ) for facilitating this visit and for helpful discussions.

Marie Doumic

INRIA Rocquencourt, projet BANG, Domaine de Voluceau, BP 105, 78153 Rocquencourt, France
e-mail: marie.doumic@inria.fr

Pedro Maia

UFRJ, Cidade Universitaria, Caixa-Postal: 68530, Rio de Janeiro, RJ 21945-970, Brazil.
e-mail: pd_maia@hotmail.com

Jorge P. Zubelli

IMPA, Est. D. Castorina 110, Rio de Janeiro, RJ 22460-320, Brazil.
e-mail: zubelli@impa.br