This is a preliminary version of the article published as:
A. Carta, M. Chaves and J.-L. Gouzé, in D. Gilbert and M. Heiner Eds, Springer-Verlag Heidelberg, LNCS 7605 ("CMBS 2012"), pp. 107-126, 2012.

A simple model to control growth rate of synthetic E. coli during the exponential phase: model analysis and parameter estimation

Alfonso Carta Madalena Chaves Jean-Luc Gouzé

BIOCORE, INRIA, 2004 Route des Lucioles, BP 93, 06902 Sophia Antipolis, France
{alfonso.carta,madalena.chaves,jean-luc.gouze}@inria.fr

Abstract

We develop and analyze a model of a minimal synthetic gene circuit, that describes part of the gene expression machinery in *Escherichia coli*, and enables the control of the growth rate of the cells during the exponential phase. This model is a piecewise non-linear system with two variables (the concentrations of two gene products) and an input (an inducer). We study the qualitative dynamics of the model and the bifurcation diagram with respect to the input. Moreover, an analytic expression of the growth rate during the exponential phase as function of the input is derived. A relevant problem is that of identifiability of the parameters of this expression supposing noisy measurements of exponential growth rate. We present such an identifiability study that we validate *in silico* with synthetic measurements.

1 Introduction

Synthetic biology has nearly emerged as a new engineering discipline. The goal of synthetic biology [1, 19, 24] is to develop and apply engineering tools to control cellular behavior—constructing novel biological circuits in the cell—to perform new and desired functions.

Most recent synthetic designs have focused on the cell transcription machinery, which includes the genes to be expressed, their promoters, RNA polymerase and transcription factors, all serving as potential engineering components. Indeed, synthetic bio-molecular circuits are typically fabricated in *Escherichia coli* (*E. coli*), by cutting and pasting together coding regions and promoters (natural and synthetic) according to designed structures and specific purposes ([11, 14, 31]).

Along these lines, synthetic biology ultimately aims at developing synthetic bio-molecular circuitry that may help in producing bio-pharmaceuticals, bio-films, bio-fuels, novel cancer treatments and novel bio-materials (see [19] for a review on synthetic biology applications).

In the present work we focus on the gene expression machinery of the bacterium *Escherichia coli*, with the aim of controlling the growth rate of the cells. *E. coli* is a model organism that is easy to manipulate and much knowledge is available about its regulatory networks.

In the presence of a carbon source—such as glucose—*E. coli* grows in an exponential manner until it exhausts the nutrient sources, and then enters a stationary phase with practically zero growth [23]. The wild-type (namely the genetically unmodified) bacteria grow at different rates in the presence of carbon sources of different types [22]. Notably, glucose is the preferred substrate because it leads to a higher growth rate in wild type. Our control objective is to force the bacterium to significantly modify its response to glucose so as to tune the cells' growth rates. To this end, we take into account the recent applications of synthetic biology which allow us to fabricate engineered promoters which in turn can be externally controlled by inducers [18].

Notably, we will study an open loop configuration of a bi-dimensional model of a mutant *E. coli* inspired by the experiments in [30]. The two basic variables of our model, which describe the gene expression machinery that is responsible for bacterial growth are (see Fig.1):

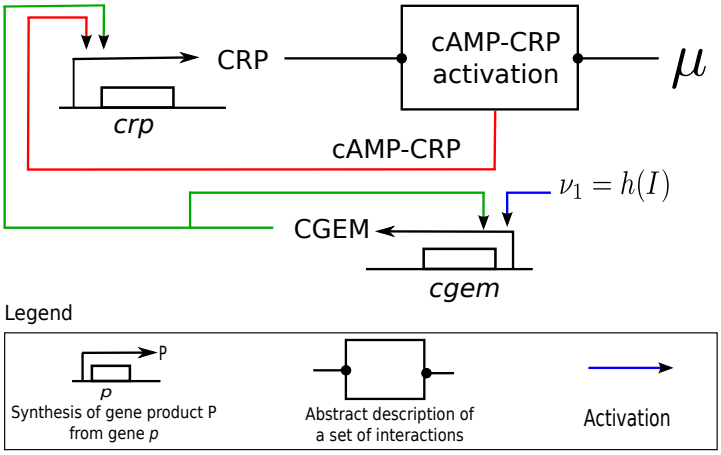

Figure 1: Regulatory network of the open-loop model in the *mutant E. coli*. The model consists of genes *crp* and synthetic-*cgem* (modified promoter of a component of the gene expression machinery (CGEM) in *E. coli*). The synthetic-*cgem* promoter is positively regulated by the inducer I—according to the input function $\nu_1 = h(I)$—and CGEM. CGEM, being responsible for the bacterial gene expression, positively regulates *crp* gene too. Moreover, *crp* transcription is induced by cAMP-CRP, a metabolite whose formation relies on CRP protein abundance and low level of bacterial growth rate μ.

1. the concentration of a *Component of the Gene Expression Machinery* (CGEM), proteins responsible for global growth (ribosomes and RNA polymerase). Without this CGEM, the bacteria cannot produce any proteins and thus cannot grow.

2. the concentration of CRP, a protein involved in the formation of the complex cAMP-CRP whose level positively correlates with less preferred carbon sources and slower growth [2].

We will assume that an engineered inducible-promoter is used to express the CGEM. Moreover it is assumed that the mutant CGEM activates its own expression. The number and location of equilibria can thus be controlled by means of an input control function of the inducer and, in particular, there can be regions of bi-stability, as observed in [30].

The type of growth rate control we present—which directly acts upon the GEM—could be useful in creating bacterial cells that divert resources used for growth towards the production of a target compound. Thus, the analysis of the simple model presented here is an attempt to help guide the construction of synthetic gene networks, which improves product yield and productivity.

This paper is structured as follow: in Section 2 we describe the open-loop model, providing some biological motivations for the terms forming the differential equations. Next, in Section 3 we qualitatively analyze the open-loop model by means of phase-plane and bifurcation diagram, showing how the steady states of the CGEM can be controlled by the external input (inducer). In Section 4 we derive a mathematical expression of the growth rate during the exponential phase as a function of the amount of the inducer. Finally, in Section 5 we present an *in silico* practical identifiability analysis of such expression.

2 The Open-loop Model

The principal modeling challenges come from incomplete knowledge of the networks, and the dearth of quantitative data for identifying kinetic parameters required for detailed mathematical models. Qualitative methods overcome both of these difficulties and are thus well-suited to the modeling and simulation of genetic networks ([27, 8]).

In this work we used a novel *piecewise non-linear* formalism—derived from piece wise affine (PWA) systems (see [4, 5, 7, 16, 17] for more details)—to model gene expression affected by dilution due to growth rate.

The open-loop model depicted in Fig. 1—similarly to PWA models of regulatory genetic networks—is built with discontinuous (step) functions. The use of step function has been motivated by the experimental observation that the activity of certain genes changes in a switch-like manner at a threshold concentration of a regulatory protein [34]. The non linearity is concentrated in the removal term of differential equations, which takes into account the protein degradation and the dilution due to growth.

The open-loop model, expressed by (1), describes the qualitative dynamics of a CGEM responsible for bacterial growth and another protein that reflects growth, such as CRP. The CGEM is assumed to be externally controlled by an inducer I (such as IPTG (Isopropil β-D-1-tiogalattopiranoside), Tc (tetracycline) etc). This model of ODE exhibits bi-stability in CGEM expression for some parameter sets, as experimentally verified in [30]. We shall take into account this bi-stability to control the model's state to the "low" or to the "high" CGEM stable steady state. Let $x_c,\ x_p \in \mathbb{R}_{\geq 0}$ be the CRP and CGEM concentrations respectively. Thus, the open-loop model graphically depicted in Fig. 1, can be mathematically translated into:

$$\begin{cases} \dot{x}_c\,(t) = k_c^0\ s^+(x_p, \theta_p^1) + k_c^1\ s^+(x_p, \theta_p^2)\ s^+(x_c, \theta_c^1)\ s^-(x_p, \theta_{\bar{\mu}}) \\ \qquad\qquad - (\bar{\mu}\ x_p(t) + \gamma_c)\ x_c(t) \\ \dot{x}_p\,(t) = \nu_1\ k_p^0\ s^+(x_p, \theta_p^1) + \nu_1\ k_p^1\ s^+(x_p, \theta_p^2) \\ \qquad\qquad - (\bar{\mu}\ x_p(t) + \gamma_p)\ x_p(t) \end{cases} \tag{1}$$

where:

- $k_i^0 > 0$ $(i = c, p)$ is the basal synthesis rate constant;
- $k_i^1 > 0$ $(i = c, p)$ is the main synthesis rate constant;
- ν_1 is a positive input accounting for the inducer I; it will be a function $\nu_1(v)$, v being the concentration of I;
- $\gamma_i > 0$ $(i = c, p)$ is the degradation rate constant;
- $\theta_i^j > 0$ $(i = c, p;\ j = 1, 2)$ is the x_i threshold concentration for activation/inhibition;
- $\theta_{\bar{\mu}} > 0$ is a growth threshold depending on which substrate is used;
- $\bar{\mu} > 0$ is a growth constant depending on which substrate is used.

and s^+, s^- denote the step-like functions, defined as

$$s^+(x_i, \theta_i^j) = \begin{cases} 1 & \text{if } x_i > \theta_i^j \\ 0 & \text{if } x_i < \theta_i^j \end{cases}; \quad s^-(x_i, \theta_i^j) = 1 - s^+(x_i, \theta_i^j)\ ,$$

which are used to model the switch-like promoters' regulation carried out by the generic protein x_i. These s^+, s^- are not defined at the threshold values so, to define solutions on the surfaces of discontinuity, i.e. $x_i = \theta_i^j$, we use the approach of Filippov [12], which extends the vector field to a differential inclusion.

In what follows, we will explain the main assumptions adopted in building the system equations (1), which were inspired by the models in [30, 27] and the literature on *E. coli*.

2.1 Growth rate

In bacteria, growth rate is intimately interwined with gene expression ([20, 28]) and with the type of substrate [22]. Hence, to keep model complexity to a minimum, we assume growth rate μ to be proportional—with a constant $\bar{\mu}$ depending on the quality of medium—to the concentration of the CGEM which is responsible for bacterial growth:

$$\mu(t) = \bar{\mu}\ x_p(t)\ . \tag{2}$$

2.2 cAMP-CRP activation

The cAMP-CRP complex is formed from cAMP, a small metabolite, which binds the protein CRP. The cAMP concentration is higher at low growth rate and rapidly decreases at high growth rate [2]. Thus, cAMP abundance in cells can be well captured by a negative step function of μ, i.e. $s^-(\mu, \theta_\mu)$. Moreover, being cAMP association with or dissociation from CRP much faster than the synthesis and degradation of proteins [27], we have assumed that as soon as CRP reaches a certain threshold, i.e. θ_c, CRP instantly binds to cAMP in a switch-like fashion . For these reasons, the positive regulation carried out by cAMP-CRP reads as:

$$b^+_{cAMP-CRP} = s^+(x_c, \theta_c)\ s^-(\mu, \theta_\mu).$$

Focusing on the negative step function $s^-(\mu, \theta_\mu)$ and taking into account the expression of μ in (2), we can rewrite $b^+_{cAMP-CRP}$ as:

$$b^+_{cAMP-CRP}(x_c, x_p) = s^+(x_c, \theta_c)\ s^-(x_p, \theta_{\bar{\mu}}) \tag{3}$$

where $\theta_{\bar{\mu}}$ is a threshold concentration of CGEM which depends on the type of carbon source.

2.3 CRP synthesis

We have assumed that a lower value of x_p, i.e. θ^1_p, induces the basal synthesis ($k^0_c\ s^+(x_p, \theta^1_p)$) of x_c while a higher value of x_p, i.e. θ^2_p, is needed to stimulate its main expression ($k^1_c s^+(x_p, \theta^2_p)$). Moreover, the *crp* gene is regulated both positively and negatively by cAMP-CRP. However, in order to simplify, we omit the negative control of *crp*, because this mechanism only plays a role when the CRP concentration is low [27][1]. Thus, only one concentration threshold of CRP, i.e. θ^1_c, is required in the model, to allow production of the cAMP-CRP complex. In conclusion, taking into account the regulation function of cAMP-CRP in (3), the CRP synthesis reads:

$$f_c(x) = k^0_c\ s^+(x_p, \theta^1_p) + k^1_c\ s^+(x_p, \theta^2_p)\ b^+_{cAMP-CRP}(x_c, x_p), \tag{4}$$

with

$$0 < \theta^1_c < max_c, \tag{5}$$

where max_c is the maximum concentration value for CRP.

2.4 CGEM synthesis

In this bi-dimensional model, since the CGEM is the main factor which determines growth of the cell, it is also responsible for its own synthesis. We have thus assumed that a low concentration (θ^1_p) is sufficient to stimulate its basal production $k^0_p\ s^+(x_p, \theta^1_p)$ while its main production $k^1_p\ s^+(x_p, \theta^2_p)$ is stimulated only above the θ^2_p threshold. Thus, we can order the thresholds for x_p as:

$$0 < \theta^1_p < \theta^2_p < max_p, \tag{6}$$

where max_p is the maximum concentration value.

Moreover, the inducer effect is modeled by input ν_1. For a general formulation of the activation of x_p by an inducer I, we will later on assume that ν_1 is a positive increasing function of I. Consequently, x_p synthesis reads:

$$f_p(x) = \nu_1\ k^0_p\ s^+(x_p, \theta^1_p) + \nu_1\ k^1_p\ s^+(x_p, \theta^2_p). \tag{7}$$

2.5 Proteins removal

The negative terms in $\dot{x}_c$ and $\dot{x}_p$ of (1) take into account the fact that cells remove proteins by two processes: degradation and dilution due to cell growth [10]. Notably, these terms can generally be expressed as $(\mu(t) + \gamma_i)x_i$ (for $i = c, p$) where $\mu(t) = \bar{\mu}\ x_p(t)$, which is the bacterial growth rate in (2), is responsible for the proteins' dilution while γ_i stands for protein's degradation.

[1] We found that a model involving the negative control of *crp* by cAMP-CRP does not have any effect on the conclusion of this study.

3 Qualitative Analysis of the Open-loop Model

In this section we will qualitatively study, by means of phase-planes and bifurcation diagrams, model (1) in the case that cells are grown in glucose. This will elucidate how qualitative dynamics—in terms of equilibria' location and their stability—is intertwined with biological phenomena. Moreover, we shall show how—through the external input ν_1—the stability of equilibria in (1) can be controlled, pointing out a reciprocal influence between growth rate and gene expression.

3.1 Open-loop model in glucose growth

If cells are grown in glucose, then parameters depending on the substrate become $\theta_{\bar{\mu}} = \theta_p^G$ and $\bar{\mu} = \mu_G$ in model (1). Moreover, in the presence of glucose or other PTS sugars, adenylate cyclase[2] activity decreases, leading to a drop in the cellular level of cAMP [21, 25]. Thus, we have modeled this effect assuming:

$$0 < \theta_p^1 < \theta_p^G < \theta_p^2 < max_p. \tag{8}$$

Therefore, during growth on glucose, the state space of model (1) can be partitioned into eight *regular domains*, where the vector field is uniquely defined:

$$\begin{aligned}
D_1^G &= \left\{x \in \mathbb{R}^2_{\geq 0} : 0 \leq x_c < \theta_c^1,\ 0 \leq x_p < \theta_p^1\right\}\\
D_2^G &= \left\{x \in \mathbb{R}^2_{\geq 0} : \theta_c^1 < x_c \leq max_c,\ 0 \leq x_p < \theta_p^1\right\}\\
D_3^G &= \left\{x \in \mathbb{R}^2_{\geq 0} : 0 \leq x_c < \theta_c^1,\ \theta_p^1 < x_p < \theta_p^G\right\}\\
D_4^G &= \left\{x \in \mathbb{R}^2_{\geq 0} : \theta_c^1 < x_c \leq max_c,\ \theta_p^1 < x_p < \theta_p^G\right\}\\
D_5^G &= \left\{x \in \mathbb{R}^2_{\geq 0} : 0 \leq x_c < \theta_c^1,\ \theta_p^G < x_p < \theta_p^2\right\}\\
D_6^G &= \left\{x \in \mathbb{R}^2_{\geq 0} : \theta_c^1 < x_c \leq max_c,\ \theta_p^G < x_p < \theta_p^2\right\}\\
D_7^G &= \left\{x \in \mathbb{R}^2_{\geq 0} : 0 \leq x_c < \theta_c^1,\ \theta_p^2 < x_p \leq max_p\right\}\\
D_8^G &= \left\{x \in \mathbb{R}^2_{\geq 0} : \theta_c^1 < x_c \leq max_c,\ \theta_p^2 < x_p \leq max_p\right\}.
\end{aligned}$$

In addition, there are also *switching domains*, where the model is defined only as a differential inclusion [12], corresponding to the segments where each of the variables is at a threshold ($x_i = \theta_i$ and $x_j \in [0, max_j]$).

In general, for any regular domain D, the synthesis rates (4) and (7) are constant for all $x \in D$, and it follows that model (1) can be written as

$$\begin{cases}
\dot{x}_c\,(t) = f_c^D - (\bar{\mu}\ x_p(t) + \gamma_c)\ x_c(t)\\
\dot{x}_p\,(t) = f_p^D - (\bar{\mu}\ x_p(t) + \gamma_p)\ x_p(t)
\end{cases} \tag{9}$$

with $f_c^D, f_p^D, \bar{\mu}, \gamma_c, \gamma_p$ positive real constants. For any initial condition $x(t_0) \in D$ the unique solution of (9) can be found explicitly by solving first the x_p-equation of (9), which is an autonomous differential equation, and then solving the x_c-equation, having substituted $x_p(t)$ into it. Thus, it can be shown that $x_c(t)$ is given by:

$$x_c(t) = \frac{1}{b(t)}\left(b(t_0)x_c(t_0) + f_c^D \int_{t_0}^{t} b(s)ds\right)$$

where $b(t) = exp\left(\int_{t_0}^{t}(\bar{\mu}\ x_p(\tau) + \gamma_p)d\tau\right)$. Moreover, defining $\Phi(D) = (\bar{x}_c, \bar{x}_p)^T$ with

$$\begin{aligned}
\bar{x}_c &= \frac{f_c^D}{\bar{\mu}\bar{x}_p + \gamma_c},\\
\bar{x}_p &= \frac{-\gamma_p + \sqrt{\gamma_p^2 + 4\bar{\mu}f_p^D}}{2\bar{\mu}},
\end{aligned} \tag{10}$$

[2]Enzyme that catalyzes the conversion of ATP to cAMP and pyrophosphate.

(it is easy to check that $\bar{x}_p$—in (10)—is the only positive solution of $\dot{x}_p = 0$) it turns out that either $x(t) \to \Phi(D)$ as $t \to \infty$ or $x(t)$ reaches the boundary of D.

Definition 1 *Given a regular domain D, the point $\Phi(D) = (\bar{x}_c, \bar{x}_p)^T$ (defined by (10)) is called the* focal point *for the flow in D.*

We will group into regions R_j those domains D_i^G where model (1)—in glucose growth— has the same dynamics and thus the same focal points. Considering Definition 1, we have the following focal points:

- $\forall x \in R_1 = \left\{x \in \mathbb{R}^2_{\geq 0} : x \in D_1^G \cup D_2^G\right\}$

$$x_c \to 0 \quad \wedge \quad x_p \to 0$$

  Thus, $\Phi_0^G = (0, 0)$ is the focal point of region R_1.

- $\forall x \in R_2 = \left\{x \in \mathbb{R}^2_{\geq 0} : x \in D_3^G \cup D_4^G \cup D_5^G \cup D_6^G\right\}$

$$x_c \to \frac{k_c^0}{\mu_G\ \bar{x}^1_{p,G} + \gamma_c} = \bar{x}^2_{c,G}$$
$$x_p \to \frac{-\gamma_p + \sqrt{\gamma_p^2 + 4\ \nu_1\ k_p^0\ \mu_G}}{2\mu_G} = \bar{x}^1_{p,G}$$

  Thus, $\Phi_1^G = (\bar{x}^2_{c,G}, \bar{x}^1_{p,G})$ is the focal point of region R_2^G.

- $\forall x \in R_3 = \left\{x \in \mathbb{R}^2_{\geq 0} : x \in D_7^G \cup D_8^G\right\}$

$$x_c \to \frac{k_c^0}{\mu_G\ \bar{x}^2_{p,G} + \gamma_c} = \bar{x}^1_{c,G}$$
$$x_p \to \frac{-\gamma_p + \sqrt{\gamma_p^2 + 4\ \nu_1(k_p^0 + \ k_p^1)\mu_G}}{2\mu_G} = \bar{x}^2_{p,G}$$

  Thus, $\Phi_2^G = (\bar{x}^1_{c,G}, \bar{x}^2_{p,G})$ is the focal point of region R_3.

The focal points Φ_i^G $(i = 1, ..., 3)$ are equilibrium points of model (1) provided that they belong to their respective regular domain, i.e. $\Phi(D) \in D$. The local stability of equilibrium points is given by the following theorem.

Theorem 1 *Let D be a regular domain and $\Phi(D)$ be the focal point of D. If $\Phi(D) \in D$, then $\Phi(D)$ is a locally stable point of model* (1)*.*

Proof: Model (1) restricted to D is given by (9). In order to assess the stability of $\Phi(D)$, we compute the Jacobian matrix of (9) calculated in $\Phi(D) = (\bar{x}_c, \bar{x}_p)^T$:

$$J(\bar{x}_c, \bar{x}_p) = \begin{pmatrix} -\bar{\mu}\bar{x}_c & -(\bar{\mu}\bar{x}_p + \gamma_p) \\ 0 & -(2\bar{\mu}\bar{x}_p + \gamma_p) \end{pmatrix}.$$

Since all the eigenvalues of J, which are the diagonal entries as J is diagonal, are negative, $\Phi(D)$ turns out to be a locally stable point. ■

Hence, there can be at most three locally stable steady states during growth on glucose.

Fig. 2 depicts the phase-plane of model (1). It can be seen that Φ_0^G, Φ_1^G, Φ_2^G, (for the parameter values used) are locally stable steady states since they are within their respective regular domains (Theorem 1). Notably, it is easy to verify that Φ_0^G is locally stable for any set of parameters. It

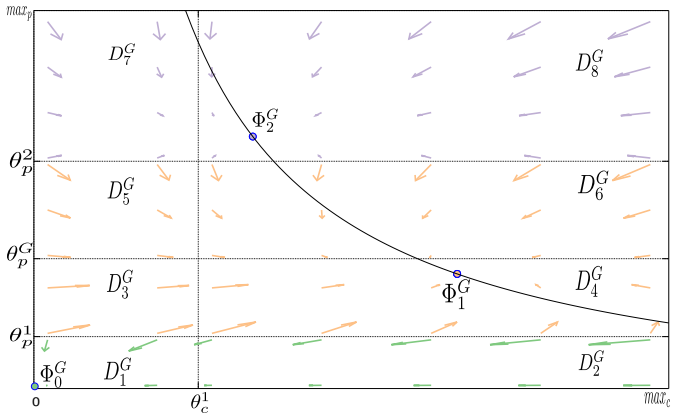

Figure 2: Phase plane of model (1) during growth in glucose. Parameter values used: $\theta_c^1 = 0.6$, $\theta_p^1 = 0.8$, $\theta_p^G = 2$, $\theta_p^2 = 3.5$, $k_c^0 = 7$, $k_c^1 = 10$, $k_p^0 = 40$, $k_p^1 = 50$, $\gamma_c = 1$, $\gamma_p = 1$, $\mu_G = 2$ e $\nu_1 = .5$. The black curve is the x_c-nullcline: $x_p = \dfrac{k_c^0}{x_c\ \mu_G} - \dfrac{\gamma_c}{\mu_G}$. Stable fixed points: Φ_0^G, Φ_1^G, Φ_2^G.

represents absence of growth and can happen when the initial condition $x_p(t_0)$, is too low—specifically $x_p(t_0) < \theta_p^1$—to initiate gene transcription or when the control input ν_1 does not sufficiently induce CGEM expression, that is when $\bar{x}_{p,G}^1 < \theta_p^1$. We refer to Φ_0^G as the *trivial* fixed point. Φ_1^G represents CGEM basal level—leading to a low growth rate (see (2))— while CRP is at a high level, which is in agreement with high *crp* gene expression (by cAMP-CRP) at lower growth rate. Thus, because of the low growth rate achieved, we refer to Φ_1^G as the *low* fixed point. Conversely, at Φ_2^G, CRP is at low level while CGEM , as well as μ, have reached their highest stable values. Thus, Φ_2^G is named the *high* fixed point.

Since $\bar{x}_{p,G}^1(\nu_1)$ and $\bar{x}_{p,G}^2(\nu_1)$ are function of ν_1, it turns out that the location of focal points Φ_1^G and Φ_2^G, and thus the number of equilibria of model (1), depend on the control input ν_1. Hence, choosing appropriate values of ν_1 it is possible to control model (1) towards Φ_1^G or Φ_2^G. To illustrate this, we have depicted in Fig. 3 the x_p-bifurcation diagram when parameter ν_1 varies from 0 to 1 while the other parameter values are the same as those used in Fig. 2.

We notice that Fig. 3 is divided into four parts in which x_p stability changes significantly. In part I, for those values of ν_1 such that $\bar{x}_{p,G}^1 < \theta_p^1$ and $\bar{x}_{p,G}^2 < \theta_p^2$, neither Φ_1^G nor Φ_2^G are stable steady states. In this case, model (1) during growth on glucose converges towards the only stable point Φ_0^G (not depicted in Fig. 3). So, in I, the control input is too small to allow CGEM to reach a basal level, and prevents bacterial growth.

In part II, when $\bar{x}_{p,G}^1(\nu_1) > \theta_p^1$ and $\bar{x}_{p,G}^2(\nu_1) < \theta_p^2$ hold, only Φ_1^G is a stable steady state (besides the trivial one) according to Theorem 1. Hence, it turns out that choosing an initial condition of CGEM $x_p(t_0) > \theta_p^1$ and ν_1 such that $\bar{x}_{p,G}^1(\nu_1) > \theta_p^1$ and $\bar{x}_{p,G}^2(\nu_1) < \theta_p^2$, we can control model (1) to the stable point Φ_1^G.

In part III, characterized by $\theta_p^1 < \bar{x}_{p,G}^1(\nu_1) < \theta_p^2$ and $\bar{x}_{p,G}^2(\nu_1) > \theta_p^2$, both Φ_1^G and Φ_2^G are stable steady states: this is a region of bi-stability. Moreover, the phase plane corresponding to this configuration is depicted in Fig. 2, where we can also observe the presence of two separatrices $x_p = \theta_p^1$ and $x_p = \theta_p^2$. Is is clear that, depending on $x_p(t_0)$, the model can converge to Φ_1^G (if $\theta_p^1 < x_p(t_0) < \theta_p^2$) or to Φ_2^G (if $x_p(t_0) > \theta_p^2$).

In part IV, when $\bar{x}_{p,G}^1(\nu_1) > \theta_p^2$ holds, only Φ_2^G is a stable steady state and thus, whenever $x_p(t_0) > \theta_p^1$, model (1) converges to Φ_2^G.

The open-loop control in glucose growth can be summarized as follows.

Proposition 1 *Consider model* (1) *with control input* ν_1 *and initial condition* $x_p(t_0)$ *such that:*

- *if* $(\bar{x}_{p,G}^1(\nu_1) < \theta_p^1 \ \wedge \ \bar{x}_{p,G}^2(\nu_1) < \theta_p^2) \ \vee \ x_p(t_0) < \theta_p^1$, *then model* (1) *converges to the* trivial focal point Φ_0^G *(region I in Fig. 3);*

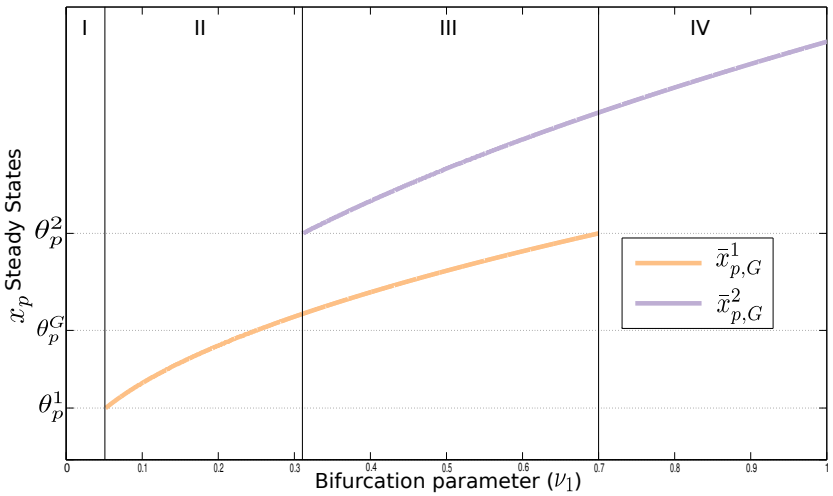

Figure 3: Bifurcation diagram for model (1) during growth in glucose, showing the non trivial locally stable steady states of x_p as a function of the control input ν_1. Other parameter values used are the same as those in Fig. 2. See Proposition 1 for more details.

- *if* $\bar{x}_{p,G}^1(\nu_1) > \theta_p^1 \ \wedge \ \bar{x}_{p,G}^2(\nu_1) < \theta_p^2 \ \wedge \ x_p(t_0) > \theta_p^1$, *then model* (1) *converges to the* low focal point Φ_1^G *(region II in Fig. 3);*
- *if* $\theta_p^1 < \bar{x}_{p,G}^1(\nu_1) < \theta_p^2 \ \wedge \ \bar{x}_{p,G}^2(\nu_1) > \theta_p^2 \ \wedge \ x_p(t_0) > \theta_p^1$, *then model* (1) *is* bistable *(region III in Fig. 3) and notably:*
  - *if* $\theta_p^1 < x_p(t_0) < \theta_p^2$, *then model* (1) *converges to the* low focal point Φ_1^G*;*
  - *if* $x_p(t_0) > \theta_p^2$, *then model* (1) *converges to the* high focal point Φ_2^G
- *if* $\bar{x}_{p,G}^1(\nu_1) > \theta_p^2 \ \wedge \ x_p(t_0) > \theta_p^1$, *then model* (1) *converges to the* high focal point Φ_2^G *(region IV in Fig. 3).*

4 Growth rate expression for exponential phase

Here, to account for different dosage of inducer, we make an assumption to analytically characterize the function $\nu_1 = h(v)$. Notably, to describe the regulation of CGEM gene expression by the inducer, we employ a function typically used in synthetic biology [18]:

$$\nu_1(v) = \alpha + (1-\alpha)\frac{v^n}{K_v^n + v^n} \tag{11}$$

where v denotes inducerconcentration and α accounts for the basal transcriptional activity. Controlled gene expression follows Hill-type dosage-response curve with promoter-activator affinity K_v and cooperative (Hill) coefficient n. During *exponential phase*—the period characterized by cell doubling— the bacterial culture shows a constant growth rate [23]. This means that, according to (2), a stable fixed point of the CGEM has to be reached. Hence, our expression of growth rate during exponential phase reads:

$$\mu = \mu_G \bar{x}_p \tag{12}$$

where $\bar{x}_p$ is the CGEM concentration at steady state, which can be either $\bar{x}_{p,G}^1$ or $\bar{x}_{p,G}^2$—depending on the amount of inducerwhich determines the level of CGEM expression. Thus, our expression of growth

rate during exponential phase can assume the two values below:

$$\mu(v) = \begin{cases} \mu_G \bar{x}^1_{p,G} = \dfrac{-\gamma_p + \sqrt{\gamma_p^2 + 4\ \nu_1\ k_p^0\ \mu_G}}{2} \\ \\ \mu_G \bar{x}^2_{p,G} = \dfrac{-\gamma_p + \sqrt{\gamma_p^2 + 4\ \nu_1\ (k_p^0 + k_p^1)\ \mu_G}}{2} \end{cases} . \tag{13}$$

Specifically, we assumed there is a particular value of inducer, i.e. v^*, such that for an appropriate choice of initial condition and for all $v \leq v^*$ the CGEM steady state is $\bar{x}^1_{p,G}$ while for all $v > v^*$ the steady state is $\bar{x}^2_{p,G}$. Thus, considering that, and substituting (11) into (13) we obtain the theoretical expression for growth rate during exponential phase:

$$\mu(v) = \begin{cases} -\dfrac{\gamma_p}{2}\left[1 - \sqrt{1 + \dfrac{4k_p^0\mu_G\alpha}{\gamma_p^2} + \dfrac{4k_p^0\mu_G(1-\alpha)}{\gamma_p^2}\dfrac{v^n}{K_v^n + v^n}}\ \right] & \text{if, } v \leq v^* \\ \\ -\dfrac{\gamma_p}{2}\left[1 - \sqrt{1 + \dfrac{4(k_p^0 + k_p^1)\mu_G\alpha}{\gamma_p^2} + \dfrac{4(k_p^0 + k_p^1)\mu_G(1-\alpha)}{\gamma_p^2}\dfrac{v^n}{K_v^n + v^n}}\ \right] & \text{if, } v > v^* \end{cases} \tag{14}$$

It is worthy to notice that expression (14) directly relates the growth rate μ during exponential phase to the amount of the inducer v. Hence, using (14) we can fine tune—by means of appropriate level of the inducer—the growth rate of the cells during the exponential phase.

5 *In silico* Identifiability Analysis of Growth Rate

Our collaborators (Jérôme Izard and Hans Geiselmann [3]) are currently performing an ongoing experiment on a synthetic *E. coli* – implementing the open-loop model depicted Fig. 1 – which relates the level of growth rate during the exponential phase to the amount of the inducer. In the future, these dose-response curves will be useful to calibrate and validate the growth rate expression (during exponential phase) (14).

Here, we used simulated data to fit the the growth rate model (14) and to study the identifiability of the parameters.

5.1 Problem Statement

Given a parametric non-linear model, such as (14), the relationship between a response variable (output) and one or more predictor variables (input) can be represented by the expression:

$$y = \eta(v, p) + \epsilon\ ,$$

where

- y is an $n \times 1$ vector of observations of the response variable,
- v is an $n \times m$ matrix of predictors,
- p is a $q \times 1$ vector of unknown parameters to be estimated,
- η is any function of v and p,
- ϵ is an $n \times 1$ vector of independent, identically distributed random disturbances.

[3] *Laboratoire Adaptation et Pathogénie des Microorganismes*, (CNRS UMR 5163), Université Joseph Fourier, Bâtiment Jean Roget, Faculté de Médecine-Pharmacie, La Tronche, France

The nonlinear regression problem consists of finding a vector $\hat{p}$ minimizing a scalar cost function $J(p)$, which is generally a measurement of the agreement of experimental data with the outputs predicted by the model. The cost function that we have considered in this work is a weighted least squares criterion:

$$J(p) = \sum_{i=1}^{n} \frac{(y_i - \eta(v_i, p))^2}{y_i^2} \tag{15}$$

where y_i denotes the i-th data-point of the observable y, measured at input-points v_i, and $\eta(v_i, p)$ the i-th observable as predicted by the parameters p. The parameters can be estimated numerically by:

$$\hat{p} = \arg\min \left[J(p) \right] \ . \tag{16}$$

Determining the parameter vector $\hat{p}$ which minimizes $J(p)$ is only a part of the parameter estimation problem. In fact, when preparing to fit a mathematical model or expression to a set of experimental data, the prior assessment of parameter identifiability is a crucial aspect [32]. However, the structural identifiability analysis for non-linear models in systems biology is still a challenging question [6]. Whether or not parameters can be estimated uniquely depends on the model structure, the parameterization of the model and the experiment used to get the data [26].

Regarding this problem, we briefly recall two important definitions on identifiability [33]:

- the parameter p_i, $i = 1, ..., q$ is **structurally globally identifiable** if assuming ideal conditions (error-free model structure and unlimited noise-free observations (v, y)) and if for almost any $p^* \in \mathcal{P}$ (admissible parametric space $\mathcal{P}$),

$$y(p, v) = y(p^*, v), \forall v \Rightarrow p_i = p_i^*.$$

- the parameter p_i, $i = 1, ..., q$ is **structurally locally identifiable** if assuming ideal conditions (error-free model structure and unlimited noise-free observations (v, y)) and if for almost any $p^* \in \mathcal{P}$ (admissible parametric space $\mathcal{P}$), there exists a neighborhood $V(p^*)$ such that

$$p \in V(p^*) \wedge y(p, v) = y(p^*, v), \forall v \Rightarrow p_i = p_i^*.$$

An important complement to the structural identifiability definitions is the notion of **practical identifiability**. Practical identifiability is indeed related to the quality of experimental data and their information content [9]. The question raised by this notion is the following: in the presence of observation errors and/or few data are reliable estimations of the parameters possible? Thus, once having determined the value of $\hat{p}$ minimizing the cost function $J(p)$, it is very important to find a realistic measure of how $\hat{p}$ is precise. To this end, the confidence intervals[4] of the estimated parameters have to be calculated.

It must be noted that, unlike for the linear case for which an exact theory exists, there is no exact theory for the evaluation of confidence intervals for systems which are nonlinear in the parameters. An approximate method based on a local linearization of the output function $\eta(v, p)$ is generally used [29, 13], thus the confidence region is evaluated as a function of the parameter covariance matrix. The applicability of such approximate method requires that the response function $\eta(v, p)$ must be continuous in its arguments (v, p), the first partial derivatives $\frac{\partial}{\partial p_i}\eta(v, p)$ must be continuous in its arguments (v, p), and the second partial derivatives $\frac{\partial^2}{\partial p_i \partial p_j}\eta(v, p)$ must be continuous in its arguments (v, p), but our model (14) does not satisfy these conditions because of the discontinuity in $v = v*$. Hence, in the remainder of the paper a computational method, based on *in silico* generated data, is suggested to argue the practical identifiability of non-linear discontinuous model such as (14).

5.2 Generation of Simulated Data Sets

In order to assess the quality of parameter estimation and thus the practical identifiability of parameters in (14), artificial data were generated by simulation of (14) from a set of pre-defined parameters (to be considered as true values). The true parameter values (Tab. 1) were chosen from physiological parameters of *E.coli* cells [20, 3] and were based on similar studies of this type [30].

[4] A confidence interval $[\sigma_i^-, \sigma_i^+]$ of a parameter estimate $\hat{p}_i$ to a confidence level α signifies that the true value p_i^* is located within this interval with probability α.

k_p^0 $[\mu M \cdot min^{-1}]$	k_p^1 $[\mu M \cdot min^{-1}]$	γ_p $[min^{-1}]$	μ_G $[(\mu M \cdot min)^{-1}]$	α	K_v $[\mu M]$	n	v^* $[\mu M]$
0.02	0.11	0.006	0.0014	0.1	30	2	50

Table 1: Nominal parameter values

Thus, the artificial growth rate values have been simulated considering a measurement error proportional to the nominal value of growth rate:

$$y = \mu(v) + \sigma\mu(v)\mathcal{N}(0,1) \tag{17}$$

where $\mathcal{N}(0,1)$ is a normally distributed random variable with zero mean and unit variance and $\sigma\mu(v)$ is the standard deviation of the observation errors. Four different types of data sets were considered to account for practical identifiability:

- data set I, with $v = [0, 5, 10, 15, ..., 295, 300, 1000]$ and $\sigma = 10^{-2}$;
- data set II, with $v = [0, 10, 20, 30, ..., 290, 300, 1000]$ and $\sigma = 10^{-2}$;
- data set III, with $v = [0, 5, 10, 15, ..., 295, 300, 1000]$ and $\sigma = 5 \cdot 10^{-2}$;
- data set IV, with $v = [0, 10, 20, 30, ..., 290, 300, 1000]$ and $\sigma = 5 \cdot 10^{-2}$;

Notably, data sets I, II, III and IV, have been generated with different number of points (N_{exp}) and different intensities of noise (σ) to study the practical identifiability of the parameters in four realistic experimental conditions. In particular, data sets I and III have the same number of data points, i.e. $N_{exp} = 62$, but different noise, $\sigma = 10^{-2}$ for data set I and $\sigma = 5 \cdot 10^{-2}$ for data set III. Data set II and IV have less number of points, i.e. $N_{exp} = 32$, while the level of noise considered is $\sigma = 10^{-2}$ for data set II and $\sigma = 5 \cdot 10^{-2}$ for data set VI.

5.3 Model Parameterization and Global Optimization

First, to avoid evident structural identifiability problems we will group together those parameters in (14) which appear as combinations of products and/or quotients between parameters. Thus, after some algebraic manipulations expression (14) reads as:

$$\mu(v) = \begin{cases} -\frac{\gamma_p}{2}\left[1 - \sqrt{1 + \frac{4k_p^0\mu_G\alpha}{\gamma_p^2}\left(1 + \frac{(1-\alpha)}{\alpha}\frac{v^n}{K_v^n + v^n}\right)}\right] & \text{if, } v \leq v^* \\ -\frac{\gamma_p}{2}\left[1 - \sqrt{1 + \frac{4(k_p^0 + k_p^1)\mu_G\alpha}{\gamma_p^2}\left(1 + \frac{(1-\alpha)}{\alpha}\frac{v^n}{K_v^n + v^n}\right)}\right] & \text{if, } v > v^* \end{cases} \tag{18}$$

Moreover, to avoid dependence on physical unit as well as to overcome possible scaling problem and to reduce the number of parameters, we decided to calculate a non-dimensional version of expression (18). Notably, the non-dimensional slope $\mu_N(v)$ is obtained by dividing $\mu(v)$ in (18) for the minimal growth rate, which is achieved at the minimum value of the inducer, i.e. at $v = v_0$, which for our data sets I, II, III, IV consists in $v_0 = 0$. Thus, considering the necessary condition $v_0 < v^*$, the non-dimensional growth rate during the exponential phase reads:

$$\mu_N(v) = \begin{cases} \dfrac{1-\sqrt{1+\dfrac{4k_p^0\mu_G\alpha}{\gamma_p^2}\left(1+\dfrac{(1-\alpha)}{\alpha}\dfrac{v^n}{K_v^n+v^n}\right)}}{1-\sqrt{1+\dfrac{4k_p^0\mu_G\alpha}{\gamma_p^2}}} & \text{if, } v \le v^* \\ \dfrac{1-\sqrt{1+\dfrac{4(k_p^0+k_p^1)\mu_G\alpha}{\gamma_p^2}\left(1+\dfrac{(1-\alpha)}{\alpha}\dfrac{v^n}{K_v^n+v^n}\right)}}{1-\sqrt{1+\dfrac{4k_p^0\mu_G\alpha}{\gamma_p^2}}} & \text{if, } v > v^* \end{cases} \tag{19}$$

Now, considering the following parameterization

$$p_1 = \frac{4k_p^0\mu_G\alpha}{\gamma_p^2};\ p_2 = \frac{(1-\alpha)}{\alpha};\ p_3 = K_v;\ p_4 = n;\ p_5 = \frac{4k_p^1\mu_G\alpha}{\gamma_p^2};\ p_6 = v^*$$

the expression (19) can be rewritten as

$$\mu_N(v,p) = \begin{cases} \dfrac{1-\sqrt{1+p_1\left(1+p_2\dfrac{v^{p_4}}{p_3^{p_4}+v^{p_4}}\right)}}{1-\sqrt{1+p_1}} & \text{if, } v \le p_6 \\ \dfrac{1-\sqrt{1+(p_1+p_5)\left(1+p_2\dfrac{v^{p_4}}{p_3^{p_4}+v^{p_4}}\right)}}{1-\sqrt{1+p_1}} & \text{if, } v > p_6 \end{cases} \tag{20}$$

where $p = [p_1, p_2, p_3, p_4, p_5, p_6]$ and, considering the true parameters values in Tab 1 we obtain the true vector of parameters p^*:

$$p^* = [0.3033, 9, 30, 2, 1.6683, 50]\ . \tag{21}$$

Similarly, the data sets I to IV will also be normalized to their minimal value, i.e., each output-point is divided by the minimal observation value, that is $y_{min} = \mu(v_0)$, where $v_0 = 0$.

Our approach in identifying the unknown parameters of model (19) consists in solving a non-linear least squares minimization problem, using a hybrid optimization approach which makes use of the functions *ga* (Genetic Algorithm [15]) and *GlobalSearch* of the *MATLAB® Global Optimization Toolbox™*. To start, we used the Genetic Algorithm (GA) for 10^4 generations to get near an optimum point. The genetic algorithm does not use derivatives to detect descent in its minimization steps. Hence, it is a good choice for non-differentiable and/or discontinuous problems. Moreover, GA does not necessarily need an user supplied initial guess, which in most case leads to local sub-optimal convergence if the initial guess is far from the global optimum. The result obtained with the genetic algorithm is then used as initial point of a hybrid function, to further improve the value of the cost function $J(p)$. We decided to use the *GlobalSearch*[5] command as hybrid function since it searches many basins of attraction near the starting point given by GA, arriving faster at an even better solution.

5.4 In Silico Practical Identifiability Analysis

The practical identifiability of model (20) has been tested using data sets I, II, III and IV, which have different values of errors' measurement and different data points. Hence, these artificial data are suitable to mimic realistic experimental set-ups.

For each data set mentioned above, parameters' confidence intervals have been computed following a *Monte Carlo*-like approach.

[5] *GlobalSearch* first runs *fmincon* from the start point you give. If this run converges, GlobalSearch records the start point and end point for an initial estimate on the radius of a basin of attraction. Then, *GlobalSearch* solver starts a local solver (*fmincon*) from multiple starting points and store local and global solutions found during the search process. Notably, the *GlobalSearch* solver first uses a scatter-search algorithm to randomly generate multiple starting points, then filters non-promising start points based upon objective and constraint function values and local minima already found, and finally runs a constrained nonlinear optimization solver to search for a local minimum from the remaining start points.

	DATA SET I $\sigma = 10^2$ $N_{exp} = 62$	DATA SET II $\sigma = 10^{-2}$ $N_{exp} = 32$	DATA SET III $\sigma = 5 \cdot 10^{-2}$ $N_{exp} = 62$	DATA SET IV $\sigma = 5 \cdot 10^{-2}$ $N_{exp} = 32$
CI_1	0.3328 ± 0.4939	0.3738 ± 0.5441	0.2631 ± 0.4220	0.32 ± 0.49
CI_2	9.23 ± 3.45	9.36 ± 3.88	8.63 ± 3.06	9.21 ± 4.67
CI_3	30.16 ± 3.55	30.00 ± 3.55	29.39 ± 5.15	30.33 ± 7.52
CI_4	2.002 ± 0.079	2.011 ± 0.089	2.006 ± 0.232	2.01 ± 0.33
CI_5	2.053 ± 4.192	2.39 ± 4.51	1.53 ± 3.59	1.93 ± 3.99
CI_6	53.32 ± 4.48	55.98 ± 6.99	53.06 ± 3.58	56.70 ± 6.79

Table 2: Confidence intervals of estimated parameters $\hat{p}_i$ when (20) is fitted to (non-dimensionalized) data sets I, II,III,IV. The confidence intervals for parameters become larger at increasing values of the measurement error and at decreasing numbers of data points, indicating possible practical identifiability problems especially for $\hat{p}_1$ and $\hat{p}_5$.

	DATA SET I $\sigma = 10^{-2}$ $N_{exp} = 62$	DATA SET II $\sigma = 10^{-2}$ $N_{exp} = 32$	DATA SET III $\sigma = 5 \cdot 10^{-2}$ $N_{exp} = 62$	DATA SET IV $\sigma = 5 \cdot 10^{-2}$ $N_{exp} = 32$
$CI_{\hat{p}_5/\hat{p}_1}$	5.29 ± 2.39	5.54 ± 2.43	4.99 ± 1.15	5.2 ± 1.3

Table 3: Confidence intervals of the ratio $\hat{p}_5/\hat{p}_1$ when (20) is fitted to (non-dimensionalized) data sets I, II,III,IV.

Notably, $N_{simul} = 200$ runs of the previously described hybrid optimization were performed. Where, at each of the N_{simul} runs, a new realization of the artificial measurements—according to the inputs and noise statistic of each data set—is considered. These N_{simul} optimization yields N_{simul} estimated values for each parameter p_i, $i = 1, \ldots, 5$. Then, for each i, an average value, $\hat{m}_i$, and a standard deviation, $\hat{s}_i$, were computed by fitting a Gaussian distribution $\mathcal{N}(\hat{m}_i, \hat{s}_i^2)$ to the histogram of the N_{simul} values of p_i. Thus, the 95% confidence interval (CI_i) for the p_i parameter is calculated as:

$$CI_i = \hat{m}_i \pm 1.96\hat{s}_i \tag{22}$$

This leads to the confidence intervals listed in Table 2.

As we can see in Table 2, parameters p_i for $i \in \{2, 3, 4, 6\}$ do not show any practical identifiability issues, as the true value is contained in the respective CI with sufficiently precision. On the contrary, the CIs of parameters $\hat{p}_1$ and $\hat{p}_5$ tend to become very large at increasing values of the measurement's errors (σ) and at decreasing numbers of data points, indicating that in real experimental conditions (that is, limited and noisy data), the precise identification of these parameters might be impracticable. Moreover, we found that the correlation coefficient (R) between the two vectors of estimated parameters parameters $\hat{p}_1$ and $\hat{p}_5$ is $R = 0.99$, for all data sets. Recall that the correlation coefficient measures the interrelationship between $\hat{p}_1$ and $\hat{p}_5$ quantifying the compensation effects of changes in the parameter values on the model output. In fact, when two parameters are highly correlated, a change in the model output caused by a change in a model parameter can be balanced by a proper change in the other parameter value. Thus, instead of considering the CIs of $\hat{p}_1$ and $\hat{p}_5$ separately—which are not significant—we have computed the confidence interval of their ratio, i.e. $\hat{p}_5/\hat{p}_1$. These results are presented in Table 3. As we can notice in Table 3, the CIs of $\hat{p}_5/\hat{p}_1$ are accurate, since they contain the true value of the ratio $p_5^*/p_1^* = 5.5$, and more precise since their relative width is smaller than the relative width of CI_1 and CI_5.

It must be noted that a further reduced model which takes into account the correlation between p_5 and p_1 can not be achieved. This because expression (20) can be rewritten in terms of the ratio and either p_5 or p_1. Fig 4 shows the fitting of model (20) to one realization of data set IV.

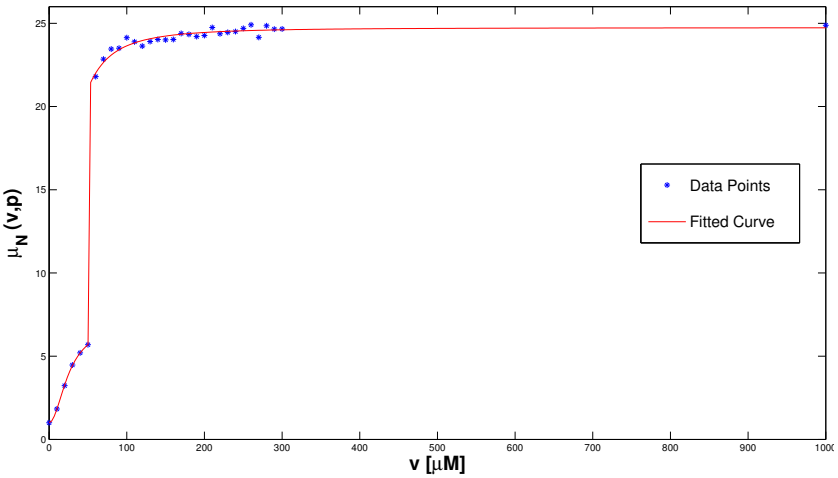

Figure 4: Fitting the growth rate function (20) using one realization of the non-dimensional data set II. The blue points are the normalized artificial data generated according to specification of data set II. The red curve is the function (20) when $\hat{p}$ is used.

6 Conclusions

In this paper, a minimal model consisting of two variables (the concentrations of two gene products) and an input (an inducer) was analyzed and used to describe one possible mechanism to control the growth rate of *E. coli* cells during exponential phase. This model is based on the piecewise affine formalism but a new, non-linear, term was added to account for the dilution effect during growth. The qualitative dynamics of the model can thus be studied, and the bifurcation diagram with respect to the input is obtained. Moreover, this mathematical formalism allows derivation of an analytic expression for the growth rate as function of the input. This expression has two applications:

- it can be directly fitted to experimental data to estimate a set of parameters (this is an advantage relative to the typical "indirect" parameter estimation by fitting to the numerical solutions of the differential equations);
- it provides an indication of how to control the growth rate to a desired value by adding a given quantity of inducer.

Finally, practical identifiability analysis based on numerical simulations is presented, which shows that some issues may arise with noisy measurements. In this case, our analysis suggests that the original growth rates' measurements should be adimensionalized and unknown parameters grouped into a new set of "lumped" parameters in order to obtain local identifiability. Notably, we found that only the ratio between the estimated parameters $\hat{p}_1$ and $\hat{p}_5$ can be estimated with sufficient precision in the case when only limited and noisy data are available. This study and the conclusions on identifiability will be most useful to help dealing with and solving parameter estimation problems with real data sets.

Acknowledgments.

This work was supported by ANR GeMCo (French national research agency).

References

[1] E. Andrianantoandro, S. Basu, D.K. Karig, and R. Weiss. Synthetic biology: new engineering rules for an emerging discipline. *Molecular systems biology*, 2(1), 2006.

[2] Katja Bettenbrock, Thomas Sauter, Knut Jahreis, Andreas Kremling, Joseph W. Lengeler, and Ernst-Dieter Gilles. Correlation between growth rates, EIIACrr phosphorylation, and Intracellular Cyclic AMP levels in Escherichia coli K-12. *J. Bacteriol.*, 189(19):6891–6900, 2007.

[3] H. Bremer, P.P. Dennis, et al. Modulation of chemical composition and other parameters of the cell by growth rate. *Escherichia coli and Salmonella: cellular and molecular biology*, 2:1553–1569, 1996.

[4] R. Casey, H. Jong, and J.L. Gouzé. Piecewise-linear models of genetic regulatory networks: Equilibria and their stability. *Journal of Mathematical Biology*, 52(1):27–56, 2006.

[5] M. Chaves and J.L. Gouzé. Piecewise affine models of regulatory genetic networks: Review and probabilistic interpretation. *Advances in the Theory of Control, Signals and Systems with Physical Modeling*, pages 241–253, 2011.

[6] O.T. Chis, J.R. Banga, and E. Balsa-Canto. Structural identifiability of systems biology models: A critical comparison of methods. *PloS one*, 6(11):e27755, 2011.

[7] H. De Jong, J.L. Gouzé, C. Hernandez, M. Page, T. Sari, and J. Geiselmann. Qualitative simulation of genetic regulatory networks using piecewise-linear models. *Bulletin of Mathematical Biology*, 66(2):301–340, 2004.

[8] Hidde de Jong, Johannes Geiselmann, Cline Hernandez, and Michel Page. Genetic network analyzer: qualitative simulation of genetic regulatory networks. *Bioinformatics*, 19(3):336–344, 2003.

[9] D. Dochain and P. Vanrolleghem. *Dynamical Modelling and Estimation in Wastewater Treatment Processes*. IWA Publishing, 2001.

[10] Eran Eden, Naama Geva-Zatorsky, Irina Issaeva, Ariel Cohen, Erez Dekel, Tamar Danon, Lydia Cohen, Avi Mayo, and Uri Alon. Proteome half-life dynamics in living human cells. *Science*, 331(6018):764–768, 2011.

[11] M.B. Elowitz, S. Leibler, et al. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, 2000.

[12] A.F. Filippov and F.M. Arscott. *Differential equations with discontinuous righthand sides*. Mathematics and its Applications Series. Kluwer Academic Publishers, 1988.

[13] A.R. Gallant. Nonlinear regression. *The American Statistician*, 29(2):73–81, 1975.

[14] T.S. Gardner, C.R. Cantor, and J.J. Collins. Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403:339–342, 2000.

[15] D.E. Goldberg. *Genetic algorithms in search, optimization, and machine learning*. Addison-wesley, 1989.

[16] J.L. Gouzé and T. Sari. A class of piecewise linear differential equations arising in biological models. *Dynamical systems*, 17(4):299–316, 2002.

[17] F. Grognard, H. De Jong, and J.L. Gouzé. Piecewise-linear models of genetic regulatory networks: theory and example. *Biology and Control Theory: Current Challenges*, pages 137–159, 2007.

[18] M. Kaern, W.J. Blake, and J.J. Collins. The engineering of gene regulatory networks. *Annual Review of Biomedical Engineering*, 5(1):179–206, 2003.

[19] A.S. Khalil and J.J. Collins. Synthetic biology: applications come of age. *Nature Reviews Genetics*, 11(5):367–379, 2010.

[20] Hwa Terence Klumpp Stefan, Zhang Zhongge. Growth rate-dependent global effects on gene expression in bacteria. *Cell*, 139(7):1366–1375, 2010.

[21] Evelyne Krin, Odile Sismeiro, Antoine Danchin, and Philippe N. Bertin. The regulation of Enzyme IIAGlc expression controls adenylate cyclase activity in Escherichia coli. *Microbiology*, 148(5):1553–1559, 2002.

[22] A G Marr. Growth rate of Escherichia coli. *Microbiological Reviews*, 55(2):316–333, 1991.

[23] J Monod. The growth of bacterial cultures. *Annual Review of Microbiology*, 3(1):371–394, 1949.

[24] S. Mukherji and A. Van Oudenaarden. Synthetic biology: understanding biological design from synthetic circuits. *Nature Reviews Genetics*, 10(12):859–871, 2009.

[25] Lucinda Notley-McRobb, Alison Death, and Thomas Ferenci. The relationship between external glucose concentration and cAMP levels inside Escherichia coli: implications for models of phosphotransferase-mediated regulation of adenylate cyclase. *Microbiology*, 143(6):1909–1918, 1997.

[26] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 2009.

[27] Delphine Ropers, Hidde de Jong, Michel Page, Dominique Schneider, and Johannes Geiselmann. Qualitative simulation of the carbon starvation response in Escherichia coli. *Biosystems*, 84(2):124 – 152, 2006.

[28] Matthew Scott, Carl W. Gunderson, Eduard M. Mateescu, Zhongge Zhang, and Terence Hwa. Interdependence of cell growth and gene expression: Origins and consequences. *Science*, 330(6007):1099–1102, 2010.

[29] G.A.F. Seber and C.J. Wild. *Nonlinear regression*, volume 503. LibreDigital, 2003.

[30] C. Tan, P. Marguet, and L. You. Emergent bistability by a growth-modulating positive feedback circuit. *Nature chemical biology*, 5(11):842–848, 2009.

[31] M. Tigges, T.T. Marquez-Lago, J. Stelling, and M. Fussenegger. A tunable synthetic mammalian oscillator. *Nature*, 457(7227):309–312, 2009.

[32] S. Vajda, H. Rabitz, E. Walter, and Y. Lecourtier. Qualitative and quantitative identifiability analysis of nonlinear chemical kinetic models. *Chemical Engineering Communications*, 83(1):191–219, 1989.

[33] É. Walter and L. Pronzato. *Identification of parametric models from experimental data*. Communications and control engineering. Springer, 1997.

[34] G. Yagil and E. Yagil. On the relation between effector concentration and the rate of induced enzyme synthesis. *Biophysical Journal*, 11(1):11–27, 1971.