Subtilin Production by *Bacillus Subtilis*: Stochastic Hybrid Models and Parameter Identification*

Eugenio Cinquemani†, Riccardo Porreca‡,
Giancarlo Ferrari-Trecate‡, and John Lygeros†

Abstract

This paper presents methods for the parameter identification of a model of subtilin production by *Bacillus subtilis*. Based on a stochastic hybrid model, identification is split in two subproblems: estimation of the genetic network regulating subtilin production from gene expression data, and estimation of population dynamics based on nutrient and population level data. Techniques for identification of switching dynamics from sparse and irregularly sampled observations are developed and applied to simulated data. Numerical results are provided to show the effectiveness of our methods.

Key words: Nonlinear estimation, switching dynamics, sparse sampling, genetic network, biochemical systems.

1 Introduction

During the last decades, different approaches to modeling of biochemical networks have been suggested in the literature. In accordance with the analysis in [9], these can be roughly classified into models with purely continuous dynamics, and discrete event models. The first class comprises, for instance, models based on smooth ordinary or partial differential equations for describing the evolution of protein concentrations. Discrete event systems, on the other hand, are exemplified by Boolean networks, Markov chains, Bayesian networks and graph models of the regulatory network interactions. However, it is by now evident that models featuring both discrete and continuous dynamics are essential to

*This research has been supported by the European Commission under project HYGEIA (NEST-4995).

†Eugenio Cinquemani and John Lygeros are with the Institut für Automatik, Eidgenössische Technische Hochschule Zürich, Switzerland.

‡Riccardo Porreca and Giancarlo Ferrari-Trecate are with the Dipartimento di Informatica e Sistemistica, Università di Pavia, Italy.

derive realistic descriptions of molecular interactions. This fact steered the attention of many researchers to the application of already existing hybrid systems techniques to biological modelling and analysis [15, 1, 10, 5, 2, 11, 4, 16, 6]. In addition, it is being recognized that many biological processes are intrinsically uncertain [28, 21]. For instance, stochastic phenomena appear to be instrumental for certain biochemical processes to improve robustness [30] or induce variability [31, 32], and to play a key role in fundamental processes such as DNA replication [7, 25]. This suggests to explicitly include stochasticity in the modelling framework, leading to the development of stochastic hybrid models of diverse biological processes [22, 17].

Despite the large amount of work on modeling, research on biochemical model identification is still insufficient, particularly for hybrid models. Literature on hybrid biological model identification is only now beginning to appear [12, 14, 26]. In [11, 27], efforts were devoted to specialize general hybrid systems identification techniques to the typical structure of genetic regulatory mechanisms, with clear advantages in terms of the quality of the reconstructed models.

In this paper we concentrate on the identification of the system that regulates the synthesis of the antibiotic subtilin by *Bacillus subtilis*. A stochastic hybrid model for this system was proposed in [19]. In this model, the continuous dynamics of the system depend on the discrete state of a genetic regulatory network. Switches among discrete state values are described in terms of both deterministic and stochastic laws. In [22], the model was further elaborated and cast into the framework of piecewise deterministic Markov processes [8]. Estimation of the subtilin production model poses a number of challenges that are common to many biological identification problems:

- *Measurements at different levels of the biological scale.* Biological systems are studied at different levels of abstraction, which may be roughly classified as shown in Figure 1. At the lowest level, attention is on the behavior of individual molecules, such as different protein complexes in DNA-related processes. At the cell level, (spatio)temporal dynamics of several molecules of the same species are abstracted into the evolution of (local) molecule concentrations. One step above in the hierarchy, focus is on the dynamics of whole populations of cells that stem out of the interaction of individual cell phenomena, e.g. population growth from single cells death or replication. Finally, at the level of species, evolutionary analysis is concerned with the ability of different cell populations to survive to competing species and to progress. Measurements pertaining to different levels of abstraction are generally taken by different experiments, and are difficult to correlate to one another. In our work, which spans the cell and population levels, the issue of correlating concentration and population data is explicitly taken into account.

- *Irregular sampling.* Even though some experimental techniques produce frequently and regularly sampled data, often this is not the case. For instance, this may be due to humans carrying out the experiments, or

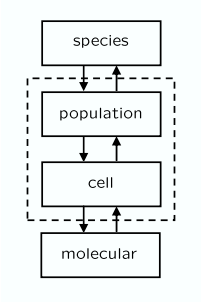

Figure 1: Hierarchy of biological abstraction levels. Dashed line highlights the levels considered in this paper.

to the process itself dictating finer sampling at some places and sparser sampling at others. Our approach allows arbitrary sampling throughout the process. Of course, performance depends on how well sampling fits the process properties.

- *Stochastic variability.* Uncertainty is often seen as an additional difficulty of biological identification. A classical approach is to eliminate data uncertainty by averaging. This can be a wasteful operation in that it reduces the amount of information available. To make better use of probabilistic data, stochastic modeling is essential.

On the other hand, there are some general principles in identification of biological systems. Our work considers all of the following.

- *Coupling of dynamic models at different levels of abstraction.* To couple models at different biological levels, it is typically necessary to introduce artificial coupling models (e.g. constant gains or simple nonlinearities). To identify the coupling model, given the different nature of its input and output variables, data from different experiments must be used. This may be a limitation in that input and output trajectories do not correspond to each other. In our approach, suitable statistics of the input and output processes are reconstructed separately from the corresponding experiments and are put into relation, thus exploiting at full the available information.

- *Decoupled identification of piecewise affine dynamics of genetic regulatory networks.* Switching linear models are often an appropriate description of protein synthesis/degradation by gene activation/inhibition. This paper illustrates how several interconnected models of this kind may be identified separately based on the relevant protein concentration profiles.

- *Use of stochastic variability for parameter estimation.* Randomness provides additional information on the system properties, which we use explicitly for the identification of the model parameters.

We will take the work in [19] as a starting point and focus on estimation of the model parameters. A preliminary study of identification for *B.subtilis* has been carried out in [23] by randomized methods, in a case where a subset of the continuous state variables of the system is observed. In analogy with [11], we specialize identification methods for piecewise affine systems (see e.g. [13], and references therein) to the identification of deterministically switching gene dynamics. Moreover, new techniques for the estimation of the dynamics and of the switching probabilities of Markovian gene expression are introduced. Finally, a modification of the prediction error identification method [24, 29] is applied to the parametric identification of nonlinear switching dynamics driven by stochastic inputs.

The paper is organized as follows. In Section 2, the *B.subtilis* subtilin production model is introduced and the identification problem is stated. Section 3 describes techniques for estimating cell-level parameters of the model from observations of the gene expression profiles. Section 4 deals with estimation of population-level parameters based on macroscopic observations such as population size and nutrient level. A short section (Section 5) is dedicated to the estimation of the threshold on the nutrient level on which the activation of subtilin production depends. In all of these sections, results from numerical simulations are reported to show the effectiveness of our identification techniques. Conclusions and perspectives of our work are given in Section 6. In order to simplify the reading, all proofs are deferred to the appendix.

2 Production of subtilin

Subtilin is an antibiotic synthesized by *B.subtilis* as an adaptive response to changes in the environment, allowing the cell to benefit optimally from the available resources. When the amount of nutrients is sufficient, *B.subtilis* population increases without a remarkable change in subtilin concentration. Subtilin production starts when the amount of nutrient falls under a threshold because of excessive population growth. The role of subtilin is to increase food supply by eliminating competing species and/or other *B.subtilis* cells. In addition to reducing the demand for nutrients, the decomposition of the cells killed by subtilin releases additional nutrients in the environment. The biosynthesis of subtilin is regulated by a positive feedback mechanism in which extracellular subtilin activates the two components regulatory system SpaK and SpaR that binds to a DNA motif promoting the expression of genes for subtilin biosynthesis (spaS and spaBTC) and immunity (spaIFEG). SpaK and SpaR react to form the complex SpaRK that will be used in our model. SpaRK expression is controlled by the sporulation transcription factor SigH. Finally, the composition of SigH is turned on whenever the nutrient concentration falls below a certain threshold. In this paper a simplified model is examined, in which spaBTC and spaIFEG are not taken into consideration.

2.1 Model

We refer to the dynamic model of subtilin production that was originally proposed in [19]. In this model, normalized population level x_1, nutrient level x_2 and the three concentrations $x_3 = [\text{SigH}]$, $x_4 = [\text{SpaRK}]$ and $x_5 = [\text{SpaS}]$ constitute the (nonnegative) continuous part $x = [x_1\ x_2\ x_3\ x_4\ x_5]^T$ of the system state. The discrete part of the state, $S = [S_3, S_4, S_5]^T$, is composed of three binary switches that account for whether the expression of SigH, SpaRK and SpaS genes (in this order) is activated ($S_i = 1$) or inhibited ($S_i = 0$).

The growth of *B.subtilis* population is governed by a logistic-type equation:

$$\dot{x}_1 = r x_1 \left(1 - \frac{x_1}{D_\infty(x_2)}\right). \tag{1}$$

For fixed D_∞, the equilibrium point $x_1 = D_\infty$ is also the asymptotic value of x_1, provided $x_1(0) > 0$. The limiting population is assumed to depend on nutrient level according to

$$D_\infty(x_2) = \min\left\{\frac{x_2}{X_0}, D_{\max}\right\}, \tag{2}$$

where constants $X_0 > 0$ and $D_{\max} > 0$ reflect certain properties of the experimental environment. Nutrient consumption and production are governed by the equation

$$\dot{x}_2 = -k_1 x_1 + k_2 x_5, \tag{3}$$

where k_1 and k_2 are the rate of nutrient consumption per unit of population and the rate of increase in nutrient availability due to the action of subtilin, respectively. The remaining continuous states follow first-order linear dynamics depending on the current value of the corresponding discrete state S_i:

$$\dot{x}_i(t) = \begin{cases} -l_i x_i(t), & S_i(t) = 0, \\ -l_i x_i(t) + k_i, & S_i(t) = 1, \end{cases} \qquad i = 3, 4, 5 \tag{4}$$

where l_i and k_i are the natural degradation and synthesis rates of the corresponding protein complexes. The rule that describes the status of SigH production is deterministic and given by

$$S_3(t) = \begin{cases} 1, & x_2(t) < \eta; \\ 0, & x_2(t) \geq \eta, \end{cases} \tag{5}$$

where threshold η indicates the nutrient level below which subtilin production mechanism is triggered. Finally, switches S_4 and S_5 are modeled as binary random processes, which makes the system a stochastic hybrid one. In [19], a discrete-time model of S_4 and S_5 is expressed in terms of the following switching probabilities: for $i = 4$ and $i = 5$,

$$\begin{aligned} \mathbb{P}[S_i(kT + T) = 1 | S_i(kT) = 0] &= a_0(x_{i-1}(kT)) \\ &= \frac{c_i x_{i-1}(kT)}{1 + c_i x_{i-1}(kT)} \end{aligned} \tag{6}$$

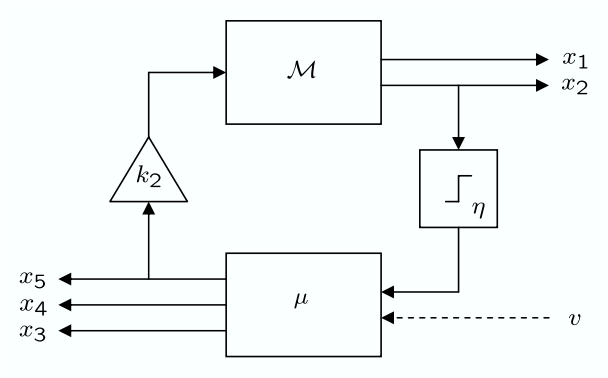

Figure 2: Schematic view of subtilin production model.

and

$$\begin{aligned}\mathbb{P}[S_i(kT+T)=0|S_i(kT)=1] &= a_1(x_{i-1}(kT)) \\ &= \frac{1}{1+c_i x_{i-1}(kT)}. \end{aligned} \tag{7}$$

The implicit assumption is that S_4 and S_5 can be considered constant between samples:

$$S_i(t) = S_i(kT), \qquad \forall t \in [kT, kT+T). \tag{8}$$

The fact that $a_0 = 1 - a_1$ has the following important consequence.

Proposition 1 *If $a_0 = 1 - a_1$ then S_i is an independent process.*

From now on, for simplicity, we shall denote $S_i(kT)$ and $x_i(kT)$ by $S_i(k)$ and $x_i(k)$, respectively. For given values of $x_3(k)$ and $x_4(k)$, S_4 and S_5 are assumed to be conditionally independent and Markovian in the following sense:

$$\begin{aligned} p\big(S_4(k+1), S_5(k+1)|S_4^-(k), S_5^-(k), x_3(k), x_4(k)\big) \\ = p\big(S_4(k+1)|S_4(k), x_3(k)\big) \\ \times p\big(S_5(k+1)|S_5(k), x_4(k)\big), \end{aligned} \tag{9}$$

where $S_i^-(k) = \{S_i(\ell) : \ \ell \leq k\}$ and $p(\cdot|\cdot)$ denotes conditional probability distribution.

This model of subtilin production can be thought of as the interconnection of two deterministic dynamical subsystems with stochastic inputs, as depicted in Fig. 2. Subsystem $\mathcal{M}$ is composed of equations (1) and (3), whereas subsystem μ implements the cascade of switching linear systems (4). Feedback interconnection is operated by two static subsystems: gain k_2 and threshold η. Process v is a fictitious stochastic input that governs the stochastic switching of S_4 and S_5. From a biological point of view, this decomposition reflects two different

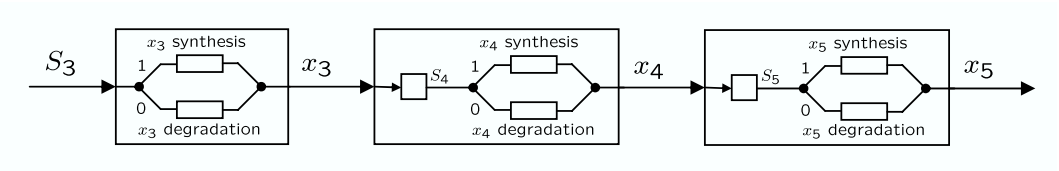

Figure 3: Cascade structure of subsystem μ.

r	X_0	$D_{\max}$	k_1	k_2
0.02	4	1	0.1	0.4

η	k_3	l_3	c_4	k_4	l_4	c_5	k_5	l_5
4	0.5	0.2	0.4	1	0.2	0.4	1	0.2

Table 1: Parameter values in all experiments.

levels in the scale of biological abstraction. Model $\mathcal{M}$ describes the evolution of macroscopic variables such as nutrient and population level. Microscopic phenomena such as (average) gene expression are accounted for by model μ. The cascade structure of model μ is depicted in Figure 3. It reflects the chain of biochemical reactions that regulate subtilin production. The interface between microscopic and macroscopic levels is formalized by converting the amount of available nutrient (x_2) into a triggering signal for the genetic expression mechanism (S_3) and, conversely, by relating the expression level of the SpaS gene to the production of new nutrient.

Fig. 4 shows the trajectories generated by a typical run of the model. Simulation is carried out with sampling time $T = 0.72$sec. The parameter values used in the experiment are shown in Table 1. Hereafter they will be regarded as the "true" parameter values. One may observe that, starting from abundance of nutrients, the nonzero initial population grows freely for about 320 minutes, consuming part of the nutrients available. When the level of nutrients falls below threshold $\eta = 4$, the subtilin production mechanism is triggered. As more subtilin is released, new nutrients become available. Conversely, when threshold η is crossed from below, subtilin production is switched off, and the nutrient level starts to decay. This explains the sustained oscillations of the nutrient level visible for $t \geq 320$min. Fast random switches of the dynamics of [SpaRK] and [SpaS] may be appreciated e.g. in the time interval $[700, 900]$min.

2.2 Discrete-time model

We will always use discrete-time versions of the models of μ and of $\mathcal{M}$ sampled at multiples of T. By assumption (8), the following discrete-time state-space model μ is found by exact integration of (4):

$$x_i(k+1) = \begin{cases} \tilde{l}_i x_i(k), & S_i(k) = 0, \\ \tilde{l}_i x_i(k) + \tilde{k}_i, & S_i(k) = 1, \end{cases} \qquad i = 3, 4, 5 \tag{10}$$

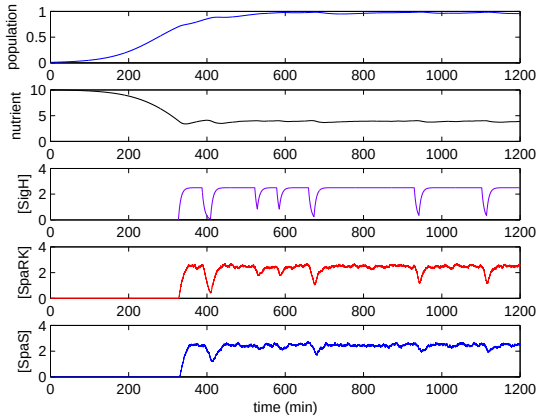

Figure 4: One run of the subtilin production model.

where $\tilde{l}_i \triangleq e^{-l_i T}$ and $\tilde{k}_i \triangleq \frac{k_i}{l_i}(1-\tilde{l}_i)$. On the other hand, integrating (1) (resp. (3)) under the approximation that x_2 (resp. x_1 and x_5) is constant between samples yields the following discrete-time state-space model for $\mathcal{M}$:

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} f\big(x_1(k), x_2(k)\big) \\ x_2(k) - \tilde{k}_1 x_1(k) \end{bmatrix} + \begin{bmatrix} 0 \\ \tilde{k}_2 \end{bmatrix} x_5(k), \tag{11}$$

with $\tilde{r} \triangleq e^{-rT}$, $\tilde{k}_1 = Tk_1$, $\tilde{k}_2 = Tk_2$ and

$$f(x_1, x_2) \triangleq \frac{D_\infty(x_2)}{\left[1 + \left(\frac{D_\infty(x_2)}{x_1} - 1\right)\tilde{r}\right]}.$$

Note that the sampling period T is not related to the time elapsed between consecutive measurements but is just a modeling constant that can be chosen sufficiently small in order to make piecewise constant approximation of the state variables acceptable.

Remark 1 *Instead of using discretization, one may redefine S_4 and S_5 in terms of continuous-time Markov chain processes. In this case, the resulting stochastic hybrid model falls into the class of Piecewise Deterministic Markov Processes (PDMP) [8]. A PDMP model of subtilin production has been presented in [22]. We believe that the identification methods presented in the sequel extend naturally to the PDMP framework, however, this is left for future work.*

2.3 Measurement model

We consider measurements of the population-level variables x_1, x_2 and of the cell-level variables x_3, x_4, x_5 separately. In other words, we allow systems $\mathcal{M}$ and μ to be observed in two different experiments (observation of nutrient and

population levels vs. observation of gene expression levels) starting from the same initial conditions but with possibly different observation times.

Definition 1 *The sequence $Z = \{Z_0, Z_1, \ldots\} \subseteq \mathbb{N}$ is a time scale if $Z_\ell < Z_{\ell+1}$, $\forall \ell \in \mathbb{N}$. For a function $g : Z \to \mathbb{R}$ and numbers $a \in \mathbb{N}$, $b \in \mathbb{N}$ such that $a \leq b$, $g|_{[a,b]}$ denotes the set $\{g(Z_a), g(Z_{a+1}), \ldots, g(Z_b)\}$. If $Z' \subseteq Z$ is a time scale, $g^{Z'}$ is the restriction of g to Z'. $Z = \mathbb{N}$ will be called the natural time scale.*

Let $\mathcal{T}$ and τ be two finite time scales. Measurements are modeled as follows:

$$y_i(\mathcal{T}_\ell) = x_i(\mathcal{T}_\ell) + e_i(\mathcal{T}_\ell), \qquad i = 1, 2, \tag{12}$$

$$y_i(\tau_\ell) = x_i(\tau_\ell) + e_i(\tau_\ell), \qquad i = 3, 4, 5, \tag{13}$$

where the time scales $\mathcal{T}$ and τ denote the observation instants in the two experiments, and the e_i are mutually independent zero-mean white Gaussian measurement noise with known variance σ_i^2. Lags between consecutive measurements will be denoted by $\Delta_\ell \triangleq \mathcal{T}_{\ell+1} - \mathcal{T}_\ell$ and $\delta_\ell \triangleq \tau_{\ell+1} - \tau_\ell$.

2.4 Identification problem

Based on measurements (12)–(13), we want to estimate the parameters of system $\mathcal{M}$, i.e. $\theta_\mathcal{M} = (r, D_{\max}, X_0, k_1)$, those of system μ, i.e. $\theta_\mu = (k_3, l_3, k_4, l_4, c_4, k_5, l_5, c_5)$, and the coupling parameters η and k_2. The idea is to split the identification problem into simpler subproblems:

(I) estimation of θ_μ given $\mathcal{Y}_\mu = \{y_3^\tau, y_4^\tau, y_5^\tau\}$;

(II) estimation of $\theta_\mathcal{M}$ given $\mathcal{Y}_\mathcal{M} = \{y_1^\mathcal{T}, y_2^\mathcal{T}\}$;

(III) estimation of the coupling parameters η and k_2.

In principle, all the above problems are mutually dependent. For instance, the evolution of μ depends on state x_2 through threshold η, but x_2 is not observed in problem (I). However, the only effect of x_2 on μ is to trigger the switches of S_3. Therefore, if $S_3^\mathbb{N}$ was known, one could solve problem (I) without using observations of x_2. In order to decouple problem (I) from (II) and (III) we will apply a segmentation procedure to isolate the switches in the time series y_3^τ. Similarly, system $\mathcal{M}$ is driven through gain k_2 by state x_5, which is not observed when problem (II) is considered. To cope with this, we shall assume that certain statistics of process x_5 are known. In practice, they can be derived separately from multiple experiments on μ. This turns problem (II) into a standard nonlinear stochastic identification problem where $\theta_\mathcal{M}$ and k_2 are the unknown parameters. Finally, a simple ad-hoc method can be used for estimating η based on $\mathcal{Y}_\mathcal{M}$. A more refined method, which makes use of multiple experiments on $\mathcal{M}$ and μ, is presented as well. To summarize, we will proceed as follows:

a. estimation of θ_μ given $\mathcal{Y}_\mu$ (Section 3);

b. estimation of $\theta_{\mathcal{M}}$ and k_2 given $\mathcal{Y}_{\mathcal{M}}$ (Section 4);

c. estimation of η given $\mathcal{Y}_{\mu}$ and $\mathcal{Y}_{\mathcal{M}}$ (Section 5).

Given that the discrete-time model parameters $\tilde{r}$, $\tilde{k}_i$ and $\tilde{l}_i$ are in one-to-one correspondence with r, l_i and k_i, when convenient, we will discuss estimation in terms of $\tilde{r}$, $\tilde{k}_i$ and $\tilde{l}_i$.

3 Identification of μ

We will exploit the cascade structure of the dynamics of μ (Figure 3) and subdivide problem (a) into the separate identification of three switching linear systems of type (10) from the corresponding output measurements y_i. For $i = 3$, switches occur at a low rate (see Figure 4) and we expect several measurements to be available between them. This allows us to set up a segmentation-based procedure that is independent of the switching mechanism. For $i = 4$ and $i = 5$, on the other hand, the switching rate is typically much larger than the measurement rate. This rules out the estimation of the discrete state sequence, rather it suggests to tackle identification based on the statistical properties of process S_i. Identification of μ is arranged in three subsequent steps:

- $i = 3$: Based on measurements y_3^τ, isolate portions of S_3^τ without switches. Within each portion, reconstruct the dynamics of x_3^τ using y_3^τ. This yields estimates of $\tilde{l}_3, \tilde{k}_3$ and of $x_3^{\mathbb{N}}$ (Section 3.1);
- $i = 4, 5$: Based on the estimates of $x_{i-1}^{\mathbb{N}}$, compute the probabilities of switching of $S_i^{\mathbb{N}}$ as a function of c_i, and the expected values $\mathbb{E}[x_i^\tau]$ as a function of $c_i, \tilde{l}_i, \tilde{k}_i$. Next, identify the dynamics of $\mathbb{E}[x_i^\tau]$ using measurements y_i^τ. This yields estimates of $c_i, \tilde{l}_i, \tilde{k}_i$ and of $x_i^{\mathbb{N}}$ (Section 3.2).

3.1 Case $i = 3$

Throughout this section we shall rely on the following simplifying assumption.

Assumption 1 *At most one switch of $S_3^{\mathbb{N}}$ may occur between two consecutive samples of y_3^τ.*

This approximation is supported by the fact that multiple switches between consecutive observations are very seldom observed in the experimental settings of our concern. In practice, the error introduced by Assumption 1 is negligible, as will be apparent from the simulations of Section 3.3.

For a fixed index ℓ, consider the case where $S_3(\tau_\ell) = S_3(\tau_{\ell+1})$, i.e. $S_3^{\mathbb{N}}\big|_{[\tau_\ell, \tau_{\ell+1}]}$ is constant. Then, one has:

$$x_3(\tau_{\ell+1}) = \begin{cases} L_3 x_3(\tau_\ell) + K_3 \ , & \text{if } S_3(\tau_\ell) = 1 \\ L_3 x_3(\tau_\ell) \ , & \text{if } S_3(\tau_\ell) = 0 \end{cases} , \tag{14}$$

where $L_3 = \tilde{l}_3^{\delta_\ell}$ and $K_3 = \tilde{k}_3(1-\tilde{l}_3^{\delta_\ell})/(1-\tilde{l}_3)$. Note that, for each value of ℓ, constants L_3 and K_3 are in one-to-one correspondence with $\tilde{l}_3$ and $\tilde{k}_3$, and hence with l_3 and k_3. This will be used implicitly to compute estimates of l_3, k_3 from estimates of L_3 and K_3. Based on model (14), we propose methods for:

- estimating S_3^τ given l_3 and k_3 (Section 3.1.1);
- estimating l_3, k_3 and $x_3^{\mathbb{N}}\big|_{[\tau_\ell,\tau_{\ell+1}]}$, given S_3^τ (Section 3.1.2);

and combine them to derive an iterative algorithm for the joint estimation of k_3, l_3 and S_3^τ (Section 3.1.3). In this process, based on Assumption 1, estimates $\hat{S}(\tau_{\ell+1}) \neq \hat{S}(\tau_\ell)$ will be flagged as potentially hiding a switch between measurements and will be dealt with accordingly. Finally, we shall describe a method for computing estimates of the whole $x_3^{\mathbb{N}}$ trajectory (Section 3.1.4), which will be used in the identification problems $i = 4, 5$.

3.1.1 Estimation of $S_3(\tau_\ell)$ given l_3 and k_3

The method is based on a statistical test between hypotheses $\mathrm{H}_0 : S_3(\tau_\ell) = 0$ and $\mathrm{H}_1 : S_3(\tau_\ell) = 1$. Consider the following quantity:

$$d_\ell = y_3(\tau_{\ell+1}) - L_3 y_3(\tau_\ell) \ .$$

Proposition 2 *It holds that:*

$$d_\ell \sim \mathcal{N}\left(0, (1+L_3^2)\sigma_3^2\right), \qquad \text{if } \mathrm{H}_0 \text{ holds;} \tag{15a}$$

$$d_\ell \sim \mathcal{N}\left(K_3, (1+L_3^2)\sigma_3^2\right), \qquad \text{if } \mathrm{H}_1 \text{ holds,} \tag{15b}$$

where $\mathcal{N}(m, \sigma^2)$ denotes Gaussian distribution with mean m and variance σ^2.

In both cases, d_ℓ has the same variance. Thus we need to discriminate the mean of d_ℓ based on data $y_3(\tau_\ell)$ and $y_3(\tau_{\ell+1})$. This is a classical statistical problem (see e.g. [18], [3]). Let $f(d_\ell|\mathrm{H}_0)$ and $f(d_\ell|\mathrm{H}_1)$ be the Gaussian density functions appearing in (15a) and (15b), respectively. According to the Neyman-Pearson lemma, an optimal strategy is to compare the likelihood ratio

$$\mathcal{L}(d_\ell) = \frac{f(d_\ell|\mathrm{H}_1)}{f(d_\ell|\mathrm{H}_0)}, \tag{16}$$

to a nonnegative threshold λ. Then, hypothesis H_0 or H_1 is accepted depending on whether $\mathcal{L}(d_\ell) < \lambda$ or $\mathcal{L}(d_\ell) \geq \lambda$. This test is optimal (i.e. most powerful) in the sense that it minimizes the probability α_0 (resp. α_1) of rejecting hypothesis H_0 (resp. H_1) when it is true, for a given value of α_1 (resp. α_0). A simple calculation reveals that λ and the error probabilities α_0 and α_1 verify

$$\begin{aligned}
\alpha_0 &= \Phi\left(-\tfrac{\sigma_3\sqrt{1+L_3^2}}{K_3}\ln(\lambda) - \tfrac{K_3}{2\sigma_3\sqrt{1+L_3^2}}\right) \ , \\
\alpha_1 &= \Phi\left(\tfrac{\sigma_3\sqrt{1+L_3^2}}{K_3}\ln(\lambda) - \tfrac{K_3}{2\sigma_3\sqrt{1+L_3^2}}\right) \ ,
\end{aligned}$$

where $\Phi(\cdot)$ denotes the standard Gaussian distribution function. In our case there is no reason to prefer smaller values of α_0 rather than α_1. Therefore, we shall make the test symmetric by choosing $\lambda = 1$. After algebraic simplifications, the estimator of $S_3(\tau_\ell)$ becomes:

$$\hat{S}_3(\tau_\ell) = \begin{cases} 1, & \text{if } y_3(\tau_{\ell+1}) - L_3 y_3(\tau_\ell) \geq K_3/2; \\ 0, & \text{if } y_3(\tau_{\ell+1}) - L_3 y_3(\tau_\ell) < K_3/2, \end{cases} \tag{17}$$

with probability of error

$$\alpha_0 = \alpha_1 = \Phi\left(-\frac{K_3}{2\sigma_3\sqrt{1+L_3^2}} \right) \ .$$

3.1.2 Estimation of l_3 and k_3 given the S_3 sequence

Assume that $S_3^\tau|_{[m,n]}$ is a known and constant sequence of discrete states. The parameters ℓ_3 and k_3 can be estimated over this window solving the nonlinear least squares problem:

$$\begin{aligned} &\min_{k_3, l_3, x_3(\tau_m)} J_3, \quad J_3 \triangleq \sum_{\ell=m}^{n} \left(y_3(\tau_\ell) - x_3(\tau_\ell) \right)^2, \\ &\text{subject to (14).} \end{aligned} \tag{18}$$

In order for the optimization to be well defined, at least as many observations as unknowns need to be considered, namely, $n-m+1 \geq 3$. We call the solution $\hat{k}_3$, $\hat{l}_3$, $\hat{x}_3(\tau_m)$ to (18) a local estimate of k_3, l_3, $x_3(\tau_m)$. From a local estimate and (14) one can also compute the estimates of $x_3^\tau|_{[m,n]}$. An approximation of the variances $\sigma_l^2 \triangleq \text{var}(\hat{l}_3)$ and $\sigma_k^2 \triangleq \text{var}(\hat{k}_3)$ may also be computed via linearization of J_3 with respect to the unknowns (see [24]). We highlight that, if $S_3^\tau|_{[m,n]} = 0$, the value of $\hat{k}_3$ is irrelevant.

Consider time-scales $\tau(1) \subset \tau, \tau(2) \subset \tau, \ldots, \tau(s) \subset \tau$ that are consecutive (i.e. if $i < j$ then $z < z'$, $\forall z \in \tau(i), z' \in \tau(j)$) and such that $S_3^{\tau(i)}$ is constant, $i = 1, \ldots, s$. By applying the previous procedure one can associate to each time-scale $\tau(i)$ the local estimates $\hat{l}(i)$ and $\hat{k}(i)$ along with their variances $\sigma_l^2(i)$ and $\sigma_k^2(i)$. We describe next an iterative procedure for combining local estimates into global estimates of l_3 and k_3. For sake of conciseness we will restrict our attention to the parameter l_3, being the procedure identical for k_3. We denote with $\hat{l}_3[j]$ the global estimate based on $\hat{l}_3(1), \ldots, \hat{l}_3(j)$ and with $\sigma_l^2[j]$ its variance. Obviously $\hat{l}_3[1] = \hat{l}_3(1)$ and $\sigma_l^2[1] = \sigma_l^2(1)$. Consider now the first two local estimates $\hat{l}_3(1)$ and $\hat{l}_3(2)$. As shown in [24], the global estimate

$$\hat{l}_3[2] = \frac{1}{\sigma_l^{-2}(1) + \sigma_l^{-2}(2)} \big(\sigma_l^{-2}(1)\hat{l}_3(1) + \sigma_l^{-2}(2)\hat{l}_3(2) \big),$$

is the one with minimal variance that is given by $\sigma_l^2[2] = \frac{1}{\sigma_l^{-2}(1)+\sigma_l^{-2}(2)}$. For $j \geq 1$, given $\hat{l}_3[j]$ and $\sigma_l^2[j]$, as a new local estimate $\hat{l}_3(j+1)$ becomes available

the updated global estimate is given by

$$\hat{l}_3[j+1] = \frac{1}{\sigma_l^{-2}(j+1) + \sigma_l^{-2}[j]} \times \big(\sigma_l^{-2}(j+1)\hat{l}_3(j+1) + \sigma_l^{-2}[j]\hat{l}_3[j]\big), \tag{19}$$

and has variance

$$\sigma_l^2[j+1] = \frac{1}{\sigma_l^{-2}(j+1) + \sigma_l^{-2}[j]}. \tag{20}$$

For avoiding to update global estimates of k_3 when $S_3^{\tau(j)} = 0$, we set $\sigma_k^2(j) = +\infty$.

3.1.3 Joint estimation

Let $\hat{l}_3[0]$ and $\hat{k}_3[0]$ be initial guesses of the values of l_3 and k_3. The iterative estimation of k_3, l_3 and S_3^τ reads as follows.

• **Initialization** Set $\ell = 0$ and $j = 0$. Recalling (17), estimate $\hat{S}_3(\tau_0)$ as

$$\hat{S}_3(\tau_\ell) = \begin{cases} 1, & \text{if } y_3(\tau_{\ell+1}) - \hat{L}_3[j]y_3(\tau_\ell) \geq \hat{K}_3[j]/2; \\ 0, & \text{if } y_3(\tau_{\ell+1}) - \hat{L}_3[j]y_3(\tau_\ell) < \hat{K}_3[j]/2, \end{cases} \tag{21}$$

where $\hat{L}_3[j] = e^{-\delta_\ell \hat{l}_3[j]T}$, $\hat{K}_3[j] = \frac{\hat{k}_3[j]}{\hat{l}_3[j]}(1 - \hat{L}_3[j])$.

• **First step** For $\ell \geq 0$ estimate $\hat{S}_3(\tau_\ell)$ according to (21) until a time τ_{ℓ_0} where $\hat{S}_3(\tau_{\ell_0})$ differs from $\hat{S}_3^\tau\Big|_{[0,\ell_0-1]}$. At this point define $\tau(1) = \{\tau_0, \ldots, \tau_{\ell_0-1}\}$, compute $\hat{x}_3(\tau_0)$, $\hat{l}_3(1)$, $\hat{k}_3(1)$ along with their variances as described above and set $\hat{l}_3[1] = \hat{l}_3(1)$, $\hat{k}_3[1] = \hat{k}_3(1)$.

• **Main iteration** Increment counter j. For $\ell \geq \ell_{j-1}$ estimate $\hat{S}_3(\tau_\ell)$ from (21) until a time τ_{ℓ_j} where $\hat{S}_3(\tau_{\ell_j})$ differs from $\hat{S}_3^\tau\Big|_{[\ell_{j-1},\ell_j-1]}$ or until $\tau_\ell = \tau_{\ell_{\max}-1}$ ($\tau_{\ell_{\max}}$ denotes the last element of τ). Define $\tau(j+1) = \{\tau_{\ell_{j-1}}, \ldots, \tau_{\ell_j-1}\}$, compute $\hat{x}_3(\tau_{\ell_{j-1}})$, $\hat{l}_3(j+1)$, $\hat{k}_3(j+1)$ along with their variances and compute $\hat{l}_3[j+1]$, $\hat{k}_3[j+1]$, $\sigma_l^2[j+1]$, $\sigma_k^2[j+1]$ according to (19)–(20). Repeat this step until all data have been processed and let $j_{\max} = j+1$.

• **State estimation** In view of (14), on each time-scale $\tau(j)$ estimate $\hat{x}_3^{\tau(j)}$ by running the model

$$\hat{x}_3(\tau_{\ell+1}) = \begin{cases} \hat{L}_3[j_{\max}]\hat{x}_3(\tau_\ell) + \hat{K}_3[j_{\max}] \ , & \text{if } \hat{S}_3(\tau_\ell) = 1 \\ \hat{L}_3[j_{\max}]\hat{x}_3(\tau_\ell) \ , & \text{if } \hat{S}_3(\tau_\ell) = 0 \end{cases}$$

where the states at the beginning of each time-scale are those computed in the first step and in the main iteration.

The procedure suffers from a limitation. Some time-scales $\tau(i)$ might have cardinality less than 3 thus making impossible to solve problem (18). This issue is circumvented by simply skipping the update of global estimates $\hat{l}_3[j]$, $\hat{k}_3[j]$, thus setting $\hat{l}_3[j+1] = \hat{l}_3[j]$ and $\hat{k}_3[j+1] = \hat{k}_3[j]$ when the cardinality of $\tau(j+1)$ is too small. Moreover, an estimate of x_3 at the beginning of such time-scales, needed in the state estimation step, can be obtained by solving (18) with the additional constraints that $l_3 = \hat{l}_3[j_{\max}]$, $k_3 = \hat{k}_3[j_{\max}]$ that make (18) solvable even if the time-scale comprises just one or two time samples. The estimation of x_3^τ over the whole time span of the experiment is obtained by concatenating local trajectory estimates.

Remark 2 *Initial parameter guesses $\hat{l}_3[0]$ and $\hat{k}_3[0]$ may be drawn from observations taken where S_3 has a known value. For instance, starting an experiment from abundance of nutrients and $x_3(0) = 0$, it is easy to detect the first portion of y_3 data where $S_3 = 1$, because these are just the earliest measurements where y_3 is above the noise level.*

3.1.4 Refinement of the x_3 estimates

Using the algorithm described in Section 3.1.3 one obtains the estimated states $\hat{x}_3^\tau$. However, for the identification of the x_4 and x_5 models estimates $\hat{x}_3^{\mathbb{N}}$ are needed. In view of Assumption 1, if $\hat{S}_3(\tau_\ell) = \hat{S}_3(\tau_{\ell+1})$, then $\hat{x}_3^{\mathbb{N}}\big|_{[\tau_\ell,\tau_{\ell+1}]}$ can be obtained by simulating model (10) with initial state $\hat{x}_3(\tau_\ell)$ and $\tilde{l}_3 = \hat{\tilde{l}}_3 = e^{-\hat{l}_3[j_{\max}]T}$, $\tilde{k}_3 = \hat{\tilde{k}}_3 = \frac{\hat{k}_3[j_{\max}]}{\hat{l}_3[j_{\max}]}(1-\hat{\tilde{l}}_3)$. More critical is the case $\hat{S}_3(\tau_\ell) \neq \hat{S}_3(\tau_{\ell+1})$ because the switch might have occurred at any time $s \in [\tau_\ell, \tau_{\ell+1}]$. In order to reconstruct s, consider the following system of equations:

$$\begin{aligned} \hat{x}_3^{\mathbb{N}}(k+1) &= \begin{cases} \hat{\tilde{l}}_3 \hat{x}_3^{\mathbb{N}}(k) + \hat{\tilde{k}}_3 \hat{S}_3(\tau_\ell), & k < s, \\ \hat{\tilde{l}}_3 \hat{x}_3^{\mathbb{N}}(k) + \hat{\tilde{k}}_3 \hat{S}_3(\tau_{\ell+1}), & k \geq s, \end{cases} \\ \hat{x}_3^{\mathbb{N}}(\tau_\ell) &= \hat{x}_3^\tau(\tau_\ell). \end{aligned} \tag{22}$$

The estimate $\hat{s}$ of the switching time is computed by minimizing the squared difference between $\hat{x}_3^\tau(\tau_{\ell+1})$ and the prediction $\hat{x}_3^{\mathbb{N}}(\tau_{\ell+1})$ that is obtained by solving (22) up to time $\tau_{\ell+1}$. This amounts to solve the finite optimization problem

$$\min_{s\in[\tau_\ell,\tau_{\ell+1}]} \left(\hat{\tilde{l}}_3^{\delta_\ell} \hat{x}_3^\tau(\tau_\ell) - \hat{x}_3^\tau(\tau_{\ell+1}) + \hat{\tilde{k}}_3 \frac{1 - \hat{\tilde{l}}_3^{\tau_{\ell+1}-s}}{1-\hat{\tilde{l}}_3} \right)^2,$$

if $\hat{S}_3(\tau_\ell) = 0$, or

$$\min_{s\in[\tau_\ell,\tau_{\ell+1}]} \left(\hat{\tilde{l}}_3^{\delta_\ell} \hat{x}_3^\tau(\tau_\ell) - \hat{x}_3^\tau(\tau_{\ell+1}) + \hat{\tilde{k}}_3 \frac{\hat{\tilde{l}}_3^{\tau_{\ell+1}-s} - \hat{\tilde{l}}_3^{\delta_\ell}}{1-\hat{\tilde{l}}_3} \right)^2,$$

if $\hat{S}_3(\tau_\ell) = 1$. Finally, estimate $\hat{x}_3^{\mathbb{N}}\big|_{[\tau_\ell,\tau_{\ell+1}]}$ is just the solution of (22) with $\hat{s}$ in place of s.

3.2 Case $i = 4, 5$

Consider the equation:

$$x(k+1) = \begin{cases} \tilde{l}x(k), & S(k) = 0; \\ \tilde{l}x(k) + \tilde{k}, & S(k) = 1, \end{cases}$$

with $\mathbb{P}[S(k+1) = j | S(k) = i] = \pi_{i,j}(z(k); c)$. This models the equations of x_4 and S_4 (resp. x_5 and S_5), with $z = x_3$ (resp. $z = x_4$). Assume that $z^{\mathbb{N}}$ is known, so that $\pi_{i,j}$ is a known function of c. We are interested in estimating parameters $\tilde{l}, \tilde{k}$ and c based on measurements $y^\tau_{[m,n]}$ of $x^\tau_{[m,n]}$, where m and n, with $n > m$, fix the observation window over which estimation is performed.

3.2.1 Mean update formulae

Define $p_i(\cdot) = \mathbb{P}[S(\cdot) = i]$, $p(k) = [p_0(k), p_1(k)]$ and

$$\pi(z(k); c) = \begin{bmatrix} \pi_{0,0}(z(k); c) & \pi_{1,0}(z(k); c) \\ \pi_{0,1}(z(k); c) & \pi_{1,1}(z(k); c) \end{bmatrix}.$$

In addition, for $\ell = m, \dots, n-1$ and $k = 0, 1, \dots, \delta_\ell$, define

$$u(\tau_\ell, k) = \sum_{j=0}^{k-1} \tilde{l}^{k-1-j} p_1(\tau_\ell + j).$$

Proposition 3 *(i) The following recursion holds:*

$$p(k+1) = p(k)\pi(z(k); c), \qquad k = \tau_m, \tau_m + 1, \dots\ , \tag{23}$$

for some initial probability distribution $p(\tau_m)$.
(ii) For any $\ell = m, \dots, n-1$, the following recursion holds:

$$u(\tau_\ell, k) = p_1(\tau_\ell + k - 1) + \tilde{l}u(\tau_\ell, k-1), \quad k = 1, \dots, \delta_\ell, \tag{24}$$

with $u(\tau_\ell, 0) = 0$.
(iii) The following recursion holds:

$$\mathbb{E}[x(\tau_{\ell+1})] = \tilde{l}^{\delta_\ell}\mathbb{E}[x(\tau_\ell)] + \tilde{k}u(\tau_\ell, \delta_\ell), \quad \ell = m, \dots, n-1, \tag{25}$$

for some initial condition $\mathbb{E}[\tau_m]$.

Formulae (23)–(25) allow to compute $\mathbb{E}[x(\tau_\ell)]$ iteratively for $\ell = m, \dots, n$, provided suitable initializations of $p(k)$ and $\mathbb{E}[x(k)]$ at $k = \tau_m$. In practice, the choice of the latter will be part of an optimization procedure, as described in Section 3.2.2. Simpler formulae may be used if switching probabilities a_0 and a_1 are defined as in Section 2.

Corollary 1 *If (6)–(7) hold, then (23)–(24) reduce to:*

$$p(k+1) = \begin{bmatrix} \dfrac{1}{1+cz(k)} & \dfrac{cz(k)}{1+cz(k)} \end{bmatrix}^T, \tag{26}$$

$$u(\tau_\ell, k) = \frac{cz(\tau_\ell + k - 2)}{1 + cz(\tau_\ell + k - 2)} + \tilde{l}u(\tau_\ell, k-1). \tag{27}$$

3.2.2 Identification procedure

The idea is to use $\mathbb{E}[x(\tau_\ell)]$ as an approximation of $x(\tau_\ell)$ and to fit it to the data. For fixed values of $\tilde{l}$, $\tilde{k}$ and c and initial conditions $p(\tau_m)$ and $\mathbb{E}[x(\tau_m)]$, the averages $\mathbb{E}[x(\tau_\ell)]$ are computed iteratively by the formulae of the previous paragraph. Given measurements $y^\tau_{[m,n]}$, estimates $\hat{\tilde{l}}, \hat{\tilde{k}}$, $\hat{c}$, $\hat{p}(\tau_m)$ and $\hat{\mathbb{E}}[x(\tau_m)]$ are then computed as follows:

$$\min_{\tilde{l},\tilde{k},c,S(\tau_m),\mathbb{E}[x(\tau_m)]} J_{4,5}(\tilde{l}, \tilde{k}, c, S(\tau_m), \mathbb{E}[x(\tau_m)]), \tag{28}$$

with

$$J_{4,5} = \frac{1}{2}\sum_{\ell=m}^{n} (y(\tau_\ell) - \mathbb{E}[x(\tau_\ell)])^2 \tag{29}$$

(note that $S(\tau_m) = i$ corresponds to $p_i(\tau_m) = 1$). The above method applies quite easily to the equation of x_4 by setting $z^{\mathbb{N}}\big|_{[\tau_m,\tau_n]} = \hat{x}_3^{\mathbb{N}}\big|_{[\tau_m,\tau_n]}$, where $\hat{x}_3^{\mathbb{N}}\big|_{[\tau_m,\tau_n]}$ can be computed with the algorithms of Section 3.1.4. Similarly, application of the method to the equation of x_5 requires the knowledge of the estimates $\hat{x}_4^{\mathbb{N}}\big|_{[\tau_m,\tau_n]}$. Based on the estimates $\hat{\tilde{l}}_4$, $\hat{\tilde{k}}_4$, $\hat{c}_4$, $\hat{\mathbb{E}}[x_4(\tau_m)]$ and $\hat{S}_4(\tau_m)$ drawn from (28), estimates $\hat{x}_4^{\mathbb{N}}\big|_{[\tau_m,\tau_n]} = \hat{\mathbb{E}}[x_4^{\mathbb{N}}]\Big|_{[\tau_m,\tau_n]}$ are computed by running the equations

$$\begin{aligned} \hat{p}(k+1) &= \hat{p}(k)\pi(\hat{x}_3(k), \hat{c}_4), \\ \hat{\mathbb{E}}[x_4(k+1)] &= \hat{\tilde{l}}_4\hat{\mathbb{E}}[x_4(k)] + \hat{\tilde{k}}_4\hat{p}_1(k) \end{aligned}$$

for $k = \tau_m, \tau_m + 1, \ldots, \tau_n - 1$. By construction, $\hat{\mathbb{E}}[x_4^\tau]\Big|_{[m,n]}$ is the trajectory that optimizes (29), hence $\hat{\mathbb{E}}[x_4^{\mathbb{N}}]\Big|_{[\tau_m,\tau_n]}$ is the best interpolation of x_4 given the available data.

3.3 Results

The methods for the estimation of model μ have been tested on simulated data from the subtilin production model of Section 2.1. Equally spaced observations of x_3, x_4 and x_5 were collected every 1.2, 6 and 12 minutes over the time span $[0, 1200]$min. These situations correspond to constant values of δ_ℓ equal to 100,

T_{obs} (min)	$\hat{l}_3$ (0.2)	$\hat{k}_3$ (0.5)	$\hat{l}_4$ (0.2)	$\hat{k}_4$ (1)	$\hat{c}_4$ (0.4)	$\hat{l}_5$ (0.2)	$\hat{k}_5$ (1)	$\hat{c}_5$ (0.4)
1.2	0.1991	0.4977	0.1900	0.9889	0.3685	0.2078	1.0061	0.4256
6	0.2088	0.5214	0.1829	0.9449	0.3747	0.2112	0.9922	0.4479

Table 2: Estimates of θ_μ in different experimental conditions. T_{obs} denotes the time between consecutive measurements.

500 and 1000, respectively, for a total of $N = 1000, 200$ and 100 observations. Measurement noise levels were set to $\sigma_3 = \sigma_4 = \sigma_5 = 0.06$. This yields noise contributes e_3, e_4 and e_5 approximately within 10% of the values of x_3, x_4 and x_5 at "regime" (e.g. values in the time interval $[700, 900]$min in Figure 4). The algorithm of Section 3.1.3 for the estimation of l_3 and k_3 was applied using minimum cardinality of the time scales equal to 6, and estimated values $\hat{x}_3(k)$ were produced according to the procedure described in Section 3.1.4. As described in Remark 2, initial guesses $\hat{l}_3[0]$ and $\hat{k}_3[0]$ were computed on the basis of the first y_3 data above noise level. Next, the optimization procedure of Section 3.2.2 has been applied to the estimation of l_4, k_4 and c_4. Estimates $\hat{x}_4^{\mathbb{N}}\big|_{[0,\tau_N]}$ were computed as described at the end of the section. Finally, estimation of l_5, k_5 and c_5 has been performed. Table 2 shows the identification results obtained for observation periods, denoted by T_{obs}, of 1.2 and 6 min. For $T_{\text{obs}} = 12$ min, as expected, observations were too sparse to capture the dynamical behavior of the three concentrations and therefore estimation could not be performed. For $T_{\text{obs}} = 6$ min, in particular, estimates of l_3 and k_3 are very similar to the true values, while estimates of l_i, k_i and c_i, for $i = 4, 5$, are less precise. This lower precision can be interpreted on the one hand as the effect of using estimates of x_{i-1} in the identification of these parameters and, on the other hand, considering that concentrations x_4 and x_5, with respect to the estimated average models, are affected by the stochastic nature of S_4 and S_5. However, all estimated parameters differ from their true value for less then 10%, except for c_5, whose relative estimation error is about 12%. Figure 5 shows the noiseless behavior of x_3, x_4, x_5 and the corresponding estimates $\hat{x}_3$, $\hat{\mathbb{E}}[x_4]$, $\hat{\mathbb{E}}[x_5]$ at observation instants with $T_{\text{obs}} = 6$ min, highlighting the good fit of the estimation results.

4 Identification of $\mathcal{M}$

In this section we are concerned with the identification of parameters $\theta_{\mathcal{M}}$ and k_2 from measurements $\mathcal{Y}_{\mathcal{M}}$. Recall that process x_5, which is the input of subsystem $\mathcal{M}$, cannot be measured at this stage.

4.1 Method

Let us make the approximation that $x_5(k)$ is an uncorrelated process. Let us further assume that its time-varying mean $\bar{x}_5(k)$ and variance $\text{var}(x_5(k))$ are known (in practice, they can be estimated from multiple experiments on μ).

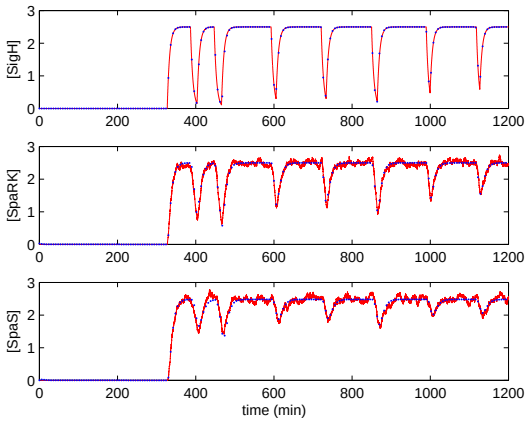

Figure 5: Comparison between concentrations x_3, x_4, x_5 (solid line) and their estimated values at observation instants (dots).

We formulate identification of $\theta_{\mathcal{M}}$ and k_2 in terms of the following optimization problem:

$$\min_{\theta_{\mathcal{M}},k_2} J(\theta_{\mathcal{M}},k_2), \qquad J_{1,2}=\sum_{\ell=0}^{n}||y_{1,2}(\mathcal{T}_\ell)-\hat{x}_{1,2}(\mathcal{T}_\ell)||^2, \tag{30}$$

where $x_{1,2}=[x_1,x_2]^T$, $y_{1,2}=[y_1,y_2]^T$ and $\hat{x}_{1,2}(\mathcal{T}_\ell)$ is an estimate of $x_{1,2}(\mathcal{T}_\ell)$ depending on the values of $\theta_{\mathcal{M}}$ and k_2. In the context of linear stochastic systems, term $\hat{x}_{1,2}(\mathcal{T}_\ell)$ is typically the Kalman predictor associated to the candidate parameter values. This is called the prediction error method (PEM, see e.g. [29]). PEM identification is proven to be asymptotically consistent. That is, provided the system is identifiable, the solution of (30) converges with probability one to the true parameter values when $n\to\infty$. In light of this property, we wish to use a similar approach to the identification of $\theta_{\mathcal{M}}$ and k_2. Because $\mathcal{M}$ is nonlinear, however, we shall write $\hat{x}_{1,2}(\mathcal{T}_\ell)$ as the extended Kalman predictor $\hat{x}_{1,2}(\mathcal{T}_\ell|\ell-1)$ associated to dynamics (11) and measurements (12). For any value $\bar{x}_{1,2}=[\bar{x}_1,\bar{x}_2]^T$ of $x_{1,2}$, let

$$F(\bar{x}_{1,2})=\frac{\partial}{\partial x_{1,2}}\begin{bmatrix}f(x_{1,2})\\ x_2-k_1x_1\end{bmatrix}\Bigg|_{\bar{x}_{1,2}}$$
$$=\begin{cases}\begin{bmatrix}\frac{\tilde{r}f(\bar{x}_{1,2})^2}{\bar{x}_1^2} & \frac{X_0(1-\tilde{r})f(\bar{x}_{1,2})^2}{\bar{x}_2^2}\\ -k_1 & 1\end{bmatrix}, & \frac{\bar{x}_2}{X_0}<D_{\max};\\ \begin{bmatrix}\frac{\tilde{r}f(\bar{x}_{1,2})^2}{\bar{x}_1^2} & 0\\ -k_1 & 1\end{bmatrix}, & \frac{\bar{x}_2}{X_0}\geq D_{\max}.\end{cases}$$

The extended Kalman estimator, $\hat{x}_{1,2}(k|\ell)$, of $x_{1,2}(k)$ given measurements $\{ y_1^{\mathcal{T}}\big|_{[0,\ell]}, y_2^{\mathcal{T}}\big|_{[0,\ell]}\}$ and the (approximate) covariance matrix of the estimation error $x_{1,2}(k) - \hat{x}_{1,2}(k|\ell)$ obey the following recursion:

- Measurement update:

$$\begin{aligned}\Sigma(\mathcal{T}_\ell|\ell) &= [I - K(\ell)]\Sigma(\mathcal{T}_\ell|\ell-1)[I-K(\ell)]^T \\ &\quad + K(\ell)\mathrm{diag}\{\sigma_1^2, \sigma_2^2\}K(\ell)^T, \\ \hat{x}_{1,2}(\mathcal{T}_\ell|\ell) &= \hat{x}_{1,2}(\mathcal{T}_\ell|\ell-1) \\ &\quad + K(\ell)[y_{1,2}(\mathcal{T}_\ell) - \hat{x}_{1,2}(\mathcal{T}_\ell|\ell-1)],\end{aligned}$$

  where $K(\ell) = \Sigma(\mathcal{T}_\ell|\ell-1)[\Sigma(\mathcal{T}_\ell|\ell-1) + \mathrm{diag}\{\sigma_1^2, \sigma_2^2\}]^{-1}$;

- Time update: for $k = \mathcal{T}_\ell, \mathcal{T}_\ell + 1, \ldots, \mathcal{T}_{\ell+1} - 1$,

$$\begin{aligned}\Sigma(k+1|\ell) &= F(\hat{x}_{1,2}(k|\ell))\Sigma(k|\ell)F(\hat{x}_{1,2}(k|\ell))^T \\ &\quad + \mathrm{diag}\{0, \mathrm{var}(x_5(k))\}, \\ \hat{x}_{1,2}(k+1|\ell) &= f(\hat{x}_{1,2}(k|\ell)) + [0, \tilde{k}_2]^T \bar{x}_5(k).\end{aligned}$$

The recursion is initialized by $\hat{x}_{1,2}(0|0) = \mathbb{E}[x_{1,2}]$ and $P(0|0) = \mathrm{Var}(x_{1,2})$. The classical derivation of the extended Kalman filter can be found e.g. in [20]. Our version is a straightforward adaptation to the discrete-time case with Δ_ℓ-steps-ahead prediction. The expression of F is discontinuous due to the presence of term $D_\infty(x_2) = \min\{x_2/X_0, D_{\max}\}$ in f. Note that f is also well regarded as switching dynamics. The complexity of the filter is partially relieved by the linearity of the x_2 dynamics and by the trivial expression of the measurement equation. Intuitively speaking, the approximation that x_5 is an uncorrelated process is especially relevant to the case where the transition probability of S_5 has the form $a_0 = 1 - a_1$, because S_5 (on which x_5 depends) turns out to be an independent process (see Proposition 1). The above algorithm allows one to evaluate the loss functional $J_{1,2}(\theta_{\mathcal{M}}, k_2)$ for arbitrary candidate parameter values. Optimization (30) may be solved by standard numerical methods.

4.2 Results

We tested our PEM identification method on synthetic data generated by the *B.subtilis* model with the true parameter values of Table 1. Equally spaced observations of x_1 and x_2 were collected every 1.2, 6 and 12 minutes, for a total of 1000, 200 and 100 observations over the time span $[0, 1200]$min. Measurement noise levels were set to $\sigma_1 = 0.03$, $\sigma_2 = 0.12$, yielding outcomes of e_1 and e_2 roughly within 10% of the average asymptotic values of x_1 and x_2 (i.e. values after time $t = 600$min in Figure 4). Mean and variance of x_5 were computed empirically at times $\tau_\ell = \ell \times 12$min, with $\ell = 1, \ldots 100$, from 50 independent experiments on μ. Their value over the natural time scale $\mathbb{N}$ was then estimated by spline interpolation. In order to make the optimization of $J_{1,2}$ more robust

r	X_0	$D_{\max}$	k_1	k_2
0.02	1	0.5	0.02	0.1
0.04	2	1	0.04	0.2
0.06	3	1.5	0.06	0.4
0.08	4	2	0.08	0.8
0.1	5	2.5	0.1	1.6

Table 3: Initial guesses of $\theta_{\mathcal{M}}$ and k_2 used to solve (30).

T_{obs} (min)	$\hat{r}$ (0.02)	$\hat{X}_0$ (4)	$\hat{D}_{\max}$ (1)	$\hat{k}_1$ (0.1)	$\hat{k}_2$ (0.4)	min $J_{1,2}$
1.2	0.0194	4.0315	1.0145	0.1013	0.4038	0.0158
6	0.0195	3.9873	1.0294	0.1036	0.4173	0.0164
12	0.0200	4.0040	1.0255	0.1024	0.4163	0.0187

Table 4: Estimates of $\theta_{\mathcal{M}}$ and k_2 in different experimental conditions.

with respect to local minima, numerical minimization was started from the 5 different parameter guesses reported in Table 3. Table 4 shows the identification results obtained for the three different observation rates (as before T_{obs} denotes the observation period), along with the minimum value found for $J_{1,2}$. There is an excellent agreement between the estimated values of r, X_0, $D_{\max}$, k_1, k_2 and their true values, reported in parenthesis in the heading row of Table 4. The increasing values of min $J_{1,2}$ indicate a small loss of performance for larger values of the observation period. Figure 6 shows the extended Kalman state predictions corresponding to the true parameters and those estimated with $T_{\text{obs}} = 12$min. It can be noticed that predictions are very similar. However, predictions with the estimated parameters tend to oversmooth the state trajectory. Note how predictions tend to settle to a constant value, although real state values oscillate. In this regime, predictions with estimated parameters also show a small bias with respect to the optimal ones. Finally, Figure 7 shows simulated runs for the true parameters $\theta_{\mathcal{M}}$, k_2 and their estimates with $T_{\text{obs}} = 12$min. True values of the remaining parameters θ_μ and η were used in both cases. Generated trajectories, including those of the hidden states x_3, x_4 and x_5, are qualitatively indistinguishable.

5 Identification of η

Due to the feedback effect of μ on $\mathcal{M}$, after an initial transient, the level of food x_2 is led to oscillate about level η. This is evident from the observation of experimental x_2 data, and provides a simple way to determine upper and lower bounds to the value of η. However, there is no evidence that x_2 settles asymptotically to the threshold value, therefore this estimation method may not be accurate enough. Better estimates may be drawn by the use of empirical statistics from multiple experiments on μ.

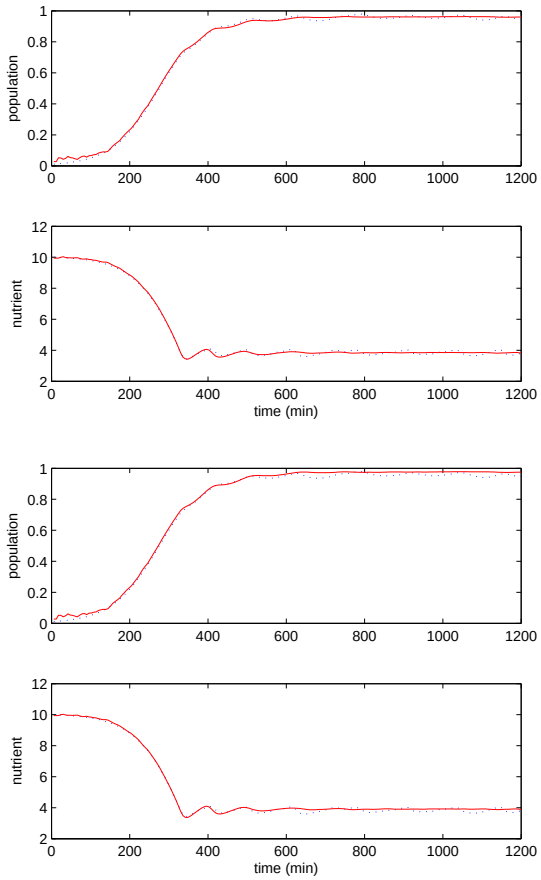

Figure 6: Extended Kalman predictions (solid lines) of the state (dashed lines) with true (top) and estimated (bottom) parameters.

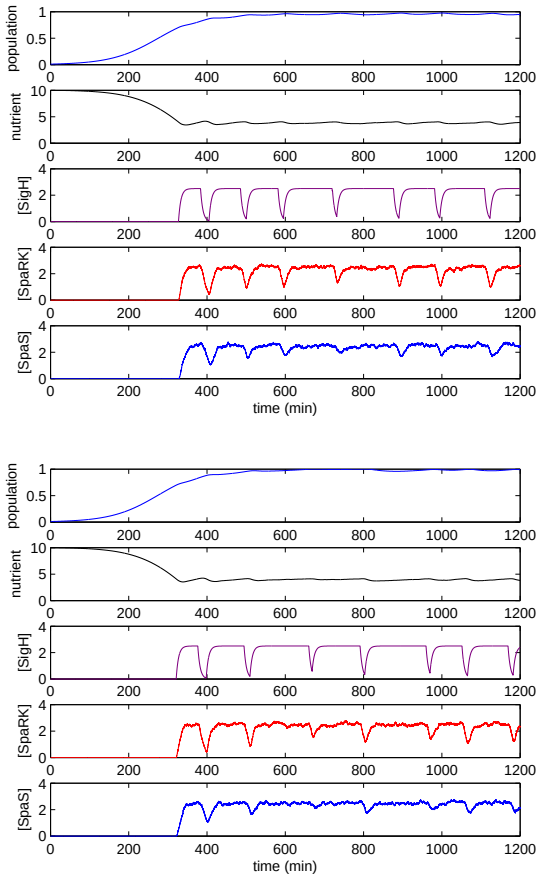

Figure 7: Simulated runs of the model with true (top) and estimated (bottom) parameters.

5.1 Method

Our approach is based on the following observation: at any time k,

$$\begin{aligned}\mathbb{E}[S_3(k)] =& \mathbb{E}[S_3(k)|x_2(k) < \eta]\mathbb{P}[x_2(k) < \eta] + \\ & \mathbb{E}[S_3(k)|x_2(k) \geq \eta]\mathbb{P}[x_2(k) \geq \eta] \\ =& \mathbb{P}[x_2(k) < \eta].\end{aligned}$$

This relation may be obviously generalized to multiple time indexes, for instance:

$$\mathbb{E}[S_3(\mathcal{T}_\ell)] = \mathbb{P}[x_2(\mathcal{T}_\ell) < \eta], \qquad \ell = 0, 2, \ldots, n.$$

Therefore, the idea is to state the identification of η in terms of the following optimization problem:

$$\min_\eta J'(\eta), \qquad J' \triangleq \sum_{\ell=0}^{n} (\hat{\mathbb{E}}[S_3(\mathcal{T}_\ell)] - \hat{\mathbb{P}}[x_2(\mathcal{T}_\ell) < \eta])^2,$$

where $\hat{\mathbb{E}}[S_3(\mathcal{T}_\ell)]$ and $\hat{\mathbb{P}}[x_2(\mathcal{T}_\ell) < \eta]$ are suitable estimates of $\mathbb{E}[S_3(\mathcal{T}_\ell)]$ and $\mathbb{P}[x_2(\mathcal{T}_\ell) < \eta]$ drawn from multiple experiments on μ and $\mathcal{M}$, respectively. A general way to compute them is the following. Consider H experiments on $\mathcal{M}$ with equal observation times $\mathcal{T}_\ell$, with $\ell = 0, \ldots, n$, and h experiments on μ. Let superscript "$^{(i)}$" denote outcomes of the i-th experiment (on $\mathcal{M}$ or on μ). For $\ell = 0, 1, \ldots, n$ one computes:

$$\hat{\mathbb{E}}[S_3(\mathcal{T}_\ell)] \triangleq \frac{1}{h} \sum_{i=1}^{h} \hat{S}_3^{(i)}(\mathcal{T}_\ell) \simeq \mathbb{E}[S_3(\mathcal{T}_\ell)],$$

$$\hat{\mathbb{P}}[x_2(\mathcal{T}_\ell) < \eta] \triangleq \frac{1}{n} \sum_{s=1}^{H} \mathbb{1}_\eta\big(y_2^{(i)}(\mathcal{T}_\ell)\big) \simeq \mathbb{P}[x_2(\mathcal{T}_\ell) < \eta],$$

where $\mathbb{1}_\eta$ is an indicator function that is equal to 1 if $y_2 < \eta$, and is zero otherwise. Note that this requires virtually no assumption on the observation times of μ. Indeed, the method of Section 3.1 provides estimates of S_3 at the natural time scale $\mathbb{N}$ regardless of the choice of τ. The estimation method is expected to converge for $H, h \to \infty$. Yet, for small values of H and h, good estimates may be obtained, provided n is large enough.

5.2 Results

Estimates of η for changing number of experiments and observation period are reported in Table 5. They were all obtained by standard numerical minimization with initial guesses 1, 5 and 10. The numbers of experiments h and H were taken to be the same in each case. Equally spaced observations were considered in the time span $t = [0, 1200]$ (min). Noise level was chosen as in Section 4.2. In the table, T_{obs} indicates the time between observations. Convergence to

T_{obs} (min)	$H = 1$	$H = 10$	$H = 50$
1.2	4.2344	4.0604	4.0365
6	4.25	4.0723	4.0544
12	4.40	4.0285	4.0301

Table 5: Estimates of η in different experimental conditions. The true value is $\eta = 4$.

reasonable estimates was verified in all cases. In general, it is confirmed that accuracy grows with the number of samples and with the number of experiments. However, improvements are limited above small values of H, which suggests that our method is suitable for use with few real-world experiments.

6 Conclusions

In this paper we have studied parameter identification for a model of subtilin production by *B.subtilis*. The model we considered was presented in [19]. In this model continuous-time dynamics for the evolution of nutrients and population are coupled to switching dynamics used to describe the genetic network that regulates the synthesis of subtilin. This results in a parametric stochastic hybrid model. We modeled experimental observation of the system by sparse, possibly irregular sampling of the system state. In addition, we considered measurements of the various components of the state separately, in accordance with their different biological nature. This setting matches the constraints of current experimental techniques in molecular biology. It brings a significant contribution to the modeling of experiments, in that it relaxes typical hypothesis of regular and simultaneous observations that are found in the literature.

We reformulated the model as the composition of a macroscopic and a microscopic dynamical systems coupled by two static systems in a feedback structure. This allowed us to derive methods for the identification of model parameters at both population-level (growth rate, nutrient consumption rate etc.) and cell-level (protein synthesis and decay rates, probabilities of gene expression) based on measurements at the same biological scale. To do this, we borrowed techniques from deterministic hybrid systems and introduced new ideas for the estimation of stochastic switching dynamics. In both cases, we tailored our methods on a typical structure of genetic regulatory chain. This provided contributions to both deterministic and stochastic hybrid identification and the theory of identification of biological systems.

There are several ways this research may be carried on. An obvious one is the validation on real-world data, which may also lead to a refinement of the model and to an optimized design of the biochemical experiments. Another direction is to reformulate *B.subtilis* identification based on the more elaborated PDMP model developed in [22]. In addition, it is our intention to extend this work to more general genetic regulatory mechanisms, e.g. networks that do not have a simple chain structure. Finally, theoretical analysis of the performance of the

estimators presented here is an aim.

This work contributes to an increasing, yet underemphasized effort to provide sound biological hybrid system modeling and identification tools. We believe that this line of research will provide important insights in the understanding of biological systems and give new stimuli to the field of system biology.

A

A.1 Proof of Proposition 1

It is immediate to verify that $\mathbb{P}[S(k+1) = j | S(k) = h]$ is independent of h. It follows that

$$\begin{aligned}\mathbb{P}[S_i(k+1) = j, S_i(k) = h] =& \mathbb{P}[S_i(k+1) = j | S_i(k) = h] \\ & \times \mathbb{P}[S_i(k) = h] \\ =& \mathbb{P}[S_i(k+1) = j] \\ & \times \mathbb{P}[S_i(k) = h],\end{aligned}$$

which is the definition of independence.

A.2 Proof of Proposition 2

Under hypothesis H_1, one has:

$$\begin{aligned}d_\ell &= x_3(\tau_{\ell+1}) + e_3(\tau_{\ell+1}) - L_3\big(x_3(\tau_\ell) + e_3(\tau_\ell)\big) \\ &= L_3 x_3(\tau_\ell) + K_3 + e_3(\tau_{\ell+1}) - L_3 x_3(\tau_\ell) - L_3 e_3(\tau_\ell) \\ &= K_3 + \begin{bmatrix}1 & -L_3\end{bmatrix} \begin{bmatrix}e_3(\tau_{\ell+1}) \\ e_3(\tau_\ell)\end{bmatrix}. \end{aligned} \tag{31}$$

Since $e_3(\tau_{\ell+1})$ and $e_3(\tau_\ell)$ are (jointly) Gaussian, so is (31). In addition:

$$\begin{aligned}\mathbb{E}[d_\ell] &= K_3 + \begin{bmatrix}1 & -L_3\end{bmatrix} \mathbb{E}\begin{bmatrix}e_3(\tau_{\ell+1}) & e_3(\tau_\ell)\end{bmatrix} = K_3, \\ \mathrm{var}(d_\ell) &= \begin{bmatrix}1 & -L_3\end{bmatrix} \mathrm{Var}\left(\begin{bmatrix}e_3(\tau_{\ell+1}) & e_3(\tau_\ell)\end{bmatrix}\right) \begin{bmatrix}1 \\ -L_3\end{bmatrix} \\ &= \sigma_3^2(1 + L_3^2).\end{aligned}$$

The proof is identical under hypothesis H_0 but with $K_3 = 0$.

A.3 Proof of Proposition 3

(i) Straightforward application of Bayes'rule. (ii) Equality $u(\tau_\ell, 0) = 0$ follows from the definition of u. Moreover, for any $k = 1, \ldots, \delta_\ell$,

$$\begin{aligned} u(\tau_\ell, k) &= \sum_{j=0}^{k-1} \tilde{l}^{k-1-j} p_1(\tau_\ell + j) \\ &= p_1(\tau_\ell + k - 1) + \sum_{j=0}^{k-2} \tilde{l}^{k-1-j} p_1(\tau_\ell + j) \\ &= p_1(\tau_\ell + k - 1) + \tilde{l} u(\tau_\ell, k-1). \end{aligned}$$

(iii) Consider the one-step difference equation for $\mathbb{E}[x(k)]$:

$$\begin{aligned} \mathbb{E}[x(k+1)] =& \mathbb{E}[x(k+1)|S(k) = 0]\mathbb{P}[S(k) = 0] \\ &+ \mathbb{E}[x(k+1)|S(k) = 1]\mathbb{P}[S(k) = 1] \\ =& \mathbb{E}[\tilde{l}x(k)|S(k) = 0]\mathbb{P}[S(k) = 0] \\ &+ \mathbb{E}[\tilde{l}x(k) + \tilde{k}|S(k) = 1]\mathbb{P}[S(k) = 1] \\ =& \tilde{l}(\mathbb{E}[x(k)|S(k) = 0]\mathbb{P}[S(k) = 0] \\ &+ \mathbb{E}[x(k)|S(k) = 1]\mathbb{P}[S(k) = 1]) \\ &+ \tilde{k}\mathbb{P}[S(k) = 1] \\ =& \tilde{l}\mathbb{E}[x(k)] + \tilde{k}p_1(k). \end{aligned}$$

By repeated substitutions one gets:

$$\mathbb{E}[x(\tau_{\ell+1})] = \tilde{l}^{\delta_\ell}\mathbb{E}[x(\tau_\ell)] + \tilde{k} \sum_{s=0}^{\delta_\ell - 1} \tilde{l}^{\delta_\ell - 1 - s} p_1(\tau_\ell + s),$$

where the summation is recognized to be equal to $u(\tau_\ell, \delta_\ell)$.

A.4 Proof of Corollary 1

Equation (26) is found from (23) as follows:

$$\begin{aligned} p_1(k+1) &= p_0(k)\pi_{0,1}(z(k)) + p_0(k)\pi_{1,1}(z(k)) \\ &= p_0(k)a_0(z(k)) + p_1(k)\big(1 - a_1(z(k))\big) \\ &= p_0(k)a_0(z(k)) + p_1(k)a_0(z(k)) \\ &= a_0(z(k)) \\ &= \frac{cz(k)}{1 + cz(k)}. \end{aligned}$$

Equation (27) is found by substitution of (26) in (24).

References

[1] R. Alur, C. Belta, F. Ivancic, V. Kumar, M. Mintz, G. Pappas, H. Rubin, J. Schug, and G.J. Pappas. Hybrid modeling and simulation of biological systems. In M. Di Benedetto and A. Sangiovanni-Vincentelli, editors, *Hybrid Systems: Computation and Control*, number 2034 in LNCS, pages 19–32. Springer-Verlag, Berlin, 2001.

[2] K. Amonlirdviman, N. A. Khare, D. R. P. Tree, W.-S. Chen, J. D. Axelrod, and C. J. Tomlin. Mathematical modeling of planar cell polarity to understand domineering nonautonomy. *Science*, 307(5708):423–426, Jan. 2005.

[3] M. Basseville and I. Nikiforov. *Detection of abrupt changes: theory and application*. Prentice-Hall, Inc., New Jersey, 1993.

[4] G. Batt, D. Ropers, H. de Jong, J. Geiselmann, R. Mateescu, M. Page and D. Schneider. Validation of qualitative models of genetic regulatory networks by model checking: Analysis of the nutritional stress response in Escherichia coli. *Bioinformatics*, 21(Suppl 1):i19-i28, 2005.

[5] G. Batt, D. Ropers, H. de Jong, J. Geiselmann, M. Page, and D. Schneider. Qualitative analysis and verification of hybrid models of genetic regulatory networks: Nutritional stress response in *Escherichia coli*. In L. Thiele and M. Morari, editors, *Hybrid Systems: Computation and Control*, number 3414 in LNCS, pages 134–150. Springer-Verlag, Berlin, 2005.

[6] M. Chaves, E.D. Sontag and R. Albert. Methods of robustness analysis for Boolean models of gene control networks. IEE Proc. Systems Biology, 153(4), pp. 154–167, 2006.

[7] J. Dai, R.Y. Chuang, and T.J. Kelly. DNA replication origins in the *schizosaccharomyces pombe* genome. *PNAS*, 102(102):337–342, January 2005.

[8] M.H.A. Davis. Piecewise deterministic Markov processes: a general class of non-diffusion stochstic models. *Journal of the Statistical Royal Society*, B, 46(3):353:388, 1984.

[9] H. de Jong. Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology*, 79:726–739, 2002.

[10] H. de Jong, J.-L. Gouze, C. Hernandez, M. Page, T. Sari, and J. Geiselmann. Hybrid modeling and simulation of genetic regulatory networks: A qualitative approach. In O. Maler and A. Pnueli, editors, *Hybrid Systems: Computation and Control*, number 2623 in LNCS, pages 267–282. Springer Verlag, Berlin, 2003.

[11] S. Drulhe, G. Ferrari-Trecate, H. de Jong, and A. Viari. Reconstruction of switching thresholds in piecewise-affine models of genetic regulatory networks. In J. Hespanha and A. Tiwari, editors, *Hybrid Systems: Computation*

and Control, volume 3927 of *Lecture Notes in Computer Science*, pages 184–199. Springer Verlag, Berlin, 2006.

[12] M.J. Dunlop, E. Franco, R.M. Murray. A multi-model approach to identification of biosynthetic pathways. To appear in *Proceedings of the 26th American Control Conference*, 2007.

[13] G. Ferrari-Trecate, M. Muselli, D. Liberati and M. Morari. A clustering technique for the identification of piecewise affine systems. *Automatica*, 39:205–217, 2003.

[14] K. Fujarewicz, M. Kimmel, and A. Swierniak. On fitting of mathematical models of cell signaling pathways using adjoint systems. *Mathematical Biosciences and Engineering*, Vol.2, No.3, pp.527534, 2005.

[15] R. Ghosh and C. Tomlin. Lateral inhibition through delta-notch signalling: A piecewise affine hybrid model. In M. Di Benedetto and A. Sangiovanni-Vincentelli, editors, *Hybrid Systems: Computation and Control*, number 2034 in LNCS, pages 232–246. Springer-Verlag, Berlin, 2001.

[16] R. Ghosh and C.J. Tomlin. Symbolic reachable set computation of piecewise affine hybrid automata and its application to biological modeling: Delta-Notch protein signaling, IEE Transactions on Systems Biology, Volume 1, Number 1, pp. 170–183, June 2004.

[17] J. Hespanha and A. Singh. Stochastic models for chemically reacting systems using polynomial stochastic hybrid systems. *International Journal on Robust Control*, 15(15):669–689, September 2005.

[18] R.V. Hogg, A. Craig. *Introduction to Mathematical Statistics.* Prentice Hall, 1994.

[19] J. Hu, W.C. Wu, and S.S. Sastry. Modeling subtilin production in *Bacillus subtilis* using stochastic hybrid systems. In R. Alur and G.J. Pappas, editors, *Hybrid Systems: Computation and Control*, No.2993 in LNCS, pages 417–431. Springer-Verlag, Berlin, 2004.

[20] A.H. Jazwinski. *Stochastic Processes and Filtering Theory.* Academic Press, Inc., New York, 1970.

[21] M. Kaern, T.C. Elston, W.J. Blake, and J.J. Collins. Stochasticity in gene expression: From theories to phenotypes. *Nature Reviews Genetics*, 6(6):451–464, 2005.

[22] P. Kouretas, K. Koutroumpas, J. Lygeros and Z. Lygerou. Stochastic hybrid modelling of biochemical processes. In C.G. Cassandras and J. Lygeros, editors, *Stochastic Hybrid Systems*, Vol.24, No.9083, Automation and Control Engineering Series, CRC press, November 2006.

[23] K. Koutroumpas, E. Cinquemani, P. Kouretas and J. Lygeros. Parameter identification for stochastic hybrid systems: An application to subtilin production by *Bacillus subtilis*. Submitted to *Control Engineering Practice*.

[24] L. Ljung. *System identification – Theory for the user.* Prentice-Hall, Inc., New Jersey, 1999.

[25] P.K. Patel, B. Arcangioli, SP. Baker, A. Bensimon, and N. Rhind. DNA replication origins fire stochastically in the fission yeast. *Molecular Biology of the Cell*, 17(1):308–316, January 2006.

[26] T.J. Perkins, M. Hallett and L. Glass. Inferring models of gene expression dynamics. *J. Theor. Biol.*, Vol.230, No.3, pp.289–299, 2004.

[27] R. Porreca, G. Ferrari-Trecate, D. Chieppi, L. Magni and O. Bernard. Switch detection in genetic regulatory networks. Technical Report 142/06, Università degli Studi di Pavia, 2006. http://sisdin.unipv.it/lab/publications/TechRepSDGRN.pdf.

[28] C.V. Rao, D.M. Wolf, and A.P. Arkin. Control, exploitation and tolerance of intracellular noise. *Nature*, 420(6912):231–237, 2002.

[29] T. Soderstrom and P. Stoica. *System identification*. Prentice Hall, UK, 1989.

[30] JM. Vilar, HY. Kueh, N. Barkai, and S. Leibler. Mechanisms of noise-resistance in genetic oscillators. *PNAS*, 99:5988–5992, 2002.

[31] D. Wolf, V.V. Vazirani, and A.P. Arkin. Diversity in times of adversity: probabilistic strategies in microbial survival games. *Journal of Theoretcial Biology*, 234(2):227–253, 2005.

[32] L. Wu, J.C. Burnett, J.E. Toettcher, A.P. Arkin, and D.V. Schaffer. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 tat fluctuations drive phenotypic diversity. *Cell*, 122(2):169–182, 2005.