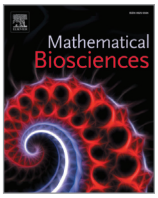
Mathematical Biosciences 279 (2016) 83–89

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs



Identification and automatic segmentation of multiphasic cell growth using a linear hybrid model

András Hartmann[a,*], Ana Rute Neves[b], João M. Lemos[c], Susana Vinga[a]

[a] *IDMEC, Instituto Superior Técnico, Universidade de Lisboa - Av. Rovisco Pais, Lisbon 1049-001, Portugal*
[b] *Discovery, Chr. Hansen A/S - Bøge Alle 10–12, Hørsholm 2970, Denmark*
[c] *INESC-ID, Instituto Superior Técnico, Universidade de Lisboa - R Alves Redol 9, Lisbon 1000-029, Portugal*

ARTICLE INFO

Article history:
Received 16 October 2015
Revised 25 June 2016
Accepted 30 June 2016
Available online 15 July 2016

Keywords:
Hybrid systems
Bacterial growth model
Diauxic growth
Multiphasic growth
System identification

ABSTRACT

This article considers a new mathematical model for the description of multiphasic cell growth. A linear hybrid model is proposed and it is shown that the two-parameter logistic model with switching parameters can be represented by a Switched affine AutoRegressive model with eXogenous inputs (SARX). The growth phases are modeled as continuous processes, while the switches between the phases are considered to be discrete events triggering a change in growth parameters. This framework provides an easily interpretable model, because the intrinsic behavior is the same along all the phases but with a different parameterization. Another advantage of the hybrid model is that it offers a simpler alternative to recent more complex nonlinear models. The growth phases and parameters from datasets of different microorganisms exhibiting multiphasic growth behavior such as *Lactococcus lactis, Streptococcus pneumoniae*, and *Saccharomyces cerevisiae*, were inferred. The segments and parameters obtained from the growth data are close to the ones determined by the experts. The fact that the model could explain the data from three different microorganisms and experiments demonstrates the strength of this modeling approach for multiphasic growth, and presumably other processes consisting of multiple phases.

1. Introduction

Many microorganisms exhibit multiphasic growth behavior. A special case of multiphasic growth caused by substrate preference is diauxic growth, which was first studied in the 40’s by Monod [14,15]. This phenomenon arises when an organism is growing on a medium consisting of two (or more) different types of carbon and energy sources. Among others, *Streptococcus pneumoniae* shows diauxic growth on mixed medium. The organism first consumes the substrate that supports the fastest growth (preferred substrate) followed by consumption of the remaining secondary carbon source(s). Theoretically, the process includes two types of phases: exponential growth and diauxic lag. The bacteria process the preferred substrate in an initial exponential growth phase. Then a diauxic lag is followed, when the bacteria do not grow significantly but synthesize enzymes in order to be able to process the next medium in another exponential growth phase. The cycle repeats until no further carbon sources are available.

Yeast (*Saccharomyces cerevisiae*) shows another type of multiphasic growth, where the diauxic behavior is caused by ethanol produced by the fungi itself. The preferred source of carbon and energy of the organism is glucose, but while metabolizing glucose, the cells release ethanol in the medium. When the glucose source becomes limiting, the cells exhibit a so called *diauxic shift* by switching their metabolism to aerobic utilization of ethanol. Compared to glycolysis, the ethanol phase is characterized by decreased growth rate [10].

Inhibitions or metabolic burdens can also cause multiphasic growth. In the *Lactococcus lactis* example considered [16], the biphasic growth behavior can be derived from a metabolic burden due to overexpression of proteins or a metabolic imbalance caused by accumulation of a toxic intermediate.

In order to infer the properties of the process, such as maximal growth, the time-series has to be segmented and a model has to be fitted to each segment. In many cases, segmentation and fitting is still done by hand (usually fitting a linear model to the logarithm of selected points) and checked visually by the experts. Despite tools for automated fitting of biological growth curves like BGFit [22] are readily available online for fitting bacterial growth

---

[*] Corresponding author.
E-mail addresses: andras.hartmann@tecnico.ulisboa.pt (A. Hartmann), dkruns@chr-hansen.com (A.R. Neves), jlml@inesc-id.pt (J.M. Lemos), susanavinga@tecnico.ulisboa.pt (S. Vinga).

http://dx.doi.org/10.1016/j.mbs.2016.06.013
0025-5564/

data, they do not consider multiple phases. In this paper a novel approach is considered for multiphasic cell growth segmentation and modeling using hybrid linear systems.

A number of different Ordinary Differential Equation(ODE) models were proposed to describe the growth of biomass on different species starting with the famous hyperbolic Michelis–Menten like equations of [15] to the more sophisticated recent developments. A simple idea to address multiphasic growth is to introduce a lag representing the time-shift between different logistic models [2]. Some approaches include additional state variables incorporating further factors in the model, like enzyme levels or substrate concentrations [5,6,8,20], however this also causes an increased number of parameters, where usually a subset of the parameters is not directly estimated but taken from the literature [5,6].

Hybrid models consist of both continuous and a discrete states (parameters) [13]. The advantage of using hybrid systems is that the dynamics typically can be modeled with continuous state evolution and the transitions between submodels are represented by the changes of discrete states. From strictly computational perspective, it is plausible to interpret the multiphasic growth phenomenon as a hybrid system, because the same model may describe all the phases of the process, while the parameters vary between the phases.

Here a switched hybrid linear model is proposed that is able to explain multiphasic growth data. The model is continuous, and the switches between the phases are modeled with discrete events. Compared to previous models, the advantages of the hybrid model is its linearity, simple and straightforward interpretability, the reduced number of parameters, yet without loss of descriptive power.

2. Methods

In this section, first the proposed model for multiphasic growth is described, then the identification algorithm is reviewed, finally the biological datasets are introduced. A sample MATLAB implementation of the methodology and the proposed model is freely available is freely available under the terms of GNU Public License (GPLv3) from the authors webpage: http://andrashartmann.info/.

The goal of modeling is to accurately describe a given process. Hereafter, a first-order Switched affine AutoRegressive model with eXogenous inputs (SARX) model [17] is proposed for multiphasic growth. SARX models are switched extensions to affine linear autoregressive models, defined as the concatenation of several submodels. Each submodel corresponds to an AutoRegressive model with eXogenous inputs (ARX) of fixed dimension. Consider the following system in input-output form

$$y(t) = \theta^{\mathrm{T}}_{\eta(t)} \begin{bmatrix} \varphi(t) \\ 1 \end{bmatrix} + \varepsilon(t) \tag{1}$$

$$\varphi(t) = [y(t-1) \dots y(t-n_a)\, u(t) \dots u(t-n_b)]^{\mathrm{T}}, \tag{2}$$

where the input $u(t)$ is observed and so is the output $y(t)$ which is corrupted by an additional noise term, $\varepsilon(t)$. When fitting the model, $y(t)$ corresponds to the actual observations. The regression vector $\varphi(t)$ of dimension $n = n_a + n_b$ consists of past measurements and inputs. The notation $\Phi(t) = [\begin{smallmatrix} \varphi(t) \\ 1 \end{smallmatrix}]$ is introduced for the extended regression vector. The time-dependent parameter vector $\theta(t) = \theta^{\mathrm{T}}_{\eta(t)} \in \mathbb{R}^n$ at each time-instance belongs to a set $\Theta = \{\theta_1, \dots, \theta_K\}$ of cardinality K, representing the submodel set. The discrete finite range function, $\eta(t) : \mathbb{R} \to \{1, \dots, K\}$ indicates which submodel generates the output at time-instance t, and is referred as switching sequence or discrete state.

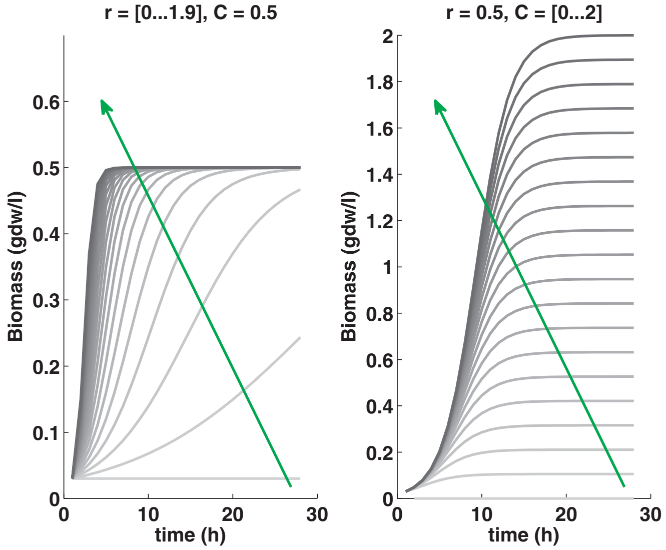


Fig. 1. Simulation example with different parameters. The arrows point towards growing parameters. Also the darker lines represent larger parameter values.

2.1. Logistic growth

As shown in [12], the discrete Verhulst logistic model for monophasic growth [24] has a linear representation, therefore a switching extension of the model can be corresponded to a SARX model. Hereafter, a first-order SARX model is proposed for multiphasic growth by showing that the continuous logistic model also can be represented by an affine AutoRegressive model with eXogenous inputs (ARX). Consequently, when considering switching parameters, the model can be represented by SARX. It is more convenient to work with the continuous model, because no discretization is needed.

The starting point of the model is the continuous two-parameter logistic model.

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \dot{x}(t) = rx(t)\left(1 - \frac{x(t)}{C}\right), \tag{3}$$

where the state variable x represents the biomass, the parameter r refers to the maximum growth rate, and C is the carrying capacity (level of saturation). To identify logistic models of longer lag time, some authors [18,24] suggest to fit the logistic model log-measurements $\log(x)$. Here, no log-transformation was applied because no long lag times were expected in the beginning of growth phases. Instead, lag times are considered to be separate phases. The differential equation in the form shown in Eq. (3) is nonlinear. However, since the biomass is strictly positive ($x > 0$), both sides of Eq. (3) can be divided by $x(t)$. Applying the chain rule for the derivation, this also corresponds to the logarithmic derivative

$$\frac{\dot{x}(t)}{x(t)} = r\left(1 - \frac{x(t)}{C}\right) = \frac{\mathrm{d}\log x}{\mathrm{d}t}, \tag{4}$$

with the substitutions of variables

$$y(t) = \frac{\dot{x}(t)}{x(t)}; \quad \theta = \begin{bmatrix} -\frac{r}{C} \\ r \end{bmatrix}, \tag{5}$$

the model may be represented by the following linear autoregressive form

$$y(t) = \theta^{\mathrm{T}} \begin{bmatrix} x(t) \\ 1 \end{bmatrix}. \tag{6}$$

The flexibility of the model is shown in Fig. 1, by simulating with different parameters r and C.