Hybrid system modeling and identification of cell biology systems: perspectives and challenges

Dirk Vries [*,**] **Peter J.T. Verheijen** [**] **Arnold J. den Dekker** [*]

[*] *Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands*
[**] *Biotechnology Department, Delft University of Technology, Delft, The Netherlands*

Abstract: Some perspectives and challenges in research on the crossing of system biology, hybrid system formulations and system identification are outlined. Emphasis is given on the hybrid, gray box modeling of interactions between the different abstract levels of organization typically recognized in (micro-)organisms, its associated identification problem and optimal experimental (input) design.

Keywords: system biology, system identification, input design, hybrid systems

1. INTRODUCTION

Since the emergence of high throughput techniques like automated DNA sequencing, but also ^{13}C-labeling techniques, a lot of data originating from several species has been generated, published and stored in databases during the last two decades. This enormous amount of data, predominantly on the genetic level, has lead to the desire to integrate the available knowledge and data in order to understand the functioning and interactions appearing in an organism or population of organisms.

The idea in the field of Systems Biology is to fulfill this desire in a systematic manner. For this purpose, but also for prediction, there is a lot of effort in putting first principle knowledge of biological mechanisms in dynamic models. Already is the systems and control community learning in and aiding to the modeling, analysis (Alur et al., 2002; Hu et al., 2004), experimental design (Sontag, 2008), identification procedures (Riel and Sontag, 2006; Drulhe et al., 2008) and control (Doyle, 2008) aspects associated with the myriad of biological interactions and processes.

Naturally, a question which arises is: *what* has to be identified? From a biology perspective, an organism can be studied on one or more organizational levels, typically ranging from the genetic, molecular interaction level, the level of cells and their functional sub-units (organelles) to the level of cell populations and organs. There are many measurement techniques associated with each set of compounds of interest. Of course, for a model describing the myriad of interacting processes at these different organizational levels, the plethora of measurement techniques complicates the identification of the model.

The nature of experimentation, i.e. characterization of a certain compound versus measurements of the quantity of a certain compound, and the available biological knowledge motivates the use of identification methods for hybrid dynamic models. Hybrid models typically describe both continuous dynamical behavior and discrete transitions between discrete modes, the latter e.g. being a switch turned from mode 'on' to 'off'.

There are many open problems when it comes to hybrid model identification of biological systems. In this contribution, a brief overview is given of to the challenges associated with identification of hybrid-type, cell microbiology systems. Additional attention is given towards ways how to profit from the degrees of freedom in experimental conditions and input signals to improve the model identification a priori.

The paper is organized as follows. Section 2 gives a brief overview of organizational levels within so-called *unicellular micro-organisms* together with commonly used measurement techniques. The use of hybrid modeling for these systems is further motivated in Section 3. Some modeling paradigms and associated identification issues are presented in Section 4. Section 5 covers some available degrees of freedom when it comes to manipulation of experiment conditions and input design. Finally, some conclusions are presented in Section 6.

2. BIOLOGICAL MECHANISMS AND MEASUREMENT TECHNIQUES

A short background on the various recognized organizational levels within microorganisms follows. A more elaborate and informative description can be found in the survey texts (Sontag, 2003; Crampin, 2006) or the textbook (Klipp et al., 2005).

2.1 Genes and their expression

At the base level, genetic information of an organism is stored in the *genome*, i.e. a collection of genes, and is encoded with the double-stranded molecules DNA (deoxyribonucleic acid). It can either be transcribed and subsequently translated [1] into proteins for further regulation or other functioning, it can have structural significance for e.g. biochemical stability or DNA can be doubled when the cell is reproducing itself. This DNA reproduction process is called *replication*. The process from gene to protein is often regulated by metabolites and generally also by other proteins, called transcription factors. Such an interacting network of regulated genes is called a *genetic regulatory network* (GRN). DNA replication, as well as the readout from gene to protein and its regulation, is shown simplified in Fig. 1.

[1] *Transcription* is the process of reading DNA and storing the information in messenger RNA which are in turn precursors of the *translation* process to proteins.

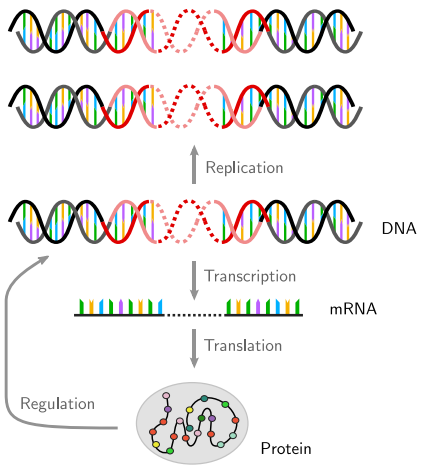

Fig. 1. Schematic view of replication and the transcoding route from a gene (shown in red) via mRNA to protein. Bars represent nucleotides, gray arrows indicate processes and dots represent a sequence not shown.

Measurement techniques. Advanced genetic engineering tools, such as gene deletions, insertions, polymerase chain reaction (PCR) [2] , but also measurement technologies like fluorescence microscopy, microarrays and blotting, have contributed to the generation of massive amounts of data typically stored in genomics and proteomics databases. These databases typically store chemical characterizations of gene or protein sequences found in species, but do not cover quantitative information. The involved measurement technologies have in common that the detection of particular compounds or fragments is based on light intensity measurements at particular wavelengths or on radioactivity intensity, see e.g. Klipp et al. (2005). Since thousands of genes may play a role, clustering techniques, singular value decomposition and truncation is typically used for static and time series analysis of microarray data (Crampin, 2006).

2.2 Metabolites and Proteins

A metabolic network is a set of chemical reactions between (bio)chemical compounds called *metabolites*, like amino and fatty acids, sugars, etc. There is a tied interplay between the metabolic and genetic regulatory network of a living cell and these interacting networks establish control mechanisms in order to maintain life, i.e. grow and reproduce, maintain their structures and respond to their environments. Proteins catalyze these reactions and can also function in the genetic regulatory network. Catalyzing proteins are called enzymes. Besides catalyzing, blocking, slowing down and regulating genes, some proteins function as receptors that endow the cell with sensing capabilities or actuators [3] . They also provide structural support of the cell, help in the transport of smaller molecules, as well as direct the breakdown and reassembly of other cellular elements such as lipids (structures composed of fatty acids) and sugars.

Measurement techniques. For *in vitro* analysis or concentration measurements [4] , a sequence of laboratory steps have to be made, i.e. (i) metabolites and proteins must be purified away from other cellular components, (ii) enzyme activity should be stopped and (iii) further chemical reactions should be prevented, see again the work of Klipp et al. (2005). Besides detection of metabolites, it is possible to measure metabolite concentrations with the aid of tagged ^{13}C-atoms and rapid sampling. Therefore, this ^{13}C–labeling technique is particularly suited for collecting time series response data of metabolites after step or pulse experiments, see e.g. Nöh et al. (2006).

2.3 Cells, Population and Environment

Cells have different phases like *cell division* with their own associated processes like DNA replication and cell built-up, *maintenance of metabolism* for repair and survival purposes and ultimately *death and cell lysis*. The average number of a set of cells, say $x(t)$ on a 'macroscopic' population level has similar definitions: growth ($\dot{x} > 0$), a stationary or adaptation phase ($\dot{x} \approx 0$) and population decay ($\dot{x} < 0$). Different growth conditions can be stimulated *in vivo* by manipulating the cultivation conditions in a bioreactor setup.

Measurement techniques. There are many techniques available for cell density (or biomass concentration) measurements. Fluorescence-tagging is the most widely used method and is mostly used as a detection method. However, it performs quantitatively acceptable under strictly controlled cultivation conditions. Moreover, fluorescence sensors for biomass determination are well developed. Other methods are dielectric spectroscopy and flow cytometry. A discussion of density measurement techniques can be found in (Madrid and Felice, 2005).

3. DATA CHALLENGES AND MOTIVATION HYBRID MODELING

3.1 Data challenges

With increasing available knowledge and the desire to understand every mechanism in an organism, non-linear network models of already large model order are expanded, especially in the metabolic network research. These are mostly white or 'light gray'–box models, i.e. models which are highly deterministic. One often tries to validate these models by new experimental findings, e.g. when studying another organism. However, identification of those large networks models is severely limited or even impossible since the biological experimental results are subject to one or more of the following:

- Different experimental sampling density, i.e. the sampling density of data can vary a lot, ranging from a high to a low sampling density. For example, real-time measurement of gene expression with reporter-gene assays or ^{13}C-labeling combined with rapid sampling provide quite accurate measurements whereas measurements of protein concentrations with so-called Western-Blot assays, only provide a few data points with low sampling density and precision. A considerable amount of data with low sampling density is collected, typically by microarray, blotting and other detection techniques.
- Low precision. Most measurement techniques are still collected by humans, introducing large measurement uncertainties. Even worse, often, different experimental conditions are considered. For a given biological system, it is

[2] PCR amplifies DNA, typically a (part of a) gene, in large enough quantities to enable the use of measurement techniques.

[3] A notable example is the *chemotaxis* process, where bacteria have sensor- and actuator-proteins which enable the movement towards nutrients or away from poisons via chemical gradient sensing.

[4] Concentration units are expressed in molar weight per volume unit. During metabolomic experiments, a time series of concentration measurements is collected by taking 'snapshots' at some time instant, thereby freezing the metabolic reactions by a sequence of chemical steps.

rare to have data obtained in the same conditions: they are usually obtained in different laboratories, with different experimental conditions and different bacterial strains.
- Irregular sampling;
- Small data sets, i.e. small amount of measurements. For example, when measuring metabolite concentrations, a typical amount of collected samples is in the order of 10 to 100;
- Few or no replicate data;
- Unobserved states. For example, not all metabolites can be measured in vivo.

3.2 Motivation hybrid modeling

The choice for a hybrid systems framework may be motivated by the following:

- *switch behavior:* a specific (genetic) switch or some signal transduction mechanism occurring in a particular organism is studied, and/or one wants to take qualitative knowledge into account, and/or
- *approximation:* the underlying dynamics, e.g. enzyme kinetics, are not known or specified *a priori* and collected data is not reliable enough to deduce a non-linear relationship and/or;
- *analysis:* the original non-linear system is to difficult to analyze and an (approximative) alternative is sought for.

The main motivation is that a hybrid modeling framework can be assessed with data having two sampling density levels: very low and high sampling density. Low sampling density is typically found with characterization experiments, e.g. to detect whether or not protein A is produced due to a gene triggering event B. Hence, a transition between discrete states [5] is associated to event B, while high sampling density data enable the reconstruction of e.g. a time series of the concentration of A. A hybrid framework also has its drawbacks. One of them is the exponential growth of discrete modes with increasing switch thresholds and the associated identification difficulty of detecting these (model) switches. Furthermore, apart from model class selection, a balance between number of identified switches and parameter uncertainty bounds, i.e. 'goodness of fit' should be found and formulated in an identification criterion. Also, parameters of the switched hybrid system may not have a deterministic interpretation. And although analysis tasks like identifiability tests may be more straightforward for linear switched systems than for non-linear systems (Vidal et al., 2002), these results are in its infancy and no implementation of these results in biological applications is yet known to us.

Thus, careful analysis is always needed, possibly after applying reduction of the original model. It is furthermore important to check and analyze the identification procedure with respect to (i) irregular sampling, (ii) post-data processing, (iii) identification criteria, (iv) unbiasedness properties when dealing with small data sets, and (v) need of excitation of the system. When considering hybrid system identification, some of these questions are still open, even with the simplest form of hybrid models: piecewise linear systems. A selected subset of hybrid model classes and their identification challenges will be discussed briefly.

[5] The restrictions in sampling density and the need of analysis tools have largely motivated researchers (e.g. de Jong (2002)) to use a qualitive, linear switched systems modeling approach for genetic regulatory networks, although in reality the assumed discrete transition or switch behaves much 'smoother'.

4. HYBRID SYSTEM BIOLOGY IDENTIFICATION

This section is divided in two parts: Section 4.1 where light-gray box, i.e. deterministic hybrid models with additive noise assumed, and their identification are treated and Section 4.2 where the use of gray-box, stochastic hybrid models in cell systems are discussed.

4.1 Deterministic, Light-gray Box Hybrid Systems

There are numerous subclasses of deterministic hybrid systems known, for example Piece-wise Affine (PWA), Mixed Logic Dynamic (MLD), (Extended) Linear Complementary ((E)LC), Max-Min-Plus Scaling (MMPS), Petri Nets and Discrete Event systems, see also (Bemporad and Morari, 1999). Most of these modeling frameworks have also been used for the modeling biological systems, see e.g. (de Jong, 2002; Alur et al., 2002). The use of hybrid modeling concepts like hybrid automata and Petri nets provide powerful descriptive tools and are often embraced by system biologists. However, the generalizing character of these descriptions exhibits an extensive syntax which complicates control, input design and/or analysis as compared to the tools available for more restrictive classes like PWA, ELC, MLD and/or MMPS hybrid systems.

Furthermore, it is shown that PWA, MLD, (E)LC and MMPS systems are equivalent up to some additional constraints in (Bemporad and Morari, 1999). Therefore, the equivalence results allow an interchange of theoretical properties and computational tools. Consequently, we restrict our attention to the identification of PWA models which have also recently gained considerable momentum on the theoretical side (Ferrari-Trecate et al., 2003; Roll et al., 2004; Rosenqvist and Karlström, 2005; Bemporad et al., 2005; Juloski et al., 2005) as well as on the biological application side (Drulhe et al., 2008; Cinquemani et al., 2008). PWA models have the advantage that the state flows are linear, but, due to the incorporation of discrete transition behavior, can still explain non-linear (switching) behaviors typically encountered in gene regulation networks (de Jong, 2002; Batt et al., 2005), or linearly approximated nonlinear metabolic network models (Musters, 2007).

Discrete-time PWA system. Consider the discrete-time formulation of a PWA linear time-invariant system (Ferrari-Trecate, 2007):

$$\begin{aligned} x(k+1) &= \theta_i^T z(k) + \eta(k) \\ y(k) &= z(k) + e(k) \end{aligned} \qquad \text{for } z(k) \in \mathcal{X}_i \qquad (1)$$

with

$$z(k) = \big[x(k)\ \ x(k-1)\ \cdots\ x(k-n_a)\ \ u(k)\ \ u(k-1)\ \cdots\ u(k-n_b)\ \ 1\big]^T$$

where $x(k) \in \mathbb{R}^n$ is the state vector, $u(k) \in \mathbb{R}^m$ is an exogenous input, $y \in \mathbb{R}^n$ the output observation vector, $\{\mathcal{X}\}_{i=1}^s$ is a region of the polytope $\mathcal{X} \subset \mathbb{R}^{n+m}$, $\theta_i \in \mathbb{R}^{n_a+n_b+1}$ are the parameter vectors, $\eta, e \in \mathbb{R}^n$ are noise sequences.

I/O representations. Several I/O representations of PWA state space models exist, differing on where the noise signal enters the system. In biological applications, the choice for a noise model is generally assumed to be independent of the process model, hence a PWA-OE form with $\eta = 0$ and e a noise sequence in (1) is more obvious. Therefore, other noise representations like PWARX ($\eta \neq 0$ a noise sequence and $e = 0$ in (1)), will not be treated here.

Identification problem. Given N data points, the identification problem is to find a model that best matches the given data $\mathcal{Z} = \{y(k), k = 1, \cdots, N\}$ under some identification criterion (not shown here, the reader is referred to the standard work of (Ljung, 1999)). When dealing with PWA systems, Roll et al. (2004) roughly categorizes two identification methods:

- all parameters, i.e. including the switch thresholds which determine the boundaries of $\{\mathcal{X}\}_{i=1}^{s}$, are identified simultaneously;
- regions $\{\mathcal{X}\}_{i=1}^{s}$ and parameters $\{\theta\}_{i=1}^{s}$ are identified iteratively.

In case the regression regions are known a priori, the identification problem reduces to s standard identification problems.

PWA-OE (PieceWise Affine Output Error). Rosenqvist and Karlström (2005) consider the identification problem when the noise model is independent of the process model and some types of I/O output-error representations when starting from different state-space forms. It is assumed that the switching function is known prior to the identification.

Identification problem for hybrid biological systems. For identification of hybrid biological systems, it is worthwile to take advantage of the available structure. However, identification of hybrid biological models is further complicated since, usually, it is a priori *not precisely* known how many modes are present. The issue of unknown number of modes is even more important when considering (linear) switched systems for approximative modeling of metabolic networks. In particular, simultaneous identification (as discussed in the previous section) can be extremely difficult, since the rate of switching may differ by different time scales and frequency spectrum of the input signal. Therefore, for these systems, iterative or segmentation based methodologies are proposed (Cinquemani et al., 2008). For instance, by assuming that gene expression or other characterized measurement profiles have been split into segments generated by a single affine mode, data classification amounts to group together segments that have been produced by the same mode (cf. discrete state of gene). In particular, this operation must be performed in a noisy setting and without using any knowledge on the number of modes excited in the experiment.

In the paper of Ferrari-Trecate (2007), but also in related works, the identification problem of genetic regulatory networks (GRNs) is split in the following iteration of tasks: (1) detection of switches in (gene expression) data by multicut algorithms; (2) classification of data to distinct modes of operation of the whole GRN; (3) reconstruction of concentration threshold variables, e.g. with a branch-and-bound algorithm for finding minimal multicuts and (4) estimation of kinetic parameters in each mode of operation.

A general assumption in system biology applications is that the state variables can be measured. By further assuming additive noise on the measurements, the state-space model result in an OE format. The following simple GRN example illustrates the use of PWA-OE modeling.

Example 1. (PWA-OE biological system). Consider the following deterministic, continuous-time state space model from (de Jong, 2002), describing a *genetic regulatory network*:

$$\dot{z} = \frac{d}{dt}\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} \kappa_1 r(z_3) - \gamma_1 z_1 \\ \kappa_2 z_1 - \gamma_2 z_2 \\ \kappa_3 z_2 - \gamma_3 z_3 \end{bmatrix}, \qquad z_0 = z(0) \qquad (2)$$

where z_1, z_2 and z_3 represent the measured concentration of mRNA, a protein and a certain metabolite, respectively, κ_i are the production constants and γ_i, $i \in \{1, 2, 3\}$, are degradation constants. Mostly, r is modeled as a so-called positive-valued sigmoid function depending on z, termed a *Hill* equation. Nevertheless, the experiments needed for characterization of r, are laborous or even impossible to implement *in vivo*. As an alternative, low-resolution data is usually generated, for instance, it is detected that mRNA is produced in either low or high quantities. In this low sampling density data case, r in (2) can be approximated by a step function, i.e. $\kappa_2 r(x_3) \approx \kappa_{2L}$, with κ_{2L} having a low value if $x_3 < \alpha$ and $\kappa_2 r(x_3) \approx \kappa_{2H}$ and $\kappa_{2H} > \kappa_{2L}$ for high values of metabolite: $x_3 \geq \alpha$, with $\alpha > 0$ a certain threshold.

After rewriting (2) in discrete-time and assuming additive white noise to the observation measurements, it can be shown that (2) turns naturally into a PWA-OE format:

$$\begin{aligned} y(k) &= x(k) + e(k) \\ &= \frac{\kappa_i}{\gamma_i} - \left(\frac{\kappa_i}{\gamma_i} - x(k_0)\right) e^{-\gamma^j (k-k_0)T} + e(k) \end{aligned}$$

where it assumed that in the interval $[k_0, k]$ only mode $j \in \{1, 2\}$ is active depending on threshold α, T is the sampling time interval, κ_i and γ_i are the parameters and $x(k_0)$ can either be treated as parameter or as a measured exogenous input. Estimation techniques for the parameters, including detection of α, are further discussed in (Ferrari-Trecate, 2007).

Note that, the transformation from continuous-time to discrete time may lead to unidentifiable parameters and requires lumping to obtain a new, affine parameter format. Furthermore, it is also assumed that one gene corresponds to one affine mode, while in practise this may not be true. The assumption that switching of genes can be modeled in a deterministical way may also be argued, see Section 4.2 for further discussion.

4.2 Stochastic, gray-box hybrid systems

The interest in stochastic hybrid formalisms has also gained momentum, especially in cases where it seems logical to model the occurence of molecular events with certain probability. Reported examples are the probability of gene switching in the antibioticum-producing bacterium *Bacillus subtilis* and DNA replication mechanism (Hu et al., 2004; Kouretas et al., 2006). Fig. 2 shows a sketch of the mechanism. The examples provide a nice showcase when embracing the modeling notion of Jump Linear Markov chains. Other hybrid model forms are possible, like stochastic Petri Nets, Bayesian networks or Agent Based models, but are not discussed here.

Identification difficulties lie again in data classification and the combinatorial nature of the number of existing modes, but also in reconstruction of the probability rate of switching. In these systems, it is worthwile to exploit the specific biological model structure and implement it by tailored estimation techniques, e.g. particle filters (Doucet et al., 2001), Maximum Likelihood (Tugnait, 1982) or Bayesian estimation (Jilkov and Li, 2004).

5. EXPERIMENT AND INPUT DESIGN PERSPECTIVES

5.1 Closed-loop experimental design

The problem of optimal input and experiment design has received ample attention in the system identification community,

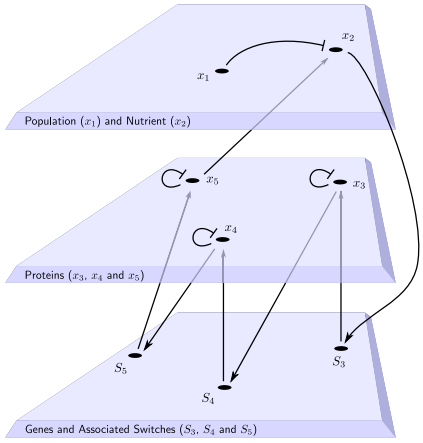

Fig. 2. Regulatory network model of subtilin defense mechanism by *B. subtilis*. Switching of the cascade of genes is represented by the arrows to S_3, S_4 and S_5. S_3 is deterministically switched on when the population (x_1) exceeds certain threshold value and nutrients become scarce. With certain probability, S_4 and S_5 trigger the proteins x_4 and x_5, which in turn aid in the production of the antibioticum subtilin (x_2). Other micro-organisms die and fall apart because of the antibioticum, thereby releasing nutrients. When x_1 falls below the threshold value, the subtilin mechanism is repressed.

see e.g. (Mehra, 1974; Walter and Pronzato, 1990), but has in the early 90s mostly focused on *open-loop* issues. During the last decade, interest in experiment design has renewed and results are extended to the closed-loop domain, with emphasis on *least-costly identification for control*, (Bombois et al., 2006; Hjalmarsson, 2005). Put briefly, this concept refers to achieving a prescribed accuracy and/or precision at the lowest possible price, measured in terms of duration of the identification experiment, the perturbation induced by the excitation signal, or any combination of these.

As far as the authors know of, experimental (input) design for biological applications has been limited to open-loop identification problems (e.g. Riel and Sontag (2006); Zak et al. (2003); Stigter et al. (2003)), while closed-loop investigation and costs of experimentation may certainly be relevant, since:

(i) the amount of samples is often limited due to labour-intensive methods and
(ii) there are I/O limitations, i.e. not all external metabolites can be monitored or manipulated.

Input design should also be feasible from a hybrid systems point of view, given the recent advancements in control laws for multi-affine systems (PWA is a subclass of multi-affine systems) such as presented in (Belta et al., 2002; Collins et al., 2006). Examining the results reported in (Stigter et al., 2003; Barenthin et al., 2005), we expect that, also for hybrid (biology) system applications, considerable improvement in parameter estimation can be achieved in a few input design iterations.

5.2 Hybrid system identifiability and information content

Key variables in experimental (input) design are *identifiability* (given the model structure, does the associated parameter estimation problem have a unique, global minimum?) and *informativeness* or 'richness' of the I/O data. The first concept is dependent on the model structure and can in principle be checked *a priori*, while the latter is usually captured in the Fisher Information Matrix (FIM) (Mehra, 1974). For PWA systems where an *a priori* known number of modes s is assumed known and the probability density function of the noise sequence e is considered to normally distributed, a 'deterministic' FIM reads:

$$\mathcal{F}(s, N, \theta) = \sum_{i=1}^{s} \sum_{k=1}^{N} \frac{1}{\sigma^2} \left(\frac{\partial y(k)}{\partial \theta} \right)^T \left(\frac{\partial y(k)}{\partial \theta} \right) \qquad (3)$$

where θ are the parameters occuring in the state space formulation, σ represents the standard deviation of the noise and $y \in \mathbb{R}$ the observations.

Note that clearly, $\mathcal{F}$ is also related to the inputs like manipulated extracellular metabolites, the initial conditions, sample strategy and amount of samples N. Furthermore, (3) does not apply to the case where the exact number of discrete modes is *not* known, which is important in the case of finding the number s of switches to hypothesize different reconstructions of regulatory networks. With respect to the data classification problem, a suitable definition for the information content and a selection criterion for the strongly related amount-of-modes problem is subject of ongoing research.

6. CONCLUDING REMARKS

It is postulated that the systems biologists' desire to model several biological mechanisms inside (micro)organisms while there is a mix of 'qualitative' data typically obtained on genetic studies and 'quantitative' data obtained with metabolomic and cell growth studies. Such a mix of data types calls for a hybrid 'light-gray' modeling approach. It is further shown that identification of these models create opportunities for both system biology and identification communities to explore the challenges when dealing with non-linear switching behavior, modeling and measurement uncertainties, parameter estimation, data classification and closed-loop experimental (input) design.

The investigation of experimental (input) design for, and identification of hybrid microbiology systems forms one of the major topics of ongoing research. To be tackled research questions are model order selection, discrete mode classification, least-costly input design and identifiability.

ACKNOWLEDGMENTS

This research is supported by the Delft Centre for Life Science and Technology. Furthermore, the authors gracefully acknowledge Sef J. Heijnen and Paul M.J. Van den Hof for their contributions to fruitful discussions.

REFERENCES

Alur, R., Belta, C., Ivančić, F., Kumar, V., Mintz, M., Pappas, G.J., Rubin, H., and Schug, J. (2002). Modeling and analyzing biomolecular networks. *Computing in Science & Engineering*, 4(1), 20.

Barenthin, M., Jansson, H., and Hjalmarsson, H. (2005). Applications of mixed H_2 and H_∞ input design in identification. In *16th IFAC World Congress on Automatic Control*. Prague.

Batt, G., Ropers, D., de Jong, H., Geiselmann, J., Mateescu, R., Page, M., and Schneider, D. (2005). Analysis and

verification of qualitative models of genetic regulatory networks: A model-checking approach. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI 2005)*. Edinburgh, Scotland.

Belta, C., Habets, L.C.G.J.M., and Kumar, V. (2002). Control of multi-affine systems on rectangles with applications to hybrid biomolecular networks. In *41st IEEE Conference on Decision and Control*, volume 1, 534–539. Las Vegas, USA.

Bemporad, A., Garulli, A., Paoletti, S., and Vicino, A. (2005). A bounded-error approach to piecewise affine system identification. *IEEE Transactions on Automatic Control*, 50(10), 1567–1580.

Bemporad, A. and Morari, M. (1999). Control of systems integrating logic, dynamics, and constraints. *Automatica*, 35(3), 407.

Bombois, X., Scorletti, G., Gevers, M., Van den Hof, P., and Hildebrand, R. (2006). Least costly identification experiment for control. *Automatica*, 42, 1651–1662.

Cinquemani, E., Porreca, R., Ferrari-Trecate, G., and Lygeros, J. (2008). Subtilin production by *Bacillus Subtilis*: Stochastic hybrid models and parameter identification. *IEEE Transactions on Automatic Control*, 53(special issue), 38–50.

Collins, P., Habets, L., Kuut, A., Nool, M., Petreczky, M., and (van Schuppen), J. (2006). ConPAHS - A Software Package for Control of Piecewise-Affine Hybrid Systems. In *IEEE Conference on Computer-Aided Control Systems Design*, 76–81. Munich, Germany.

Crampin, E. (2006). System identification challenges from systems biology. In *14th IFAC Symposium on System Identification*, 81–93. Newcastle, Australia.

de Jong, H. (2002). Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology*, 9(1), 67.

Doucet, A., Gordon, N., and Krishnamurthy, V. (2001). Particle filters for state estimation of jump markov linear systems. *IEEE Transactions on Signal Processing*, 49(3), 613–624.

Doyle, F. (2008). Robust control in biology: From genes to cells to systems. In *Proceedings of the 17th World Congress of The International Federation of Automatic Control*, 3470–3479. Seoul, Korea.

Drulhe, S., Ferrari-Trecate, G., and de Jong, H. (2008). The switching threshold reconstruction problem for piecewise-affine models of genetic regulatory networks. *IEEE Transactions on Automatic Control*, 53(Special Issue), 153–165.

Ferrari-Trecate, G. (2007). Hybrid identification methods for the reconstruction of genetic regulatory networks. In *Proceedings of the European Control Conference 2007*, 4845–4852. Kos, Greece.

Ferrari-Trecate, G., Muselli, M., Liberati, D., and Morari, M. (2003). A clustering technique for the identification of piecewise affine systems. *Automatica*, 39(2), 205–217.

Hjalmarsson, H. (2005). From experiment design to closed-loop control. *Automatica*, 41(3), 393–438.

Hu, J., Wu, W.C., and Sastry, S. (2004). Modeling subtilin production in *Bacillus subtilis* using stochastic hybrid systems. In R. Alur and G. Pappas (eds.), *HSCC*, *LNCS*, volume 2293, 417–431. Springer-Verlag, Berlin, Germany.

Jilkov, V. and Li, X. (2004). Online bayesian estimation of transition probabilities for markovian jump systems. *IEEE Transactions on Signal Processing*, 52(6), 1620–1630.

Juloski, A., Heemels, W., Ferrari-Trecate, G., Vidal, R., Paoletti, S., and Niessen, J. (2005). Comparison of four procedures for the identification of hybrid systems. In M. Morari and L. Thiele (eds.), *Hybrid Systems: Computation and Control*, volume 3414 of *LNCS*, 354–369. Springer-Verlag Berlin.

Klipp, E., Herwig, R., Kowald, A., Wierling, C., and Lehrach, H. (2005). *Systems Biology in Practice. Concepts, Implementation and Application*. Wiley-VCH Verlag.

Kouretas, P., Koutroumpas, K., Lygeros, J., and Lygerou, Z. (2006). *Stochastic Hybrid Systems*, chapter Stochastic Hybrid Modeling of Biochemical Processes. CRC Press.

Ljung, L. (1999). *System Identification: Theory for the User (2nd Edition)*. Prentice Hall, New Jersey.

Madrid, R.E. and Felice, C.J. (2005). Microbial biomass estimation. *Critical Reviews in Biotechnology*, 25(3), 97–112.

Mehra, R. (1974). Optimal input signals for parameter estimation in dynamic systems–survey and new results. *IEEE Transactions on Automatic Control*, 19(6), 753–768.

Musters, M. (2007). *Qualitative Modeling in Computational Systems Biology: Applied to Vascular Aging*. Ph.D. thesis, Eindhoven University of Technology.

Nöh, K., Gronke, K., Luo, B., Takors, R., Oldiges, M., and Wiechert, W. (2006). Metabolic flux analysis at ultra short time scale: Isotopically non-stationary 13c labeling experiments. *Journal of Biotechnology*, 129(2), 249–267.

Riel, N.v. and Sontag, E.D. (2006). Parameter estimation in models combining signal transduction and metabolic pathways: the dependent input approach. In *IEE Proceedings Systems Biology*, volume 153(4), 263–274.

Roll, J., Bemporad, A., and Ljung, L. (2004). Identification of piecewise affine systems via mixed-integer programming. *Automatica*, 40(1), 37–50.

Rosenqvist, F. and Karlström, Å. (2005). Realisation and estimation of piecewise-linear output-error models. *Automatica*, 41(3), 545–551.

Sontag, E.D. (2003). Molecular systems biology and control. *European Journal of Control*.

Sontag, E.D. (2008). For differential equations with r parameters, $2r + 1$ experiments are enough for identification. *Journal of Nonlinear Science*, 12, 553–583.

Stigter, J., Vries, D., and Keesman, K. (2003). On adaptive optimal input design. In *Proceedings of the European Conference on Control*. Cambridge, U.K.

Tugnait, J. (1982). Adaptive estimation and identification for discrete systems with markov jump parameters. *IEEE Transactions on Automatic Control*, 27(5), 1054–1065.

Vidal, R., Chiuso, A., and Soatto, S. (2002). Observability and identifiability of jump linear systems. In *Proceedings of the 41st IEEE Conference on Decision and Control*, volume 4, 3614–3619.

Walter, E. and Pronzato, L. (1990). Qualitative and quantitative experiment design for phenomenological models—a survey. *Automatica*, 26(2), 195–213.

Zak, D.E., Gonye, G.E., Schwaber, J.S., and Doyle, Francis J., I. (2003). Importance of input perturbations and stochastic gene expression in the reverse engineering of genetic regulatory networks: Insights from an identifiability analysis of an in silico network. *Genome Research*, 13(11), 2396–2405.