Mathematical model library for recombinant *e.coli* cultivation process

Mantas Butkus
Kaunas University of Technology,
Department of Automation,
Kaunas, Lithuania
mantas.butkus@ktu.edu

Vytautas Galvanauskas
Kaunas University of Technology,
Department of Automation,
Kaunas, Lithuania
vytautas.galvanauskas@ktu.lt

***Abstract*—Biotechnological processes are among the most complicated control objects that require deep knowledge about the process. These systems have nonlinear relationships between process variables and properties that vary over time. Usually such processes are hard to model and require exceptional knowledge and experience in this field. In this review article studies conducted within the last five years in the biotechnology field, that used various model types (mechanistic models, neural networks, fuzzy models) to model cultivation processes were analyzed. Recommendations on what type of models should be used taking into account available process knowledge and experimental data were provided. Mechanistic models are best suited if there is a lack in experience in this field, advanced models like neural networks, fuzzy logic or hybrid models should be used if there is enough experimental data and process knowledge since these models tend to model the process more precisely and take in to account parameters or phenomena that cannot be described by mechanistic models.**

Keywords—biotechnological processes, neural networks, fuzzy logic, cell growth modeling.

I. Introduction

Biotechnological processes are among the most complicated control objects that are characterized by all the properties complicating control: nonlinear relationships between the process variables, dynamic properties of such processes significantly change with time, the processes lack in reliable sensors for state monitoring [13]. Therefore, development of effective control systems is a relevant bioengineering task. Most of the control systems these days rely on mathematical models that are well-known but not always describe the process well or simplify the process. *E. coli* is mostly used in biotechnology, since it is well-known and researched [13]. However, there are no clear recommendations what kind of models should be used in different cases. In order to enrich the understanding of biotechnological modeling and selecting the best suited model the authors compiled a review on the methods used to model *E. coli* cultivation.

The aim of this article is to present various kinetic models for recombinant cultivation processes and recommendations on what kind of models to use depending on the process and gathered data. In Section **II** the process how studies were selected and analyzed is presented. In Section **III,** an explanation how, biotechnological processes are modeled and various models that have been used in the selected researches are presented. Section **IV** provides recommendations on which models should be used depending on the process knowledge and availability of experimental data.

II. Methods

This review was conducted using Google Scholar database. Google Scholar is an open access scholarly search engine that consists of full-text journal articles, books, and other scholarly documents. Even though this database has been criticized by many scholars because of its shortcomings on bibliometric purposes [15, 16], it is still one of the mainly used databases because of its broader coverage. Relevant articles were filtered out and processed according to the following rules and criteria:

- „Modeling“ is mentioned in the topic of the article.
- The article was published after 2014.
- Biotechnological cultivation processes are only analyzed.
- *E.coli* cultivation processes are preferred.
- Article is an open access resource.
- Article is within the first 30 pages of Google Scholar search.

In Figure 1 the structure of the selection of articles is described.

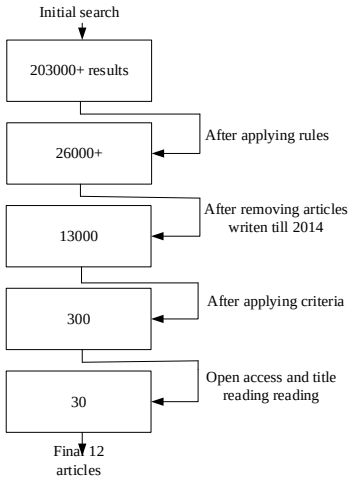

Fig. 1. Flow diagram of literature search

After selection of the relevant papers twelve articles [1-12] were selected and analyzed to determine what kind of models are used to model recombinant *E.coli* cultivation processes within the last five years.

III. Biotechnological process modelling

In order to model biotechnological processes, mass and energy balance equations for the modeled process should be created [13]. The balance equations are created in accordance with the mass conservation law. This means that the mass change in the bioreactor occurs due to:

- chemical reactions that occur in the bioreactor thus creating new products;
- quantity of material supplied by external material flows;
- the amount of culture medium containing the material in question is removed from the bioreactor.

The equation for mass balance of materials is described by:

$$\frac{d(C_1V)}{dt} = q_1C_2V + C_{in1}F_{in} - C_1F_{out}, \quad (1)$$

where, C_1 is the concentration of the material in the reactor, V is the volume of the medium. The amount of material in the medium will be equal to the product of these two variables. q_1 is the specific reaction rate relative to the concentration of C_2 material, in other words, this value indicates the amount of material C_1 formed per unit of mass C_2 per one-time unit. F_{in} and F_{out} are the input and output flows. The change in volume of the medium can only occur due to the flows into and out of the reactor. It can be described by the following differential equation:

$$\frac{dV}{dt} = F_{in} - F_{out}F_{out} \quad (2)$$

After the transformations of the equations (1)-(2) one gets final differential equation for the mass balance:

$$\frac{dC_1}{dt} = q_1C_2 + \frac{F_{in}}{V}(C_{in1} - C_1) \quad (3)$$

The change in concentration is not directly dependent on the outflow flow, and after taking a small sample, the concentration of the substance C_1 will not change drastically, However, the outflow determines the volumetric variation of the medium, while the volume is already included in the equation.

The specific reaction rates in the previously discussed mass balance equations can be modeled by different types of models. The authors will further cover the mechanistic models of these reaction rates. The main growth indicator for microorganisms is the growth rate. For example, a new *E. coli* cell, using substrate, is generated in about 40 minutes if the temperature is 37 degrees Celsius, and some other types of bacteria divide even faster [13]. Naturally, the question is how to measure the number of cells that are formed. It is possible to estimate their number, but as the cells grow and divide, it is decided that the best way to characterize the number of cells is to determine their total mass, i.e. to calculate biomass amount. Growing biomass creates new cells that utilize nutrients and release vital products. Therefore, it is common to express these specific rates for biomass. In the 1930s, Monod described the growth of biomass at the specific rate of biomass growth, which is expressed by [13]:

$$\mu = \frac{1}{xV}\frac{d(xV)}{dt} = \frac{1}{X}\frac{dX}{dt}, \quad (4)$$

where X is the biomass amount, μ is defined as the relative increase in biomass per unit time. This quantity is not constant during the process and depends on various parameters:

- physiological state of microorganism culture,
- biomass concentration in the medium,
- concentration of substrates,
- pH of medium,
- temperature,
- pressure, etc.

The equation (4) can be used to determine the experimental biomass measurement data, but the modeling of the biomass balance equation is usually a function of certain variables. Below, the most often used kinetic models are presented.

A. Monod kinetics

Monod kinetics is the most commonly used μ relationship in biotechnological process modelling. The specific reaction rate depends on the concentration of the main substrate and is described by the formula:

$$\mu = \mu_{max}\frac{s}{K_s+s} \quad (5)$$

where μ is the specific growth rate of the microorganisms, μ_{max} is the maximum specific growth rate of the microorganisms, s is the concentration of the limiting substrate for growth, K_s is the "half-velocity constant" — the value of s when $\mu/\mu_{max} = 0.5$ μ_{max} and K_s are empirical coefficients to the Monod equation. They will differ between species and based on the ambient environmental conditions [1]. This kinetic model is usually used if the kinetics of the process is not well-known. In a study conducted by Papic *et al.* [2] Monod kinetics was used since the relationship between the produced dsRNA and biomass are unknown. The results showed a 37% increase in the process productivity. Similary the Monod kinetics was used by S. Limoes [3] when modelling recombinant cellulase cultivation. In [7] several Monod kinetic models were used to model a multi-substrate environment. In all presented studies the model was sufficient and fit the experimental data.

B. Moser kinetics

Another well-known Monod kinetic modification is the model proposed by Hermann Moser [4]. Moser added another variable *n*, which integrates the microorganism mutation.

$$\mu = \mu_{max} \frac{s^n}{K_s + s^n} \quad (8)$$

d

In the analyzed studies [5, 6] Moser model was used to study the kinetic behavior of the culture since the microorganism was not well-known. Results showed, that the Moser model is inferior compared with other classical kinetic models.

C. Powell kinetics

The original Monod equation was modified by Powell, introducing the terms of maintenance rate *m* which takes into account some of the limitations of Monod model. The Powell kinetic model is described by the equation:

$$\mu = (\mu_{max} + m) \frac{s}{K_s + s} - m \quad (10)$$

All of the described models are mostly used where no additional data from the process is gathered and are considered as "classical" models that should be used if the processes are not well-known and there is not much experience gathered.

D. Blackbox and hybrid kinetics

Hybrid modelling techniques have emerged as an alternative to classical modelling techniques. Recently, these models are particularly widely used in the field of biotechnological process optimization [10,14]. Hybrid models include mechanistic models, artificial neural networks, fuzzy systems, and expert knowledge-based models into a single system, based on principled process management rules and new information. Mechanistic models are based on the application of fundamental principles and the use of certain simplistic assumptions to model phenomena in the process. Using engineering correlations, one can create different types of empirical models that describe well the nonlinear process properties. Using artificial neural networks, it is possible to successfully model functional relationships when there is a lot of measurement to identify a data model, and fundamental functional relationships between individual modelled state variables are not completely clear. In hybrid models, different parts of biotechnological processes are modelled in different ways. The main goal of modelling is to improve both process management and quality. Therefore, the aim is to model each process parameter as best as possible. Because process parameters are described in a variety of relationships, one way to model nonlinear relationships is to use artificial neural networks. An artificial neural network can be understood as a set of certain nonlinear mathematical relationships such as hyperbolic tangents, logarithmic or sigmoidal functions.

Another method, that is widely used, is the ensemble method [8]. It consists on building an ensemble of alternative models that comply with experimental observations. In particular, models with different complexity are generated and compared with respect to their ability to reproduce key features of the data. To overcome data scarcity and inaccuracies (noise), sampling-based approaches have become popular to yield surrogates for missing knowledge in parameter values [8]. In one of the studies [9] the researchers used random forest and neural networks for biomass and recombinant protein modeling in *Escherichia coli* fed-batch fermentations. The applicability of two machine learning methods, random forest and neural networks, for the prediction of cell dry mass and recombinant protein based on online available process parameters and two-dimensional multi-wavelength fluorescence spectroscopy was investigated. The researched models solely based on routinely measured process variables gave a satisfying prediction accuracy of about ± 4% for the cell dry mass, while additional spectroscopic information allows for an estimation of the protein concentration within ±12% [9]. These studies showed that hybrid models are capable of modeling complex biotechnological systems. According to [10] hybrid models have the following advantages over classical models:

- potentially fewer experiments required for process development and optimization;
- allow to study impact of certain variables without the execution of experiments, e.g., for the initial biomass concentration;
- may provide good extra- and interpolation properties.

E. Fuzzy logic models

An important feature of fuzzy logic is that it is possible to divide information into vague areas using non-specific sets [12]. In contrast to the classical set theory, where, according to a defined feature, the element is strictly assigned to one of the sets, the non-expressive set provides an opportunity to define a gradual transition from one set to another using membership functions. A model of fuzzy sets usually associates input and output variables by compiling if-rules such as:

> ***IF*** *the substrate concentration is low*
> ***AND*** *the specific rate of biomass growth is medium*
> ***AND*** *the concentration of dissolved oxygen is low*
> ***THEN*** *the speed of product production is medium.*

These kinds of models can also be used to model the cell specific growth rate or can be used for model identification. In a study conducted by Ilkova [11] fuzzy logics were used to develop a structural and parametric identification of an *E. coli* fed-batch laboratory process. In this study the authors presented an approach for multicriteria decision-making – InterCriteria Analysis to mathematical modelling of a fermentation process. It is based on the apparatus of index matrices and intuitionistic fuzzy sets. The approach for multicriteria analysis makes it possible to compare certain criteria or estimated by them objects. Basic relationships between different criteria in fed batch fermentation – biomass, substrate, oxygen and carbon dioxide were explored. This allowed to create an adequate model that was able to predict the experimental data.

In a study conducted by Liu [12] fuzzy stochastic Petri nets for modeling biotechnological systems with uncertain kinetic parameters were analyzed. In this research the authors

applied fuzzy stochastic Petri nets by combining the strength of stochastic Petri nets to model stochastic systems with the strength of fuzzy sets to deal with uncertain information, taking into account the fact that in biological systems some kinetic parameters may be uncertain due to incomplete, vague or missing kinetic data, or naturally vary, e.g., between different individuals, experimental conditions, etc.. An application of fuzzy stochastic Petri nets was demonstrated. In summary, their approach is useful to integrate qualitative experimental findings into a quantitative model and to explore the system under study from the quantitative point of view. Fuzzy stochastic Petri nets provide a good means to consider parameter uncertainties in a model and to efficiently analyze how uncertain parameters affect the outputs of a model.

IV. Conclusions

After the analysis, the following recommendations can be taken into account when modeling biotechnological processes. It can be concluded, that the best suited model depends on the experience of the researcher and available measurement data:

1. If there is little experience and lack of knowledge about the process, then mechanistic models should be used to model the process and its dynamics. Monod kinetics are usually used to model biotechnological process biomass growth.
2. If there is sufficient experimental data, hybrid models that implement machine learning methods like neural networks and classical mechanistic models to model the researched process can be used, since these models consider processes parameters or dynamics that are not described or left out in mechanistic models. This type of models requires large sets of experimental data.
3. If there are experts, that have very high process knowledge, fuzzy models can be also used, since they consider atypical process behavior. By assessing the verbal knowledge of the experts complicated systems can be modeled and researched.

These recommendations can be used while deciding what kind of methods to use creating a biotechnological process model.

Acknowledgment

This research was funded by the European Regional Development Fund according to the supported activity "Research Projects Implemented by World-class Researcher Groups" under Measure No. 01.2.2-LMT-K-718.

References

[1] C. P. Jr. Grady, L. J. Harlow and R. R. Riesing, "Effects of the Growth Rate and Influent Substrate Concentration on Effluent Quality from Chemostats Containing Bacteria in Pure and Mixed Culture", Biotechnol. Bioeng., vol. 14, no. 3, pp. 391–410, 1972.

[2] L. Papić, J. Rivas, S. Toledo and J. Romero, "Double-stranded RNA production and the kinetics of recombinant *Escherichia coli* HT115 in fed-batch culture", Biotechnol. Rep., e00292, 2018

[3] S. Limoes, S. F. Rahman, S. Setyahadi and M. Gozan, "Kinetic study of *Escherichia coli* BPPTCC-EgRK2 to produce recombinant cellulase for ethanol production from oil palm empty fruit bunch", IOP Conf. Ser. Earth Environ. Sci., vol. 141, no. 1, 2018.

[4] Y. Pomerleau and M. Perrier, "Estimation of multiple specific growth rates in bioprocesses", AlChE J., vol. 36, no. 2, pp. 207–215, 1990.

[5] F. Ardestani, F. Rezvani and G. D. Najafpour, "Fermentative lactic acid production by lactobacilli: Moser and gompertz kinetic models", Journal of Food Biosciences and Technology, vol. 7, no. 2, pp. 67–74, 2017.

[6] J. C. Leyva-Díaz, L. M. Poyatos, P. Barghini, S. Gorrasi and M. Fenice, "Kinetic modeling of Shewanella baltica KB30 growth on different substrates through respirometry", Microb. Cell Fact., vol. 16, no. 1, 189, 2017.

[7] M. E. Poccia, A. J. Beccaria, R. G. Dondo, "Modeling the microbial growth of two *Escherichia coli* strains in a multi-substrate environment", Braz. J. Chem. Eng., vol. 31, no. 2, pp. 347–354, 2014.

[8] E. Vasilakou, D. Machado, A. Theorell, I. Rocha, K. Nöh, et al., "Current state and challenges for dynamic metabolic modeling", Curr. Opin. Microbiol., vol. 33, pp. 97–104, 2016.

[9] M. Melcher, T. Scharl, B. Spangl, M. Luchner, M. Cserjan, at el., "The potential of random forest and neural networks for biomass and recombinant protein modeling in *Escherichia coli* fed-batch fermentations", Biotechnol. J., vol. 10, no. 11, pp. 1770–1782, 2015.

[10] M. von Stosch, S. Davy, K. Francois, V. Galvanauskas, J. Hamelink, et al., "Hybrid modeling for quality by design and PAT-benefits and challenges of applications in biopharmaceutical industry", Biotechnol. J., vol. 9, no. 6, pp. 719–726, 2014.

[11] T. S. Ilkova and M. M. Petrov, "Intercriteria analysis for identification of *Escherichia coli* fed-batch mathematical model", J. Int. Sci. Publ.: Mater., Meth. Technol, vol. 9, pp. 598–608. (2015).

[12] F. Liu, M. Heiner and M. Yang, "Fuzzy stochastic petri nets for modeling biological systems with uncertain kinetic parameters", PLoS One, vol. 11, no. 2, e0149674, 2016.

[13] V. Galvanauskas and D. Levišauskas, "Biotechnologinių procesų modeliavimas, optimizavimas ir valdymas", Vilniaus pedagoginio universiteto leidykla, ISBN 978-9955-20-261-5, pp. 1–112, 2008.

[14] V. Galvanauskas, R. Simutis and A. Lübbert, "Hybrid process models for process optimisation, monitoring and control", Bioprocess Biosyst. Eng., vol. 26, no. 6, pp. 393–400, 2004.

[15] Napoli, C., Pappalardo, G. and Tramontana, E., 2016. A mathematical model for file fragment diffusion and a neural predictor to manage priority queues over BitTorrent. International Journal of Applied Mathematics and Computer Science, 26(1), pp.147-160.

[16] P. Jacsó, "Google Scholar duped and deduped–the aura of "robometrics"." *Online Information Review* 35.1pp. 154-160, 2011.

[17] I.F. Aguillo, "Is Google Scholar useful for bibliometrics? A webometric analysis." *Scientometrics* 91.2, pp. 343-351, 2012.